Multimodal Artificial Intelligence for Histopathology & Genomics Fusion

by

Areej Alsaafin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2024

© Areej Alsaafin 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

Supervisor:	Hamid Tizhoosh, Professor Department of Artificial Intelligence and Informatics Mayo Clinic, Rochester, MN, USA
External Member:	Metin Nafi Gurcan, Professor General Internal Medicine Wake Forest University, Winston-Salem, NC, USA
Internal-External Member:	Karim Sallaudin Karim, Professor Department of Electrical and Computer Engineering University of Waterloo, Waterloo, ON, Canada
Internal-External Member:	Denise Hileeto, Clinical Associate Professor School of Optometry and Vision Science University of Waterloo, Waterloo, ON, Canada
Internal Member:	Tais Sigaeva, Assistant Professor Department of Systems Design Engineering University of Waterloo, Waterloo, ON, Canada

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The field of medical diagnostics has witnessed a transformative convergence of artificial intelligence (AI) and healthcare data, offering promising avenues for enhancing patient care and disease comprehension. However, this integration of multimodal data, specifically histopathology images and genetic sequencing data, presents unique challenges due to modality disparities and the need for scalable computational solutions. This thesis endeavors to bridge this gap by introducing a comprehensive multimodal framework that harmonizes these indispensable data sources, with a focus on its potential in cancer diagnosis, particularly by leveraging immunogenomic data. The thesis addresses the scarcity of multimodal solutions, primarily centered around unimodal data solutions, thus limiting the realization of the rich insights that can be derived from integrating images and genomic data.

The first stage of this research introduces a novel patching algorithm called "Sequential Patching Lattice for Image Classification and Enquiry" (SPLICE). With the ever-growing scale of whole slide images (WSIs), computerized analysis becomes essential. SPLICE's approach of sequentially analyzing and selecting representative samples from WSIs results in a compact "collage" representation, enabling efficient processing and preserving diagnostically relevant regions. SPLICE addresses the dearth of holistic representation learning solutions for high-resolution WSIs and emphasizes the importance of unsupervised approaches in a landscape dominated by labeled data.

In the second stage, the thesis presents MarbliX "Multimodal Association and Retrieval with Binary Latent Indexed matriX," an innovative multimodal framework that integrates histopathology images with immunogenomic sequencing data, encapsulating them into a concise binary barcode, referred to as "monogram." This binary representation facilitates the establishment of a comprehensive archive, enabling clinicians and pathologists to retrieve and match similar cases. MarbliX empowers healthcare professionals with in-depth insights, leading to more precise diagnoses, reduced variability, and expanded personalized treatment options, particularly in the context of cancer.

This thesis stands at the intersection of AI, medical diagnostics, and multimodal data integration. It addresses the critical need for comprehensive solutions in this evolving field, offering a two-fold approach through SPLICE and MarbliX to unlock the potential of combining diverse data sources. By merging macroscopic tissue information with microscopic (molecular) insights, it aims to advance our understanding of complex diseases and elevate the standard of patient care in the ever-evolving landscape of healthcare, with a specific emphasis on harnessing the power of immunogenomic data in the diagnostic process.

Acknowledgements

At the culmination of this challenging yet rewarding journey, I am grateful to have received guidance and support from various quarters that have shaped this academic endeavor.

First and foremost, I thank God, who overwhelmed me with His mercy and grace, and has given me strength throughout all the challenging moments of completing this dissertation. My success comes from none but God. In Him alone, I have placed my trust, and to Him alone, I turn in humbleness.

I would like to express my gratitude to my supervisor, Prof. Hamid Tizhoosh, for his invaluable guidance and constant encouragement throughout my Ph.D. I also thank Dr. Morteza Babaie for his valuable inputs and feedback at key stages of the research process. Furthermore, I extend my thanks to the external examiner, Prof. Metin Gurcan, and the other examination committee members—Prof. Karim S. Karim, Dr. Denise Hileeto, Dr. Tais Sigaeva—who thoroughly reviewed this dissertation and provided valuable suggestions and feedback. I would like to thank my colleagues for all their collaboration in conducting the research and for the cherished time spent together.

The warmest thank you goes to my lovely mom, Sanaa, the source of my faith and positive energy, for her endless encouragement and sincere prayers that illuminated my path during challenging times. A special thank you to my dad, Dr. Saleh Alsaafin, for raising me to value education and for his constant support to achieve more in my academic journey. I am thankful for the supportive presence of my sisters and brother that relieved my stressful moments with joy and pure compassion. Last but not least, my heartfelt thanks go to my friends and loved ones overseas for uplifting me with their peaceful energy and unspoken love throughout my academic journey.

Dedication

This is dedicated to my beloved parents.

Table of Contents

List of Figures ix

List of Tables xiv

List of Abbreviations xv

1 Introduction 1

1.1 Motivation . . . 4

1.1.1 Machine Learning in Cancer Research . . . 4

1.1.2 Multimodal Data Integration . . . 5

1.1.3 Cancer Immunogenomics . . . 6

1.2 Objectives & Contributions . . . 6

1.3 Thesis Organization . . . 7

2 Background & Literature Review 9

2.1 Digital Pathology & Image Search . . . 10

2.2 Histopathology Image Representation . . . 12

2.3 Artificial Intelligence in Molecular Biology . . . 13

2.4 Immunogenomics in Cancer Research . . . 16

2.5 Multimodal Machine Learning . . . 18

3 SPLICE - A Framework for Histopathology Image Representation **21**
3.1 Introduction . . . 21
3.2 Methodology . . . 23
3.2.1 SPLICE: Color Characterization . . . 23
3.2.2 SPLICE Collage . . . 24
3.3 Collage Indexing: Deep Feature Extraction . . . 28
3.4 Experiments & Results . . . 29
3.4.1 Performance Evaluation on Private Datasets . . . 33
3.4.2 Performance Evaluation on Public Datasets . . . 41
3.5 Discussion . . . 46

4 MarbliX - A Framework for Multimodal Integration & Representation **48**
4.1 Introduction . . . 48
4.2 Methodology . . . 52
4.2.1 Unimodal Transformation . . . 53
4.2.2 Multimodal Latent Association . . . 57
4.2.3 MarbliX Monogram . . . 59
4.3 Experiments . . . 62
4.3.1 Patient Representation Evaluation . . . 65
4.3.2 Multimodal Search Evaluation . . . 73
4.3.3 Monogram Analysis . . . 74
4.4 Discussion . . . 76

5 Summary & Conclusion **78**
5.1 Future Work . . . 80

References **81**

List of Figures

2.1 Cellular processes of transcription and translation, enabling protein production from specific genes. 15

2.2 Depicting the interaction between a T cell receptor (TCR) and a tumor cell antigen (peptide), showing the TCR's remarkable specificity for a particular antigen type. 17

3.1 SPLICE approach for patch selection. SPLICE selects representative patches (*collage*) by analyzing their color characteristics through sequential comparisons, including only patches with unique features in the WSI collage. . . . 24

3.2 Simplified Example for SPLICE *collage* applied on a WSI region (From top to bottom) – Starting from an image, we define a lattice. In every pass, the reference patch (pointer) is marked with a pink square while the similar patches to the reference patch are overlaid. **First Pass:** Patch 1 selected and compared to all other patches, similar patches are marked by 1 and excluded, **Second Pass:** Patch 2 is compared to all remaining patches; similar patches are marked by 2 and excluded, **Third Pass:** Patch 3 is compared to all remaining patches; similar patches are marked by 3 and excluded, **Fourth Pass:** Patch 4 is compared to all remaining patches; similar patches are marked by 4 and excluded, **Fifth Pass:** Patch 5 is compared to all remaining patches; similar patches are marked by 5 and excluded, **Last Pass:** the *collage* of selected patches is generated. 25

3.3 Yottixel's mosaic approach for patch selection. Yottixel selects patches from a WSI by applying a two-step k-means clustering, beginning with color clustering based on color histograms, followed by spatial clustering based on patch coordinates. Yottixel selects a certain percentage of patches (centroids of spatial clusters) from each color cluster. The final set of selected patches is called the *mosaic* that represents the WSI. 29

3.4 Representative patch selection using SPLICE in histopathology whole slide image from Mayo-CRC dataset. These diverse patches, extracted by SPLICE, showcase various patterns within the histopathology image of a colon specimen. The patches are selected by applying sequential analysis, leveraging the color characteristics of the patches for their distinctive representation. . 30

3.5 Distribution of patches across various *collage* sizes, ranging from the 10th to the 50th percentile, employing whole-slide images sourced from the Mayo-CRC dataset. The histogram provides a visual representation of the patch distribution, while the rug plot at the bottom offers a detailed view of individual data points, enhancing the understanding of the data's spread and density. 32

3.6 Macro F1-scores for top-k image retrieval in the Mayo-CRC dataset, showcasing the impact of varying *collage* and *mosaic* sizes. The diagrams illustrate the outcomes derived from leave-one-out and majority vote (MV) techniques for three specific scenarios: (a) Top-1, (b) MV@3, and (c) MV@5. The *mosaic* thresholds employed include 9 patches, or 9 centroids (the smallest possible number of patches by mosaic), 5%, and 10%, while the *collage* thresholds encompass the 10th, 20th, 30th, 40th, and 50th percentiles. . . . 34

3.7 Distribution of macro F1-scores, comparing SPLICE and Yottixel's mosaic, across various majority-vote scenarios, including top-1, top-3, and top-5, for each class within the Mayo-CRC dataset. These classes consist of (a) Cancer-Adjacent Polyps, (b) Non-Recurrent Polyps, and (c) Recurrent Polyps. 36

3.8 Confusion matrices for search and match in the Mayo-CRC dataset with SPLICE's *collage* representation (a), Yottixel's *mosaic* representation with Mayo-CRC (b), SPLICE on Mayo-Liver (c), and Yottixel's mosaic with Mayo-Liver (d). These matrices depict the search results obtained through leave-one-out validation and top-3 majority vote techniques for colorectal cancer and liver cancer datasets. 40

3.9 Visual representation of TCGA dataset counts for diverse cancer types. . . 42

3.10 Chord diagram illustrating the TCGA dataset search results using leave-one-out and majority vote of top 3 matches with *collage* representations generated using SPLICE. The diagram shows the interplay of different cancer classes within the TCGA dataset. Each chord connection depicts the relationships between cancer classes when employing search and match with a majority vote of the top 3 similar whole slide images. 43

4.1 Diagram illustrating MarbliX framework for multimodal data fusion. MarbliX integrates histopathology images and immunogenomic data into a concise binary representation known as "monogram." The workflow encompasses unimodal transformation, multimodal latent association, and the generation of the monogram representation. [this figure is property of Mayo Clinic, Rochester, MN, USA] 51

4.2 The pipeline for utilizing MarbliX in a multimodal pathology search, aiming to identify similar cases to a given query case—exemplified here with a liver cancer patient. The process begins with specimen collection for histopathology staining and genomics analysis. In the second stage, the tissue slide is digitized, and RNA sequencing is conducted to obtain the patient's whole slide image and genomics profile, respectively. In the third stage, following unimodal transformations of the histopathology image and genomics profile, MarbliX integrates the two modalities into a binary multimodal representation, referred to as "monogram". The final stage involves a multimodal search using the generated monogram as a query to identify the top n (in this case, n=3) similar matrices in the archive. The search is performed through Hamming distance comparison. The retrieved matrices can be utilized to identify the similar patients' medical information, including histopathology images, genomics data, treatment history, and risk stratification. 63

4.3 Analysis of reconstructed features from multimodal association learning through hybrid autoencoders trained using histopathology images and immunogenomic data of (a) lung dataset and (b) kidney dataset. The violin plot illustrates the distribution of cosine similarity values between the original and reconstructed embeddings by the hybrid autoencoders, with the dashed line indicating the median. The accompanying bar plot provide a comprehensive overview of the mean squared error (MSE) across all original embeddings and their corresponding reconstructed counterparts. 66

4.4 (a) The triplet loss during the training of MarbliX to learn a multimodal representation, referred to as "monogram," of cases within each of lung and kidney datasets. (b) A heatmap illustrating the distinctions in the generated monogram representations among cases with various subtypes of lung and kidney cancers. The dissimilarity is expressed as the XOR of each monogram with every other monogram in the datasets. 68

4.5 The t-SNE map depicting the distribution of modalities within a high-dimensional space. The visual representation showcases embeddings post-PCA transformation, specifically retaining the initial 64 components. The maps shows the relationship between cancer subtypes within lung (first row) and kidney (second row) datasets. The distributions of uni-modal embeddings are presented in (a) and (e) for the image modality, while (b) and (f) illustrate the immunogenomics modality. Maps (c) and (g) show the distribution of real-value monogram representations, while (d) and (h) display binary monograms, both generated by MarbliX. 70

4.6 Comparative analysis of multimodal search performance - MarbliX versus unimodal representations of histopathology images and immunogenomic data. Mean and standard deviation of macro average F1-score and accuracy for lung (a, c) and kidney (b, d) datasets. Leave-one-out and majority vote criteria are applied to retrieve top-1, MV@3, MV@5, and MV@10 for cases in the test folds. 71

4.7 Multimodal search performance of real and binary MarbliX representations compared to unimodal representations of histopathology images and immunogenomic data. Diamond-shaped points represent the mean of macro-average precision and recall for lung (a, c) and kidney (b, d) test folds. Horizontal bars denote the standard deviation range. Leave-one-out and majority vote (MV) criteria are applied to retrieve top-1, MV@3, MV@5, and MV@10 for cases in the test folds. 72

4.8 Monogram representations generated using MarbliX, for randomly selected patients with Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC). The figure presents matrix bitwise XOR results: the left side illustrates $\{\text{LUAD}_{set} \oplus \text{LUAD}_{set}\}$, the right side $\{\text{LUSC}_{set} \oplus \text{LUSC}_{set}\}$, and the middle $\{\text{LUAD}_{set} \oplus \text{LUSC}_{set}\}$. Yellow pixels highlight features transitioning from 0 to 1, while purple pixels indicate the shift from 1 to 0. This color scheme serves to distinguish the presence or absence of features. In the context of $\{\text{LUAD}_{set} \oplus \text{LUSC}_{set}\}$, yellow pixels signify features unique to LUAD matrices, absent in LUSC, and purple pixels denote features specific to LUSC matrices, absent in LUAD. 75

List of Tables

3.1 Comparison of the performance of SPLICE's *collage*, and Yottixel's *mosaic* using CRC dataset from Mayo Clinic. The evaluation is conducted using majority vote and leave-one-out validation. The table displays the results for the majority vote of *top-1*, *top-3*, and *top-5* in terms of accuracy, precision, recall, and macro F1-score metrics. 38

3.2 Comparison of the performance of SPLICE's *collage*, and Yottixel's *mosaic* using Liver dataset from Mayo Clinic. The evaluation is conducted using majority vote and leave-one-out validation. The table displays the results for the majority vote of *top-1*, *top-3*, and *top-5* in terms of accuracy, precision, recall, and macro F1-score metrics. 39

3.3 Comparison of the performance of SPLICE's *collage*, and Yottixel's *mosaic* using TCGA dataset. The evaluation is conducted using majority vote and leave-one-out validation. The table displays the results for the majority vote of *top-1*, *top-3*, and *top-5* in terms of accuracy, precision, recall, and macro F1-score metrics. 42

3.4 Comparison of the number of selected patches for SPLICE's *collage* and Yottixel's *mosaic*. The search time varies for each method depending on the number of patches used to represent each WSI. A higher number of selected patches may lead to longer processing time for searching and retrieving similar cases to a query image. The storage requirement also increases for a larger number of patche embeddings. 44

List of Abbreviations

WSI	Whole slide image
TCR	T cell receptor
BCR	B cell receptor
SPLICE	Sequential Patching Lattice for Image Classification and Enquiry
TCGA	The Cancer Genome Atlas
MV	Majority Vote
CRC	Colorectal cancer
MSE	Mean squared error
MarbliX	Multimodal Association and Retrieval with Binary Latent Indexed matriX
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
KIRC	Kidney renal clear cell carcinoma
KIRP	Kidney renal papillary cell carcinoma
KICH	Kidney chromophobe

Chapter 1

Introduction

The traditional diagnosis of cancer primarily relies on the microscopic examination of tissue slides. However, over recent decades, the diagnostic process has undergone a transformation with the integration of molecular tests. This evolution has led to more precise grading and prognosis for cancer patients. While molecular assays have made remarkable strides in various aspects, the intricate task of assessing tissue morphology still falls within the realm of skilled pathologists. The landscape of cancer diagnosis is currently experiencing a rapid shift, largely driven by the advent of digital pathology. This transformation is especially attributable to the availability of advanced automated computer vision tools specifically designed for the analysis of histopathology images, commonly known as *whole slide images* (WSIs). Leveraging expansive datasets, the development of sophisticated algorithms capable of analyzing high-resolution WSIs has now become feasible. Among these cutting-edge techniques, deep learning emerges as a powerful tool for image analysis for streamlining the responsibilities of pathologists, significantly expediting the entire diagnostic process.

Modern computer vision models have the potential to revolutionize the histopathology workflow. These models possess the remarkable ability to process vast volumes of data, enabling pathologists to swiftly navigate and evaluate tissue slides. This not only expedites the diagnostic process but also minimizes errors in interpretation and classification.

Deep networks inherently capture the textural essence of histopathology image content, encapsulating intricate patterns that a computer discerns from the image. They have the potential to uncover novel patterns and morphological features that might elude human observation [1]. These learned patterns, often referred to as *feature representations*, are acquired during training and subsequently harnessed for diverse downstream applications. For instance, these feature representations can greatly facilitate the identification of similar histopathology images [2] through advanced image search and retrieval tools. This capability empowers pathologists to achieve enhanced diagnostic accuracy and reliability by relying on computational pathology to explore archives of previously diagnosed cases. However, most approaches developed for digital pathology are of supervised structure, requiring guidance from pathologists to obtain some representative samples from tumor regions that can be used to train a deep model. This is one of the main obstacles slowing down the development of machine learning in medical imaging as manual data annotation is a tedious and expensive process. Pathologists may spend hours annotating a few histopathology images, which in many cases are not enough to train a robust deep model. In addition, inter-observer variability is essentially one of the main issues in cancer diagnosis [3, 4]. Therefore, multiple experts may be needed to annotate and analyze the same histopathology image to achieve more accurate labels for training deep networks.

Machine learning has proven to be a powerful tool not only for analyzing histopathological images but also for significantly improving the performance of pattern recognition tasks involving molecular sequencing data. While traditional approaches may focus solely on morphological features when studying tumors in a given sample case, recent advancements in machine learning techniques have expanded our ability to extract valuable insights from complex molecular data. This development is especially important given the ever-increasing volume of biological sequence data generated in research and clinical settings. Generally, relying solely on morphological features to understand diseases like cancer may provide an incomplete picture. However, with the advent of deep learning, researchers have gained the ability to delve deeper into the molecular and genetic aspects of these diseases. Deep learning models can analyze extensive biological sequence databases, un-

covering intricate patterns and associations that might have been overlooked using conventional methods. Nonetheless, it is crucial to recognize that not all genomic data is directly relevant to cancer research, and not every piece of information contributes to the identification of cancerous cells. Some genomic elements may have little or no bearing on cancer development or progression [5, 6]. This highlights the importance of carefully curating and filtering the data used in deep learning models to ensure that the insights derived are both meaningful and clinically relevant.

In the domain of cancer immunogenomics, the adaptive immune system has demonstrated its remarkable ability to detect tumor antigens at an early stage, initiating a defense against cancer cells [7, 8]. This immune response involves the proliferation of tumor antigen-specific T lymphocytes (T cells) and B lymphocytes (B cells), which play pivotal roles in the fight against cancer [9, 9]. Similarities found in both T cell receptor (TCR) and B cell receptor (BCR) sequences, which are protein complexes located on the surface of T cells and B cells, signify shared antigen specificity among receptors, offering a promising avenue for the discovery of novel therapeutic targets [6, 9, 10]. The quantifiable measures of the adaptive immune system's diversity, as provided by immunogenomic data, contribute to the improved categorization of patients who would derive optimal benefits from specific treatments [9, 11]. This addresses the prevalent observation that patients sharing the same clinical diagnosis or symptoms often respond differently to identical treatments [12, 13]. The analysis of immune cell sequencing data holds promise in predicting the effectiveness of treatments on specific cancerous cells, thereby curbing costs and enhancing the sustainability and efficiency of healthcare systems. Recent research underscores the capacity of deep neural networks to glean meaningful insights from complex immunogenomic data patterns. This encompasses tasks such as outcome prediction or clustering of TCR and BCR sequences to uncover their underlying similarities [6, 14–16].

The integration of diverse data streams derived from various sources, including molecular profiling and histopathology slides, provides pathologists and clinicians with a holistic perspective on cancer's complex manifestations. Each modality illuminates unique facets of tumor biology, collectively offering crucial insights into a patient's prognosis and treat-

ment options. However, the inherently high-dimensional nature of some of these data modalities poses a formidable challenge for manual interpretation, making it arduous for clinicians to extract meaningful information from these multimodal biomedical datasets in order to guide treatment decisions and estimate prognosis. Patient heterogeneity further underscores the need for personalized tumor characterization to optimize treatment strategies [17]. Moreover, this heterogeneity can reveal underlying commonalities between tumors of different cancer types [17]. Leveraging machine learning and deep learning tools to uncover hidden patterns within multimodal data is pivotal in constructing robust models capable of more accurately gauging tumor aggressiveness and predicting treatment outcomes. Despite the availability of multimodal data, their integration into a cohesive computerized framework remains rare; most current automated solutions are tailored to a single data modality. This underscores the untapped potential for harnessing the synergy of diverse data sources in cancer diagnosis and treatment planning.

1.1 Motivation

At the core of this research endeavor, an intersection emerges among three dynamic domains: the latent capabilities of machine learning in cancer research, the essential need for the harmonious integration of diverse multimodal data sources, and the burgeoning frontier of cancer immunogenomics. These convergent factors compel an exploration into the multifaceted landscape of cancer research, with an overarching goal to enhance the comprehension of this intricate disease, refine established diagnostic and therapeutic methodologies, and, ultimately, improve the outcomes experienced by patients.

1.1.1 Machine Learning in Cancer Research

The traditional methods of cancer diagnosis have been significantly enhanced by the integration of molecular tests, leading to more precise grading and prognosis for cancer

patients. However, the visual assessment of tissue morphology, a critical aspect of cancer diagnosis, still relies on skilled pathologists. The emergence of digital pathology, powered by advanced automated computer vision tools and deep models, offers a promising research landscape. These technologies may not only expedite the diagnostic process but also have the potential to uncover hidden patterns and morphological features that may elude human observation. Moreover, machine learning has extended its reach beyond histopathological images to molecular sequencing data, providing a more comprehensive understanding of cancer. Deep learning models can analyze vast biological sequence databases, uncovering intricate patterns that traditional methods might overlook. Nevertheless, the challenge of data annotation and curation, often requiring extensive manual effort, remains a bottleneck. Addressing this challenge is crucial to harness the full potential of machine learning in cancer research.

1.1.2 Multimodal Data Integration

The integration of diverse data streams, including molecular profiling and histopathology slides, offers a more holistic perspective on cancer biology. Each modality provides unique insights into tumor behavior, prognosis, and treatment response. However, the high-dimensional nature of these data sources poses a daunting challenge for manual interpretation. Machine learning techniques have the potential to discover the hidden patterns within these multimodal biomedical datasets. By developing robust models that can effectively fuse and analyze data from various sources, researchers can empower clinicians to make more informed decisions and personalize treatment strategies. This approach becomes especially relevant in addressing patient heterogeneity, which underscores the need for tailored tumor characterization. Additionally, the integration of multimodal data can reveal shared characteristics among tumors of different cancer types, opening up new avenues for research and treatment development.

1.1.3 Cancer Immunogenomics

The field of cancer immunogenomics presents a promising avenue of exploration, highlighting the critical role played by the adaptive immune system in detecting tumor antigens and initiating a defense against cancer cells. Recent research into T cell receptor (TCR) and B cell receptor (BCR) sequences has unveiled the potential for shared antigen specificity among receptors to lead us towards novel therapeutic targets [9, 14, 15]. Furthermore, the application of deep neural networks has showcased their ability to extract valuable insights from intricate immunogenomic data patterns, enabling the prediction of treatment outcomes [6]. This integration of advanced computational techniques not only enhances our understanding of the complex interplay between the immune system and cancer but also promises to revolutionize the field of oncology. It offers a glimpse into a future where tailored immunotherapies, guided by a deeper understanding of immunogenomic nuances, could usher in a new era of precision medicine, offering hope and improved outcomes to patients battling this formidable disease.

1.2 Objectives & Contributions

The primary objective of this thesis is to address the critical need for comprehensive multimodal solutions within the intricate landscape of medical settings, with a specific focus on revolutionizing cancer diagnosis. Recognizing the inherent complexity in harmonizing histopathology images and genomic data, this research aspires to bridge this divide and enhance the standard of patient care through the fusion of these indispensable data modalities.

To achieve this objective, the contributions of this thesis are summarized as follows:

- **Introducing SPLICE**: a novel and unsupervised patching algorithm to address the challenge of processing whole slide images (WSIs). SPLICE employs a sequential analysis approach to process WSIs, generating a compact, representative patch set

coined *collage*. This *collage* efficiently encapsulates the entire image while ensuring comprehensive coverage of relevant tissue regions, reducing computational complexity and storage requirements.

- **Introducing MarbliX**: Multimodal Association and Retrieval with Binary Latent Indexed matriX (MarbliX) is a cutting-edge approach for the fusion of histopathology images with genetic sequencing data. This innovative approach condenses these complex medical data sources into a concise binary representation known as *monogram*. This representation captures hidden patterns shared across patients with the same tumor subtype. By emphasizing the significance of multimodal data integration in medical settings, MarbliX bridges a substantial gap in the literature and addresses the scarcity of multimodal solutions.

- **Learning hidden multimodal patterns**: this thesis employs self-supervised triplet learning to unveil concealed patterns from morphological structure in histopathology images and immunogenetic signature in latent space. By generating similar *monogram* representations for patients sharing the same tumor subtype, it offers valuable insights into disease progression and treatment strategies.

- **Multimodal search and retrieval**: this research introduces a new multimodal search framework for retrieving similar cases based on their multimodal data, histopathology images and immunogenomic data, which has the potential to minimize variability in diagnosis and broaden personalized treatment options.

1.3 Thesis Organization

The thesis unfolds through five chapters, which are structured as follows:

Chapter 2 serves as a comprehensive introduction to essential definitions, fundamental concepts, and a thorough review of related work. It provides an overview of current ap-

proaches, with a focus on multimodality and machine learning applications in medical settings, laying the foundation for the subsequent chapters.

Chapter 3 introduces SPLICE, an novel unsupervised approach designed for the representation of histopathology images. This chapter also demonstrates the practical application of SPLICE in image search.

Chapter 4 introduces MarbliX, an innovative framework engineered for the multimodal integration of histopathology and genomic data through a compact binary representation. This chapter also illuminates the versatile application of MarbliX in the domain of multimodal search and retrieval, showcasing its capability in unlocking insights from diverse data sources.

Chapter 5 brings the thesis to a conclusion, while also illuminating the path ahead with a glimpse into future research endeavors.

Chapter 2

Background & Literature Review

This chapter serves as a foundation for the research undertaken in this dissertation, as it comprehensively surveys essential concepts and explores the existing literature across diverse fields. These fields include digital pathology and image search, representing the transformative shift from traditional microscopic tissue examination to digital technology that offers advanced techniques for image retrieval; histopathology whole slide image representation, which plays a pivotal role in enabling detailed analysis of tissue samples at a macroscopic level by processing gigapixel images; artificial intelligence in molecular biology, which delves into the role of AI in deciphering the intricate processes of biological molecules, aiding in the understanding of molecular mechanisms and interactions; immunogenomics in cancer research, an emerging interdisciplinary domain at the confluence of immunology and genomics in the context of cancer, shedding light on the complex interplay between the immune system and cancer cells; and multimodal machine learning, the fusion of various data sources to empower more comprehensive analysis, allowing for a holistic approach to data integration and interpretation. Together, these domains underline the interdisciplinary nature of the data and tools harnessed in this dissertation, illuminating their collective importance in advancing our comprehension of cancer research and paving the way for innovative approaches to diagnosis and treatment.

2.1 Digital Pathology & Image Search

Pathology is the science of diagnosing diseases by examining tissue biopsies, typically mounted on glass slides and scrutinized through visual inspection via a microscope by skilled pathologists [18, 19]. However, the process can be time-consuming and may require second opinions, potentially delaying patient care until the physical glass slide or specimen reaches a sub-specialized pathologist. In recent years, a transformative shift has occurred in the field of pathology through the accelerated adoption of digital imaging. This involves the conversion of glass tissue slides into high-resolution digital representations known as *whole slide images* (WSIs), achieved by scanning the entire glass slide at a level of detail equivalent to traditional light microscopy [18, 19]. This technological leap has revolutionized the concept of laboratory workflows, allowing histopathology images to be instantly shared, viewed, and analyzed worldwide via computer networks.

The term "digital pathology" encompasses a broader spectrum of processes that maximize the practical utility of tissue images. It includes aspects such as storage, visualization, annotation, and their applications in fields like education, research, and clinical practice [19–21]. The recent surge in digital pathology adoption is closely tied to the advancements in artificial intelligence (AI) within the field of pathology [20–22]. AI tools, including deep learning networks and machine learning algorithms, have showcased remarkable capabilities when applied to digital histopathology images, greatly enhancing the value of digital pathology [20]. Recent research has demonstrated that quantitative analysis of histopathology images using computer vision techniques can reveal insights beyond the human vision capabilities [20]. Remarkably, from a computational perspective, a histopathology image can contain an almost infinite number of patterns [22]. Deep learning, with its automatic pattern recognition capabilities, has gained prominence as the preferred method for image processing in medical applications [1, 23].

Histopathology, in particular, benefits significantly from machine learning algorithms due to the intricate nature of diagnostic decisions, the abundance of data generated through routine patient examinations, digital data archiving, and the continuous advancements in

imaging technologies. Deep neural networks, or simply deep models, have proven their efficiency in various clinical tasks using histopathology images, such as metastasis detection [24, 25], mitosis detection [26–28], tumor grading [29, 30], and classification of cancer subtypes [31, 32]. Furthermore, they provide pathologists and clinicians with valuable insights, contributing to more accurate diagnoses and better patient outcomes [3, 33–36].

In the realm of digital pathology, the predominant focus of AI and machine learning solutions has traditionally revolved around classification. These solutions employ machine learning models designed to make diagnostic decisions on behalf of pathologists [1, 37–41]. However, a paradigm shift is underway towards the development of models that facilitate the retrieval of relevant diagnosed cases sharing similar histopathological features, thereby ushering in a new era of decision support tools [42].

Image search and retrieval systems represent a pivotal part of this transformation. They seek to locate similar images within vast biomedical archives using a query image rather than relying on textual descriptions (e.g., pathology reports). While image retrieval has been a topic of investigation in medical imaging for over two decades [43–45], the advent of digital pathology and machine learning has kindled a renewed interest in this area [2, 46–51]. In image search, histopathology images are translated into feature vectors composed of numerical values. These feature vectors are then employed to search for and retrieve images resembling the characteristics of a given query image. The ultimate goal is to furnish decision support tools capable of aiding experts in rendering more dependable diagnoses or, in some cases, enabling a form of "virtual peer review" [42, 52, 53].

However, despite the potential intelligence and efficiency these systems may bestow upon the pathology ecosystem, they face a multitude of challenges that demand attention to foster the development of robust image search solutions. For instance, several studies have underscored the necessity of histopathological datasets offering comprehensive labeling for various diseases [54–56]. Another formidable obstacle is the gigapixel size of histopathology images. Often, a single image boasts exceedingly large dimensions, presenting computational and storage challenges in processing these images in their entirety.

Despite these formidable challenges, recent research endeavors have yielded promising results in the retrieval of similar images. These breakthroughs hinge on the extraction of deep morphological features from histopathology images using convolutional neural networks (CNNs) [2, 34, 48–50, 52, 57]. As the field of image search and retrieval in digital pathology continues to evolve, addressing these challenges will be pivotal in the development of robust and effective solutions. Digital pathology and the integration of AI technologies are revolutionizing disease diagnosis, accelerating the delivery of critical information, and enhancing the capabilities of pathologists and clinicians. This transformative approach holds great promise for improving patient care and advancing the field of medicine.

2.2 Histopathology Image Representation

Due to the complex nature of pathology images, applying deep learning to analyze WSIs is a challenge. To illustrate, a single WSI can be as massive as $100,000 \times 100,000$ pixels in size. Consequently, conventional computing devices, such as GPUs, struggle to handle such enormous images due to their substantial computational and storage demands. As a result, researchers frequently resort to dividing images into smaller "patches" or tiles, analyzing these sections or subsets instead of the entire image [2]. Several techniques are commonly employed, like thresholding [58] and segmentation networks [59, 60], to segment tissue regions from the background before patching a WSI.

The selection of patches can be done randomly or through more sophisticated methods. For example, "Yottixel" [2] generates a **mosaic** of patches by employing a two-stage k-means clustering approach, with the first stage based on staining and the second stage considering patch locations. On the other hand, many researchers seem to skip patch selection and rather opt for patch-level training followed by fusion at a higher level to achieve a single WSI representation [61, 62]. For instance, in [31], patch-level labels are predicted and then combined by considering the spatial relationships between patches to identify the most discriminative ones. A two-step method presented in [63] employs CNNs for WSI

classification. In the first stage, unsupervised learning is used to compress image patches, preserving their spatial information. In the second stage, the compressed WSI representations are fed into a CNN to predict WSI-level labels. However, training CNN models at the patch level and obtaining multiple representations per WSI can pose challenges for certain tasks, such as image search, which involves comparing two WSIs by measuring the distance between their sets of representations [64]. In [65], an attention-based method was introduced, showing effectiveness in medical imaging. This method considers the contribution of each patch to a certain class, and the weighted vectors are then aggregated to produce one vector for each WSI. More recently, an end-to-end set representation learning approach was proposed in [66], which relies on a permutation-invariant network architecture employing universal set function approximators in a neural network to generate a single representation. Furthermore, the CNN-DS method was introduced for end-to-end WSI representation learning, combining CNN and Deep Sets architecture [64]. In this approach, CNN is applied before the Deep Sets model, with a reshape layer connecting the two models. Additionally, two dense layers are used to preprocess the second input before feeding it into the Deep Sets model to generate a global representation for each WSI.

In summary, although there are not many patch selection methods, the field of WSI representation is evolving rapidly, with various strategies and approaches emerging to tackle the complexities of pathology images, ultimately enabling more accurate and efficient analysis in medical and research settings.

2.3 Artificial Intelligence in Molecular Biology

Molecular biology [67] explores the inner workings of living cells, focusing on the complex machinery that sustains life. At its core, the field centers around proteins, which play a pivotal role in a wide array of essential cellular functions, including metabolism, cell structure maintenance, intracellular and extracellular regulation, cellular movement, and defense mechanisms. Due to their multifaceted responsibilities, proteins are the most

abundant macromolecules in all living organisms [68, 69].

Proteins are constructed from amino acids, the fundamental building blocks of these versatile molecules. Remarkably, there are precisely 20 distinct amino acids that serve as the elemental components of proteins [67]. These amino acids come together in a specific sequence, forming a protein's primary structure, which is often represented as a sequence of letters corresponding to the amino acids [68, 69]. The blueprint for creating proteins resides within the cells' DNA, a long-chain molecule composed of nucleic acids. DNA consists of four primary nucleic acids: Adenine, Thymine, Guanine, and Cytosine[69]. In addition to DNA, cells also produce RNA, which, like DNA, comprises nucleic acids. RNA features a slightly different set of nucleic acids, including Adenine, Uracil, Guanine, and Cytosine, with three nucleic acids shared in common with DNA[68]. Much like amino acids in proteins, nucleic acids are symbolically represented by letters A, U (or T in DNA), G, and C, allowing for the coding and storage of genetic information [67]. Genes are pivotal segments of DNA, functioning as the keys that unlock the cell's ability to produce specific proteins [67]. Each gene contains the instructions for creating a particular protein, ensuring the cell's precision in protein synthesis [68, 70].

To synthesize a protein, the cell executes two essential processes: transcription and translation [67]. During transcription, the gene responsible for producing the desired protein is copied into a short strand of RNA. This RNA molecule effectively holds the genetic code or recipe for synthesizing the required protein [70]. In the subsequent translation process, the RNA molecule is translated from its four-letter nucleic acid code (A, U, G, and C) into the twenty-letter amino acid sequence that defines the protein. Once translation is complete, the newly formed protein undergoes folding into its most stable configuration, shaping its overall structure. This structural conformation, combined with the specific amino acids it comprises, ultimately determines the protein's function within the cell [68, 69]. Figure 2.1 illustrates the intricate relationship between DNA, RNA, and proteins, highlighting the pivotal steps of transcription and translation in protein synthesis.

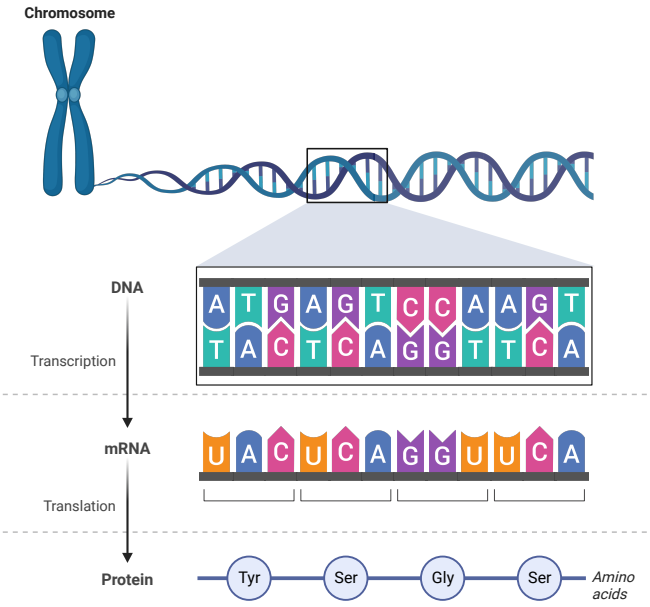


Figure 2.1: Cellular processes of transcription and translation, enabling protein production from specific genes.

The integration of artificial intelligence (AI) and machine learning technologies has brought about a revolution in molecular biology research. These advancements have automated data preprocessing and feature extraction tasks, enhancing data quality and reducing the workload of researchers [71, 72]. AI-driven predictive modeling is being used extensively in genomics and proteomics to predict protein interactions [73–75], gene expressions [76, 77], and disease outcomes [78, 79], aiding in early diagnosis and personalized treatment strategies. In structural biology, AI has achieved remarkable accuracy in predicting protein structures, benefiting drug design and understanding protein functions [80, 81]. Additionally, AI methods like clustering and dimensionality reduction have uncovered hidden patterns in molecular datasets, leading to discoveries in cancer genomics and systems biology [73, 82]. AI tools are driving scientific innovation across various molecular domains, promising ongoing advancements in medicine and biotechnology.

2.4 Immunogenomics in Cancer Research

Immunogenetics[83], a branch of immunology, deals with the interplay between genetics and immune system functioning, driven by genetic sequencing and advanced techniques. This knowledge not only enhances our comprehension of various diseases but also fuels the development of targeted therapies and personalized medicine. The adaptive immune system is a remarkable defense mechanism, characterized by its specificity and ability to recognize and combat invading pathogens or abnormal cells individually [5, 8, 12]. Furthermore, it has the unique capacity to establish a long-lasting protective memory specific to encountered pathogens, enhancing its effectiveness upon subsequent encounters [10]. The cell-mediated immune response, regulated by lymphocytes including B cells and T cells, is pivotal in achieving this immune specificity. T cells and B cells are specialized cells that recognize specific molecular components known as *antigens*, rather than the entire pathogen itself [83]. They achieve this recognition through the presence of specific receptors: the T cell receptor (TCR) for T cells and the B cell receptor (BCR) for B cells [9, 10]. These receptor complexes are situated on the surface of T cells and B cells, respectively, and are at the core of the adaptive immune system. The uniqueness and precision of T cells and B cells in recognizing antigens are established through genetic events known as V(D)J recombination during their development. This process results in the generation of millions of distinct cell variants, each equipped with different antigen-recognizing receptors [11]. Among these receptors, the most variable regions are found in the "complementarity-determining regions 3" (CDR3) within the antigen-binding domains [83]. These CDR3 regions are the primary points of contact with antigens, and the diversity within them is fundamental for the immune system's ability to identify a wide array of antigens [11, 84].

TCRs have the task of identifying and binding to antigens presented on major histocompatibility complex (MHC) molecules, facilitating direct cell-to-cell interactions (see Figure 2.2) [9, 10]. In contrast, BCRs can directly bind to antigens without the need for antigen presentation and employ the variable regions of their immunoglobulin struc-

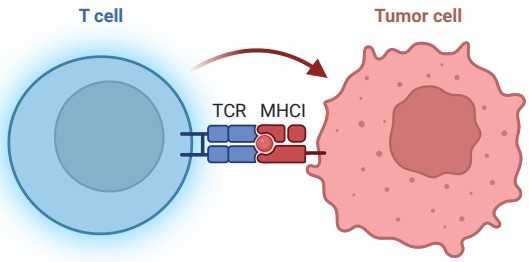


Figure 2.2: Depicting the interaction between a T cell receptor (TCR) and a tumor cell antigen (peptide), showing the TCR's remarkable specificity for a particular antigen type.

ture for precise recognition [12, 84, 85]. Both TCRs and BCRs play pivotal roles in the adaptive immune response, enabling the immune system to recognize a diverse range of antigens [86, 87]. Exploring and analyzing the immune cell repertoire, including TCR and BCR sequences, is crucial for understanding the adaptive immune response. This knowledge is essential for early-stage diagnosis [88–90], treatment planning [91, 92], personalized medicine [89, 93–97], and prognosis prediction [98–100]. Additionally, profiling TCR and BCR sequences can reveal abnormalities or malignancies [101]. For example, a study demonstrated the ability to detect malignant clones in certain blood disorders, achieving higher accuracy and sensitivity compared to conventional diagnostic methods [101].

Our immune system orchestrates a multitude of intricate interactions among cells, organs, and proteins to combat diseases. Nonetheless, cancer frequently evades the body's innate defenses, enabling malignant cells to persist [102]. The ability to monitor variations in both TCR and BCR profiles before or after a viral infection proves indispensable for unraveling the intricacies of the immune response [90, 103]. Enter "immunotherapy" – a revolutionary cancer treatment approach that empowers the immune system to wage a relentless battle against cancer [7, 8, 102]. Given that TCRs and BCRs serve as the primary tools our immune system employs to identify foreign antigens, such as cancer cells, analysis of patterns in both healthy and malignant tissues (or blood samples) can unveil TCRs and

BCRs linked to specific tumors. This groundbreaking discovery opens up new avenues for advancing immunotherapy research [89, 95, 96, 104–106].

The immune system's role in identifying and eradicating cancerous cells hinges on genetic factors shaping immune responses. Variations in immune genes can bolster or hinder the effectiveness of an individual's immune response against cancer cells, enabling tailored cancer immunotherapy approaches aligned with each patient's unique genetic profile [104]. In recent years, significant research endeavors have arisen to pinpoint antigen-specific TCRs and BCRs by harnessing analytical tools to extract profound insights from their sequencing data [107–110]. Nonetheless, the need for more sophisticated and specialized analytical tools remains pressing, as antigen-specific T cells and B cells are dispersed among a multitude of T cells and B cells with characteristics unrelated to the tumor under investigation [6, 16]. Despite the generation of numerous immunogenomic data in recent years, the interpretation of TCR and BCR data remains hampered by the scarcity of known antigen-specificities [111]. Consequently, the clustering of similar TCR or BCR sequences has evolved into a pivotal unsupervised approach for identifying antigen-specific receptors, circumventing the requirement for labeled data [111–114]. Recent efforts have begun to underscore the efficacy of deep learning as a potent tool for extracting invaluable insights from immune cell sequences [115–124].

2.5 Multimodal Machine Learning

Multimodal processing represents a cutting-edge paradigm that is transforming the way we extract insights from an array of diverse data sources. Gone are the days of relying solely on single-modal datasets, which have traditionally been the cornerstone of research. Multimodal processing takes us deep into the intricate interplay between various data sources or modalities, offering a holistic perspective. It stitches together information from different angles, such as histopathology slides, molecular tests, and medical reports, unveiling hidden connections and patterns that would otherwise remain concealed [125, 126]. Each

data modality acts as a unique lens through which we can scrutinize the state of diseases. For example, histopathology slides provide visual data about tissue morphology, while molecular data expose intricate genetic & molecular signatures. The integration of diverse modalities holds immense promise for deeper understanding of complex diseases [127, 128].

Nonetheless, despite its potential, the integration of specific molecular insights with morphological patterns in tissue images remains a relatively uncharted territory. This is where deep learning emerges as a game-changer, poised to ignite this research direction and facilitate the development of algorithms capable of gleaning invaluable information from multimodal clinical data [129–133]. Recent advances in this field have witnessed deep learning models leveraging deep features extracted from histopathology slides. These models transcend mere tissue type identification and expand their horizons to uncover associations with traits that extend beyond mere tissue categorization. They achieve this through transfer learning, a technique that empowers them to discern molecular traits, including genomic alterations, transcriptomic changes, tumor immune infiltration, and even survival outcomes, which were previously not linked to tissue morphology [134, 135].

By autonomously learning the hidden associations between image features and genetic attributes, deep learning equips researchers to navigate the challenge of limited labeled data, a common roadblock in medical research [125]. A noteworthy recent milestone involves the creation of deep learning models that predict gene expression from histopathology images, utilizing gene counts as labels for the images themselves [136]. Alternatively, some researchers adopt a diverse strategy, harnessing labeled data to train individual models for each data modality. This multifaceted approach encompasses clinical reports, radiology images, histopathology images, and DNA/RNA/protein sequencing data, allowing for tailored model selection for each data type [137–140]. These unimodal models are then harmoniously integrated by combining the feature vectors extracted from each one. It's important to note, however, that this approach can yield an abundance of data, necessitating substantial storage capacity and computational power for effective processing. The fusion of data across multiple modalities requires careful consideration of merging techniques to ensure that the integrated information remains coherent and insightful [133].

In a distinctive approach, authors in [141] introduced a methodology to explore the connection between multiple modalities. They employed a graph-based technique to evaluate Autism Spectrum Disorder (ASD) in individuals, integrating eye-tracking data and EEG recordings. This approach involves the creation of relationship between these dual-modal features through a variant of Multi-Layer Perception (MLP) and their subsequent refinement using Deep Graph Convolutional Networks (D-GCN). Another recent study [142] introduced a multimodal deep learning model for predicting tumor mutational burden (TMB) status in colorectal cancer (CRC) patients using histopathology images and clinical data from The Cancer Gene Atlas (TCGA). The study focused on image preprocessing and the fusion of image and clinical features to predict TMB status. In the realm of personalized prostate cancer therapy, a multimodal deep learning approach [143], utilizing clinical data and digital histopathology, outperformed traditional methods in predicting patient outcomes across clinical trials, potentially revolutionizing therapy decisions.

In essence, multimodal processing, driven by machine learning and innovative data fusion methods, holds the potential to unlock new frontiers in our understanding of complex diseases. It bridges the gap between disparate data modalities, empowering researchers to make groundbreaking discoveries.

Chapter 3

SPLICE - A Framework for Histopathology Image Representation

3.1 Introduction

The process of indexing whole slide images (WSIs) involves two primary tasks, each of which is of paramount importance: patching and deep feature extraction. In image processing, the patching stage involves dividing a WSI into patches, also referred to as "tiles," which represent small-scale sub-images that are amenable to comprehensive computer analysis. The concept of "Divide & Conquer" is a venerable strategy routinely used to address complex problems. It provides for the initial division of a complex problem into smaller, more manageable sub-problems, followed by conquering these component parts (in our case patches) in order to build back up a solution to solve the initial complex problem (in our case process a WSI).

WSIs represent an enormous computational challenge due to their gigapixel dimensions and complex patterns, making it infeasible for computers to process them in their entirety.

Hence, one has to split the WSI into more manageable patches to enable computerized processing, which is essentially the *Divide* stage. However, this process has to be *unsupervised* to be applicable to all types of tissue in all organs. As well, reliable patching should ensure comprehensive WSI coverage without inadvertently omitting any diagnostically relevant regions. This *inclusivity* is essential to guarantee that the subsequent phase, the *Conquer* stage, wherein the set of patches is fed into a deep network to represent the whole image. The accuracy of the diagnosis crucially depends on the quality of the WSI representation. All these have to be achieved by selecting as few patches as possible to reduce the computation time and storage requirements.

This chapter introduces a novel patching algorithm called *Sequential Patching Lattice for Image Classification and Enquiry* (SPLICE). At its core, SPLICE aims to sequentially analyze the patches of a WSI at low magnification and only keep one unique sample for each distinctive tissue morphology encountered (3.1). This strategy leads to forming a compact set of representative patches that we name a *collage*.

The design of the SPLICE framework is motivated by several crucial observations that collectively underscore the significance of its development. Firstly, there is a notable scarcity of solutions dedicated to representation learning and retrieval for entire high-resolution WSIs, with the majority of prior work focusing primarily on patch-level processing [48, 57, 144, 145]. Additionally, the preponderance of research efforts has centered around labeled repositories, where malignancy regions are laboriously annotated by pathologists [40, 146, 147]. However, this fixation on labeled data can limit the potential of utilizing unlabelled, real-world data [148]. Furthermore, the practicality of relying on real-valued features for image indexing has posed challenges due to their substantial storage and computational requirements [149, 150]. Similarly, while hashing-based approaches can expedite search operations, they may not facilitate data exchange among diverse repositories [145, 151, 152]. Lastly, the inherent complexity of histopathology images, characterized by diverse edge shapes, intricate structures, stain variations, and high gradient changes, presents a formidable challenge for conventional computer vision algorithms [20, 21]. Considering these observations, the SPLICE framework is designed to provide an unsupervised

solution, addressing the intricacies of high-resolution WSIs while promoting efficiency and versatility in real-world applications.

3.2 Methodology

In this section, the design and implementation of the SPLICE framework is described. This work unveiled two pivotal practical contributions. Firstly, an innovative approach is introduced, aiming to condense the entirety of a histopathology whole slide image into a unique set of patches, referred to as the *collage*. This novel concept forms the bedrock of the methodology, enabling the efficient processing of gigapixel WSIs. Secondly, an end-to-end solution for the indexing and retrieval of WSIs based on their *collage* representation is proposed, providing a powerful and robust solution for image classification and retrieval with significant efficiency improvement compared to state-of-the-art patching solutions.

SPLICE is a framework that follows a systematic process to transform histopathology images into a compact *collage* representation, which preserves essential diagnostic information in a histopathology image while effectively mitigating both computational and analytical challenges. SPLICE can be used to index a queue of WSIs by calculating a compact representative collage for every WSI in the queue. The collage serves as a compact representation of the entire WSI that can be used to perform all downstream tasks, which in turn significantly alleviates the computation burden.

SPLICE processes a WSI by performing (i) color characterization, and (ii) *collage* generation. The stages are described below in more detail.

3.2.1 SPLICE: Color Characterization

In the initial stage of SPLICE, the histopathology image is divided into a number of small patches. This step allows for the efficient handling of the image, breaking it down into more manageable, localized components that can be processed by available computational

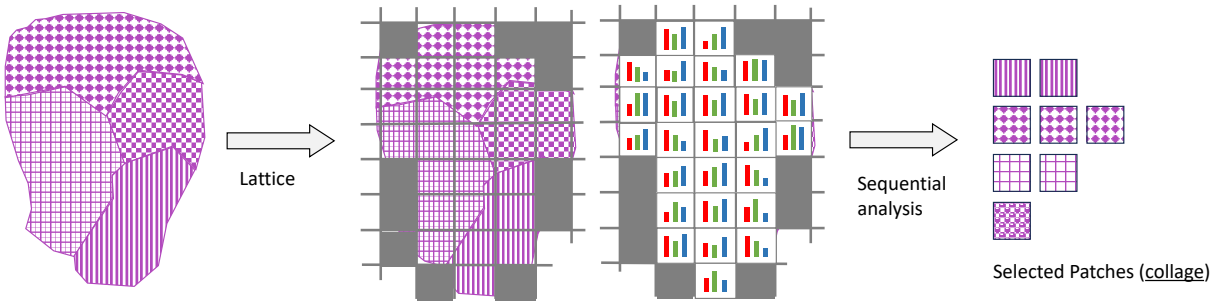


Figure 3.1: SPLICE approach for patch selection. SPLICE selects representative patches (*collage*) by analyzing their color characteristics through sequential comparisons, including only patches with unique features in the WSI collage.

devices. Only patches containing tissue are included in subsequent processing steps (image background excluded). This is done by gathering patches from all segmented tissue areas of a given histopathology image (refer to Algorithm 1, Lines 2-3). This selective inclusion ensures that the analysis focuses on tissue regions while disregarding areas with minimal or no tissue content.

Within this selected subset of patches (that contain tissue), the color distribution within the RGB channels (red, green, and blue) of each individual patch is quantified (see Figure 3.1). Staining is of paramount importance, as it often serves as an indicator of variations in tissue composition. The analysis of color features allows us to identify patches that may exhibit monotonous or redundant color characteristics. Color features can be computed at a lower resolution, such as a 1.25× magnification, as higher resolutions do not necessarily provide a significant advantage for some tasks involving histopathology whole slide images [1, 153, 154]. However, we can change to higher magnification after patch selection.

3.2.2 SPLICE Collage

The second stage of SPLICE introduces a sequential analysis of color features, aiming to filter out patches with repetitive color characteristics (refer to Algorithm 1, Lines 4-17).

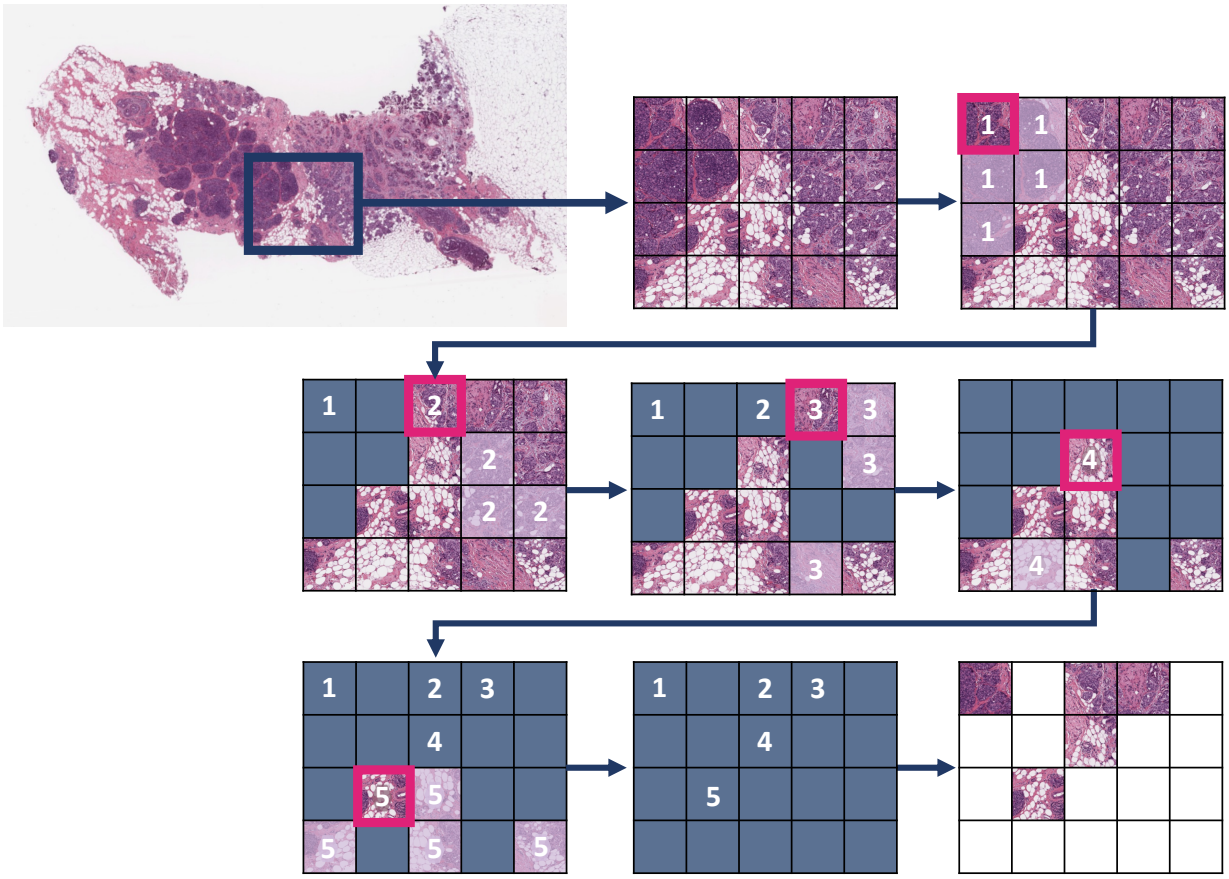


Figure 3.2: Simplified Example for SPLICE *collage* applied on a WSI region (From top to bottom) – Starting from an image, we define a lattice. In every pass, the reference patch (pointer) is marked with a pink square while the similar patches to the reference patch are overlaid. **First Pass:** Patch 1 selected and compared to all other patches, similar patches are marked by 1 and excluded, **Second Pass:** Patch 2 is compared to all remaining patches; similar patches are marked by 2 and excluded, **Third Pass:** Patch 3 is compared to all remaining patches; similar patches are marked by 3 and excluded, **Fourth Pass:** Patch 4 is compared to all remaining patches; similar patches are marked by 4 and excluded, **Fifth Pass:** Patch 5 is compared to all remaining patches; similar patches are marked by 5 and excluded, **Last Pass:** the *collage* of selected patches is generated.

As described in Figure 3.2, this iterative process assesses the similarity between the color features of patches across a series of sequential scans. The color feature vector from the previous stage along with the standard deviation of every color channel is used for each scan: a single patch assumes the role of a reference point when a new scan cycle is initiated, against which the Euclidean distance to every other still non-excluded patch is calculated.

The essence of this sequential color feature analysis lies in its capacity to identify and subsequently exclude patches that exhibit the color characteristics of a reference patch. This step is crucial because regions featuring redundant color attributes typically contribute limited additional information to the analysis. By systematically comparing patches through sequential scans, SPLICE ensures the systematic elimination of patches that do not significantly contribute to the overall representation of the histopathology image.

In each scan, the k^{th} percentile of the Euclidean distances between the reference patch and the remaining patches is computed. A small k value results in a more lenient similarity threshold and thus more patches included in the *collage*. In contrast, a large k value means a more strict similarity threshold to the reference patch, and therefore fewer selected patches. The k^{th} percentile serves as a dynamic threshold, governing the exclusion of patches with color features closely mirroring those of the reference patch. The dynamic nature of this threshold adapts to the specific color distribution of each reference patch, facilitating nuanced and adaptive patch selection. Patches with a distance smaller than the percentile are systematically excluded. This iterative process continues until no further patches can be compared, resulting in the generation of the final SPLICE *collage*, a representative view of the histopathology image. A standard *collage* obtained after sequential analysis can be considerably smaller than the specimen area depicted in a given WSI.

Algorithm 1 SPLICE for unsupervised Divide of a histopathology WSI to generate a "collage" for indexing

```
function SPLICE(image)
    tissue_area ← segment_tissue_area(image)
    patches ← patch_tissue_area(tissue_area)
    collage ← ∅
    excluded_patches ← ∅
    percentile_thresh ← k
    for patch_i in patches do
        if patch_i not in excluded_patches then
            reference_patch ← patch_i
            color_features ← calculate_color_features(reference_patch)
            for patch_j in patches do
                if patch_j not in excluded_patches then
                    distance ← euclidean_distance(color_features, patch_j.color_features)
                    threshold ← k^th_percentile_distance(collage)
                    if distance < threshold then
                        excluded_patches.append(patch_j)
            collage.append(reference_patch)
    return collage
```

3.3 Collage Indexing: Deep Feature Extraction

After SPLICE has divided WSIs into a small set of representative and less redundant patches, namely the collage, the indexing process can be applied to each individual patch within the generated collage utilizing a pre-trained deep neural network. This operation should yield distinctive deep features, effectively encapsulating the intricate high-level tissue characteristics inherent in the collage. Of course, this will directly depend on the quality of embeddings acquired from a pre-trained or fine-tuned network. These collage representations emerge as pivotal elements for an array of downstream analytical tasks, including the facilitation of image search and classification. The utilization of the SPLICE collage for comparing and matching histopathology images enables expeditious and resource-efficient image comparisons, leading to substantial reductions in search time and storage requirements due to the compact nature of the collage. This can be further magnified when *binary* representations, like barcodes proposed by the Yottixel search engine [2], are used. The binarization method known as *MinMax* [155, 156] can be employed. This method converts the rich collage features into more efficient binary codes, enabling rapid real-time searches using the Hamming distance metric. This approach replaces the need for exhaustive k-nearest neighbor searches employing distance metrics like l_2, significantly optimizing search performance, particularly if implemented using logarithmic data types. For an average WSI file of approximately 700 MB, the binary codes can be compressed to as small as around 10 KB, resulting in a remarkable reduction in size, approximately 70,000 times smaller than the original file [2].

With a sufficiently extensive archive of indexed WSI collages, SPLICE can emerge as a powerful and interactive interface, facilitating the generation and indexing of query WSIs while adeptly matching them with the most similar WSIs in the archive, all guided by the collage representations. This transformative process not only streamlines research and analysis but also empowers users to navigate and extract insights from vast repositories of WSIs with unprecedented efficiency and efficacy.



Figure 3.3: Yottixel's mosaic approach for patch selection. Yottixel selects patches from a WSI by applying a two-step k-means clustering, beginning with color clustering based on color histograms, followed by spatial clustering based on patch coordinates. Yottixel selects a certain percentage of patches (centroids of spatial clusters) from each color cluster. The final set of selected patches is called the *mosaic* that represents the WSI.

3.4 Experiments & Results

In this section, the efficacy and efficiency of SPLICE are evaluated for image search and retrieval. This evaluation entails a comparison of search results achieved with SPLICE against those obtained using Yottixel search engine [2]. Only a few works in the literature have addressed the issue of patching WSIs. Yottixel is the state-of-the-art approach for patch selection and WSI matching. The primary objective of this evaluation is to facilitate rapid and efficient searching and retrieval of similar cases based on a given query image. As shown in Figure 3.3, Yottixel employs a two-level k-means clustering approach, considering both color histogram and spatial proximity of patches. Consequently, a certain percentage of patches from each cluster are selected taking into account the intra-cluster heterogeneity. In contrast, SPLICE prioritizes patches with uniquely representative features. Yottixel represents a WSI image by generating a set of patches, referred to as a *mosaic*. Whereas Yottixel's mosaic requires several parameters to be set (number of clusters, sample percentage), SPLICE works with only one parameter and does not need to make assumptions about the number of clusters.

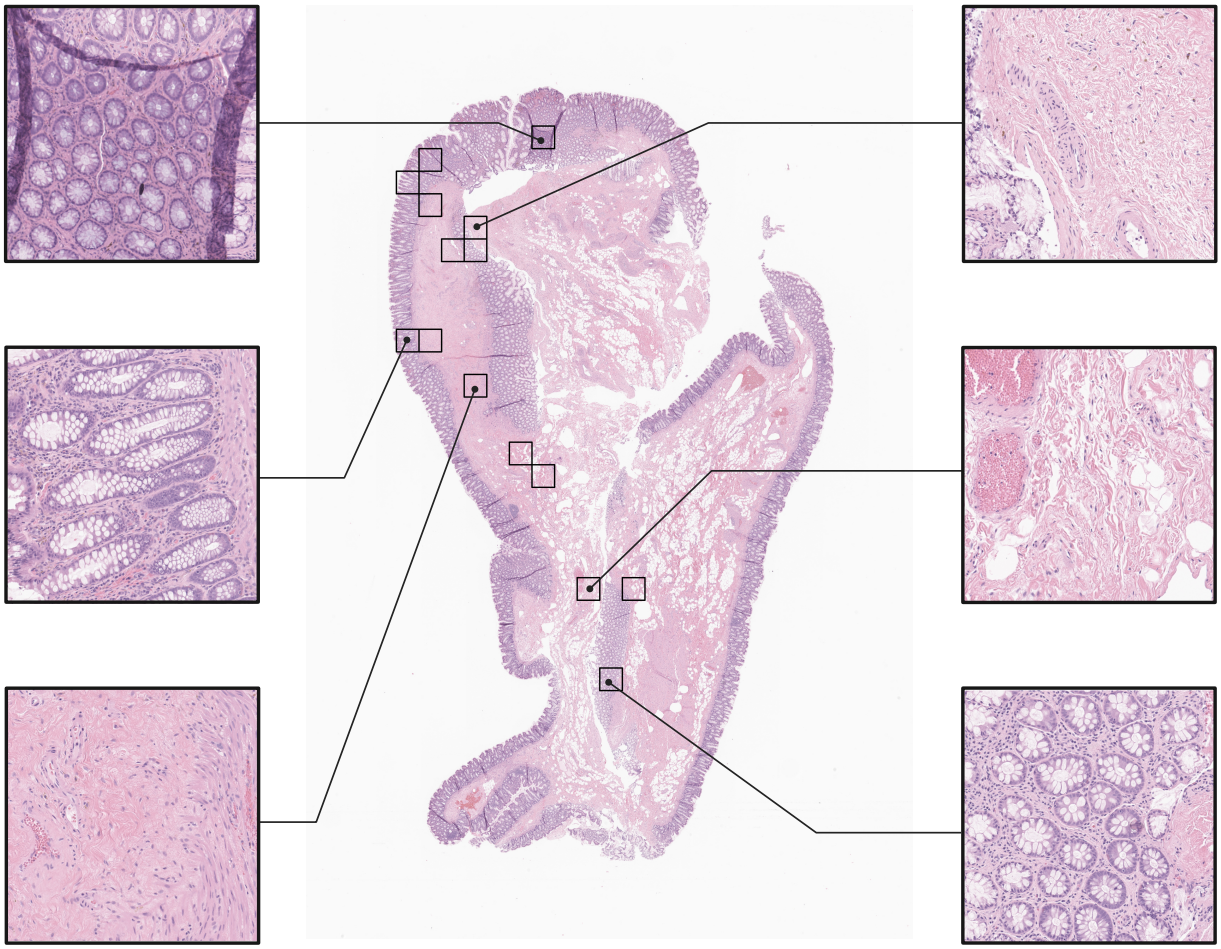


Figure 3.4: Representative patch selection using SPLICE in histopathology whole slide image from Mayo-CRC dataset. These diverse patches, extracted by SPLICE, showcase various patterns within the histopathology image of a colon specimen. The patches are selected by applying sequential analysis, leveraging the color characteristics of the patches for their distinctive representation.

To assess the performance of SPLICE in search scenarios, various datasets were employed as detailed below. In each test scenario, SPLICE was employed for patching alongside Yottixel, yielding a *mosaic* and a *collage* for each WSI. In the case of the collage, the sequential analysis of WSI patches was performed utilizing their color histograms at a very low magnification level of 0.625x, using small patch dimensions of 32×32 pixels. This choice is based on empirical evidence indicating that higher magnifications do not significantly enhance performance for stain/color comparisons. Lower magnifications are favored as high resolutions primarily impact the shape and structure (morphological characteristics) in histopathology images. Subsequently, features are extracted from the patches belonging to both mosaic and collage using KimiaNet, a deep model with a DenseNet topology, trained on all diagnostic WSIs from TCGA (The Cancer Genome Atlas). The patch size employed in these experiments for feature extraction was set to 1024×1024 at 20x magnification as reported in literature [2, 52].

After generating the collage, comprising a low-redundant representative set of patches for each WSI, these patches were fed into KimiaNet to extract deep features. KimiaNet generates a feature vector of size 1024 for every patch in a given collage, resulting in n feature vectors per WSI. For Yottixel, the mosaic was generated for each WSI using the configurations described in the original Yottixel paper [2]. These generated mosaic patches are subsequently processed in KimiaNet to extract features for each patch, yielding a set of mosaic feature vectors. After feature extraction, the resulting feature vectors were binarized using the *MinMax* method [155, 156]. The *MinMax* algorithm maps each feature value to a binary string of 0s and 1s by applying an approximation of 1D derivatives of feature values. This approach enables fast Hamming distance search by comparing the binary barcodes, instead of the real values of feature vectors through Euclidean distance. A set of binary barcodes is obtained for each WSI to represent its mosaic/collage feature vectors. Hence, each WSI in the archive is represented by a set of binary barcodes.

To perform the evaluation, the leave-one-out validation was employed to compare the set of binary barcodes of every image (as a query image) to those of the images stored in the archive. The Hamming distance between the barcodes of the query image and the

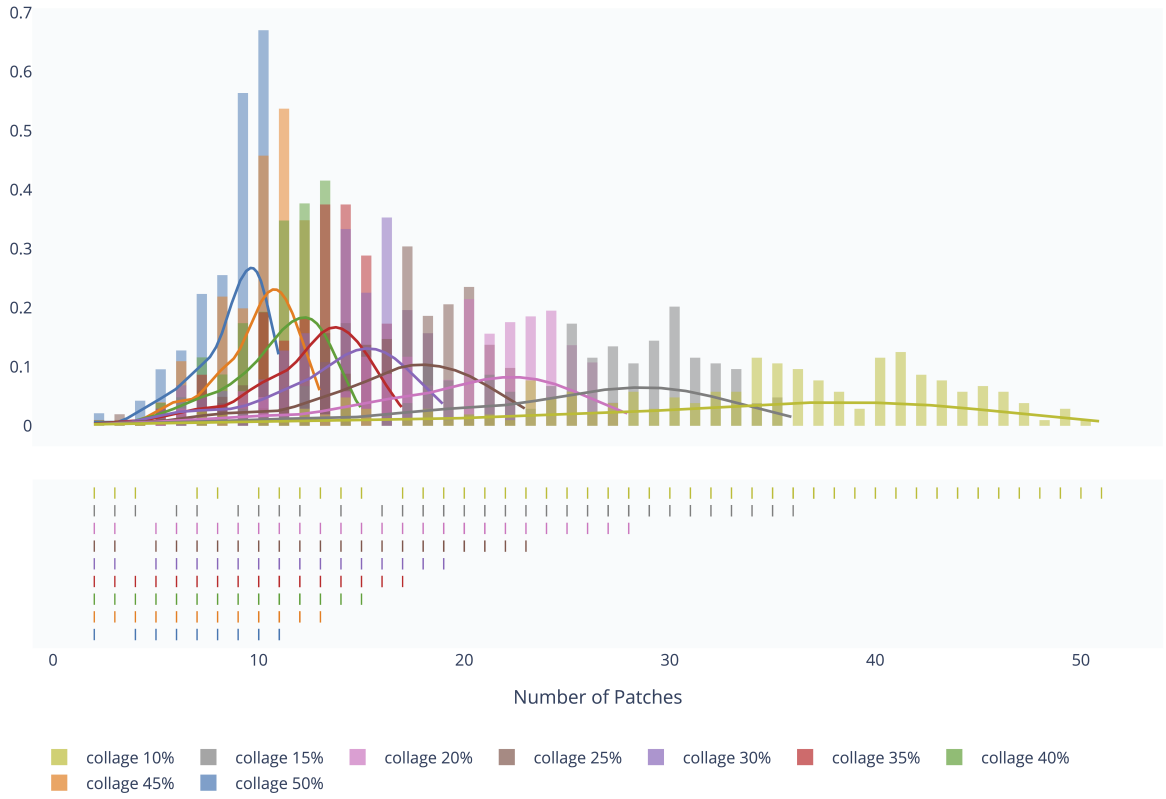


Figure 3.5: Distribution of patches across various *collage* sizes, ranging from the 10th to the 50th percentile, employing whole-slide images sourced from the Mayo-CRC dataset. The histogram provides a visual representation of the patch distribution, while the rug plot at the bottom offers a detailed view of individual data points, enhancing the understanding of the data's spread and density.

barcodes of every image in the archive was calculated. To compare sets of patches, the "median of minimum" approach, introduced in Yottixel [2], was performed to enable WSI-to-WSI matching. This method computes the distance between two WSIs based on the median of the minimum Hamming distances among the barcodes of their patch sets. In essence, the distance between a query WSI (I_q) and another WSI (I) is determined by the median of the minimum Hamming distances between each barcode in the patch set of I_q and the patch set of I. The best match is identified as the one with the median value of the minimum Hamming distances among the patches (i.e., the barcodes) of the two sets. Based on the calculated distances, the top-1 match and the *majority vote* of top-3 (MV@3) and top-5 (MV@5) matches were retrieved for each WSI. *Majority vote* criterion indicates that at least $n/2 + 1$ of the top-n images should have the same class as the query image.

The following sections detail the experiments conducted on three different datasets to evaluate the performance of SPLICE compared to Yottixel's mosaic. Two of the datasets are private datasets obtained from Mayo Clinic: Mayo-CRC and Mayo-Liver. The third dataset is the public TCGA with different primary sites.

3.4.1 Performance Evaluation on Private Datasets

Two private datasets were obtained from Mayo Clinic and were utilized to evaluate the performance of SPLICE in image search applications. The first dataset, Mayo-CRC, comprises a collection of histopathology images from colorectal cancer cases treated at Mayo Clinic. In the Mayo-CRC dataset, 63 WSIs were labeled as Cancer-Adjacent Polyps (CAP), another 63 were labeled as Non-Recurrent Polyps (POP-NR), and 83 were labeled as Recurrent Polyps (POP-R). Note that labels are at WSI level not at pixel level. Within pathology workflows, a patient may have multiple WSIs. However, in this evaluation, every patient has a single diagnostic WSI.

The second dataset, referred to as Mayo-Liver, encompasses histopathology images from liver cancer cases at Mayo Clinic. Mayo-Liver includes three classes: 150 WSIs belonging

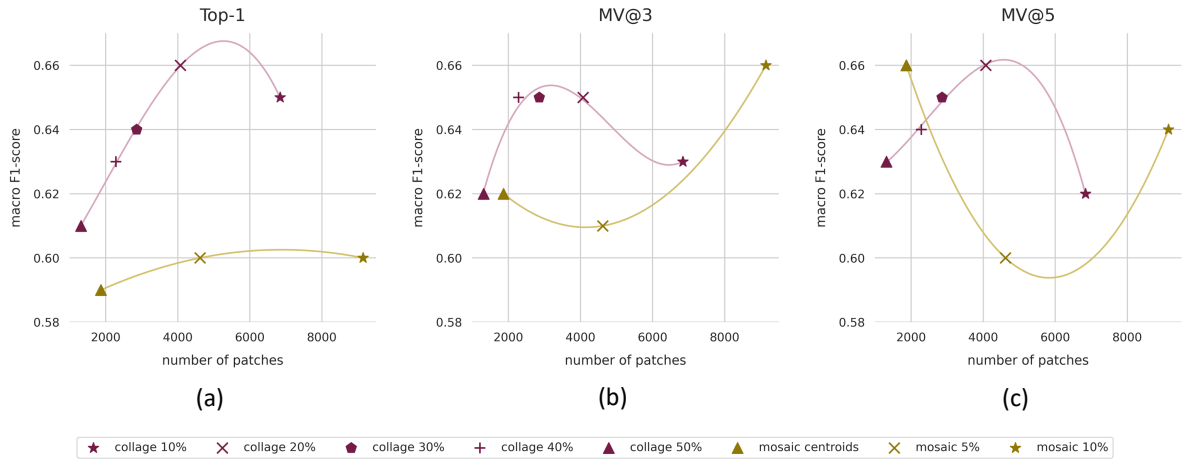


Figure 3.6: Macro F1-scores for top-k image retrieval in the Mayo-CRC dataset, showcasing the impact of varying *collage* and *mosaic* sizes. The diagrams illustrate the outcomes derived from leave-one-out and majority vote (MV) techniques for three specific scenarios: (a) Top-1, (b) MV@3, and (c) MV@5. The *mosaic* thresholds employed include 9 patches, or 9 centroids (the smallest possible number of patches by mosaic), 5%, and 10%, while the *collage* thresholds encompass the 10th, 20th, 30th, 40th, and 50th percentiles.

to Alcoholic Steatohepatitis (ASH), 158 WSIs attributed to Non-Alcoholic Steatohepatitis (NASH), and 18 WSIs categorized as Normal.

For both the Mayo-CRC and Mayo-Liver datasets, high-throughput slide scans were conducted at a magnification of 40x within the facilities of Mayo Clinic Rochester Pathology Research Core, utilizing scanning equipment such as the Aperio AT Turbo Scanner, Aperio AT2 Scanner, and Aperio GT450 Scanner. All slide scans are for slides that contain Formalin-Fixed Paraffin-Embedded (FFPE) tissue specimens.

Yottixel's mosaic and SPLICE take distinct approaches to the patch selection process. Yottixel's mosaic prioritizes well-distributed patch selection from all regions of a WSI, whereas SPLICE places a higher emphasis on the uniqueness of the selected patches. Yottixel's mosaic original configuration, as detailed in [2], employs a k value of 9 for color k-means clustering, while the spatial k-means clustering is set at 0.05, which corresponds

to 5% of patches in every color cluster. Increasing these values results in the selection of a larger number of patches per WSI. When it comes to the *collage* generation, the threshold is dynamically determined based on the population of patches within each WSI. This threshold adapts continuously during each iteration of the sequential scanning process, depending on the remaining patches in the pool. Therefore, the threshold is expressed as a percentile of the calculated distances in each iteration, rather than as a fixed numerical value. Figure 3.5 provides insight into the distribution of the number of patches per WSI across different percentiles, ranging from the 10th to the 50th percentile (the median). A higher percentile corresponds to a more stringent threshold, resulting in fewer patches included in the collage of each WSI. Conversely, a lower percentile implies a more lenient threshold, accommodating more patches. However, as demonstrated in Figure 3.5, using a strict threshold, such as the 10th percentile, may lead to a small number of patches for some WSIs, either due to a limited number of patches (e.g., for small tissue fragments) or because of the high similarity between the patches, resulting in a reduced number of representative patches.

A visual representation of the selected patches is shown in Figure 3.4. The figure shows the selected patches for a WSI with a colon specimen from Mayo-CRC. The patches were selected by applying sequential analysis with the 25th percentile as a similarity threshold for the color characteristics of the patches.

An extensive experiment was conducted to investigate the impact of varying percentiles for SPLICE compared to adjusting the spatial parameter of Yottixel's mosaic. In this experimental analysis, for SPLICE, five thresholds, representing the 10th, 20th, 30th, 40th, and 50th percentiles, were explored. The three thresholds considered for Yottixel's mosaic encompass 9 patches (color cluster centroids), 5% (following the original work's configuration), and 10% to facilitate the inclusion of a higher number of patches per WSI.

Figure 3.6 presents the macro F1-scores for retrieving the top-1, majority vote top-3 (MV@3), and majority vote top-5 (MV@5) similar WSIs for each WSI in the Mayo-CRC dataset, employing leave-one-out evaluation. The figure also shows the number of

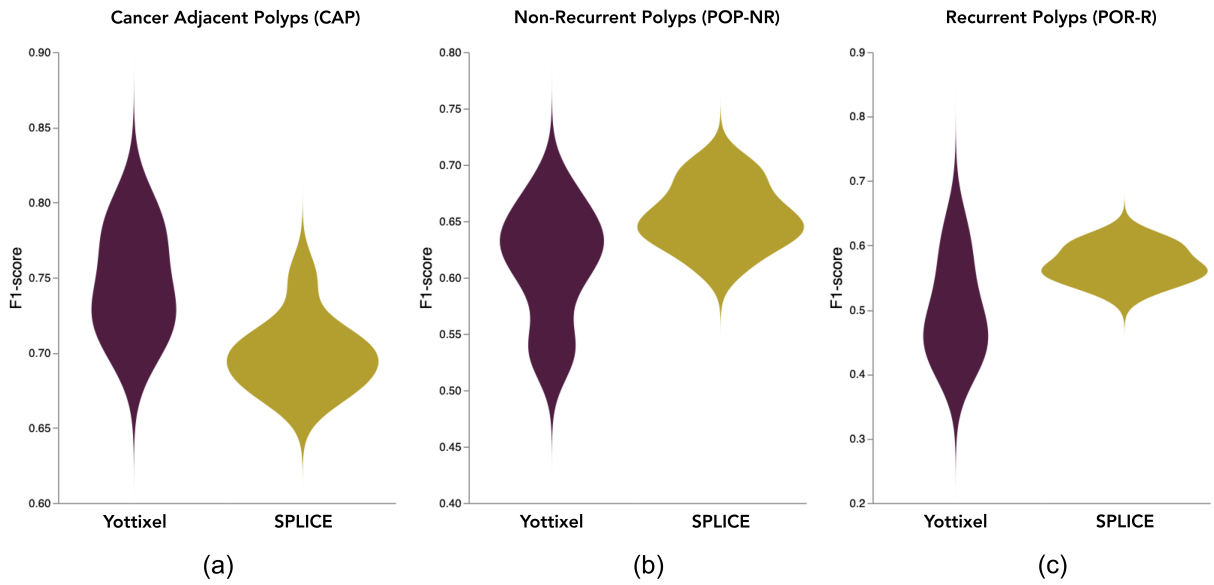


Figure 3.7: Distribution of macro F1-scores, comparing SPLICE and Yottixel's mosaic, across various majority-vote scenarios, including top-1, top-3, and top-5, for each class within the Mayo-CRC dataset. These classes consist of (a) Cancer-Adjacent Polyps, (b) Non-Recurrent Polyps, and (c) Recurrent Polyps.

patches selected for generating the *collage* and *mosaic* for each WSI within the dataset. The highest macro F1-score achieved by SPLICE for top-1 was 0.66 when using the 20th percentile for *collage* creation, with 4,073 patches extracted from 209 WSIs in the Mayo-CRC dataset. In comparison, Yottixel's mosaic attained a macro F1-score of 0.60 for top-1 when employing 5% and 10% thresholds, with a total number of patches amounting to 4,619 and 9,143, respectively. Remarkably, both SPLICE and Yottixel's mosaic exhibited similar performance with a macro F1-score of 0.66 for MV@3 under the strictest thresholds in this experiment (generating the fewest patches for both methods), with 1,814 and 1,868 total patches for SPLICE and Yottixel, respectively. In all measures shown in Figure 3.6, the *collage* with the 20th percentile and the *mosaic* with 5% thresholds result in a comparable number of patches, but SPLICE displayed better performance across top-1, MV@3, and

MV@5 by 4% to 6%. It's essential to consider that when employing MV@5, mosaic utilized more than four times as many patches (9,143) as SPLICE (2,071) to achieve the same level of performance. Overall, SPLICE demonstrated better performance across different percentiles compared to the various thresholds used for mosaic. The general trend for Yottixel across different thresholds is that more patches tend to improve performance, especially for top-1 and MV@3. Additionally, using more patches to generate the *mosaic* had a smaller impact on the retrieval results of top-1 compared to MV@3 and MV@5, indicating that more patches per WSI were needed to achieve a correct vote for the retrieved similar WSIs. On the other hand, including more patches to generate the *collage* resulted in better results in identifying the most similar WSI (top-1), suggesting that SPLICE selects representative and less redundant patches. The performance for MV@3 and MV@5 also improved with more patches until the 20th percentile, when the macro F1-score began to decrease as the criteria to include or exclude patches from a *collage* became less strict. In general, there is a trade-off between performance on one side, and computation and storage requirements on the other side. However, SPLICE mitigated the impact of this trade-off by achieving satisfactory performance with strict thresholds that allow a small number of patches to be used in generating WSI *collages*.

The violin plots in Figure 3.7 depict the distribution of F1-scores for the top-1, top-3, and top-5 majority vote scenarios across three distinct thresholds applied to each class within the Mayo-CRC dataset. For mosaic, 9-patches, 5%, and 10% were considered, while for SPLICE, 20th, 30th, and 4th percentiles were included in this experiment. A noticeable pattern emerges as the distribution of F1-scores for Yottixel exhibits a more pronounced vertical spread, reflecting sensitivity to varying threshold values. In contrast, SPLICE's distribution is characterized by a more horizontal orientation, indicating a more consistent performance, with the threshold having a less pronounced impact. This difference underscores the quality of images incorporated in the *collage* generated by SPLICE, which contributes to stable and reliable performance across a range of threshold values. The Yottixel's mosaic showed better performance in identifying images from the CAP class when a higher number of patches were included. In contrast, SPLICE demonstrated constant

Table 3.1: Comparison of the performance of SPLICE's *collage*, and Yottixel's *mosaic* using CRC dataset from Mayo Clinic. The evaluation is conducted using majority vote and leave-one-out validation. The table displays the results for the majority vote of *top-1*, *top-3*, and *top-5* in terms of accuracy, precision, recall, and macro F1-score metrics.

Method	**Metric**	**Top-1**	**MV@3**	**MV@5**
SPLICE	Accuracy	0.62	0.65	0.63
	Precision	0.63	0.65	0.64
	Recall	0.63	0.66	0.64
	F1-score	0.63	0.65	0.64
Yottixel's mosaic	Accuracy	0.60	0.60	0.60
	Precision	0.62	0.61	0.61
	Recall	0.62	0.62	0.63
	F1-score	0.60	0.61	0.60

performance in categorizing images from the POP-NR and POP-R classes.

Table 3.1 provides a comprehensive performance comparison between collage and mosaic, with a focus on various metrics spanning top-1, top-3, and top-5 majority vote scenarios, assessed using leave-one-out evaluation on the Mayo-CRC dataset. These results were generated by employing the 40th percentile threshold for SPLICE and a 5% threshold for Yottixel's mosaic. The findings depicted in Table 3.1 demonstrate that SPLICE outperforms Yottixel's mosaic across all metrics for each of the three criteria: top-1, top-3, and top-5. For instance, in the context of top-3 majority vote, SPLICE attains an accuracy of 65%, surpassing mosaic's search accuracy of 60%. Furthermore, SPLICE exhibits recall of 66% and precision of 65% in comparison to mosaic's recall of 62% and precision of 61% for the same top-3 criterion.

The assessment was extended to the Mayo-Liver dataset, with results presented in Table 3.2, again employing leave-one-out evaluation across top-1, top-3, and top-5 majority vote scenarios. These results were generated by employing the 35th percentile threshold for SPLICE and a 5% threshold for Yottixel's mosaic. SPLICE showcases a markedly

Table 3.2: Comparison of the performance of SPLICE's *collage*, and Yottixel's *mosaic* using Liver dataset from Mayo Clinic. The evaluation is conducted using majority vote and leave-one-out validation. The table displays the results for the majority vote of *top-1*, *top-3*, and *top-5* in terms of accuracy, precision, recall, and macro F1-score metrics.

Method	**Metric**	**Top-1**	**MV@3**	**MV@5**
SPLICE	Accuracy	0.83	0.79	0.77
	Precision	0.83	0.87	0.85
	Recall	0.74	0.63	0.66
	F1-score	0.77	0.68	0.71
Yottixel's mosaic	Accuracy	0.76	0.79	0.80
	Precision	0.76	0.87	0.88
	Recall	0.59	0.63	0.62
	F1-score	0.62	0.67	0.65

better macro F1-score across all criteria, exhibiting a substantial 15% difference for top-1, achieving an impressive 77% in contrast to mosaic's 62%. In terms of accuracy for top-1, SPLICE outperforms Yottixel's mosaic by 7%, achieving an accuracy rate of 83% as compared to mosaic's 76%. However, the mosaic demonstrates higher accuracy when the top-5 criterion is applied, reaching 80%, surpassing SPLICE's 77%. However, it's interesting to observe that collage and mosaic exhibit remarkably similar performance when assessed under the top-3 majority vote scenario. Of particular interest is the observation that mosaic's performance seems to improve as the number of retrieved images increases, while SPLICE excels at retrieving the most similar top-1 image, demonstrating robust performance across accuracy, recall, precision, and macro F1-score metrics.

Figure 3.8 presents the confusion matrices of the search results achieved by SPLICE's *collage* and Yottixel's *mosaic* representations, specifically when applied to the Mayo-CRC and Mayo-Liver datasets. The matrices were generated through leave-one-out validation. Additionally, the matrices illustrate the outcomes of the top-3 majority vote for each image in both colorectal and liver cancer datasets.

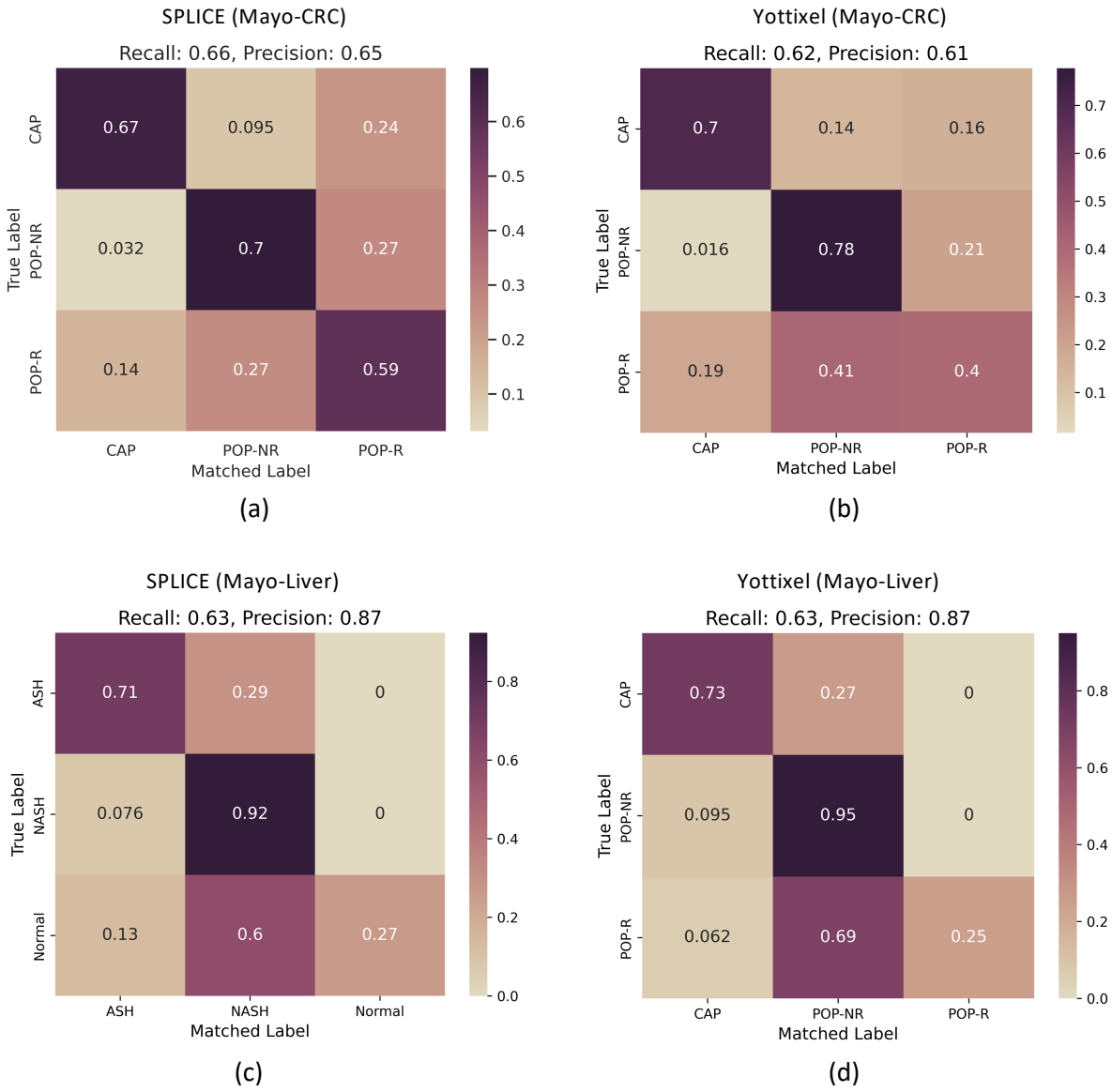


Figure 3.8: Confusion matrices for search and match in the Mayo-CRC dataset with SPLICE's *collage* representation (a), Yottixel's *mosaic* representation with Mayo-CRC (b), SPLICE on Mayo-Liver (c), and Yottixel's mosaic with Mayo-Liver (d). These matrices depict the search results obtained through leave-one-out validation and top-3 majority vote techniques for colorectal cancer and liver cancer datasets.

3.4.2 Performance Evaluation on Public Datasets

SPLICE's performance was further assessed using public datasets obtained from The Cancer Genome Atlas (TCGA) project, which is publicly available through the National Cancer Institute (NCI). As depicted in Figure 3.9, this assessment included a total of 711 WSIs derived from 24 TCGA datasets originating from 20 primary sites. These WSIs contained Formalin-Fixed Paraffin-Embedded (FFPE) tissue specimens. These WSIs from TCGA have been utilized for KimiaNet's testing. Therefore, to ensure an unbiased assessment, all cases that were part of the training set of KimiaNet were excluded from the evaluation.

Table 3.3 presents a comparative analysis of the performance between SPLICE's *collage* and Yottixel's *mosaic* using the TCGA dataset. In this experiment, a 5% threshold was applied for Yottixel's mosaic, while the 30th percentile was employed for SPLICE. The evaluation was carried out using majority vote and leave-one-out validation. Table 3.3 provides a comprehensive overview of the results for top-1, MV@3, and MV@5 in terms of accuracy, precision, recall, and macro F1-score metrics. The results in Table 3.3 demonstrate that collage and mosaic generated comparable results, utilizing 11,663 patches for SPLICE and 25,679 patches for Yottixel's mosaic. Across most evaluation metrics, SPLICE outperforms Yottixel's mosaic, displaying higher accuracy, precision, recall, and macro F1-scores. However, for MV@3, both methods achieve an identical accuracy of 0.78. This comparison underscores the effectiveness of SPLICE, even when using significantly fewer patches than Yottixel's mosaic. In fact, less than half of the patches used by Yottixel's mosaic were used by SPLICE to generate *collage* representations for WSIs.

Figure 3.10 presents a chord diagram illustrating the TCGA dataset search results using leave-one-out and majority vote of top 3 matches with *collage* representations generated using SPLICE. The diagram shows the interplay of different cancer classes within the TCGA dataset. Each chord connection depicts the relationships between cancer classes when employing search and match with a majority vote of the top 3 similar whole slide images. This visualization provides insight into the performance of SPLICE in categorizing and retrieving WSIs from a diverse set of cancer classes within the TCGA dataset.

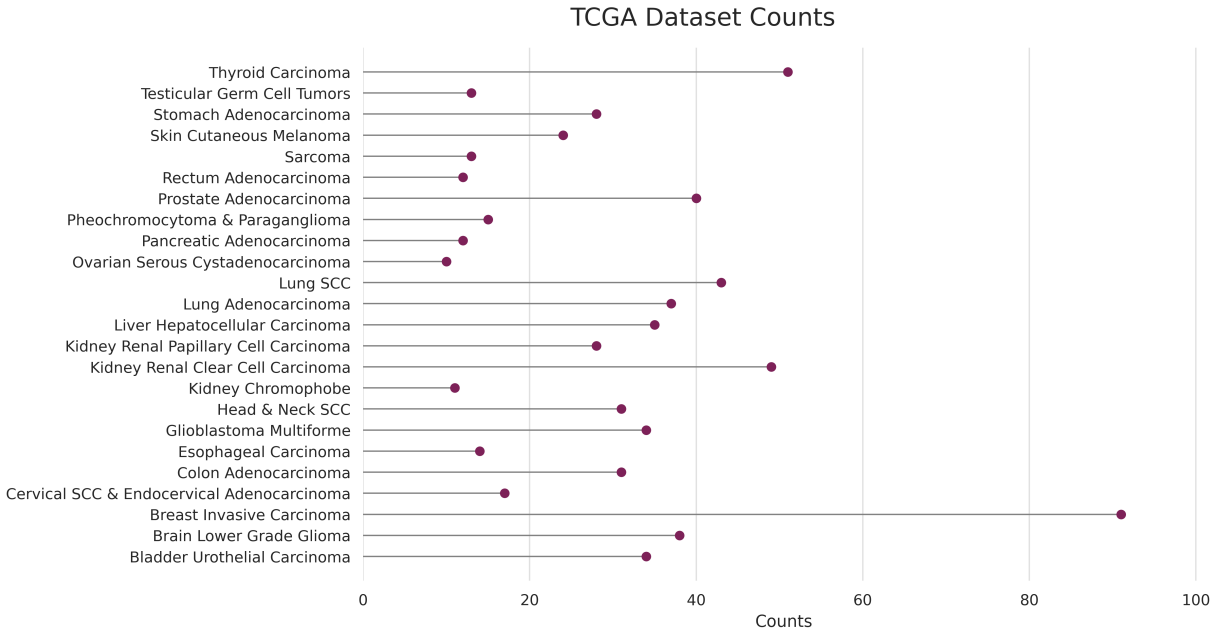


Figure 3.9: Visual representation of TCGA dataset counts for diverse cancer types.

Table 3.3: Comparison of the performance of SPLICE's *collage*, and Yottixel's *mosaic* using TCGA dataset. The evaluation is conducted using majority vote and leave-one-out validation. The table displays the results for the majority vote of *top-1*, *top-3*, and *top-5* in terms of accuracy, precision, recall, and macro F1-score metrics.

Method	**Metric**	**Top-1**	**MV@3**	**MV@5**
SPLICE	Accuracy	0.77	0.78	0.78
	Precision	0.76	0.79	0.79
	Recall	0.72	0.73	0.71
	F1-score	0.74	0.74	0.73
Yottixel's mosaic	Accuracy	0.76	0.78	0.77
	Precision	0.73	0.75	0.74
	Recall	0.71	0.71	0.69
	F1-score	0.71	0.73	0.71

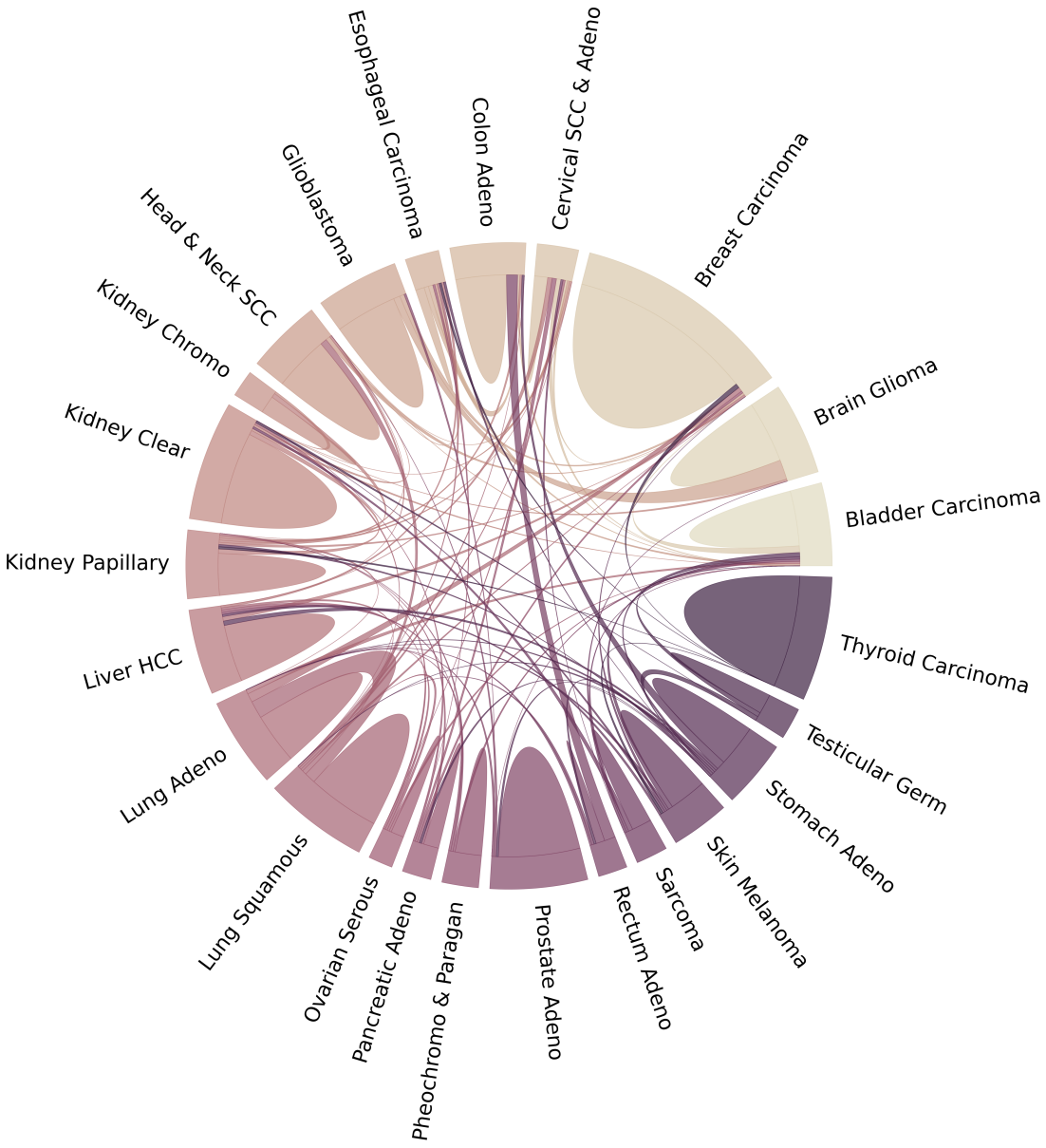


Figure 3.10: Chord diagram illustrating the TCGA dataset search results using leave-one-out and majority vote of top 3 matches with *collage* representations generated using SPLICE. The diagram shows the interplay of different cancer classes within the TCGA dataset. Each chord connection depicts the relationships between cancer classes when employing search and match with a majority vote of the top 3 similar whole slide images.

Table 3.4: Comparison of the number of selected patches for SPLICE's *collage* and Yottixel's *mosaic*. The search time varies for each method depending on the number of patches used to represent each WSI. A higher number of selected patches may lead to longer processing time for searching and retrieving similar cases to a query image. The storage requirement also increases for a larger number of patche embeddings.

	patching	**# patches**	**# patches/WSI**	**storage (KB)**	**time (sec)**
Mayo-CRC	collage	2,282	10 ± 2.43	43 ± 9.71	6.01
	mosaic	4,619	22 ± 14.89	96 ± 51.29	16.89
Mayo-Liver	collage	2,709	8 ± 2.12	34 ± 9.01	6.96
	mosaic	2,974	9 ± 2.41	48 ± 10.97	9.83
TCGA	collage	11,663	15 ± 2.78	74 ± 14.28	88.66
	mosaic	25,679	36 ± 20.39	157 ± 83.29	233.64

The primary objective of SPLICE is to select representative patches from WSIs, exhibiting low redundancy, for *collage* generation, striking a balance between a minimal number of patches and high performance. In contrast, Yottixel excels in retrieving similar WSIs through the generation of *mosaics*, which inherently involve a higher number of patches even when a small threshold (like 5%) is used. This distinction is prominently evident in Table 3.4, which provides insights into the total number of patches required to generate *collages* and *mosaics* across the three datasets considered in the experiments: Mayo-CRC, Mayo-Liver, and TCGA. It's important to note that the configurations for both Yottixel's mosaic and SPLICE to obtain these results align with those used in Table 3.1, Table 3.2, and Table 3.3.

Table 3.4 illustrates the significant storage and time factors that are contingent on the number of patches used in generating *collages* and *mosaics* within the context of image search. For Mayo-CRC and TCGA datasets, SPLICE consistently demonstrates

its efficiency by utilizing less than half of the patches needed by Yottixel's mosaic. On average, only 10±2.43 and 15±2.78 patches per WSI were required for *collage* generation in Mayo-CRC and TCGA, respectively, compared to 22±14.89 and 36±20.39 patches per WSI on average for *mosaic* generation in Mayo-CRC and TCGA, respectively. The lower standard deviation for *collage* generation highlights the reliability of SPLICE's approach by limiting the inclusion of redundant patches. In terms of storage requirements, using SPLICE for *collage*-based image search results in a significantly reduced storage utilization. On average, the storage required for *collage*-based search is 43±9.71 KB and 74±14.28 KB for Mayo-CRC and TCGA embeddings, respectively, as opposed to 96±51.29 KB and 157±83.29 KB for *mosaic*-based search in Mayo-CRC and TCGA embeddings, respectively. These storage disparities reflect the efficiency of SPLICE in minimizing the data needed for effective image retrieval. The implications of these efficiency gains extend to search times. Yottixel's mosaic, due to its reliance on a larger pool of patches, necessitates more time for image retrieval. For instance, when retrieving similar images to 209 images in Mayo-CRC, Yottixel's mosaic requires a total search time of 16.89 seconds, compared to the 6.01 seconds needed by SPLICE. In the case of the TCGA dataset, using *mosaic* for search consumes 3.9 minutes, while employing *collage* reduces the retrieval time to just 1.5 minutes. Table 3.4 also provides insights into the storage and time requirements for the Mayo-Liver dataset. Here, the difference between SPLICE and Yottixel is less pronounced, primarily because the available tissue area in the WSIs from Mayo-Liver is limited, making it not feasible to extract a large number of patches even when looser thresholds are applied. Nevertheless, SPLICE continues to exhibit its efficiency in balancing performance and resource consumption.

In summary, the experiments in this section underscore SPLICE's ability to efficiently represent histopathology images with a reduced number of patches, resulting in significant reductions in storage and search times, particularly in the context of large datasets. This efficiency positions SPLICE as a valuable tool in image search and retrieval, capable of delivering high-performance results with minimal resource utilization.

3.5 Discussion

The results of the experiments conducted to evaluate the performance of SPLICE in image search and retrieval, compared to the state-of-the-art method, namely Yottixel's mosaic [2], have provided valuable insights into the efficacy and efficiency of SPLICE. The primary objective of these experiments was to compare the performance of SPLICE and Yottixel's mosaic in their ability to retrieve similar WSIs based on a query image. The critical differences between the two methods lie in their approach to patch selection. Yottixel's mosaic focuses on well-distributed patch selection, while SPLICE prioritizes the quality and uniqueness of selected patches, differences that significantly impact the results.

SPLICE's emphasis on selecting representative and less redundant patches was a key factor in its effective performance. This was evident in the experiments, where SPLICE consistently achieved better or comparable performance as Yottixel's mosaic with significantly fewer patches. Furthermore, one of the standout advantages of SPLICE is its efficiency. It requires significantly fewer patches for *collage* generation compared to Yottixel's *mosaic* approach. This not only reduces storage requirements but also leads to faster search times. The efficiency gains are particularly pronounced in larger datasets, making SPLICE an attractive option for applications where resource efficiency is crucial.

SPLICE's utilization of percentiles as thresholds for *collage* generation offers a notable advantage in terms of flexibility when it comes to balancing performance and resource consumption. This approach requires only one parameter to be determined, the percentile, while Yottixel's mosaic, as the state-of-the-art method, necessitates adjusting two distinct parameters: the k values for both color and spatial k-means clustering. This design choice simplifies the implementation and tuning process for its ease of use and reproducibility. Moreover, the impact of the SPLICE parameter on the performance is manageable, as the experiments demonstrated that SPLICE achieves satisfactory performance even with strict thresholds, allowing for a minimal number of patches. This flexibility can be valuable in various applications.

In conclusion, SPLICE's efficient and effective approach to image representation for search and retrieval has promising implications for various applications where histopathology images need to be processed. The experiments conducted in this study have demonstrated that SPLICE is a powerful representation for image search and retrieval. Its focus on patch quality, efficiency in resource utilization, and fewer parameters to adjust makes it a valuable tool for a wide range of applications. SPLICE's consistent and robust performance across diverse datasets underscores its potential to enhance histopathology image representation with a minimal representative set of patches. Beyond Histopathology, SPLICE has the potential to be used as an efficient technique in fields that require dealing with large images such as astrophysics, satellite imaging, and climate imagery.

Chapter 4

MarbliX - A Framework for Multimodal Integration & Representation

4.1 Introduction

In recent years, the integration of artificial intelligence (AI) techniques with healthcare data has emerged as a beacon of hope for the advancement of medical diagnostics and treatment. In particular, the fusion of multimodal data, combining histopathology images with genetic sequencing data, holds tremendous promise for revolutionizing our comprehension of rare and complex diseases and, more importantly, elevating the standard of patient care. This dissertation endeavors to fill a critical void by introducing a comprehensive multimodal framework tailored to harmonize these two indispensable data sources, with a particular focus on its transformative potential in the realm of cancer diagnosis. The driving force behind this research is rooted in the profound challenges and untapped opportunities inherent in the fusion of histopathology images and genetic sequencing data.

Multimodal medical data, a diverse amalgamation of images, genomic data, clinical

records, and more, inherently embodies complexity. Meaningful integration of these diverse data sources are imperative for comprehensive patient care and in-depth disease understanding [157]. However, the convergence of histopathology images and genetic sequencing data presents distinct hurdles due to their intrinsic differences in data modalities. Histopathology images, capturing macroscopic tissue-level information, and genetic sequencing data, uncovering microscopic molecular-level insights, often exhibit substantial disparities in data structure, scale, and dimensionality [158]. In addition to this modality mismatch, multimodal medical data has other challenges like data heterogeneity, pervasive noise, and the imperative for scalable computational solutions [158]. These factors hinder the straightforward application of AI and machine learning tools in the medical field, making the integration of such data a complex undertaking.

While notable strides have been taken in the domain of multimodal data analysis, a significant proportion of existing research focuses on combining natural images with textual data, even within the medical field [159, 160]. It is often assumed that textual data integration is simpler. However, this perspective overlooks the potential for valuable insights that can be unlocked through the integration of images and genomic data. In medical diagnosis, physicians frequently rely on analyzing various data modalities before arriving at a conclusive diagnosis [157]. This is particularly true for complex diseases like cancer, where pathologists generate reports based on tissue slides, but supplementary tests such as immunohistochemistry (IHC) images and RNA-sequencing data are requested when there is a high inter-pathologist variability or at advanced tumor stages like metastasis [161]. In medical settings, these data are typically analyzed by different experts, each specializing in distinct fields. This practice can result in a fragmentation of information and miscommunication between experts, potentially impacting the quality of diagnosis and treatment [162]. Additionally, the analysis of vast datasets like genetic sequencing data often exceeds human capacity to discern the intricate patterns they may contain [136]. This further complicates the task of linking discriminating patterns in genomic data to morphological characteristics in histopathology images, preventing the full realization of the rich information stored within these data sources. Furthermore, an underutilized resource lies in the vast archives

of biomedical data, which contain valuable insights from previous cases. Neglecting this wealth of information limits progress in disease diagnosis and treatment, perpetuating the repetition of errors over time. By leveraging historical data and connecting a given case to similar cases with comparable tumor behavior patterns, we can make more informed diagnostic decisions and advance personalized medicine, reducing the suffering of future patients on their treatment journeys.

In the scope of this research, a novel multimodal framework known as Multimodal Association and Retrieval with Binary Latent Indexed matriX (MarbliX) is introduced. This innovative framework harnesses the capabilities of AI to merge histopathology images with genetic sequencing data, condensing them into a concise binary barcode referred to as "monogram." This compact binary representation encapsulates the intricate patterns found within extensive multimodal data files, encompassing both histopathology images and genomic data. Moreover, MarbliX serves as an efficient interactive tool for establishing a comprehensive "monogram"-based archive. It enables the retrieval and matching of similar cases to a given query case, thus empowering clinicians and pathologists with an in-depth comprehension of cases. This facilitates more precise and timely diagnoses, reduces variability in assessments, and augments personalized treatment options.

The motivation for this research stems from the significant challenges and unexplored opportunities that arise due to the conspicuous scarcity of multimodal solutions adept at merging histopathology images with genetic sequencing data (i.e., immune repertoire). Furthermore, as far as knowledge extends, the fusion of histopathology images with immunogenomic data remains an untapped area. By fusing these data modalities within the MarbliX framework, the objective is to effectively bridge critical research gaps and contribute to a more comprehensive understanding of complex diseases, while also unlocking new avenues for potential research in this domain.

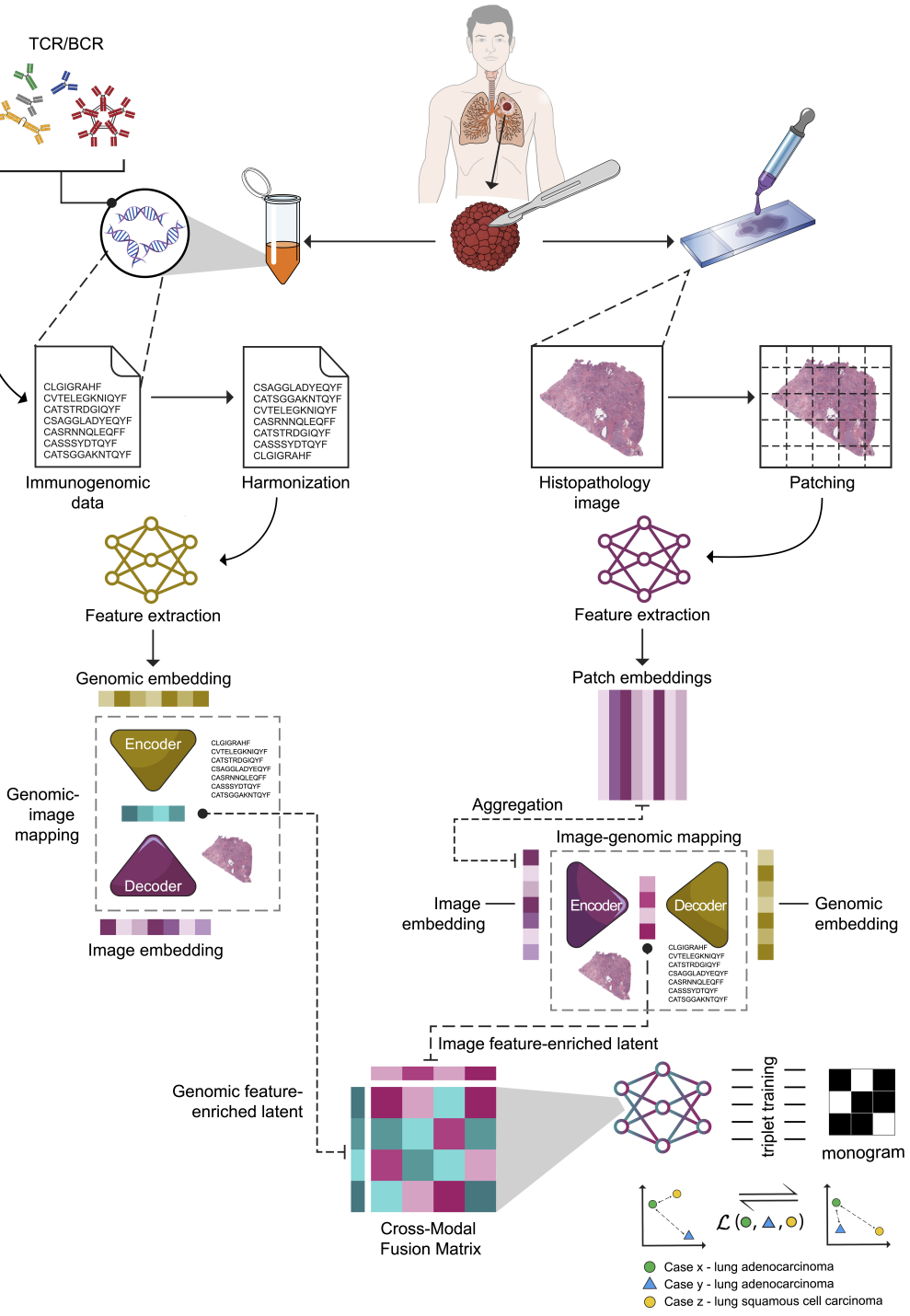


Figure 4.1: Diagram illustrating MarbliX framework for multimodal data fusion. MarbliX integrates histopathology images and immunogenomic data into a concise binary representation known as "monogram." The workflow encompasses unimodal transformation, multimodal latent association, and the generation of the monogram representation. [this figure is property of Mayo Clinic, Rochester, MN, USA]

4.2 Methodology

This section describes the details of MarbliX's design and the intricate process involved in generating a unique *monogram* representation. The work at hand brings forth four substantial contributions to the burgeoning field of multimodal research.

First and foremost, MarbliX introduces a novel approach that condenses diverse data modalities into a concise representation, referred to as *monogram*. This innovative concept enables the generation of *monogram* for each individual patient, serving as a tangible representation of their multimodal data. This representation encapsulates the intricate patterns unearthed from various data sources, thereby affording each patient a highly informative imprint or signature. This approach empowers researchers and medical practitioners to extract rich insights from a patient's multimodal clinical data, facilitating personalized and data-driven healthcare.

The second contribution of MarbliX lies in its ability to create a binary barcode of multimodal patient data, which often comprises voluminous files like histopathology whole slide images (WSIs) and RNA-seq data. This compactness holds profound implications for time efficiency, computational complexity, and storage requirements. Employing *monograms* in diverse downstream tasks such as search and retrieval not only streamlines these operations but also unlocks a powerful tool that can handle data in varying formats, sizes, complexities, and sources.

Furthermore, the third significant contribution of MarbliX is its backbone independence, a feature that distinguishes it as a versatile multimodal AI framework. The capability to map different modalities within a latent space not only contributes to the model's training efficiency but also enables the extraction of associated patterns from data modalities originating from diverse sources within the medical field. This adaptability is paramount for addressing the heterogeneous nature of medical data.

Lastly, but no less important, MarbliX serves as a valuable asset in facilitating multimodal search, a crucial task in medical research and diagnosis. With the ability to construct

a monogram-indexed archives, MarbliX empowers practitioners with an interactive tool to multimodal search for cases similar to a given query. This process involves generating a monogram for the query case and then scouring the archive to retrieve the most analogous monograms that share similar patterns with the query. This feature enhances the discoverability of relevant cases and expedites the decision-making process in healthcare.

Within this dissertation, MarbliX framework is designed and implemented with the purpose of harmonizing the subtle morphological characteristics found in histopathology images with the intricate immune response patterns within immune cell sequencing profiles. The underlying rationale for this choice is deeply rooted in the profound clinical implications of these modalities. Histopathology data, with its potentials to unveil the subtleties of tissue structures, and immune cell sequencing data, capable of unraveling the molecular nuances of immune responses. Furthermore, the absence of any pre-existing multimodal AI solutions that incorporate both histopathology and immunogenomic data serves as a compelling incentive to dedicate this research endeavor to unveiling the underlying correlations between these informative modalities.

An overview of MarbliX is illustrated in Figure 4.1. The design of MarbliX involves three main phases: unimodal transformation, multimodal latent association, and monogram representation. The details of every phase is described below.

4.2.1 Unimodal Transformation

The integration of histopathology images and immune cell sequencing data into a shared computational framework requires an alignment step to transform them into a common format. This enables joint manipulation and integration within a unified model. Hence, as a first stage, each modality, is processed and transformed into a single feature vector or *embedding*. For simplicity, in the context of the notation $(\boldsymbol{I}, S)$, $\boldsymbol{I}$ represents the histopathology image, while S represents the immunogenomic data (a set of immune cell sequences) of a given case.

Histopathology image processing: histopathology images, denoted as $\boldsymbol{I}$, are invaluable high-resolution visual snapshots of tissue samples captured through microscopy or digital scanners. These images provide insights into the cellular morphology, tissue structure, and spatial relationships within the tumor microenvironment. Extracting meaningful information from histopathology images is a complex endeavor due to their sheer size and the richness of visual data they contain. In this thesis, SPLICE (introduced in Chapter 3) is employed to tackle this challenge and generate an informative *collage* for image $\boldsymbol{I}$. This collage is a condensed representation, composed of a select set of representative patches extracted from $\boldsymbol{I}$, capturing the crucial tissue characteristics that define the image. The process of generating these collages is elaborated upon in Section 3.2.1 and Section 3.2.2 of Chapter 3, highlighting its pivotal role in simplifying and condensing the visual information embedded within $\boldsymbol{I}$.

Once the *collage* is generated for $\boldsymbol{I}$, the next step involves extracting deep features from the individual patches that compose the *collage*. This process is achieved by leveraging a pre-trained deep neural network $\mathcal{F}$, which possesses the ability to extract patterns and meaningful information within these patches. As a result of this indexing, a feature vector $\boldsymbol{f}_i$ of patch $\boldsymbol{P}_i$ within the collage is generated by applying $\boldsymbol{f}_i = \mathcal{F}(\boldsymbol{P}_i)$, where $\boldsymbol{f}_i \in \mathbb{R}^{l \times 1}$.

To craft an all-encompassing feature vector that encapsulates the entirety of image $\boldsymbol{I}$ and effectively represents its rich content, a widely adopted practice involves computing the average of the patch-level feature vectors [137], resulting in a single feature vector $\boldsymbol{f} \in \mathbb{R}^{l \times 1}$ that serves as a holistic representation of the entire image $\boldsymbol{I}$.

Sequencing data processing: immunogenomic data offers a wealth of information about the intricate diversity of immune cells present in tissue samples. This data encompasses T-cell receptor (TCR) and B-cell receptor (BCR) sequences, which play a pivotal role in elucidating immune responses in various disease and tumor contexts. However, deciphering the profound insights embedded within these sequences is a formidable task. Their varying lengths, complex and implicit patterns, and the inherent specificity of TCRs and BCRs, which naturally generate a multitude of sequences, present significant chal-

lenges in processing this data to extract meaningful information, particularly related to tumors. These complexities also hinder the effective application of machine learning and deep learning techniques. Moreover, working at the level of patient profiles, rather than individual sequences, further compounds the difficulty. Most research in the field traditionally focuses on sequence-level analysis, such as mapping specific sequences to tumor classes [16, 118, 163]. In contrast, *MarbliX operates at the patient data level*, where both histopathology images and the corresponding immune cell profiles are processed.

Algorithm 2 Harmonization Approach for Unifying Genetic Sequencing Data

Initialize λ as a dictionary with arbitrary weights for each amino acid
procedure COMPUTE HARMONIZATION WEIGHT(S, λ)
    for each sequence $s_i \in S$ **do**
        Initialize $w_{s_i}) \leftarrow 0$
        $l \leftarrow$ Length of s_i
        for each base b in s_i **do**
            $\lambda_b \leftarrow \lambda[b]$
            $x_b \leftarrow$ Position of base b in s_i
            $w_{s_i} \leftarrow w_{s_i} + \left(\lambda_b \times \frac{1}{x_b}\right)$
        $l_{\max} \leftarrow$ Maximum sequence length in S
        $w_{s_i} \leftarrow w_{s_i} + \frac{l}{l_{\max}}$
    $S_h \leftarrow \text{descending\_sort}(S, \text{key} = w_{s_i})$
    return S_h
procedure FEATUREEXTRACTION($S_h, \mathcal{G}$)
    $\boldsymbol{g} \leftarrow \mathcal{G}(S_h)$ ▷ Apply pre-trained deep learning model $\mathcal{G}$ to obtain embedding $\boldsymbol{g}$
    return $\boldsymbol{g}$

To address these challenges, a "harmonization" approach (Algorithm 2) tailored to genetic sequencing data is proposed. This method serves as a crucial preprocessing step, aimed at preparing these sequences for analysis using deep models designed for textual data. Deep models often rely on the sequential order of elements, a characteristic shared with the

sequences in the profiles. However, the sequences themselves do not inherently carry order-related meaning within each patient's profile. To address this issue, the harmonization method unifies diverse lists of sequences derived from various patient profiles, aligning them into a unified representation before proceeding with deep feature extraction. The ultimate goal is to create a single, coherent embedding that encapsulates vital information while negating the effects of varying sequence orders within each patient's profile.

This harmonization approach hinges on two pivotal factors for calculating the harmonization weight of each sequence within a given profile: **sequence length** and **base position** within the sequence. Sequence length represents the number of bases within a sequence, while base position denotes the order of a base (nucleotide in DNA/RNA and amino acid in protein) within a given sequence. The first step of harmonization is to assign a weight for every base. Since TCR and BCR sequences consist of 20 amino acids (bases), arbitrary but fixed weights are assigned to each amino acid (Algorithm 2, Lines 1). These factors are used to calculate the harmonization weight w for each sequence through

$$w = \sum_{b=1}^{l} \left(\lambda_b \times \frac{1}{x_b} \right) + \frac{l}{l_{\max}} \tag{4.1}$$

In this context, λ_b represents the arbitrary weight assigned to each amino acid base b, while x_b indicates the position of base b within the sequence, starting from 1. The variable l stands for the length of the sequence, and $l_{\max}$ signifies the maximum sequence length within the set of sequences. Equation 4.1 encapsulates the process of computing the weighted sum of the base factors for every base within a sequence, taking into account their positions. This sum, combined with the ratio of sequence length to the maximum sequence length, ultimately determines the harmonization weight.

As illustrated in Algorithm 2, Lines 2-11, the harmonization weight is computed for sequence s_i, where $s_i \in S$. Once these harmonization weights are calculated for the entire set, the sequences are sorted in descending order based on these weight values to generate the harmonized set S_h (Algorithm 2, Lines 12-13). This sorting process is critical to

harmonization, effectively arranging the sequences so that those with higher harmonization weights occupy prominent positions in the resulting list. With the sequences harmonized, they are now ready for consistent and generalizable feature extraction. A pre-trained deep learning model $\mathcal{G}$ can then be employed to distill features from the sequences by applying $\boldsymbol{g} = \mathcal{G}(S_h)$, where $\boldsymbol{g} \in \mathbb{R}^{l \times 1}$ represents a feature vector extracted from the harmonized sequences set S_h (Algorithm 2, Lines 15-16).

4.2.2 Multimodal Latent Association

Following the transformation of each modality, histopathology image $\boldsymbol{I}$ into embedding $\boldsymbol{f}$ and immune cell sequence profile S into embedding $\boldsymbol{g}$, the association between these embeddings is learned by projecting them into a shared latent space. In this shared space, the embeddings from both modalities coalesce to form a concise patient representation. This step addresses the issue of non-relevant features that may exist in the uni-modal data obtained from a pretrained network. By merging these embeddings into an encoded representation, we want to extract and consolidate the pertinent features from each modality, enhancing their combined synergy and informative value in subsequent analyses.

To accomplish this, as shown in Algorithm 3, two deep neural networks with an encoder-decoder (autoencoder) architecture are employed, each tailored to emphasize the salient features from its corresponding modality while suppressing extraneous information. This is achieved by training each of these hybrid models to generate an encoded latent representation, highlighting the relevant features. Specifically, each autoencoder model is designed to take one modality and reconstruct the other, resulting in a latent representation that embodies the dominant features of its respective modality.

In the first stage (illustrated in Algorithm 3, Lines 4-9), an autoencoder denoted as $\mathcal{A}_I$ is designed, where it takes the histopathology image embedding $\boldsymbol{f}$ as input, and reconstructs the immune cell sequence profile embedding $\boldsymbol{g}$. During training, $\mathcal{A}_I$ learns to focus on critical features present in histopathology images that offer insights into immune cell

Algorithm 3 Multimodal Latent Association

1: **Input:**
2: Histopathology image embedding $\boldsymbol{f}$
3: Immune cell sequence profile embedding $\boldsymbol{g}$
4: **Training Stage:**
5: Initialize $\mathcal{A}_I$ (Autoencoder for Histopathology):
6: **while** not converged **do**:
7: Forward pass: $\boldsymbol{f} \rightarrow \mathcal{E}_I(\boldsymbol{f}) \rightarrow \mathcal{D}_I(\mathcal{E}_I(\boldsymbol{f}))$
8: Compute loss: $l_I = \text{MSE}(\boldsymbol{g}, \mathcal{D}_I(\mathcal{E}_I(\boldsymbol{f})))$
9: Backpropagate and update weights
10: Initialize $\mathcal{A}_S$ (Autoencoder for Immune Cell Sequences):
11: **while** not converged **do**:
12: Forward pass: $\boldsymbol{g} \rightarrow \mathcal{E}_S(\boldsymbol{g}) \rightarrow \mathcal{D}_S(\mathcal{E}_S(\boldsymbol{g}))$
13: Compute loss: $l_S = \text{MSE}(\boldsymbol{f}, \mathcal{D}_S(\mathcal{E}_S(\boldsymbol{g})))$
14: Backpropagate and update weights
15: **Latent Representation:**
16: Encode using the encoder of $\mathcal{A}_I$:
17: $\boldsymbol{u} \leftarrow \mathcal{E}_I(\boldsymbol{f})$
18: Encode using the encoder of $\mathcal{A}_S$:
19: $\boldsymbol{v} \leftarrow \mathcal{E}_S(\boldsymbol{g})$
20: **return** $\boldsymbol{u}$, $\boldsymbol{v}$

sequence patterns. These features are encapsulated within the bottleneck layer, situated just before the first decoder layer of model $\mathcal{A}_I$.

Conversely, in the second stage (described in Algorithm 3, Lines 10-14), another autoencoder, denoted as $\mathcal{A}_S$, is employed, which takes the immune cell sequence profile embedding $\boldsymbol{g}$ as input and reconstructs the histopathology image embedding $\boldsymbol{f}$. This design empowers the model to emphasize essential characteristics inherent to immune cell sequence data that are relevant to the histopathological context, also embedded within the bottleneck layer of model $\mathcal{A}_S$.

Following the training of $\mathcal{A}_I$ and $\mathcal{A}_S$, the encoder from each hybrid model is employed to generate blended latent representations (Algorithm 3, Lines 15-20). For instance, $\boldsymbol{u}$, characterized by a strong emphasis on histopathological features, is derived through the application of $\boldsymbol{u} = \mathcal{E}_I(\boldsymbol{f})$, where $\boldsymbol{u} \in \mathbb{R}^{l \times 1}$. Likewise, $\boldsymbol{v}$, accentuating immune cell characteristics, is derived as $\boldsymbol{v} = \mathcal{E}_S(\boldsymbol{g})$, where $\boldsymbol{v} \in \mathbb{R}^{l \times 1}$. These latent representations effectively capture the essential information from each modality while eliminating noise and irrelevant features. Consequently, they facilitate the integration of both modalities into a more informative and harmonious representation within the shared latent space.

Upon completion of the fusion process, the resulting compact representation not only reduces dimensionality but also encapsulates critical aspects of both modalities. This representation serves as a powerful encoding of joint information extracted from histopathology images and immune cell sequences, facilitating profound integration in the subsequent phase.

4.2.3 MarbliX Monogram

After generating the two latent representations $\boldsymbol{u}$ and $\boldsymbol{v}$, the next crucial step within the MarbliX framework is the projection and indexing of these representations into a 2D binary matrix, referred to as "monogram." This matrix serves as an effective representation for capturing the intricate relationships and correlations that exist between the two modalities. The process of learning the *monogram* representation is the core of the MarbliX framework, enabling a comprehensive exploration of the joint information encoded within $\boldsymbol{u}$ and $\boldsymbol{v}$.

To accomplish this task, a deep neural network, denoted as $\mathcal{Q}$, is designed to uncover the correlations between histopathology and immunogenomic features based on diagnosis. This process aims to unveil the underlying structure that interlinks histopathological characteristics with immune cell behavior among cases within the same diagnostic class. In other words, the model is engineered to capture the commonality in multimodal relationships among cases that share similar diagnoses, embedding these features within their respective *monograms*. Simultaneously, it strives to discern the distinctions between cases

of different diagnostic classes and accentuate these disparities within their corresponding monograms.

Algorithm 4 Learning Multimodal Monogram Representation

1: **Input:**
2: $\boldsymbol{u}$: image latent representation
3: $\boldsymbol{v}$: sequencing latent representation
4: **Training Stage:**
5: Initialize $\mathcal{Q}$ model with three branches $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$
6: $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3 \leftarrow$ shared weights
7: **for each triplet** $\mathcal{T}_1(\boldsymbol{u}, \boldsymbol{v})$, $\mathcal{T}_2(\boldsymbol{u}^+, \boldsymbol{v}^+)$, $\mathcal{T}_3(\boldsymbol{u}^-, \boldsymbol{v}^-)$ **do**:
8: $M \leftarrow \boldsymbol{u} \otimes \boldsymbol{v}$
9: $\bar{M} \leftarrow \mathcal{T}(M) \leftarrow M$
10: $\bar{M}_{\text{binary}} \leftarrow \{$**if** $w > 0.5$ **then** 1 **else** 0 **for each** w **in** $\bar{M}\}$
11: Calculate triplet loss:
12: $d(\boldsymbol{a}, \boldsymbol{p}) \leftarrow \text{distance}(\bar{M}, \bar{M}^+)$
13: $d(\boldsymbol{a}, \boldsymbol{n}) \leftarrow \text{distance}(\bar{M}, \bar{M}^-)$
14: $\mathcal{L}_{\text{triplet}} \leftarrow \max\{(d(\boldsymbol{a}, \boldsymbol{p}) - d(\boldsymbol{a}, \boldsymbol{n})) + \alpha, 0\}$
15: $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3 \leftarrow$ update weights using gradient descent

As described in Algorithm 4, The $\mathcal{Q}$ model comprises three branches with shared weights, each taking a pair of latent representations, $\boldsymbol{u}$ and $\boldsymbol{v}$. Thus, the model is provided with triplet pairs as input $\mathcal{Q}(\{\boldsymbol{u}, \boldsymbol{v}\}, \{\boldsymbol{u}^+, \boldsymbol{v}^+\}, \{\boldsymbol{u}^-, \boldsymbol{v}^-\})$, consisting of an anchor pair $(\boldsymbol{u}, \boldsymbol{v})$, a positive pair $(\boldsymbol{u}^+, \boldsymbol{v}^+)$, and a negative pair $(\boldsymbol{u}^-, \boldsymbol{v}^-)$. The anchor case serves as the reference for which the model endeavors to generate a representative monogram. The positive case shares the same diagnosis as the anchor, reinforcing common features. In contrast, the negative case differs in diagnosis from the anchor, shedding light on the discrepancies between diagnostic classes.

Within each branch (Algorithm 4, Lines 11-15), the pair of $\boldsymbol{u}$ and $\boldsymbol{v}$ is projected into a matrix through the computation of the outer product between their respective layers.

Subsequently, this matrix undergoes several dense layers to learn the intricate patterns it encodes. The triplet loss function (Equation 4.2) calculates the distances between the anchor's predicted matrix and both the positive and negative matrices. This loss function is designed to minimize the distance between matrices of cases within the same class (anchor and positive), while simultaneously maximizing the distance between cases of differing classes (anchor and negative).

The triplet loss function can be given as

$$\mathcal{L}_{\text{triplet}}(\boldsymbol{a}, \boldsymbol{p}, \boldsymbol{n}) = \max\{d(\boldsymbol{a}, \boldsymbol{p}) - d(\boldsymbol{a}, \boldsymbol{n}) + \alpha, 0\} \tag{4.2}$$

In Equation 4.2, $\boldsymbol{a}$ represents the anchor case, $\boldsymbol{p}$ signifies a positive case (sharing the same diagnosis as the anchor), and $\boldsymbol{n}$ denotes a negative case (with a different diagnosis from the anchor). $d(\boldsymbol{a}, \boldsymbol{p})$ calculates the distance between the anchor and positive case matrices, while $d(\boldsymbol{a}, \boldsymbol{n})$ computes the distance between the anchor and negative case matrices. The margin parameter α ensures a minimum separation between the positive and negative cases.

Notably, the $\mathcal{Q}$ model incorporates a binary representation branch in the final dense layer. This branch plays a pivotal role by transforming the weights in the last layer into binary values. This process is crucial as it enables the generation of compact binary representations, highly efficient for subsequent indexing and storage.

After training the $\mathcal{Q}$ model, it can be utilized to generate binary monogram representations for new cases. This can be accomplished by applying monogram $= \mathcal{Q}(\{\boldsymbol{u}, \boldsymbol{v}\})$, where monogram represents an 8×8 binary matrix (with encoding capability to cover $2^{64} = 1.8 \times 10^{19}$ combinations) derived from the latent representations $\boldsymbol{u}$ and $\boldsymbol{v}$ of the histopathology image and immune cell sequences, respectively.

4.3 Experiments

MarbliX was implemented and evaluated using datasets obtained from The Cancer Genome Atlas (TCGA), which included histopathology images and genomic data. Specifically, cases from two primary sites, **lung** and **kidney**, were sourced from TCGA. Only cases featuring both Whole Slide Images (WSI) and genomic profiles were considered. This resulted in a dataset comprising 535 lung adenocarcinoma (LUAD) cases and 510 lung squamous cell carcinoma (LUSC) cases for the lung primary site. For the kidney primary site, the dataset consisted of three classes: 508 kidney renal clear cell carcinoma (KIRC) cases, 248 kidney renal papillary cell carcinoma (KIRP) cases, and 38 kidney chromophobe (KICH) cases.

To represent the WSIs, SPLICE (as detailed in Chapter 3) was employed to select representative patches, forming a *collage* for each WSI after segmenting the tissue region from the background using Otsu thresholding. The *collage* of every WSI in SPLICE was generated by setting the similarity threshold to the 30th percentile, striking a balance between performance and computational/storage requirements. Subsequently, deep features were extracted from the patches within the *collage* of every WSI using DINO ViT [164] as a pre-trained model, resulting in a 768-dimensional embedding for every patch within the WSI collages. To represent each WSI as a single vector, the patch embeddings were averaged, resulting in a 768-dimensional embedding for the entire WSI.

For the immunogenomic data, raw RNA-seq files from TCGA were utilized to reconstruct the immune repertoire of every patient. RNA-seq provides information about the transcriptome, including the expression levels of genes, and can capture the variable regions of TCR and BCR genes [165]. In RNA-seq, the expressed TCR and BCR sequences can be identified by analyzing the reads that align to these specific genomic regions [165]. Various computational methods have been proposed in the literature for extracting immune cell sequences from RNA-seq data. In this work, TRUST4 [165] was employed to obtain the TCR and BCR sequences of each patient from their RNA-seq profiles. Rare sequences and artifacts were filtered out by excluding those that were not common to a sufficient proportion of patients within each subtype class. Through experimentation, for

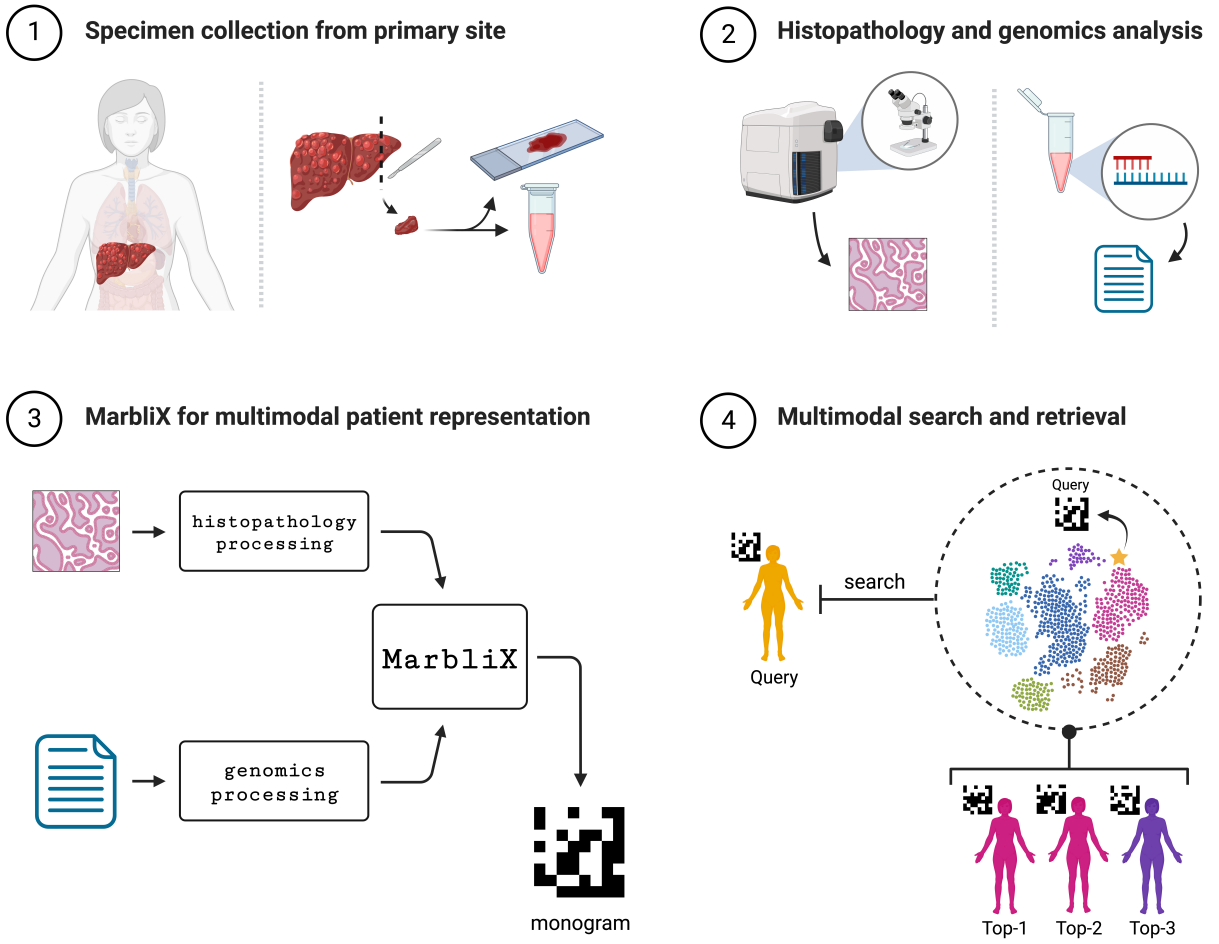


Figure 4.2: The pipeline for utilizing MarbliX in a multimodal pathology search, aiming to identify similar cases to a given query case—exemplified here with a liver cancer patient. The process begins with specimen collection for histopathology staining and genomics analysis. In the second stage, the tissue slide is digitized, and RNA sequencing is conducted to obtain the patient's whole slide image and genomics profile, respectively. In the third stage, following unimodal transformations of the histopathology image and genomics profile, MarbliX integrates the two modalities into a binary multimodal representation, referred to as "monogram". The final stage involves a multimodal search using the generated monogram as a query to identify the top n (in this case, n=3) similar matrices in the archive. The search is performed through Hamming distance comparison. The retrieved matrices can be utilized to identify the similar patients' medical information, including histopathology images, genomics data, treatment history, and risk stratification.

the lung dataset, sequences that were not common to at least 30% of the patients within the subtype class were excluded, while a lower threshold of 15% was applied to kidney cases due to the limited number of samples. Subsequently, the sequences of each patient were harmonized using the method described in Section 4.2.1 and Algorithm 2 to unify the order of the sequences across patients before conducting feature extraction. A pretrained model, BERT [166], was employed to extract features from the harmonized immune cell profiles, resulting in a 768-dimensional embedding for each patient.

The sequencing and image embeddings were then utilized to perform multimodal latent association. However, as these embeddings come from different models, they have different ranges. Therefore, min-max rescaling was performed to bring the embeddings to a common scale before learning the association between them. After normalization, hybrid autoencoders were designed to take one modality as input and reconstruct the other. For every dataset, two hybrid autoencoders were trained: one for image embeddings as input and one for genomic embeddings as input. Each autoencoder comprised an encoder with two dense layers of size 512 and 256, followed by a bottleneck layer of size 128, and finally, a decoder with two dense layers of size 256 and 512. All autoencoders were trained using the Adam optimizer and mean square error (MSE) as the loss function. The image-genomics autoencoder was trained for 150 epochs with a learning rate of 1×10^{-5}, while the genomics-image autoencoder was trained for 50 epochs with a learning rate of 1×10^{-4}.

The trained hybrid autoencoders were then utilized to extract the bottleneck layer for each sample, resulting in two encoded vectors: image features-enriched latent and genomic features-enriched latent. These vectors are of size 128 and highlight relevant features within their respective modalities in the latent space. The extracted latents were then used to train the final model in the MarbliX approach to generate *monogram* representation. Self-supervised training was conducted using triplet loss to minimize the distance between patients with the same primary diagnosis and maximize the distance between patients with different primary diagnoses. This was achieved by ensuring that the model generates similar embeddings for patients within the same class (anchor and positive samples) and different embeddings for patients from different classes (anchor and negative samples).

Anchor and positive samples represent pairs of image and genomic data with the same diagnosis, while the negative sample features a pair of image and genomic embeddings with a different diagnosis. The positive and negative samples were selected for each anchor sample by calculating pairwise Euclidean distances and identifying the farthest positive sample and the closest negative sample for each anchor. This approach was implemented to ensure robust training by introducing hard triplets to the model, guiding toward learning the similarities between samples belonging to the same class, despite eventual dissimilarity between them. Analogously, this approach guides the model to generate different embeddings for cases that share common features in their data but belong to different classes. The triplet model was trained for 150 epochs using the *tanh* activation function and Adam optimizer with a learning rate of 1×10^{-5}. The model first combines the image and genomic embeddings into a single tensor by finding the outer product between them. This tensor is then flattened and passed through three consecutive dense layers of size 1024, 256, and 64 to learn the deep multimodal relationship. The last layer of the model has a binary branch that generates a binary representation of the last layer. As the tanh function results in values within the range of $[-1, 1]$, the binary dense layer sets positive values to 1 and negative values to 0. Figure 4.4 (a) shows the triplet loss during the training of the models for the lung and kidney datasets. MarbliX was evaluated for lung dataset using 5-fold cross validation while 2-fold cross validation was used to evaluate kidney dataset due to the limited number of cases in the KICH subtype where enough cases from every subtype is needed for meaningful training.

4.3.1 Patient Representation Evaluation

Different experiments were conducted to evaluate the quality of the patient representation. Before evaluating the monogram generated by the final triplet model, an experiment was performed to analyze the quality of the multimodal latent association applied through the two hybrid autoencoders to map histopathology features to immunogenomic features, and vice versa. As shown in Figure 4.3, the cosine similarity was calculated between

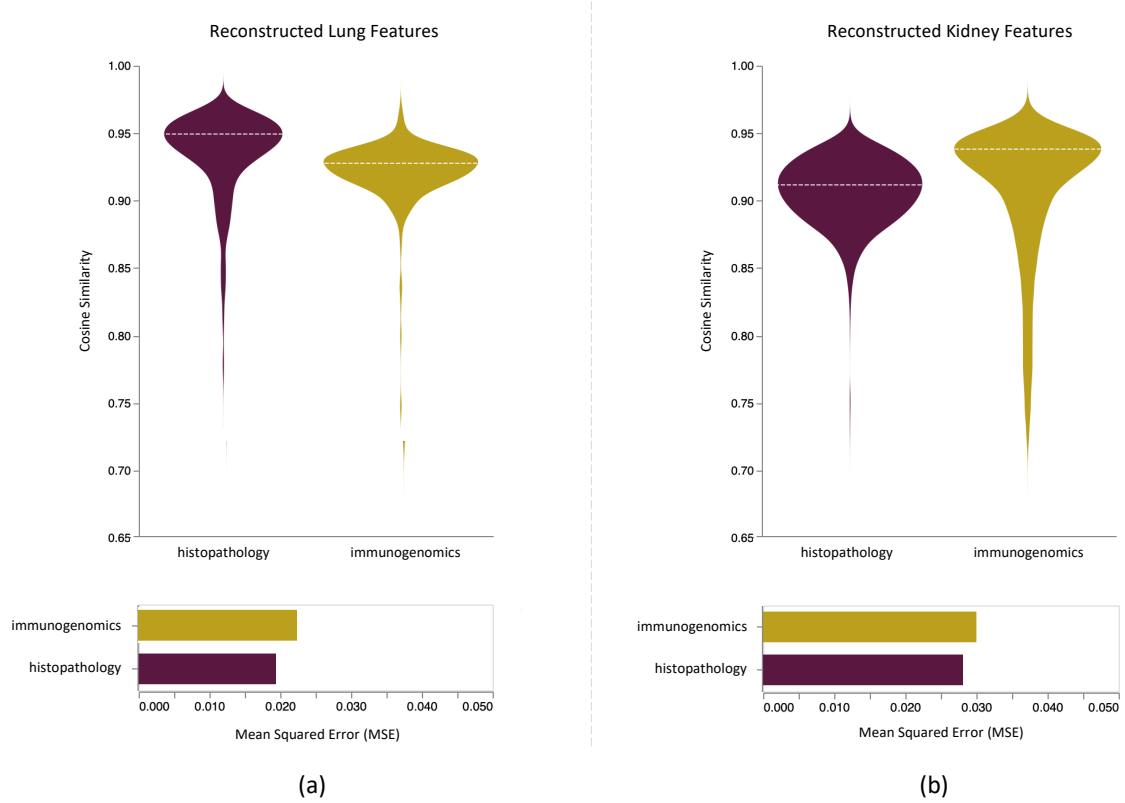


Figure 4.3: Analysis of reconstructed features from multimodal association learning through hybrid autoencoders trained using histopathology images and immunogenomic data of (a) lung dataset and (b) kidney dataset. The violin plot illustrates the distribution of cosine similarity values between the original and reconstructed embeddings by the hybrid autoencoders, with the dashed line indicating the median. The accompanying bar plot provide a comprehensive overview of the mean squared error (MSE) across all original embeddings and their corresponding reconstructed counterparts.

every histopathology and immunogenomic embedding extracted by the pretrained model with the reconstructed histopathology and immunogenomic, respectively, embedding by the trained hybrid autoencoders. The violin plots in Figure 4.3 show the distribution of the cosine similarities between the original embedding and the reconstructed embedding of every case in the test sets of (a) lung dataset and (b) kidney dataset. The median cosine similarities of 0.95 for lung histopathology and 0.91 for kidney histopathology indicate a strong overall resemblance between the original and reconstructed embeddings, suggesting effective feature retention. Similarly, in the immunogenomic distribution, with median cosine similarities of 0.93 for lung and 0.94 for kidney, the model maintains a high degree of similarity on average during the reconstruction process. The tightly packed quartiles in both lung and kidney histopathology plots imply consistent performance across the majority of cases. However, the wider spread in the upper quartile of the immunogenomic kidney plot suggests challenges in accurately reconstructing complex patterns, potentially attributed to the limited representation of the KICH subtype in the training set. This limitation may lead to deviations in reconstructions when confronted with new, less familiar patterns in the test set.

In Figure 4.3, the accompanying bar plot provides a quantitative assessment of the reconstruction quality by illustrating the average mean squared error (MSE) between the original embeddings and their reconstructed counterparts for both lung and kidney datasets. Remarkably, the MSE values exhibit consistency across both histopathology and immunogenomics reconstructions, ranging between 0.020 and 0.030 for both datasets. The MSE values indicate that the model performs well in minimizing the squared differences between the original and reconstructed embeddings, reaffirming the reliability of the reconstruction process for both histopathological and immunogenomic features in the lung and kidney datasets.

The training curves shown in Figure 4.4 (a) provide an overview of the learning curves of the two MarbliX models trained on triplet loss, aiming to learn monogram representations for lung and kidney cases. The curves represent the evolution of the loss function over training epochs for each model. Initially, both models exhibit a sharp decrease in the loss,

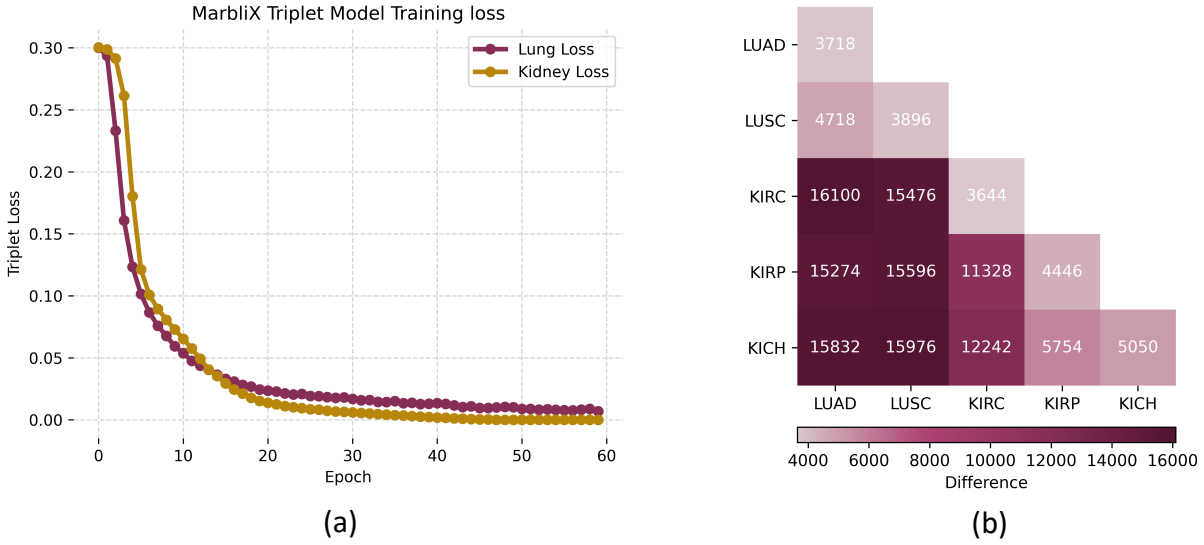


Figure 4.4: (a) The triplet loss during the training of MarbliX to learn a multimodal representation, referred to as "monogram," of cases within each of lung and kidney datasets. (b) A heatmap illustrating the distinctions in the generated monogram representations among cases with various subtypes of lung and kidney cancers. The dissimilarity is expressed as the XOR of each monogram with every other monogram in the datasets.

indicative of effective learning and convergence tendency. However, as the epochs progress, subtle differences emerge. The model trained on the kidney dataset demonstrates a slightly more gradual decline in the loss compared to its lung counterpart. This might suggest that the kidney dataset poses greater complexity or variability, requiring the model to adapt more cautiously.

The binary monograms generated by MarbliX were analyzed to assess both inter and intra-dissimilarity between the representations. To address the limited number of KICH cases, a set of 19 random cases was selected from the test set of each subtype. Quantifying the dissimilarity between matrices involved performing a bitwise XOR operation to identify differing bits between any two matrices. The resulting heatmap, displaying calculated dissimilarities, is presented in 4.4 (b). Notably, the heatmap's diagonal reflects the small-

est dissimilarity, indicating that MarbliX generates similar representations for patients with shared diagnoses. However, intra-dissimilarity among patient representations within LUAD, LUSC, and KIRC was observed to be lower compared to the intra-dissimilarity within KIRP and KICH. This can be attributed to the limited number of cases in the latter two subtypes, leading to constrained coverage of subtype features during the training process. Furthermore, the heatmap illustrates that KIRP and KICH shows a lower degree of dissimilarity in their representations than the dissimilarity observed between either of them and KIRC. This could stem from more common patterns between KIRP and KICH or, alternatively, the limited available cases belonging to KIRP and KICH during training. The dissimilarity between lung monograms (LUAD and LUSC) was smaller than the dissimilarity between lung monograms and kidney monograms (KIRC, KIRP, and KICH). Similarly, the dissimilarity between representations of kidney cases was smaller than the dissimilarity between kidney and lung cases. Additionally, the dissimilarity between representations of kidney subtypes was higher than the dissimilarity between representations of lung subtypes.

In assessing the efficacy of MarbliX in generating high quality multimodal patient representations, an experiment was conducted, comparing them to unimodal representations derived from histopathology and immunogenomics data. PCA transformation was applied to extract the top 64 components from each embedding in the lung and kidney test sets, using folds not employed during training. These components were then projected into a high-dimensional space, forming the t-SNE maps presented in Figure 4.5. The maps provide comprehensive insights into feature properties, highlighting distinctions among various subtype classes. In Figure 4.5 (a), lung histopathology image embeddings of LUAD and LUSC are intermixed, lacking a clear discriminative pattern. Figure 4.5 (b) demonstrates a more discriminative pattern in immunogenomic embeddings, although some LUSC features intertwine with LUAD features. Merging histopathology and immunogenomics features into a monogram using MarbliX enhances distribution in Figure 4.5 (c) and (d), showcasing real and binary features. Notably, the t-SNE maps reveal improved separation of LUAD and LUSC, particularly using binary monogram representations. Transitioning to kidney

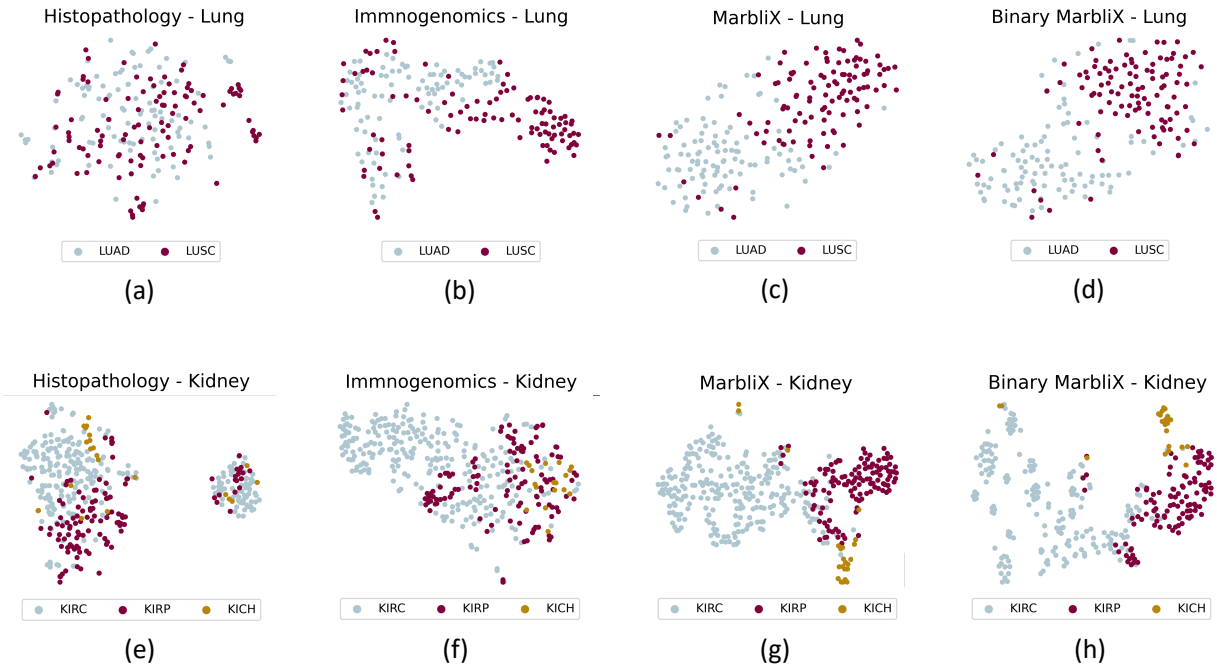


Figure 4.5: The t-SNE map depicting the distribution of modalities within a high-dimensional space. The visual representation showcases embeddings post-PCA transformation, specifically retaining the initial 64 components. The maps shows the relationship between cancer subtypes within lung (first row) and kidney (second row) datasets. The distributions of uni-modal embeddings are presented in (a) and (e) for the image modality, while (b) and (f) illustrate the immunogenomics modality. Maps (c) and (g) show the distribution of real-value monogram representations, while (d) and (h) display binary monograms, both generated by MarbliX.

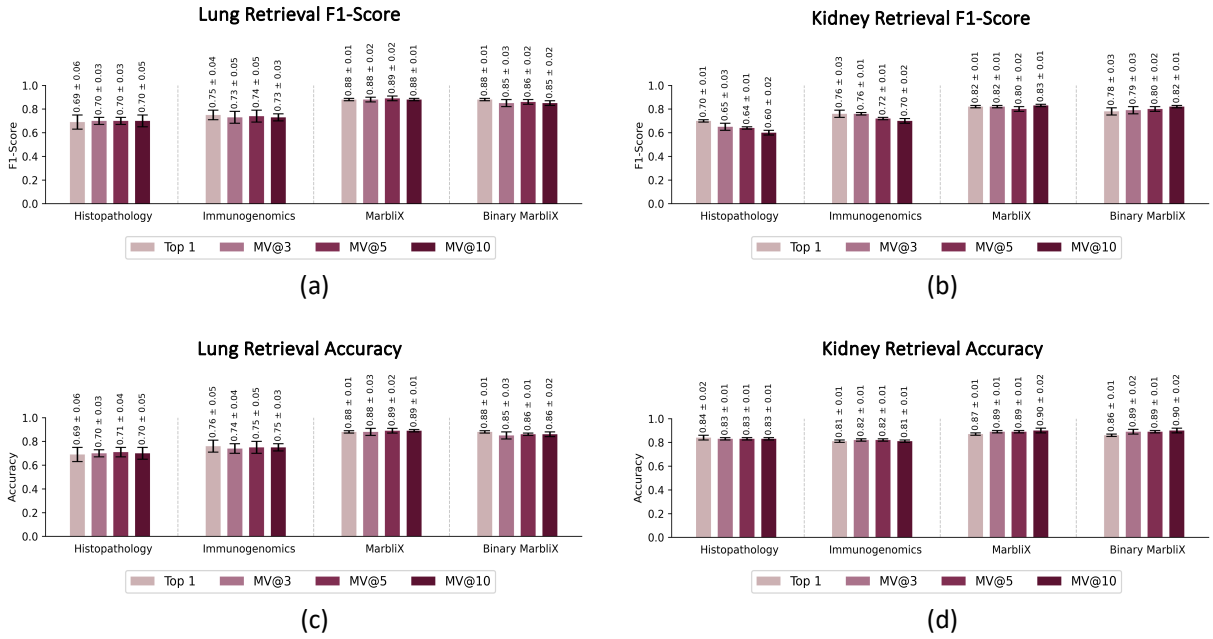


Figure 4.6: Comparative analysis of multimodal search performance - MarbliX versus unimodal representations of histopathology images and immunogenomic data. Mean and standard deviation of macro average F1-score and accuracy for lung (a, c) and kidney (b, d) datasets. Leave-one-out and majority vote criteria are applied to retrieve top-1, MV@3, MV@5, and MV@10 for cases in the test folds.

maps, Figure 4.5 (e) shows two well-separated clusters of mixed cased from the three kidney subtypes in the histopathology t-SNE map. Analysis reveals that this separation is influenced by different hospitals, indicating the dominance of WSI technical properties in the embeddings. Immunogenomic t-SNE map (Figure 4.5 (f)) exhibits a more meaningful distribution than histopathology, with KIRC being the most discriminated class due to the available number of cases during training compared to KIRP and KICH. Despite this, MarbliX application refines the distribution of patient representations (Figure 4.5 (g) and (h)), creating distinct groups for the three kidney cancer subtypes, highlighting the ability of MarbliX to derive compact yet discriminative representations from diverse modalities.

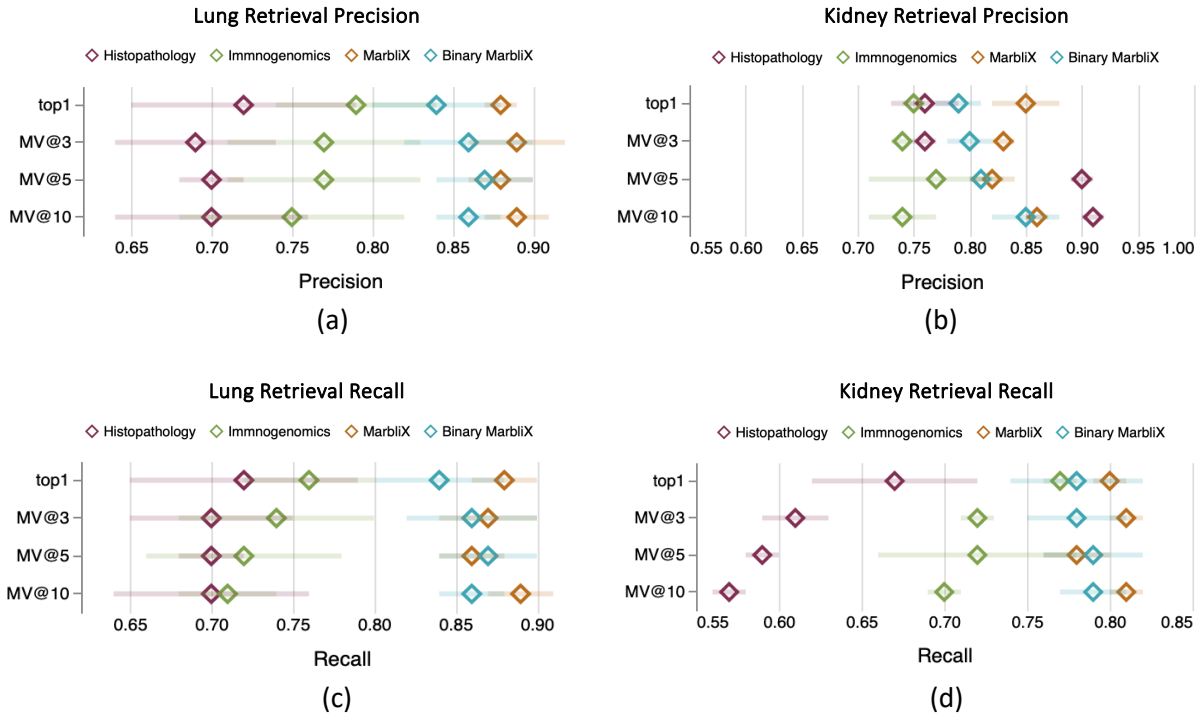


Figure 4.7: Multimodal search performance of real and binary MarbliX representations compared to unimodal representations of histopathology images and immunogenomic data. Diamond-shaped points represent the mean of macro-average precision and recall for lung (a, c) and kidney (b, d) test folds. Horizontal bars denote the standard deviation range. Leave-one-out and majority vote (MV) criteria are applied to retrieve top-1, MV@3, MV@5, and MV@10 for cases in the test folds.

4.3.2 Multimodal Search Evaluation

MarbliX demonstrated its primary utility through the application of the *monogram* representation for conducting efficient multimodal search and retrieval (Figure 4.2. The *monogram*, a condensed representation, integrates diverse features extracted from multimodal data, thereby streamlining the process of multimodal search and retrieval. MarbliX was employed to fuse each patient's histopathology image and immunogenomic data into a monogram. To assess MarbliX's performance, "leave-one-out" validation was performed, utilizing monograms generated by MarbliX for cases in the test folds of lung and kidney datasets. Cases in the test folds were not part of the training set. The evaluation involved searching for similar cases within the monograms archive, treating each case as a query case by excluding it from the search process. The *majority vote* criterion was applied for top-3 (MV@3), top-5 (MV@5), and top-10 (MV@10) retrievals, requiring at least $n/2 + 1$ of the top-n retrieved cases to belong to the same diagnosis class as the query case.

Figure 4.6 presents the mean and standard deviation of macro average F1-score and accuracy for the test folds. The results of MarbliX representations, real and binary monograms are compared against unimodal representations of histopathology and immunogenomics. Analyzing the lung dataset results in Figure 4.6 (a) and (c), multimodal search using monogram yields more accurate and similar retrieved cases compared to unimodal searches. MarbliX consistently achieves performance between 85% and 89% across all criteria from top-1 to MV@10, surpassing the 69%-71% and 73%-76% ranges of histopathology and immunogenomics, respectively. The integration of distinctive patterns from both modalities into a compact multimodal representation notably enhances the quality of retrieved cases. Analyzing the performance on the kidney dataset shown in Figure 4.6 (b) and (d), MarbliX representation also outperforms unimodal representations. The best performance was achieved using real monograms, with F1-score ranging between 80% and 83%, and accuracy between 87% and 90%. However, employing binary monograms results in a slight decrease in the F1-score range (78%-82%). Immunogenomic representations outperform histopathology representations in terms of F-score (70%-76%

vs. 60%-70%), while accuracy remains comparable. Notably, all kidney representations exhibit lower macro average F1-score values compared to accuracy values, reflecting dataset imbalance.

In Figure 4.7, the precision and recall results are illustrated for the same search evaluation on the lung and kidney datasets. The mean precision and mean recall of the folds are represented with the diamond shape, while the standard deviation of each mean value is shown as a horizontal bar. As shown in the figure, MarbliX representation maintains more stable results across the folds (different test sets) compared to the unimodal representations. This is evident from the generally shorter standard deviation bars for both lung and kidney results, indicating a more consistent performance. Furthermore, MarbliX maintains higher precision and recall compared to both lung histopathology and immunogenomic representations. For the kidney dataset, although the histopathology representation achieved the highest precision when a top-5 and top-10 majority vote was applied (90%-91%), it also achieved the lowest recall for the same measures (56%-59%). Binary MarbliX representation resulted in approximately the same recall as immunogenomics over the top-1 measure but demonstrated higher precision. Overall, MarbliX consistently maintained precision and recall above 78% for top-1, MV@3, MV@5, and MV@10 measures.

In summary, MarbliX demonstrates robust performance in multimodal search and matching, leveraging monograms to fuse histopathology and immunogenomic data effectively. Overall, MarbliX maintained more accurate and stable performance compared to unimodal representations.

4.3.3 Monogram Analysis

The primary goal of MarbliX is to generate multimodal representations that are similar for patients with similar characteristics and distinct for those with different attributes. In this study, similarity is defined based on a shared diagnosis, specifically the commonality of cancer subtypes. Figure 4.8 illustrates MarbliX's efficacy in achieving this objective. The figure presents four LUAD monograms and four LUSC monograms generated for

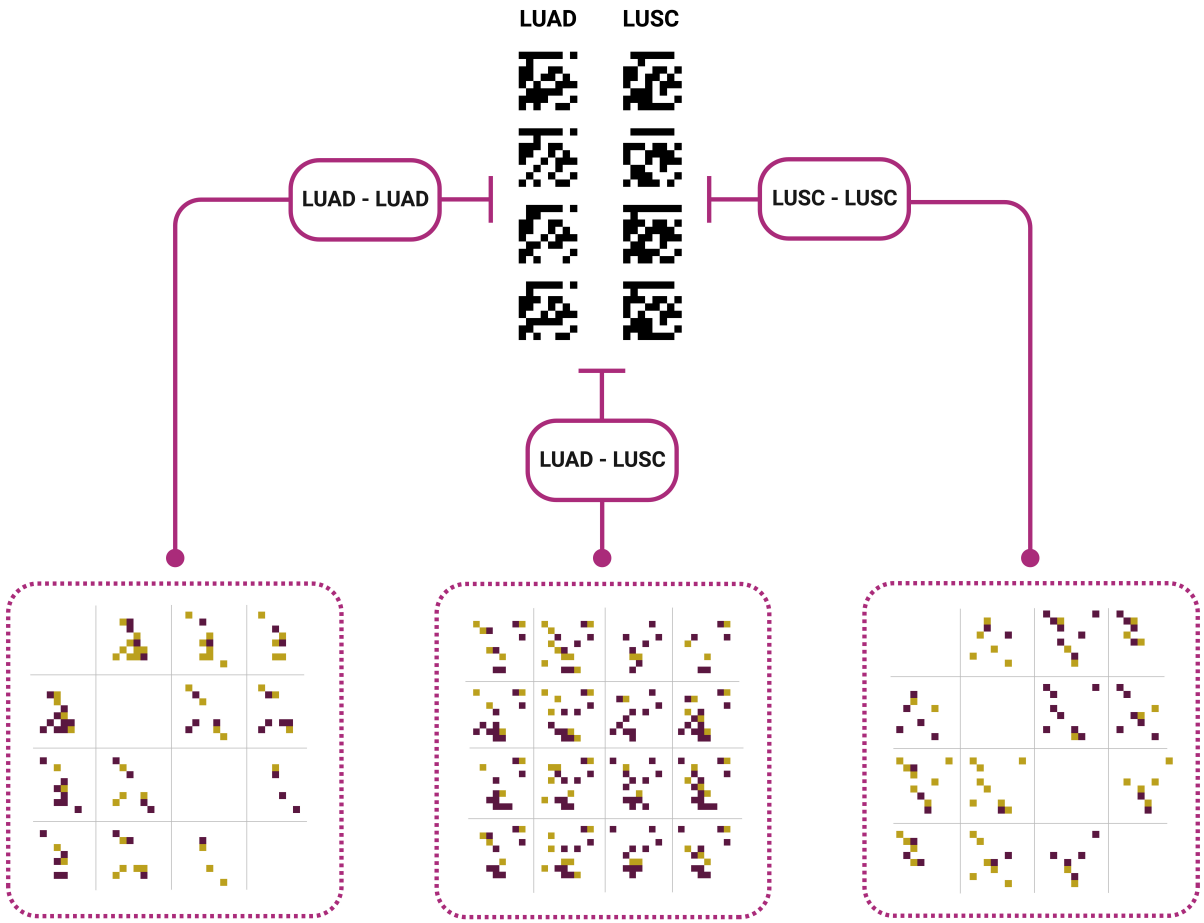


Figure 4.8: Monogram representations generated using MarbliX, for randomly selected patients with Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC). The figure presents matrix bitwise XOR results: the left side illustrates $\{\text{LUAD}_{set} \oplus \text{LUAD}_{set}\}$, the right side $\{\text{LUSC}_{set} \oplus \text{LUSC}_{set}\}$, and the middle $\{\text{LUAD}_{set} \oplus \text{LUSC}_{set}\}$. Yellow pixels highlight features transitioning from 0 to 1, while purple pixels indicate the shift from 1 to 0. This color scheme serves to distinguish the presence or absence of features. In the context of $\{\text{LUAD}_{set} \oplus \text{LUSC}_{set}\}$, yellow pixels signify features unique to LUAD matrices, absent in LUSC, and purple pixels denote features specific to LUSC matrices, absent in LUAD.

randomly selected samples from each class (refer to Figure4.4 (b) for a quantified monogram difference). It is evident that matrices from patients with the same subtype exhibit a consistent pattern (intra-similarity).

To highlight this intra-similarity and inter-dissimilarity, the figure displays the result of applying bitwise XOR on each LUAD matrix with every other matrix within the LUAD set. The same process was applied to the LUSC set, resulting in $\{\text{LUSC}_{set} - \text{LUSC}_{set}\}$. Additionally, the set of LUAD was XORed with the set of LUSC to identify the differences between each LUAD matrix and every LUSC matrix in the sets. Yellow pixels represent values turning from 0 to 1, while purple indicates the shift from 1 to 0. This color scheme is particularly informative when analyzing the results of $\{\text{LUAD}_{set} - \text{LUSC}_{set}\}$, where yellow pixels signify features present in LUAD matrices but absent in LUSC matrices, and purple denotes features present in LUSC matrices but absent in LUAD matrices.

As shown in Figure 4.8, the dissimilarity within LUAD or LUSC matrices is notably lower than the dissimilarity resulting from XORing the two sets. This is evident from the greater number of white empty areas in the intra-XOR results, as opposed to the more densely colored squares in the $\{\text{LUAD}_{set} - \text{LUSC}_{set}\}$ difference.

4.4 Discussion

The evaluation of MarbliX has demonstrated its remarkable success in capturing multimodal patient representations by integrating histopathology images and immunogenomic data to generate *monogram* representations. These monograms serve as condensed binary representations of patients' multimodal features, achieved through unimodal processing and transformation, multimodal latent association by training hybrid autoencoders, and self-supervised training with triplet loss. The assessment of patient representation demonstrated the efficacy of MarbliX in capturing distinctive features from both histopathology and immunogenomics modalities.

The similarity analysis of generated monograms showed intra-similarity for patients

with the same primary diagnosis and inter-dissimilarity between those with different primary diagnoses. Both visual and statistical analyses underscored MarbliX's capability to generate meaningful compact binary barcodes, effectively discriminating between patients with distinct primary diagnoses. Furthermore, the evaluation of MarbliX in multimodal search and retrieval demonstrated its superiority over unimodal representations. Monograms consistently resulted in accurate and similar retrieved cases across various criteria, highlighting the model's capability in capturing informative patterns from both histopathology and immunogenomic data. The t-SNE maps provided visual insights into MarbliX's ability to create discriminative representations, notably enhancing the separation of cancer subtypes.

In summary, MarbliX has effectively achieved its objectives by generating informative multimodal monograms that capture and represent the complex features of patients with different cancer subtypes. Beyond demonstrating its robustness in patient representation through multimodal search, MarbliX emerges as a versatile framework with applications extending beyond search and retrieval. Its capability to efficiently integrate and encapsulate distinct data modalities, each inherently possessing large sizes and complexity, into a compact binary matrix is remarkable. This resulting binary matrix serves as a powerful bridge between different modalities, with each entry encoding whether specific features from one modality align with or correspond to certain features in the other. The binary nature of the representations simplifies storage requirements and computational complexity needed for various applications. Multimodal data fusion through MarbliX paves the way for more informed and data-driven decision-making in complex biomedical analyses.

Chapter 5

Summary & Conclusion

This thesis has addressed the critical need for comprehensive multimodal solutions in medical settings, particularly focusing on advancing cancer diagnosis. The primary objective was to bridge the gap between histopathology images and genomic data, recognizing the complexity inherent in aligning these two essential data modalities. The research has resulted in several contributions with the potential of a paradigm shift in the landscape of medical data integration. As medical diagnostics become increasingly reliant on diverse data sources, the demand for innovative approaches that integrate these disparate modalities has grown exponentially. However, each modality presents distinct challenges that must be tackled to facilitate the development of multimodal solutions.

SPLICE, a novel patching algorithm, has played a crucial role in effectively addressing the challenge of processing entire whole slide images (WSIs). Employing a sequential analysis approach, SPLICE processes WSIs to generate a compact and representative patch set known as a *collage*. The *collage* efficiently encapsulates the entire image while ensuring comprehensive coverage of representative tissue regions, thereby reducing computational complexity and storage requirements for downstream tasks. The experiments comparing SPLICE with the state-of-the-art method have demonstrated its efficiency, effectiveness, and flexibility in balancing performance and resource consumption. SPLICE's emphasis

on patch quality and its minimal degrees of freedom make it a valuable tool for diverse applications, particularly in resource-intensive scenarios.

MarbliX, a novel approach for fusing histopathology images with genetic sequencing data, has demonstrated remarkable performance in capturing multimodal patient representations through the generation of *monogram*. These monograms serve as compact binary representations, effectively discriminating between patients with distinct primary diagnoses. MarbliX employs self-supervised triplet learning to unveil concealed correlations between morphological features in histopathology images and complex immunogenetic patterns in latent space, generating similar *monogram* representations for patients sharing the same tumor subtype. This provides valuable insights into disease progression and treatment strategies. MarbliX's success in multimodal search and retrieval, along with its ability to create discriminative representations, highlights its versatility and robustness in multimodal patient representation. Beyond search and retrieval, MarbliX emerges as a powerful framework with applications extending to various biomedical analyses, simplifying storage requirements, and computational complexity. This innovative approach condenses complex medical data sources into a concise binary representations, referred to as *monogram*, emphasizing the significance of multimodal data integration in medical settings. MarbliX bridges a substantial gap in the literature and addresses the scarcity of multimodal solutions.

The significance of this work lies not only in its conceptual contributions but also in the practical implications for medical practitioners seeking more comprehensive insights into patient conditions. By successfully developing and validating SPLICE and MarbliX, this research lays a foundation for a new era in medical data analysis, where the fusion of imaging and genetic data opens up unprecedented possibilities for personalized and effective healthcare solutions. The potential impact of this research extends to pushing the frontiers of medical data integration, with a clear aim to improve diagnostic accuracy and elevate the standards of patient care.

5.1 Future Work

Exploring the integration of additional data modalities, such as gene expression data and immunohistochemistry (IHC) images, could further enhance the comprehensiveness of patient representations. This expansion would contribute to a more profound comprehension of disease progression and treatment response.

Additionally, investigating the scalability of SPLICE and MarbliX to even larger datasets and diverse medical conditions would provide insights into their applicability across a broader spectrum of scenarios. The robustness of these approaches in handling real-world variations and challenges is a critical aspect that requires further exploration.

Furthermore, the development of user-friendly interfaces and tools for clinicians to interact with and interpret the generated representations could facilitate the adoption of these technologies in real-world medical settings. This has the potential to contribute significantly to informed decision-making, leveraging the wealth of information from previously diagnosed and treated patients, thus reducing variability among pathologists.

References

[1] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L Moreira, Narges Razavian, and Aristotelis Tsirigos. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature medicine*, 24(10):1559–1567, 2018.

[2] Shivam Kalra, Hamid R Tizhoosh, Charles Choi, Sultaan Shah, Phedias Diamandis, Clinton JV Campbell, and Liron Pantanowitz. Yottixel–an image search engine for large archives of histopathology whole slide images. *Medical Image Analysis*, 65:101757, 2020.

[3] Zizhao Zhang, Pingjun Chen, Mason McGough, Fuyong Xing, Chunbao Wang, Marilyn Bui, Yuanpu Xie, Manish Sapkota, Lei Cui, Jasreman Dhillon, et al. Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. *Nature Machine Intelligence*, 1(5):236–245, 2019.

[4] Hamid R Tizhoosh, Phedias Diamandis, Clinton JV Campbell, Amir Safarpoor, Shivam Kalra, Danial Maleki, Abtin Riasatian, and Morteza Babaie. Searching images for consensus: can ai remove observer variability in pathology? *The American journal of pathology*, 191(10):1702–1708, 2021.

[5] Bo Li, Taiwen Li, Jean-Christophe Pignon, Binbin Wang, Jinzeng Wang, Sachet A Shukla, Ruoxu Dou, Qianming Chen, F Stephen Hodi, Toni K Choueiri, et al.

Landscape of tumor-infiltrating t cell repertoire of human cancers. *Nature genetics*, 48(7):725–732, 2016.

[6] John-William Sidhom, H Benjamin Larman, Drew M Pardoll, and Alexander S Baras. Deeptcr is a deep learning framework for revealing sequence concepts within t-cell repertoires. *Nature communications*, 12(1):1–12, 2021.

[7] Sin Yee Gun, Sharon Wei Ling Lee, Je Lin Sieow, and Siew Cheng Wong. Targeting immune cells for cancer therapy. *Redox biology*, 25:101174, 2019.

[8] John F Beausang, Amanda J Wheeler, Natalie H Chan, Violet R Hanft, Frederick M Dirbas, Stefanie S Jeffrey, and Stephen R Quake. T cell receptor sequencing of early-stage breast cancer tumors identifies altered clonal structure of the t cell repertoire. *Proceedings of the National Academy of Sciences*, 114(48):E10409–E10417, 2017.

[9] Mikhail V Pogorelyy, Alla D Fedorova, James E McLaren, Kristin Ladell, Dmitri V Bagaev, Alexey V Eliseev, Artem I Mikelov, Anna E Koneva, Ivan V Zvyagin, David A Price, et al. Exploring the pre-immune landscape of antigen-specific t cells. *Genome medicine*, 10(1):1–14, 2018.

[10] Ruslan Medzhitov and Charles A Janeway Jr. Innate immunity: impact on the adaptive immune response. *Current opinion in immunology*, 9(1):4–9, 1997.

[11] David Jung and Frederick W Alt. Unraveling v (d) j recombination: insights into gene regulation. *Cell*, 116(2):299–311, 2004.

[12] Wolf Herman Fridman, Catherine Sautès-Fridman, Jérôme Galon, et al. The immune contexture in human tumours: impact on clinical outcome. *Nature Reviews Cancer*, 12(4):298–306, 2012.

[13] Robert D Leone and Jonathan D Powell. Metabolism of immune cells in cancer. *Nature reviews cancer*, 20(9):516–531, 2020.

[14] Daria Beshnova, Jianfeng Ye, Oreoluwa Onabolu, Benjamin Moon, Wenxin Zheng, Yang-Xin Fu, James Brugarolas, Jayanthi Lea, and Bo Li. De novo prediction of cancer-associated t cell receptors for noninvasive cancer detection. *Science translational medicine*, 12(557), 2020.

[15] Jan A Burger and Adrian Wiestner. Targeting b cell receptor signalling in cancer: preclinical and clinical advances. *Nature Reviews Cancer*, 18(3):148–167, 2018.

[16] David S Fischer, Yihan Wu, Benjamin Schubert, and Fabian J Theis. Predicting antigen specificity of single t cells based on tcr cdr 3 regions. *Molecular systems biology*, 16(8):e9416, 2020.

[17] Ash A Alizadeh, Victoria Aranda, Alberto Bardelli, Cedric Blanpain, Christoph Bock, Christine Borowski, Carlos Caldas, Andrea Califano, Michael Doherty, Markus Elsner, et al. Toward understanding and exploiting tumor heterogeneity. *Nature medicine*, 21(8):846–853, 2015.

[18] Sanjay Mukhopadhyay, Michael D Feldman, Esther Abels, Raheela Ashfaq, Senda Beltaifa, Nicolas G Cacciabeve, Helen P Cathro, Liang Cheng, Kumarasen Cooper, Glenn E Dickey, et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). *The American journal of surgical pathology*, 42(1):39, 2018.

[19] Mark D Zarella, Douglas Bowman, Famke Aeffner, Navid Farahani, Albert Xthona, Syeda Fatima Absar, Anil Parwani, Marilyn Bui, and Douglas J Hartman. A practical guide to whole slide imaging: a white paper from the digital pathology association. *Archives of pathology & laboratory medicine*, 143(2):222–234, 2019.

[20] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis*, 33:170–175, 2016.

[21] Daisuke Komura and Shumpei Ishikawa. Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, 16:34–42, 2018.

[22] Hamid Reza Tizhoosh and Liron Pantanowitz. Artificial intelligence and digital pathology: challenges and opportunities. *Journal of pathology informatics*, 9, 2018.

[23] Dmitrii Bychkov, Nina Linder, Riku Turkki, Stig Nordling, Panu E Kovanen, Clare Verrill, Margarita Walliander, Mikael Lundin, Caj Haglund, and Johan Lundin. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific reports*, 8(1):1–11, 2018.

[24] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019.

[25] Douglas Joseph Hartman, Jeroen AWM Van Der Laak, Metin N Gurcan, and Liron Pantanowitz. Value of public challenges for the development of pathology deep learning algorithms. *Journal of pathology informatics*, 11, 2020.

[26] Haibo Wang, Angel Cruz Roa, Ajay N Basavanhally, Hannah L Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio Gonzalez, and Anant Madabhushi. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, 1(3):034003, 2014.

[27] Maschenka CA Balkenhol, David Tellez, Willem Vreuls, Pieter C Clahsen, Hans Pinckaers, Francesco Ciompi, Peter Bult, and Jeroen AWM van der Laak. Deep learning assisted mitotic counting for breast cancer. *Laboratory investigation*, 99(11):1596–1606, 2019.

[28] Monjoy Saha, Chandan Chakraborty, and Daniel Racoceanu. Efficient deep learning model for mitosis detection using breast histopathology images. *Computerized Medical Imaging and Graphics*, 64:29–40, 2018.

[29] Wouter Bulten, Hans Pinckaers, Hester van Boven, Robert Vink, Thomas de Bel, Bram van Ginneken, Jeroen van der Laak, Christina Hulsbergen-van de Kaa, and Geert Litjens. Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study. *The Lancet Oncology*, 21(2):233–241, 2020.

[30] Eirini Arvaniti, Kim S Fricker, Michael Moret, Niels Rupp, Thomas Hermanns, Christian Fankhauser, Norbert Wey, Peter J Wild, Jan H Rueschoff, and Manfred Claassen. Automated gleason grading of prostate cancer tissue microarrays via deep learning. *Scientific reports*, 8(1):1–11, 2018.

[31] Le Hou, Dimitris Samaras, Tahsin M Kurc, Yi Gao, James E Davis, and Joel 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*, pages 2424–2433, 2016.

[32] Heather D Couture, Lindsay A Williams, Joseph Geradts, Sarah J Nyante, Ebonee N Butler, JS Marron, Charles M Perou, Melissa A Troester, and Marc Niethammer. Image analysis with deep learning to predict breast cancer grade, er status, histologic subtype, and intrinsic subtype. *NPJ breast cancer*, 4(1):1–8, 2018.

[33] Miriam Hägele, Philipp Seegerer, Sebastian Lapuschkin, Michael Bockmayr, Wojciech Samek, Frederick Klauschen, Klaus-Robert Müller, and Alexander Binder. Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. *Scientific reports*, 10(1):1–12, 2020.

[34] Yoichiro Yamamoto, Toyonori Tsuzuki, Jun Akatsuka, Masao Ueki, Hiromu Morikawa, Yasushi Numata, Taishi Takahara, Takuji Tsuyuki, Kotaro Tsutsumi,

Ryuto Nakazawa, et al. Automated acquisition of explainable knowledge from unannotated histopathology images. *Nature communications*, 10(1):1–9, 2019.

[35] Philipp Seegerer, Alexander Binder, René Saitenmacher, Michael Bockmayr, Maximilian Alber, Philipp Jurmeister, Frederick Klauschen, and Klaus-Robert Müller. Interpretable deep neural network to predict estrogen receptor status from haematoxylin-eosin images. In *Artificial Intelligence and Machine Learning for Digital Pathology*, pages 16–37. Springer, 2020.

[36] TRACERx Consortium et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nature Medicine*, 2020.

[37] Muhammed Talo. Automated classification of histopathology images using transfer learning. *Artificial Intelligence in Medicine*, 101:101743, 2019.

[38] Heechan Yang, Ji-Ye Kim, Hyongsuk Kim, and Shyam P Adhikari. Guided soft attention network for classification of breast cancer histopathology images. *IEEE transactions on medical imaging*, 39(5):1306–1315, 2019.

[39] Shallu Sharma and Rajesh Mehra. Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight. *Journal of digital imaging*, 33(3):632–654, 2020.

[40] Baris Gecer, Selim Aksoy, Ezgi Mercan, Linda G Shapiro, Donald L Weaver, and Joann G Elmore. Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks. *Pattern recognition*, 84:345–356, 2018.

[41] R Krithiga and P Geetha. Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review. *Archives of Computational Methods in Engineering*, 28(4):2607–2619, 2021.

[42] Daisuke Komura, Akihiro Kawabe, Keisuke Fukuta, Kyohei Sano, Toshikazu Umezaki, Hirotomo Koda, Ryohei Suzuki, Ken Tominaga, Mieko Ochi, Hiroki Konishi, et al. Universal encoding of pan-cancer histology by deep texture representations. *Cell Reports*, 38(9), 2022.

[43] Liang Zheng, Yi Yang, and Qi Tian. Sift meets cnn: A decade survey of instance retrieval. *IEEE transactions on pattern analysis and machine intelligence*, 40(5):1224–1244, 2017.

[44] Mohamed Maher Ben Ismail. A survey on content-based image retrieval. *Int. J. Adv. Comput. Sci. Appl*, 8(5):159–170, 2017.

[45] Uzma Sharif, Zahid Mehmood, Toqeer Mahmood, Muhammad Arshad Javid, Amjad Rehman, and Tanzila Saba. Scene analysis and search using local features and support vector machine for effective content-based image retrieval. *Artificial Intelligence Review*, 52:901–925, 2019.

[46] Pengshuai Yang, Yupeng Zhai, Lin Li, Hairong Lv, Jigang Wang, Chengzhan Zhu, and Rui Jiang. A deep metric learning approach for histopathological image retrieval. *Methods*, 179:14–25, 2020.

[47] Yun Gu and Jie Yang. Multi-level magnification correlation hashing for scalable histopathological image retrieval. *Neurocomputing*, 351:134–145, 2019.

[48] Xiyue Wang, Yuexi Du, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Retccl: clustering-guided contrastive learning for whole-slide image retrieval. *Medical image analysis*, 83:102645, 2023.

[49] Yushan Zheng, Bonan Jiang, Jun Shi, Haopeng Zhang, and Fengying Xie. Encoding histopathological wsis using gnn for scalable diagnostically relevant regions retrieval. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, pages 550–558. Springer, 2019.

[50] Tingying Peng, Melanie Boxberg, Wilko Weichert, Nassir Navab, and Carsten Marr. Multi-task learning of a deep k-nearest neighbour network for histopathological image classification and retrieval. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, pages 676–684. Springer, 2019.

[51] Edson D Carvalho, OC Antonio Filho, Romuere RV Silva, Flavio HD Araujo, Joao OB Diniz, Aristofanes C Silva, Anselmo C Paiva, and Marcelo Gattass. Breast cancer diagnosis from histopathological images using textural features and cbir. *Artificial intelligence in medicine*, 105:101845, 2020.

[52] Shivam Kalra, Hamid R Tizhoosh, Sultaan Shah, Charles Choi, Savvas Damaskinos, Amir Safarpoor, Sobhan Shafiei, Morteza Babaie, Phedias Diamandis, Clinton JV Campbell, et al. Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence. *NPJ digital medicine*, 3(1):31, 2020.

[53] David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, and Lars Petersson. A survey on graph-based deep learning for computational histopathology. *Computerized Medical Imaging and Graphics*, 95:102027, 2022.

[54] Raphaël Marée, Loïc Rollus, Benjamin Stévens, Renaud Hoyoux, Gilles Louppe, Rémy Vandaele, Jean-Michel Begon, Philipp Kainz, Pierre Geurts, and Louis Wehenkel. Collaborative analysis of multi-gigapixel imaging data using cytomine. *Bioinformatics*, 32(9):1395–1401, 2016.

[55] Ezgi Mercan, Selim Aksoy, Linda G Shapiro, Donald L Weaver, Tad T Brunyé, and Joann G Elmore. Localization of diagnostically relevant regions of interest in whole slide images: A comparative study. *Journal of digital imaging*, 29(4):496–506, 2016.

[56] Michael Nalisnik, Mohamed Amgad, Sanghoon Lee, Sameer H Halani, Jose Enrique Velazquez Vega, Daniel J Brat, David A Gutman, and Lee AD Cooper. Inter-

active phenotyping of large-scale histology imaging data with histomicsml. *Scientific reports*, 7(1):1–12, 2017.

[57] Narayan Hegde, Jason D Hipp, Yun Liu, Michael Emmert-Buck, Emily Reif, Daniel Smilkov, Michael Terry, Carrie J Cai, Mahul B Amin, Craig H Mermel, et al. Similar image search for histopathology: Smily. *NPJ digital medicine*, 2(1):1–9, 2019.

[58] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[59] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[60] Abtin Riasatian, Maral Rasoolijaberi, Morteza Babaei, and Hamid R Tizhoosh. A comparative study of u-net topologies for background removal in histopathology images. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.

[61] Kevin Faust, Quin Xie, Dominick Han, Kartikay Goyle, Zoya Volynskaya, Ugljesa Djuric, and Phedias Diamandis. Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction. *BMC bioinformatics*, 19(1):1–15, 2018.

[62] Wafa Chenni, Habib Herbi, Morteza Babaie, and Hamid R Tizhoosh. Patch clustering for representation of histopathology images. In *European Congress on Digital Pathology*, pages 28–37. Springer, 2019.

[63] David Tellez, Geert Litjens, Jeroen van der Laak, and Francesco Ciompi. Neural image compression for gigapixel histopathology image analysis. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):567–578, 2019.

[64] Sobhan Hemati, Shivam Kalra, Cameron Meaney, Morteza Babaie, Ali Ghodsi, and Hamid Tizhoosh. Cnn and deep sets for end-to-end whole slide image representation learning. In *Medical Imaging with Deep Learning*, pages 301–311. PMLR, 2021.

[65] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018.

[66] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

[67] Bruce Alberts. *Molecular biology of the cell*. Garland science, 2017.

[68] David L Nelson, Albert L Lehninger, and Michael M Cox. *Lehninger principles of biochemistry*. Macmillan, 2008.

[69] Donald Voet, Judith G Voet, and Charlotte W Pratt. *Fundamentals of biochemistry: life at the molecular level*. John Wiley & Sons, 2016.

[70] Jiannan Guo. Transcription: the epicenter of gene expression, 2014.

[71] Hideyuki Shimizu and Keiichi I Nakayama. Artificial intelligence in oncology. *Cancer science*, 111(5):1452–1460, 2020.

[72] Alex Zhavoronkov, Polina Mamoshina, Quentin Vanhaelen, Morten Scheibye-Knudsen, Alexey Moskalev, and Alex Aliper. Artificial intelligence for aging and longevity research: Recent advances and perspectives. *Ageing research reviews*, 49:49–66, 2019.

[73] Feifei Cui, Zilong Zhang, Chen Cao, Quan Zou, Dong Chen, and Xi Su. Protein–dna/rna interactions: Machine intelligence tools and approaches in the era of artificial intelligence and big data. *Proteomics*, 22(8):2100197, 2022.

[74] Ashwin Dhakal, Cole McKay, John J Tanner, and Jianlin Cheng. Artificial intelligence in the prediction of protein–ligand interactions: recent advances and future directions. *Briefings in Bioinformatics*, 23(1):bbab476, 2022.

[75] Hansaim Lim, Fatma Cankara, Chung-Jung Tsai, Ozlem Keskin, Ruth Nussinov, and Attila Gursoy. Artificial intelligence approaches to human-microbiome protein–protein interactions. *Current Opinion in Structural Biology*, 73:102328, 2022.

[76] Sreya Vadapalli, Habiba Abdelhalim, Saman Zeeshan, and Zeeshan Ahmed. Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine. *Briefings in bioinformatics*, 23(5):bbac191, 2022.

[77] Gherman Novakovsky, Nick Dexter, Maxwell W Libbrecht, Wyeth W Wasserman, and Sara Mostafavi. Obtaining genetics insights from deep learning via explainable artificial intelligence. *Nature Reviews Genetics*, 24(2):125–137, 2023.

[78] Evan Calabrese, Javier E Villanueva-Meyer, and Soonmee Cha. A fully automated artificial intelligence method for non-invasive, imaging-based identification of genetic alterations in glioblastomas. *Scientific reports*, 10(1):11852, 2020.

[79] Nour Eldeen M Khalifa, Mohamed Hamed N Taha, Dalia Ezzat Ali, Adam Slowik, and Aboul Ella Hassanien. Artificial intelligence technique for gene expression by tumor rna-seq data: a novel optimized deep learning approach. *IEEE Access*, 8:22874–22883, 2020.

[80] Irene Lena Hudson. Data integration using advances in machine learning in drug discovery and molecular biology. *Artificial Neural Networks*, pages 167–184, 2021.

[81] Shubham Vishnoi, Himani Matre, Prabha Garg, and Shubham Kumar Pandey. Artificial intelligence and machine learning for protein toxicity prediction using proteomics data. *Chemical Biology & Drug Design*, 96(3):902–920, 2020.

[82] Michael W Dorrity, Lauren M Saunders, Christine Queitsch, Stanley Fields, and Cole Trapnell. Dimensionality reduction by umap to visualize physical and genetic interactions. *Nature communications*, 11(1):1537, 2020.

[83] H Hugh Fudenberg, JRL Pink, A-C Wang, Steven D Douglas, et al. *Basic immunogenetics.* Number Ed. 2. Oxford University Press., 1978.

[84] Anastasia Minervina, Mikhail Pogorelyy, and Ilgar Mamedov. T-cell receptor and b-cell receptor repertoire profiling in adaptive immunity. *Transplant International*, 32(11):1111–1123, 2019.

[85] Igor Mandric, Jeremy Rotman, Harry Taegyun Yang, Nicolas Strauli, Dennis J Montoya, William Van Der Wey, Jiem R Ronas, Benjamin Statz, Douglas Yao, Velislava Petrova, et al. Profiling immunoglobulin repertoires across multiple human tissues using rna sequencing. *Nature communications*, 11(1):1–14, 2020.

[86] Yen-Jung Chiu, Yi-Hsuan Hsieh, and Yen-Hua Huang. Improved cell composition deconvolution method of bulk gene expression profiles to quantify subsets of immune cells. *BMC medical genomics*, 12(8):1–17, 2019.

[87] Francesca Finotello and Zlatko Trajanoski. Quantifying tumor-infiltrating immune cells from transcriptomics data. *Cancer Immunology, Immunotherapy*, 67(7):1031–1040, 2018.

[88] Adele De Masson, John T O'Malley, Christopher P Elco, Sarah S Garcia, Sherrie J Divito, Elizabeth L Lowry, Marianne Tawa, David C Fisher, Phillip M Devlin, Jessica E Teague, et al. High-throughput sequencing of the t cell receptor β gene identifies aggressive early-stage mycosis fungoides. *Science translational medicine*, 10(440), 2018.

[89] Shi Xie, Rong Yan, Anqi Zheng, Mengfen Shi, Longqing Tang, Xueying Li, Jiabang Liu, Yifan Gan, Yu Wang, Deke Jiang, et al. T cell receptor and b cell receptor

exhibit unique signatures in tumor and adjacent non-tumor tissues of hepatocellular carcinoma. *Frontiers in Immunology*, 14:1161417, 2023.

[90] Xianliang Hou, Wujun Wei, Junning Zhang, Zhenyu Liu, Guangyu Wang, Xueli Yang, and Yong Dai. Characterisation of t and b cell receptor repertoire in patients with systemic lupus erythematosus. *Clinical and Experimental Rheumatology*, 2023.

[91] Nikolaos Zacharakis, Harshini Chinnasamy, Mary Black, Hui Xu, Yong-Chen Lu, Zhili Zheng, Anna Pasetto, Michelle Langhan, Thomas Shelton, Todd Prickett, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nature medicine*, 24(6):724–730, 2018.

[92] Yoshiro Nakahara, Takaji Matsutani, Yuka Igarashi, Norikazu Matsuo, Hidetomo Himuro, Haruhiro Saito, Kouzo Yamada, Kenta Murotani, Tomoaki Hoshino, Koichi Azuma, et al. Clinical significance of peripheral tcr and bcr repertoire diversity in egfr/alk wild-type nsclc treated with anti-pd-1 antibody. *Cancer Immunology, Immunotherapy*, pages 1–12, 2021.

[93] Fokhrul Hossain, Samarpan Majumder, and Lucio Miele. Immunogenomics: steps toward personalized medicines. In *Clinical Precision Medicine*, pages 73–90. Elsevier, 2020.

[94] Hongyu Zhang, Xianchao Zhou, Dongliang Liu, Yufei Zhu, Qingyang Ma, and Yuchao Zhang. Progress and challenges of personalized neoantigens in the clinical treatment of tumors. *Medicine in Drug Discovery*, 6:100030, 2020.

[95] Guangyu Wang, Weiguo Sui, Wen Xue, Junning Zhang, Xueli Yang, Chune Mo, Xiaoping Pan, Minglin Ou, and Xianliang Hou. Comprehensive analysis of b and t cell receptor repertoire in patients after kidney transplantation by high-throughput sequencing. *Clinical Immunology*, 245:109162, 2022.

[96] Jiangping Li, Zhiwen Xiao, Donghui Wang, Lei Jia, Shihong Nie, Xingda Zeng, and

Wei Hu. The screening, identification, design and clinical application of tumor-specific neoantigens for tcr-t cells. *Molecular Cancer*, 22(1):1–56, 2023.

[97] Yuwei Zhou, Ziru Huang, Wenzhen Li, Jinyi Wei, Qianhu Jiang, Wei Yang, and Jian Huang. Deep learning in preclinical antibody drug discovery and development. *Methods*, 218:57–71, 2023.

[98] David B Page, Jianda Yuan, David Redmond, Y Hanna Wen, Jeremy C Durack, Ryan Emerson, Stephen Solomon, Zhiwan Dong, Phillip Wong, Christopher Comstock, et al. Deep sequencing of t-cell receptor dna as a biomarker of clonally expanded tils in breast cancer after immunotherapy. *Cancer immunology research*, 4(10):835–844, 2016.

[99] Jinhua He, Jian Shen, Wenfeng Luo, Zeping Han, Fangmei Xie, Ting Pang, Liyin Liao, Zhonghui Guo, Jianhao Li, Yuguang Li, et al. Research progress on application of single-cell tcr/bcr sequencing technology to the tumor immune microenvironment, autoimmune diseases, and infectious diseases. *Frontiers in Immunology*, 13:969808, 2022.

[100] Mattia Rediti, David Venet, Françoise Rothé, Tao Qing, Marion Maetens, Ian Bradbury, Miguel A Izquierdo, Serena Di Cosimo, Florentine Hilbers, Mohammed Bajji, et al. Predictive and prognostic role of t-and b-cell receptor repertoire in her2-positive breast cancer: An analysis of the neoaltto clinical trial. *Cancer Research*, 80(16_Supplement):1998–1998, 2020.

[101] VI Nazarov, AA Minervina, AY Komkov, MV Pogorelyy, MA Maschan, YV Olshanskaya, IV Zvyagin, DM Chudakov, YB Lebedev, and IZ Mamedov. Reliability of immune receptor rearrangements as genetic markers for minimal residual disease monitoring. *Bone marrow transplantation*, 51(10):1408–1410, 2016.

[102] Sofia Farkona, Eleftherios P Diamandis, and Ivan M Blasutig. Cancer immunotherapy: the beginning of the end of cancer? *BMC medicine*, 14(1):1–18, 2016.

[103] Kilian Schober, Florian Voit, Simon Grassmann, Thomas R Mueller, Joel Eggert, Sebastian Jarosch, Bianca Weißbrich, Patrick Hoffmann, Lisa Borkner, Enzo Nio, et al. Reverse tcr repertoire evolution toward dominant low-affinity clones during chronic cmv infection. *Nature immunology*, 21(4):434–441, 2020.

[104] Si-Yi Chen, Tao Yue, Qian Lei, and An-Yuan Guo. Tcrdb: a comprehensive database for t-cell receptor sequences with powerful search function. *Nucleic Acids Research*, 49(D1):D468–D474, 2021.

[105] Thomas D Wu, Shravan Madireddi, Patricia E de Almeida, Romain Banchereau, Ying-Jiun J Chen, Avantika S Chitre, Eugene Y Chiang, Hina Iftikhar, William E O'Gorman, Amelia Au-Yeung, et al. Peripheral t cell expansion predicts tumour infiltration and clinical response. *Nature*, 579(7798):274–278, 2020.

[106] Kroopa Joshi, Martina Milighetti, and Benjamin M Chain. Application of t cell receptor (tcr) repertoire analysis for the advancement of cancer immunotherapy. *Current opinion in immunology*, 74:1–8, 2022.

[107] Anja Mösch, Silke Raffegerst, Manon Weis, Dolores J Schendel, and Dmitrij Frishman. Machine learning for cancer immunotherapies based on epitope recognition by t cell receptors. *Frontiers in Genetics*, 10:1141, 2019.

[108] Pierre G Coulie, Benoît J Van den Eynde, Pierre Van Der Bruggen, and Thierry Boon. Tumour antigens recognized by t lymphocytes: at the core of cancer immunotherapy. *Nature Reviews Cancer*, 14(2):135–146, 2014.

[109] Daniel T Harris and David M Kranz. Adoptive t cell therapies: a comparison of t cell receptors and chimeric antigen receptors. *Trends in pharmacological sciences*, 37(3):220–230, 2016.

[110] Swapnil Mahajan, Randi Vita, Deborah Shackelford, Jerome Lane, Veronique Schulten, Laura Zarebski, Martin Closter Jespersen, Paolo Marcatili, Morten Nielsen,

Alessandro Sette, et al. Epitope specific antibodies and t cell receptors in the immune epitope database. *Frontiers in immunology*, 9:2688, 2018.

[111] Hongyi Zhang, Xiaowei Zhan, and Bo Li. Giana allows computationally-efficient tcr clustering and multi-disease repertoire classification by isometric transformation. *Nature communications*, 12(1):1–11, 2021.

[112] Neerja Thakkar and Chris Bailey-Kellogg. Balancing sensitivity and specificity in distinguishing tcr groups by cdr sequence similarity. *BMC bioinformatics*, 20(1):1–14, 2019.

[113] Pradyot Dash, Andrew J Fiore-Gartland, Tomer Hertz, George C Wang, Shalini Sharma, Aisha Souquette, Jeremy Chase Crawford, E Bridie Clemens, Thi HO Nguyen, Katherine Kedzierska, et al. Quantifiable predictive features define epitope-specific t cell receptor repertoires. *Nature*, 547(7661):89–93, 2017.

[114] Jacob Glanville, Huang Huang, Allison Nau, Olivia Hatton, Lisa E Wagar, Florian Rubelt, Xuhuai Ji, Arnold Han, Sheri M Krams, Christina Pettus, et al. Identifying specificity groups in the t cell receptor repertoire. *Nature*, 547(7661):94–98, 2017.

[115] Alan M Luu, Jacob R Leistico, Tim Miller, Somang Kim, and Jun S Song. Predicting tcr-epitope binding specificity using deep metric learning and multimodal learning. *Genes*, 12(4):572, 2021.

[116] Vanessa Isabell Jurtz, Leon Eyrich Jessen, Amalie Kai Bentzen, Martin Closter Jespersen, Swapnil Mahajan, Randi Vita, Kamilla Kjærgaard Jensen, Paolo Marcatili, Sine Reker Hadrup, Bjoern Peters, et al. Nettcr: sequence-based prediction of tcr binding to peptide-mhc complexes using convolutional neural networks. *BioRxiv*, page 433706, 2018.

[117] Yicheng Gao, Yuli Gao, Yuxiao Fan, Chengyu Zhu, Zhiting Wei, Chi Zhou, Guohui Chuai, Qinchang Chen, He Zhang, and Qi Liu. Pan-peptide meta learning for t-cell

receptor–antigen binding recognition. *Nature Machine Intelligence*, 5(3):236–249, 2023.

[118] Dan Hudson, Ricardo A Fernandes, Mark Basham, Graham Ogg, and Hashem Koohy. Can we predict t cell specificity with digital biology and machine learning? *Nature Reviews Immunology*, pages 1–11, 2023.

[119] Emily K Makowski, Hsin-Ting Chen, and Peter M Tessier. Simplifying complex antibody engineering using machine learning. *Cell Systems*, 14(8):667–675, 2023.

[120] Jonathan J Park, Kyoung A V Lee, Stanley Z Lam, Katherine S Moon, Zhenhao Fang, and Sidi Chen. Machine learning identifies t cell receptor repertoire signatures associated with covid-19 severity. *Communications Biology*, 6(1):76, 2023.

[121] Bei Zhang, Zhiyao Ren, Jianfu Zhao, Yue Zhu, Boya Huang, Chanchan Xiao, Yan Zhang, Jieping Deng, Lipeng Mao, Lei Tang, et al. Global analysis of hla-a2 restricted mage-a3 tumor antigen epitopes and corresponding tcrs in non-small cell lung cancer. *Theranostics*, 13(13):4449, 2023.

[122] Julian Gan, Jong Hyun Cho, Ryan Lee, Alireza Naghizadeh, Ling Yue Poon, Ethan Wang, Zachary Hui, and Dongfang Liu. Methods of machine learning-based chimeric antigen receptor immunological synapse quality quantification. In *The Immune Synapse: Methods and Protocols*, pages 493–502. Springer, 2023.

[123] Yuepeng Jiang, Miaozhe Huo, and Shuai Cheng Li. Teinet: a deep learning framework for prediction of tcr–epitope binding specificity. *Briefings in Bioinformatics*, 24(2):bbad086, 2023.

[124] Lidia Gil and Maksymilian Grajek. Artificial intelligence and chimeric antigen receptor t-cell therapy. *Acta Haematologica Polonica*, 53(3):176–179, 2022.

[125] Anika Cheerla and Olivier Gevaert. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics*, 35(14):i446–i454, 2019.

[126] Zhaoyi Sun, Mingquan Lin, Qingqing Zhu, Qianqian Xie, Fei Wang, Zhiyong Lu, and Yifan Peng. A scoping review on multimodal deep learning in biomedical images and texts. *Journal of Biomedical Informatics*, page 104482, 2023.

[127] Shangran Qiu, Matthew I Miller, Prajakta S Joshi, Joyce C Lee, Chonghua Xue, Yunruo Ni, Yuwei Wang, Ileana De Anda-Duran, Phillip H Hwang, Justin A Cramer, et al. Multimodal deep learning for alzheimer's disease dementia assessment. *Nature communications*, 13(1):3404, 2022.

[128] Jialiang Yang, Jie Ju, Lei Guo, Binbin Ji, Shufang Shi, Zixuan Yang, Songlin Gao, Xu Yuan, Geng Tian, Yuebin Liang, et al. Prediction of her2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Computational and structural biotechnology journal*, 20:333–342, 2022.

[129] Andrew J Schaumberg, Mark A Rubin, and Thomas J Fuchs. H&e-stained whole slide image deep learning predicts spop mutation state in prostate cancer. *BioRxiv*, page 064279, 2017.

[130] P Chang, J Grinband, BD Weinberg, M Bardis, M Khy, G Cadena, M-Y Su, S Cha, CG Filippi, D Bota, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *American Journal of Neuroradiology*, 39(7):1201–1207, 2018.

[131] Pooya Mobadersany, Safoora Yousefi, Mohamed Amgad, David A Gutman, Jill S Barnholtz-Sloan, José E Velázquez Vega, Daniel J Brat, and Lee AD Cooper. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences*, 115(13):E2970–E2979, 2018.

[132] Summaira Jabeen, Xi Li, Muhammad Shoib Amin, Omar Bourahla, Songyuan Li, and Abdul Jabbar. A review on methods and applications in multimodal deep learn-

ing. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2s):1–41, 2023.

[133] Sören Richard Stahlschmidt, Benjamin Ulfenborg, and Jane Synnergren. Multimodal deep learning for biomedical data fusion: a review. *Briefings in Bioinformatics*, 23(2):bbab569, 2022.

[134] Joel Saltz, Rajarsi Gupta, Le Hou, Tahsin Kurc, Pankaj Singh, Vu Nguyen, Dimitris Samaras, Kenneth R Shroyer, Tianhao Zhao, Rebecca Batiste, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell reports*, 23(1):181–193, 2018.

[135] Jinru Shia, Nikolaus Schultz, Deborah Kuk, Efsevia Vakiani, Sumit Middha, Neil H Segal, Jaclyn F Hechtman, Michael F Berger, Zsofia K Stadler, Martin R Weiser, et al. Morphological characterization of colorectal cancers in the cancer genome atlas reveals distinct morphology–molecular associations: clinical and biological implications. *Modern pathology*, 30(4):599–609, 2017.

[136] Benoît Schmauch, Alberto Romagnoni, Elodie Pronier, Charlie Saillard, Pascale Maillé, Julien Calderaro, Aurélie Kamoun, Meriem Sefta, Sylvain Toldo, Mikhail Zaslavskiy, et al. A deep learning model to predict rna-seq expression of tumours from whole slide images. *Nature communications*, 11(1):1–15, 2020.

[137] Luís A Vale-Silva and Karl Rohr. Long-term cancer survival prediction using multimodal deep learning. *Scientific Reports*, 11(1):1–12, 2021.

[138] Janani Venugopalan, Li Tong, Hamid Reza Hassanzadeh, and May D Wang. Multimodal deep learning models for early detection of alzheimer's disease stage. *Scientific reports*, 11(1):1–13, 2021.

[139] Nathaniel Braman, Jacob WH Gordon, Emery T Goossens, Caleb Willis, Martin C Stumpe, and Jagadish Venkataraman. Deep orthogonal fusion: Multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data.

In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 667–677. Springer, 2021.

[140] Dongdong Sun, Minghui Wang, and Ao Li. A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. *IEEE/ACM transactions on computational biology and bioinformatics*, 16(3):841–850, 2018.

[141] Yun Gu, Khushi Vyas, Mali Shen, Jie Yang, and Guang-Zhong Yang. Deep graph-based multimodal feature embedding for endomicroscopy image retrieval. *IEEE Transactions on Neural Networks and Learning Systems*, 32(2):481–492, 2020.

[142] Kaimei Huang, Binghu Lin, Jinyang Liu, Yankun Liu, Jingwu Li, Geng Tian, and Jialiang Yang. Predicting colorectal cancer tumor mutational burden from histopathological images and clinical information using multi-modal deep learning. *Bioinformatics*, 38(22):5108–5115, 2022.

[143] Andre Esteva, Jean Feng, Douwe van der Wal, Shih-Cheng Huang, Jeffry P Simko, Sandy DeVries, Emmalyn Chen, Edward M Schaeffer, Todd M Morgan, Yilun Sun, et al. Prostate cancer therapy personalization via multi-modal deep learning on randomized phase iii clinical trials. *NPJ Digital Medicine*, 5(1):71, 2022.

[144] Roger Schaer, Sebastian Otálora, Oscar Jimenez-del Toro, Manfredo Atzori, and Henning Müller. Deep learning-based retrieval system for gigapixel histopathology cases and the open access literature. *Journal of pathology informatics*, 10(1):19, 2019.

[145] Yushan Zheng, Zhiguo Jiang, Fengying Xie, Jun Shi, Haopeng Zhang, Jianguo Huai, Ming Cao, and Xiaomiao Yang. Diagnostic regions attention network (dra-net) for histopathology wsi recommendation and retrieval. *IEEE transactions on medical imaging*, 40(3):1090–1103, 2020.

[146] Osamu Iizuka, Fahdi Kanavati, Kei Kato, Michael Rambeau, Koji Arihiro, and

Masayuki Tsuneki. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Scientific reports*, 10(1):1504, 2020.

[147] Jocelyn Barker, Assaf Hoogi, Adrien Depeursinge, and Daniel L Rubin. Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles. *Medical image analysis*, 30:60–71, 2016.

[148] Noorul Wahab, Islam M Miligy, Katherine Dodd, Harvir Sahota, Michael Toss, Wenqi Lu, Mostafa Jahanifar, Mohsin Bilal, Simon Graham, Young Park, et al. Semantic annotation for computational pathology: Multidisciplinary experience and best practice recommendations. *The Journal of Pathology: Clinical Research*, 8(2):116–128, 2022.

[149] Shiv Ram Dubey. A decade survey of content based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5):2687–2704, 2021.

[150] Lie-fu Ai, Jun-qing Yu, Yun-feng He, and Tao Guan. High-dimensional indexing technologies for large scale content-based image retrieval: a review. *Journal of Zhejiang University SCIENCE C*, 14(7):505–520, 2013.

[151] Menglin Jiang, Shaoting Zhang, Junzhou Huang, Lin Yang, and Dimitris N Metaxas. Scalable histopathological image analysis via supervised hashing with multiple features. *Medical image analysis*, 34:3–12, 2016.

[152] Huei-Fang Yang, Kevin Lin, and Chu-Song Chen. Supervised learning of semantics-preserving hash via deep convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 40(2):437–451, 2017.

[153] Mark D Zarella, Matthew R Quaschnick, David E Breen, and Fernando U Garcia. Estimation of fine-scale histologic features at low magnification. *Archives of pathology & laboratory medicine*, 142(11):1394–1402, 2018.

[154] Mesut Toğaçar, Kutsal Baran Özkurt, Burhan Ergen, and Zafer Cömert. Breastnet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. *Physica A: Statistical Mechanics and its Applications*, 545:123592, 2020.

[155] Hamid R Tizhoosh. Barcode annotations for medical image retrieval: A preliminary investigation. In *2015 IEEE international conference on image processing (ICIP)*, pages 818–822. IEEE, 2015.

[156] Meghana Dinesh Kumar, Morteza Babaie, and Hamid R Tizhoosh. Deep barcodes for fast retrieval of histopathology scans. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

[157] Luis R Soenksen, Yu Ma, Cynthia Zeng, Leonard Boussioux, Kimberly Villalobos Carballo, Liangyuan Na, Holly M Wiberg, Michael L Li, Ignacio Fuentes, and Dimitris Bertsimas. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ digital medicine*, 5(1):149, 2022.

[158] Kevin M Boehm, Pegah Khosravi, Rami Vanguri, Jianjiong Gao, and Sohrab P Shah. Harnessing multimodal data integration to advance precision oncology. *Nature Reviews Cancer*, 22(2):114–126, 2022.

[159] Shih-Cheng Huang, Anuj Pareek, Saeed Seyyedi, Imon Banerjee, and Matthew P Lungren. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ digital medicine*, 3(1):136, 2020.

[160] Babita Pandey, Devendra Kumar Pandey, Brijendra Pratap Mishra, and Wasiur Rhmann. A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions. *Journal of King Saud University-Computer and Information Sciences*, 34(8):5083–5099, 2022.

[161] Fatemeh Behrad and Mohammad Saniee Abadeh. An overview of deep learning methods for multimodal medical data mining. *Expert Systems with Applications*, 200:117006, 2022.

[162] Madhura Jayaratne, Dinithi Nallaperuma, Daswin De Silva, Damminda Alahakoon, Brian Devitt, Kate E Webster, and Naveen Chilamkurti. A data integration platform for patient-centered e-healthcare and clinical decision support. *Future Generation Computer Systems*, 92:996–1008, 2019.

[163] Tianshi Lu, Ze Zhang, James Zhu, Yunguan Wang, Peixin Jiang, Xue Xiao, Chantale Bernatchez, John V Heymach, Don L Gibbons, Jun Wang, et al. Deep learning-based prediction of the t cell receptor–antigen binding specificity. *Nature machine intelligence*, 3(10):864–875, 2021.

[164] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021.

[165] Li Song, David Cohen, Zhangyi Ouyang, Yang Cao, Xihao Hu, and X Shirley Liu. Trust4: immune repertoire reconstruction from bulk and single-cell rna-seq data. *Nature Methods*, 18(6):627–630, 2021.

[166] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[167] Mark M Davis and Pamela J Bjorkman. T-cell antigen receptor genes and t-cell recognition. *Nature*, 334(6181):395–402, 1988.

[168] Sugata Banerji and Sushmita Mitra. Deep learning in histopathology: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(1):e1439, 2022.