Representation Learning for Image Search in Histopathology

by

Abubakr Shafique

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

Examining Committee Membership

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

External:	Dr. Ulysses G. J. Balis Director, Division of Informatics, Department of Pathology, University of Michigan, MI, USA
Internal-External:	Prof. Oleg Michailovich Associate Professor, Dept. of Electrical & Computer Engineering, University of Waterloo, ON, Canada
Internal:	Prof. Tais Sigaeva Assistant Professor, Dept. of Systems Design Engineering, University of Waterloo, ON, Canada
	Prof. Kumaraswamy Ponnambalam Professor, Dept. of Systems Design Engineering, University of Waterloo, ON, Canada
Supervisor:	Prof. Hamid R. Tizhoosh Professor, Artificial Intelligence and Informatics Mayo Clinic, Rochester, MN, USA

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of Contributions

All experimental data, graphical illustrations, and written material included here are the result of my personal work conducted during my Ph.D. research. They collectively represent the contributions I have made over the course of my doctoral studies. The extent to which this thesis incorporates the papers I have authored varies:

1. **Shafique, A.**, Alfasly, S., Alsaafin, A., Nejat, P., Khan, J. A., & Tizhoosh, H. R. (2023). Selection of Distinct Morphologies to Divide & Conquer Gigapixel Pathology Images. arXiv preprint arXiv:2311.09902. (Under Review)

2. Alfasly, S., **Shafique, A.**, Nejat, P., Khan, J., Alsaafin, A., Alabtah, G., & Tizhoosh, H. R. (2023). Rotation-Agnostic Image Representation Learning for Digital Pathology. arXiv preprint arXiv:2311.08359. (Under Review)

3. Lahr, I., Alfasly, S., Nejat, P., Khan, J., Kottom, L., Kumbhar, V., Alsaafin, A., **Shafique, A.**, Hemati, S., Alabtah, G. and Comfere, N., 2024. Analysis and Validation of Image Search Engines in Histopathology. arXiv preprint arXiv:2401.03271. (Under Review)

4. Alfasly, S., Nejat, P., Hemati, S., Khan, J., Lahr, I., Alsaafin, A., **Shafique, A.**, Comfere, N., Murphree, D., Meroueh, C. and Yasir, S., 2024. Foundation Models for Medicine – Fanfare or Flair. Mayo Clinic Proceedings: Digital Health. (Under Review)

5. **Shafique, A.**, Gonzalez, R., Pantanowitz, L., Tan, P. H., Machado, A., Cree, I. A., & Tizhoosh, H. R. (2024). A Preliminary Investigation into Search and Matching for Tumor Discrimination in World Health Organization Breast Taxonomy Using Deep Networks. Modern Pathology, 37(2), 100381.

6. **Shafique, A.**, Babaie, M., Gonzalez, R., Batten, A., Sikdar, S., & Tizhoosh, H. R. (2023) Composite Biomarker Image for Advanced Visualization in Histopathology. In 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE.

7. **Shafique, A.**, Babaie, M., Gonzalez, R., & Tizhoosh, H. R. (2023). Immunohistochemistry Biomarkers-Guided Image Search for Histopathology. In 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE.

8. Hosseini, M., **Shafique, A.**, Babaie, M., & Tizhoosh, H. (2023). Class-imbalanced Unsupervised and Semi-Supervised Domain Adaptation for Histopathology Images. In 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE.

9. **Shafique, A.**, Babaie, M., Sajadi, M., Batten, A., Sikdar, S., & Tizhoosh, H. R. (2021, November). Automatic multi-stain registration of whole slide images in histopathology. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 3622-3625). IEEE.

Abstract

Advancements in the field of Machine Learning (ML) have shown significant promise in complementing the endeavors of healthcare professionals. However, the widespread acceptance and trust in clinical applications necessitate the creation of state-of-the-art algorithms characterized by superior accuracy and performance. Digital Pathology (DP) and Whole Slide Image (WSI) technologies present an innovative pathway for image-based diagnosis in the field of histopathology. DP's advantages offer a unique opportunity to delve into vast archives of medical images using Content-based Image Retrieval (CBIR). CBIR, by enabling pathologists to access information from previously diagnosed cases, can serve as a virtual second opinion, empowering physicians to make confident diagnoses.

The representation of whole slide images (WSIs) plays a pivotal role in various domains, notably in pathology and medicine. However, this task is particularly challenging due to the vast dimensions of WSIs, making comprehensive processing a formidable undertaking within the constraints of existing hardware resources. To confront the complexities associated with processing and searching within expansive repositories of gigapixel WSIs, akin to numerous other substantial big-data challenges, there emerges a compelling need to employ a fundamental computer science methodology known as the "Divide and Conquer" strategy. It is employed to break down WSIs into smaller, meaningful patches. Accurate representation of these patches is vital, especially in medical image analysis for tasks like search and matching. In this thesis, I address these challenges by dividing WSIs into significant patches and creating distinct representations for different tissue types.

Regarding the "divide" process, I have introduced an unsupervised method known as the Selection of Distinct Morphologies (SDM). This approach aims to identify and select all unique patches from the WSI, which we refer to as a "montage". The creation of this montage serves as a pivotal element essential for enabling a variety of applications, including image search. The primary objective of this methodology is to construct a montage consisting of a smaller number of patches that display diversity while retaining their meaningfulness within the framework of the WSI. Furthermore, for the "conquer" aspect, a novel method for learning representations that discriminate between different morphological features has been developed, employing a ranking loss mechanism specifically designed for image retrieval tasks. This metric learning strategy effectively attracts representations of similar morphological attributes closer together in the latent space, while concurrently distancing those that are dissimilar by a predefined margin. The cumulative research efforts during the Ph.D. program have culminated in a comprehensive and pragmatic framework. This framework is designed to facilitate the acquisition of meaningful representations for WSI in the field of DP, with a specific focus on applications related to image search.

Acknowledgements

Throughout the journey of completing this thesis, I have been fortunate to receive guidance, support, and a feeling of belonging from many different sources.

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor H.R. Tizhoosh. His unwavering support, wise counsel, and patience have been crucial to my progress. Working under his guidance has been not only a privilege but also an immensely transformative experience.

I am immensely grateful to the dynamic groups at KIMIA Lab at the University of Waterloo, and Rhazes Lab at the Mayo Clinic. A special mention goes to Mitra and Morteza for their exceptional commitment. My friends Peyman, Saghir, Jibran, Yalda, Areej, Ghazal, Parinaz, Sobhan H., Milad, Danial, Shivam, Sohaila, and Maryam have also played a significant role; the camaraderie, memorable times, and shared challenges we've experienced together hold a dear place in my heart.

I owe a great deal of gratitude to my committee members: Prof. Michailovich, Prof. Sigaeva, Prof. Ponnambalam, and Prof. Balis. Their incisive feedback, expert advice, and academic expertise were pivotal in my development. I also want to extend a big thank you to Prof. Rahnamayan, who served as a committee member during my comprehensive exam but had to be replaced later due to his Adjunct status. I am appreciative of his support and understanding throughout that period.

Filled with deep emotions and heartfelt warmth, I offer my sincerest gratitude to my mother, Khalida Nasreen, my beloved sister, Komal Samreen, my cherished niece, Eshal Fatima, and my brothers, Umar Shafique and Usama Arshad. Together, they have intricately woven a tapestry of values into my essence, teaching me the importance of hard work, patience, maintaining dignity in keeping promises and valuing the enduring gift of knowledge. Their unwavering faith and wisdom have been the guiding stars throughout my journey of exploration.

A special acknowledgment goes to my social group, particularly Saqib, Usman, Gohram, Zaeem, Ujwal, Ali, Anas, and Salem. They provided comfort and a welcome break from the demands of my research endeavors.

To conclude, I offer my respects to the spiritual energies that blessed me with the resilience, determination, and endless inspiration needed to surmount each obstacle.

Dedication

This is dedicated to my mother, elder brother, and elder sister.

Table of Contents

Examining Committee ii

Author's Declaration iii

Statement of Contributions iv

Abstract vi

Acknowledgements vii

Dedication viii

List of Figures xii

List of Tables xxi

List of Abbreviations xxiv

1 Introduction 1

1.1 Histopathology 2

1.2 Digital Pathology & Whole Slide Images (WSIs) 3

1.3 Motivation 4

1.4 Thesis Objectives and Contributions 6

1.5 Thesis Organization 7

2 Related Work **8**
2.1 Deep Models & Architectures . . . 8
2.2 Metric Learning . . . 10
2.3 Content-Based Image Retrieval (CBIR) . . . 11
2.3.1 Opportunities . . . 12
2.3.2 Challenges . . . 13
2.3.3 CBIR in Digital Pathology . . . 15
2.4 Tissue Segmentation . . . 17
2.5 Patching & WSI Representation . . . 18
2.6 Summary . . . 19

3 Selection of Distinct Morphologies to Divide & Conquer the Whole Slide Images **21**
3.1 Introduction . . . 21
3.2 Methodology . . . 23
3.2.1 Selection of Distinct Morphologies (SDM) . . . 24
3.2.2 Atlas for WSI Matching . . . 27
3.3 Evaluation & Results . . . 28
3.3.1 Public – The Cancer Genome Atlas (TCGA) . . . 31
3.3.2 Private – Colorectal Cancer (CRC) . . . 39
3.3.3 Private – Breast Cancer (BC) . . . 44
3.4 Discussion & Conclusion . . . 49

4 NeXtPath – Representation Learning for Image Search **53**
4.1 Methodology . . . 53
4.1.1 NeXtPath: Fine-Tuned ConvNeXt on TCGA FFPE Slides . . . 54
4.1.2 Metric Learning for Image Search . . . 56
4.2 Evaluation & Results . . . 60

4.2.1 The Cancer Genome Atlas (TCGA) . . . 60
4.2.2 BReAst Carcinoma Subtyping (BRACS) . . . 76
4.3 Discussion & Conclusion . . . 84

5 Summary and Conclusions **88**

References **90**

APPENDICES **105**

A **106**

A.1 Extended Results for the Proposed SDM Framework . . . 106
A.1.1 Public – The Cancer Genome Atlas (TCGA) . . . 106
A.1.2 Public – BReAst Carcinoma Subtyping (BRACS) . . . 109
A.1.3 Public – Prostate cANcer graDe Assessment (PANDA) . . . 116
A.1.4 Private – Colorectal Cancer (CRC) . . . 123
A.1.5 Private – Liver ASH vs. NASH . . . 125

B **131**

B.1 Extended Results for TCGA Retrieval Evaluation . . . 131
B.1.1 KimiaNet . . . 131
B.1.2 NeXtPath . . . 133
B.2 Extended Results for BRACS Retrieval Evaluation . . . 136

Glossary **138**

List of Figures

1.1 The steps from biopsy sample collection, slide preparation up to tissue analysis under microscope (images taken from [1, 2] to create this figure). . . . 2

1.2 A high-resolution histopathological digital slide with two highlighted parts at various magnification levels; the low magnified region reveals several tissue types, while the highly magnified area shows individual nuclei of a single tissue type. 5

2.1 CBIR in digital pathology (image re-created using idea from [3]). 12

2.2 U-Net for tissue segmentation (image copied from [4]). 18

2.3 Patching from the binary mask and grid. 19

3.1 **Conceptual Overview.** The overall process to generate a montage from the WSI using SDM. 23

3.2 **The overall SDM process.** Commencing with the extraction of all patches from the WSI at low magnification (say at 2.5x), these patches subsequently undergo processing through a deep network (say DenseNet [5]), resulting in the generation of embeddings for each patch. After obtaining all embeddings, k-means clustering is applied around a *single centroid*, resulting in the calculation of the Euclidean distance of each patch from the centroid. Patches exhibiting similar Euclidean distances are organized into distinct Euclidean bins. Finally, one patch is selected from each bin to build the montage. 24

3.3 **Discrete Euclidean bins within SDM.** The bar chart visually represents the distribution of patches from the WSI across various Euclidean bins. Patches grouped within the same Euclidean bin exhibit similarity. Randomly selected patches (displayed at the top of each bin) represent the montage. 27

3.4 **WSI-Level Search**. The process involves matching one WSI to another using the *median of minimum distances* [3]. For each query WSI, its patch embeddings are compared with the patch embeddings of every WSI in the archive. 29

3.5 Feature extraction to generate the atlas (indexed archive) after pushing the selected patches (Yottixel's Mosaic and SDM's Montage) through the KimiaNet [6]. 30

3.6 Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the TCGA dataset. The diagram shows comparable performance of SDM montage against Yottixel when comparing top-1 and MV@3 retrievals. Hoverver, SDM performs marginally better than Yottixel when comparing MV@5 retrievals. 32

3.7 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the TCGA dataset. 37

3.8 The boxplot illustrates the distribution of patches selected for each WSI in the TCGA dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 33 ± 21. Conversely, for the SDM Montage, the median number of selected patches is 24 ± 4. Here, SDM selects significantly fewer patches than Yottixel. . . . 38

3.9 The t-SNE projection displays the embeddings of all patches extracted from the TCGA dataset using Yottixel's mosaic (left) and SDM's montage (right). 38

3.10 Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the CRC dataset. The diagram shows that SDM montage significantly outperforms Yottixel's mosaic. . . . 40

3.11 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the CRC dataset. 41

3.12 The boxplot illustrates the distribution of patches selected for each WSI in the CRC dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 17 ± 15. On the other hand, for the SDM Montage, the median number of selected patches is 21 ± 4. Here, SDM selects significantly fewer patches than Yottixel. 42

3.13 The t-SNE projection displays the embeddings of all patches extracted from the CRC dataset using Yottixel's mosaic (left) and SDM's montage (right). 43

3.14 Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval in the Breast Cancer dataset. The diagram shows that SDM montage significantly outperforms Yottixel's mosaic. 45

3.15 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval from the BC dataset. 47

3.16 The boxplot illustrates the distribution of patches selected for each WSI in the Breast Cancer (BC) dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 11 ± 9. Conversely, for the SDM Montage, the median number of selected patches is 27 ± 5. Here, SDM selects slightly more patches than Yottixel's mosaic. 48

3.17 The t-SNE projection displays the embeddings of all patches extracted from the BC dataset using Yottixel's mosaic (left) and SDM's montage (right). . 48

3.18 A comprehensive ranking scheme was devised to evaluate the performance of the two methods: Yottixel mosaic and SDM montage, across multiple datasets using various metrics. In this scheme, a rank of '1' signifies superior performance of a method relative to the other, a rank of '2' indicates inferior performance, and identical ranks of '1' for both methods denote comparable performance. After aggregating the results across all metrics, Yottixel mosaic achieved an average rank of 1.64, while SDM montage recorded a more favorable score of 1.09. 50

3.19 The collective accuracy, both macro and weighted averages, at top-1, MV@3, and MV@5 using both Yottixel mosaic and SDM montage methods across all datasets employed for evaluation. 52

4.1 ConvNeXt-tiny [7] architecture for NeXtPath. 54

4.2 (left) Training and validation accuracy for the NeXtPath (ConvNeXt fine-tuned with TCGA diagnostic slides over a span of 20 epochs), exhibited a peak training accuracy of 99.99% and a maximum validation accuracy of 81.45%. (right) Training loss over a span of 20 epochs using cross-entropy loss function. 55

4.3 Triplet-loss is a loss function used primarily in metric learning. The goal of this loss function is to ensure that an anchor (query) pulls the positive (similar) sample closer and pushes away the negative (dissimilar) sample by some margin. 56

4.4 Ranking loss for image search algorithm is illustrated in this image. It is an iterative process with batch embeddings as an input. Within each iteration, a point from the batch is designated as the anchor, and distances to all other points are computed and subsequently arranged in ascending order. From this order, the first dissimilar embedding from the anchor is chosen as a negative embedding, and the last similar embedding as the anchor is chosen as a positive embedding. Subsequently, generating anchor-positive, and anchor-negative pairs. This procedure progresses iteratively until the final anchor point is paired with both its positive and negative counterparts. 57

4.5 Multi-class learning process using "ranking loss for image search" to pull the similar class embeddings closer and push the dissimilar embeddings away. 58

4.6 (left) Training and validation accuracy for the KimiaNet + Ranking (fine-tuned with the proposed ranking loss), exhibited a peak training accuracy of 94.25% and a maximum validation accuracy of 68.88%. (right) Training loss over a span of 10 epochs using the proposed ranking loss function tailored specifically for search and matching. 61

4.7 t-SNE projections for all the embeddings of 240,527 training patches, 24,492 validation patches, and 110,032 test patches from KimiaNet and KimiaNet + Ranking. 63

4.8 Accuracy, macro average of f1-scores, and weighted average of F1-scores are shown for the patch matching when using features from KimiaNet (fine-tuned using cross-entropy loss) [6], and KimiaNet + Ranking (fine-tuned using the proposed ranking loss). The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA validation patches used as query and test patches as a reference atlas. KimiaNet trained with the proposed ranking loss performs slightly better than the KimiaNet trained with cross-entropy loss. 64

4.9 Confusion matrices and chord diagrams from KimiaNet (left column), and KimiaNet + Ranking (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the TCGA Patch dataset. 68

4.10 (left) Training and validation accuracy for the NeXtPath + Ranking (fine-tuned with the proposed ranking loss), exhibited a peak training accuracy of 99.25% and a maximum validation accuracy of 68.95%. (right) Training loss over a span of 10 epochs using the proposed ranking loss function tailored specifically for search and matching. 69

4.11 t-SNE projections for all the embeddings of 240,527 training patches, 24,492 validation patches, and 110,032 test patches from NeXtPath and NeXtPath + Ranking. 71

4.12 Accuracy, macro average of f1-scores, and weighted average of f1-scores are shown for the patch matching when using features from NeXtPath (fine-tuned using cross-entropy loss), and NeXtPath + Ranking (fine-tuned with the proposed ranking loss). The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA test and validation patches combined when used as query and training patches as a reference atlas. NeXtPath trained with the proposed ranking loss performs slightly better than the NeXtPath trained with cross-entropy loss. 72

4.13 Confusion matrices and chord diagrams from KimiaNet (left column), and NeXtPath (right column) trained with the proposed ranking loss. The evaluations are based on the majority of the top 5 retrievals when evaluating the TCGA Patch dataset. 76

4.14 (left) Training and validation accuracy for the NeXtPath fine-tuned with BRACS ROI images over a span of 10 epochs, exhibited a peak training accuracy of 96.71% and a maximum validation accuracy of 65.06%. (right) Training loss over a span of 10 epochs using cross-entropy loss function. . . 77

4.15 (left) Training and validation accuracy for the autoencoder fine-tuned with BRACS ROI embeddings from NeXtPath over a span of 50 epochs, exhibited a peak training accuracy of 92.00% and a maximum validation accuracy of 47.81%. (right) Training loss over a span of 50 epochs using the proposed ranking loss function. 79

4.16 Accuracy, macro average of f1-scores, and weighted average of f1-scores are shown for the patch matching when using features from NeXtPath, and NeXtPath trained with the proposed ranking loss. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the BRACS test and validation patches combined when used as query and training patches as a reference atlas. NeXtPath trained with the proposed ranking loss performs slightly better than the NeXtPath trained with cross-entropy loss. 79

4.17 The architecture of the autoencoder used to learn distinct representations. In this study, RELU is used as the activation function 80

4.18 t-SNE projections for all the embeddings of 3657 training patches, and 882 test & validation patches combined from NeXtPath and NeXtPath + Ranking (bottleneck of autoencoder trained using the proposed ranking loss). . . 82

4.19 Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) autoencoder trained with the proposed ranking loss. The evaluations are based on the majority of the top 5 retrievals when evaluating the BRACS ROI dataset. 84

4.20 85% of the variance explained by the top principle components (PCs). Left: The top 26 PCs from the KimiaNet embeddings. Right: The top 138 PCs from the NeXtPath embeddings contributed 85% of the variance. 86

A.1 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the TCGA dataset. 107

A.2 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the TCGA dataset. 108

A.3 Accuracy, macro average of F1-scores, and weighted average of F1-scores are reported from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the BRACS dataset. 110

A.4 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the BRACS dataset. 112

A.5 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the BRACS dataset. 113

A.6 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the BRACS dataset. 114

A.7 The boxplot illustrates the distribution of patches selected for each WSI in the BRACS dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 21 ± 16. On the other hand, for the SDM Montage, the median number of selected patches is 30 ± 5. 115

A.8 The t-SNE projection displays the embeddings of all patches extracted from the BRACS dataset using Yottixel's mosaic (left) and SDM's montage (right). 115

A.9 Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals in the PANDA dataset. 117

A.10 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the PANDA dataset. 119

A.11 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the PANDA dataset. 120

A.12 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the PANDA dataset. 121

A.13 The boxplot illustrates the distribution of patches selected for each WSI in the PANDA dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 9 ± 2. On the other hand, for the SDM Montage, the median number of selected patches is 12 ± 3. 122

A.14 The t-SNE projection displays the embeddings of all patches extracted from the PANDA dataset using Yottixel's mosaic (left) and SDM's montage (right). 122

A.15 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the CRC dataset. 123

A.16 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the CRC dataset. 124

A.17 Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals in the Liver dataset. 126

A.18 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the Liver dataset. 127

A.19 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the Liver dataset. 128

A.20 Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the Liver dataset. 129

A.21 The boxplot illustrates the distribution of patches selected for each WSI in the Liver dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 9 ± 3. Conversely, for the SDM Montage, the median number of selected patches is 17 ± 4. 130

A.22 The t-SNE projection displays the embeddings of all patches extracted from the Liver dataset using Yottixel's mosaic (left) and SDM's montage (right). 130

B.1 Confusion matrices and chord diagrams from KimiaNet (left column), and KimiaNet + Ranking (right column). The evaluations are based on the top 1 retrieval when evaluating the TCGA Patch dataset. 132

B.2 Confusion matrices and chord diagrams from KimiaNet (left column), and KimiaNet + Ranking (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the TCGA Patch dataset. 133

B.3 Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) trained with the proposed ranking loss. The evaluations are based on the top 1 retrieval when evaluating the TCGA Patch dataset. 134

B.4 Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) trained with the proposed ranking loss. The evaluations are based on the majority of the top 3 retrievals when evaluating the TCGA Patch dataset. 135

B.5 Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) trained with the proposed ranking loss. The evaluations are based on the top 1 retrieval when evaluating the BRACS ROI dataset. 136

B.6 Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) trained with the proposed ranking loss. The evaluations are based on the majority of the top 3 retrievals when evaluating the BRACS ROI dataset. 137

List of Tables

3.1 Comprehensive details regarding the TCGA dataset utilized in this study, encompassing the corresponding acronyms and the number of slides attributed to each primary diagnosis. 34

3.2 Detailed precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using the Yottixel mosaic. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the TCGA dataset. 35

3.3 Detailed precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using the SDM Montage. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA dataset. 36

3.4 Comprehensive dataset particulars pertaining to the Colorectal Cancer dataset utilized in this experiment, encompassing relevant acronyms and the number of slides attributed to each primary diagnosis. 39

3.5 Precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the CRC dataset. 40

3.6 Detailed information related to the BC dataset, inclusive of the respective acronyms and the number of slides associated with each primary diagnosis. 44

3.7 Precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval in the Breast Cancer dataset. . 46

4.1 Detailed precision, recall, f1-score, and the number of patches processed for each subtype are shown in this table using the validation patches when matched against the test patches using KimiaNet for feature extraction. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA Patch dataset. 66

4.2 Detailed precision, recall, f1-score, and the number of patches processed for each subtype are shown in this table using the validation patches when matched against the test patches using KimiaNet (trained with the proposed ranking loss) for feature extraction. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA Patch dataset. 67

4.3 Detailed precision, recall, f1-score, and the number of patches processed for each subtype are shown in this table using the validation patches when matched against the test patches using NeXtPath for feature extraction. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA Patch dataset. 74

4.4 Detailed precision, recall, f1-score, and the number of patches processed for each subtype are shown in this table using the validation patches when matched against the test patches using NeXtPath (trained with the proposed ranking loss) for feature extraction. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA Patch dataset. 75

4.5 Information concerning the BRACS dataset employed in this experiment, inclusive of the respective acronyms and the number of slides associated with each primary diagnosis. 78

4.6 Precision, recall, F1-score, and the number of slides processed for each subtype are reported in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the BRACS dataset. 83

A.1 Information concerning the BRACS dataset employed in this experiment, inclusive of the respective acronyms and the number of slides associated with each primary diagnosis and group. 109

A.2 Precision, recall, F1-score, and the number of slides processed for each subtype are reported in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the BRACS dataset. 110

A.3 Comprehensive dataset particulars pertaining to the Prostate cANcer graDe Assessment (PANDA) dataset, encompassing relevant ISUP grade and the number of slides attributed to each grade. 116

A.4 Precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals in the PANDA dataset. 118

A.5 Information related to the Liver dataset, inclusive of the respective acronyms and the number of slides associated with each primary diagnosis. 125

A.6 Precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals in the Liver dataset. 126

List of Abbreviations

ASH Alcoholic Steatohepatitis 30

BC Breast Cancer 30

BRACS BReAst Carcinoma Subtyping 29, 60

CAD Computer-aided Diagnosis 1, 4, 24

CBIR Content-based Image Retrieval 6, 8, 11–15, 22, 31

CNN Convolutional Neural Network 9, 16, 54, 85

CPU Central Processing Unit 14

CRC Colorectal Cancer 30

CV Computer Vision 9

DL Deep Learning 6, 11, 13, 21, 22, 53

DNN Deep Neural Network 6, 57

DP Digital Pathology 1, 3, 4, 11, 13, 15, 16, 89

FFPE Formalin-fixed Paraffin-embedded Tissue 54

GPU Graphical Processing Unit 13, 14

H&E Haematoxylin and Eosin 3, 16, 31

IHC Immunohistochemistry 3, 15, 16

ISUP International Society of Urological Pathology 116

ML Machine Learning 5, 10, 12–14, 21, 88

NASH Non-alcoholic Steatohepatitis 30

NCI National Cancer Institute 31

NIH National Institutes of Health 31

PANDA Prostate cANcer graDe Assessment (PANDA) 29, 116

RUMC Radboud University Medical Center 116

SDM Selection of Distinct Morphologies 23, 24, 32, 33, 45, 49, 50, 110, 111, 118

SIFT Scale-invariant Feature Transform 15

SMILY Similar Image Search for Histopathology 15

t-SNE t-distributed Stochastic Neighbor Embedding 33, 39, 45, 85, 86, 109, 117, 125

TCGA The Cancer Genome Atlas 16, 29, 31, 60, 85

ViT Vision Transformer 8, 9

WSI Whole Slide Image 1, 3, 4, 7, 8, 15–19, 22–24, 44, 45, 49, 50, 109, 117, 125

Chapter 1

Introduction

Pathologists examine tissue slides under a microscope on a regular basis and produce diagnostic and prognostic reports based on their visual inspection. Because of the growing quantity of tissue slides and the relevance of this type of inspection in clinical care and biological research, this visual task has become monotonous and inefficient [8]. A biopsy is a process in which biological samples are taken from the human body to study under a microscope for the detection of abnormalities. It is the gold standard procedure for cancers and tumour diagnosis [9, 10]. Figure 1.1 shows the biopsy sample collection up to the analysis process. With the advancement in Digital Pathology (DP), glass slides with fixed tissue can now be transformed into digital files called Whole Slide Image (WSI). Pathologists can examine the digital slides using a range of image processing tools to improve disease diagnosis [11]. Medical images are complex in nature and need specialized and trained human experts for interpretation. Computerized pathology slides and automated methods, which are generally termed as Computer-aided Diagnosis (CAD) have the potential to transform present pathology diagnosis and prognosis methods by assisting physicians in making faster and more accurate diagnoses [12, 13].

Obtaining a second opinion in the context of cancer diagnosis is of paramount importance from a clinical and scientific perspective [13, 14, 15]. It serves as a vital mechanism for validating the accuracy of the initial diagnosis, mitigating the potential for misdiagnosis (variability among the pathologists), and ensuring the appropriateness of the chosen treatment modality [16, 17]. Furthermore, seeking a second opinion permits the exploration of diverse therapeutic approaches, which can include the consideration of emerging clinical trials and experimental treatments, thereby broadening the spectrum of therapeutic options available to patients [18]. In essence, a second opinion contributes significantly to the scientific rigor and precision of cancer diagnosis and treatment by facilitating in-

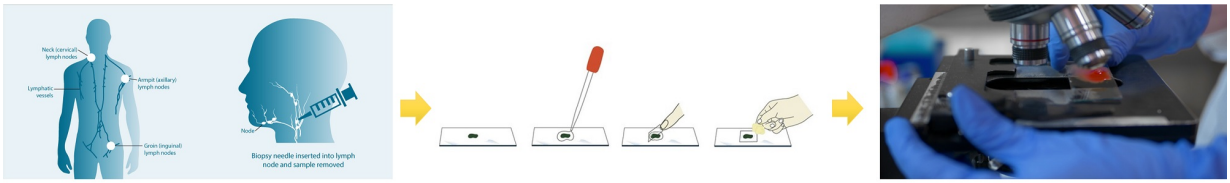

Figure 1.1: The steps from biopsy sample collection, slide preparation up to tissue analysis under microscope (images taken from [1, 2] to create this figure).

formed, evidence-based decision-making and optimizing the patient's overall quality of life throughout their cancer care journey [19].

This Ph.D. research is founded on the hypothesis that the utilization of image search technology has the potential to improve the pronounced variability observed both within and between observers in the context of medical diagnosis. This improvement can be achieved by enabling pathologists to conduct searches within a high-quality repository of previously diagnosed cases, thereby harnessing the collective knowledge and expertise of pathologists who have previously diagnosed cases exhibiting similar tissue patterns.

1.1 Histopathology

Histopathology is the study of disease manifestation by inspecting and interpreting varied forms, sizes, and architectural patterns of cells and tissues, which can be paired with a patient's clinical records and other diagnostic modalities [20]. The term "histopathology" is derived from the combining of two scientific disciplines: histology and pathology [21]. pathology is the identification of diseases through microscopic studies of tissue specimens (acquired through different types of biopsy), whereas histology is the study of microscopic structures of tissues [22, 23, 24]. One of the most important specialties in the healthcare delivery systems is histopathology. Pathologists are the ones who study and practice it. Pathologists' primary clinical responsibility is to conduct microscopic analysis of glass slides containing tissue specimens in order to generate pathology reports. Pathology reports are used to make a variety of clinical decisions, including illness screening, formulating diagnostic programs, monitoring disease development, and managing various medicines and their prognosis [25, 9]. Before moving to microscopy, it is critical to comprehend the pathology glass slide preparation process. Before glass slides are digitized or used for diagnosis, there are four phases involved in their preparation which include collection,

dehydration, embedding, sectioning, and staining [26, 27, 28]. First of all, in the collection, tissue samples (specimens) are taken from a patient's affected body area, e.g., via surgery or needle biopsy, and placed in a fixative. The purpose of fixation is to prevent degradation and stabilize the tissue and cell components in order to retain a cellular structure that is as close to its natural condition as feasible. Formalin is the most often used fixative for light microscopy and Immunohistochemistry (IHC). Secondly, The dehydration of a sample is accomplished by adding ethanol. It removes water from the sample and hardens the tissue even more in preparation for light microscopy. Following the application of ethanol and the completion of tissue dehydration, the ethanol is removed with xylene. Thirdly, in embedding, tissue samples are fixed in paraffin wax so that tiny slices can be taken out for the details of the structures of the tissues and individual cells to be clearly visible through the microscope (sometimes frozen sections are used, e.g., for surgical pathology). Furthermore, in sectioning, A specific piece of equipment called a "microtome" is used to cut embedded material into thin slices. Finally, in staining, to emphasize distinct components of the "sectioned" tissue, different stains and dyes are utilized to colorize tissue structures. Haematoxylin and Eosin (H&E) staining is the most prevalent form of staining procedure [29].

The core of histopathology is the interpretation of high-resolution images of tissues and cells. The **light microscope**, in various forms, has been the only device accessible for this task for centuries, enabling live images at increasing resolution through ever-improving optics [30]. The microscope revolutionized disease treatment by shifting the focus from complete organs to cells; it permitted the practice of histopathology and created a slew of technical advancements required for present practice [31]. Histopathology is leading the way in using digital imaging technology as a "digital-age" alternative to conventional light microscopy, as clinical practices become more digitized. Now, through a technology known as WSI or virtual microscopy, robotic microscopic scanners are utilized to digitize glass slides into gigapixel images. The gigapixel WSI are digital slides that can be displayed on computer screens like any other digital image.

1.2 Digital Pathology & Whole Slide Images (WSIs)

For pathologists, digital slides simulate a light microscope: on computer screens, digital slides combined with software programs provide the same capability as a microscope and more [32]. The WSI scanner can digitize glass tissue slides, making image interpretation in pathology much easier. DP is a branch of pathology that focuses on data management and processing from digital specimen slides. Digital pathology employs virtual microscopy

through the use of computer-based technology. Figure 1.2 shows a sample digital slide obtained from a WSI scanner. In DP, digitizing specimens is one of the recent significant achievements in the integration of modern computational practices within conventional medicine [33]. With the advanced technology, digital slides and the impact of the COVID-19 pandemic have sparked a revolution in diagnostic pathology [32]. Image analysis, which uses computer techniques to interpret pathology images, is rapidly gaining traction as a useful tool for investigating a wide range of pathology operations [9, 34, 32, 35, 36]. Numerous studies have shown that high-throughput analysis can greatly reduce pathologists' workload while reducing the inherent subjectivity of visual analysis [37, 38]. Due to ongoing improvements in the capability and throughput of WSI scanners, the development of user-friendly software systems for organizing and viewing digital slides, and vendor-supplied storage options, WSI technology has grown quickly in recent years [9, 34].

The enormous dimensionality of images in DP makes their processing and storage difficult. For this reason, understanding regions of interest in images aids in quicker diagnosis and detection when using digital-computing approaches [39]. Tissue attributes such as cell nuclei, glands, and lymphocytes are discovered to have notable traits that serve as markers for recognizing malignant cells, particularly in histopathology [9]. Researchers also believe that pathologists will be able to connect histological patterns with protein and gene expressions, undertake exploratory histopathology image analysis, and perform CAD to help them make better decisions [9]. The concept of using CAD to quantify spatial histopathological features has been investigated by a number of works since 1990s [40, 41, 42].

1.3 Motivation

The Institute of Medicine (IOM)[1] released a report titled "To Err is Human: Building a Safer Health System" which states that "...as many as 98,000 people die in any given year from medical errors. that's more than deaths from motor vehicle accidents, breast cancer, or "AIDS" [43]. According to the report, "error" is the third leading cause of death in the United States. Humans, by their very nature, make mistakes, and healthcare is no exception. What matters most is that we learn from our mistakes and use the information we have to avoid or minimize future misdiagnoses.

Physicians order a variety of diagnostic tests spanning many modalities to guide patients through the diagnosis, therapy, and monitoring phases [44]. Although, histopathologic examination is still considered the most reliable method for diagnosis. However, some

[1]Since 2015, it is known as National Academy of Medicine (NAM)

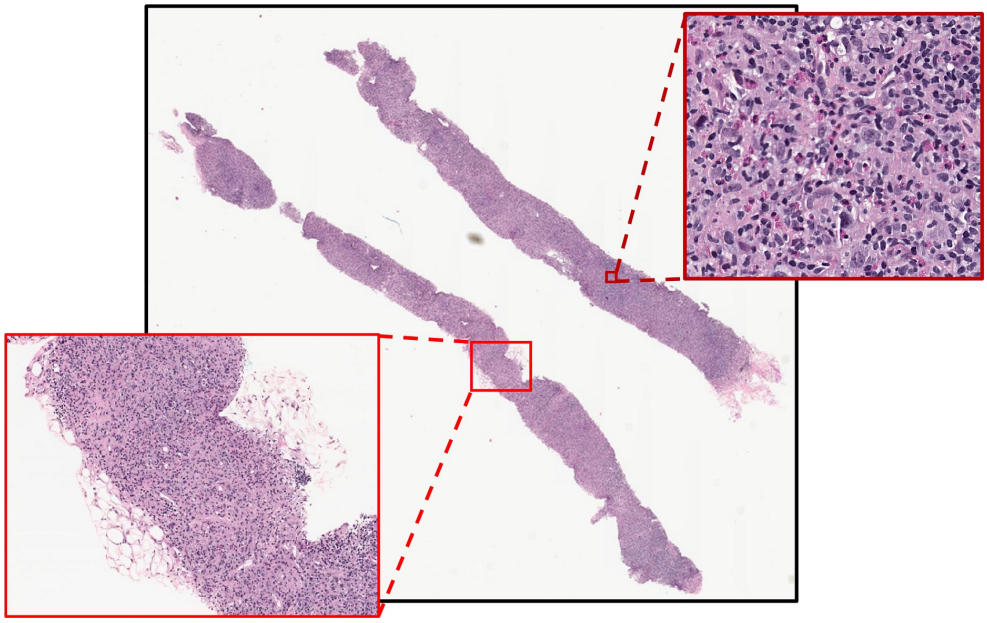

Figure 1.2: A high-resolution histopathological digital slide with two highlighted parts at various magnification levels; the low magnified region reveals several tissue types, while the highly magnified area shows individual nuclei of a single tissue type.

diseases pose a diagnostic challenge due to their complexity, and discordance among experts [45, 46]. Obtaining second opinions is a common practice when dealing with complex cases that exhibit histopathological criteria bridging two or more taxonomic categories. These additional assessments can be valuable in such borderline cases [47]. In a research study involving 6,900 separate case assessments, pathologists indicated that they intended to seek a second opinion, either as mandated by institutional policy or at their own discretion, in 70% of the cases, specifically 4,827 out of 6,900 cases [45]. Obtaining a *second opinion* from fellow pathologists who may be in another hospital located in another part of the world is a cumbersome, expensive, and time-consuming task. Utilizing machine learning, applied on digital images, can enhance the consistency and objectivity of these evaluations via image search. When searching for previously diagnosed cases that closely resemble a new case, valuable insights regarding the factors influencing tumor progression may be gleaned by pathologists.

In the domain of histopathology, Machine Learning (ML) training approaches encompass fully-supervised, semi-supervised, and unsupervised methodologies, each offering distinct advantages [48]. Fully-supervised learning ensures precise diagnostic classification by

training on labeled datasets. Semi-supervised learning, utilizing both labeled and unlabeled data, proves advantageous in histopathology [48, 36], where obtaining labeled samples can be resource-intensive. This approach enhances generalization and performance, particularly when labeled data is limited. Unsupervised learning, on the other hand, allows the algorithm to uncover inherent patterns within unlabeled data, potentially leading to the discovery of novel insights and biomarkers. Collectively, these ML approaches contribute to enhanced diagnostic accuracy, optimal resource utilization, and the potential discovery of previously unrecognized patterns, thereby advancing efficient histopathological analysis. The process of manual segmentation and annotation by experts introduces the potential for intra-observer variability in diagnoses, highlighting the need for supervised learning and its application in Content-based Image Retrieval (CBIR) within the scope of this PhD study.

Another challenge that necessitates this Ph.D. research is to overcome the challenges of processing and representing gigapixel histopathology images such that one can learn the discrete representation of different tumour subtypes. Conventional methods for image analysis relied on handcrafted, domain-specific features to describe attributes like color, shape, and texture in images. Nonetheless, creating and adapting such features for new images (i.e., various organs and diseases) proved challenging. Consequently, deep learning has emerged as a dominant paradigm, particularly with the effectiveness of Deep Neural Network (DNN) in image characterization. However, the majority of recent breakthroughs have been focused on processing relatively small images, such as natural images, using Deep Learning (DL) techniques (many cases using 224 by 224 pixel images [49, 50]). Extending these methodologies is imperative to address the unique challenges posed by gigapixel histopathology images and the subtle distinctions required for diagnostic interpretations in this context.

1.4 Thesis Objectives and Contributions

The primary objective of this thesis is to establish representation learning frameworks that can effectively extract distinguishing feature vectors for different tumour types for entire digital pathology whole slide image. These representations are intended for the development of specialized tools, such as image indexing and search systems, aimed at aiding clinicians in image-based diagnosis. The experiments conducted in this research are designed to quantitatively assess the efficacy of these representations in their capacity to search through the archives of histopathology slides and retrieve slides with the correct primary diagnosis. This thesis contributes to the overarching and persistent goal of the

biomedical community, which aims to incorporate machine learning as a supportive tool in the realm of medical image analysis. The thesis delivers two contributions:

1. A novel unsupervised approach to "divide & conquer" the WSI. The key contribution is to capture all diverse aspects of the unlabeled WSI using a fewer number of patches in an unsupervised fashion. The details are discussed in Chapter 3.

2. A Ranking loss to learn distinct subtype representation in a specific archive of cases for image search. It is further discussed in detail in Chapter 4

1.5 Thesis Organization

The thesis is organized as follows: In Chapter 2, related works are discussed to introduce essential definitions, fundamental concepts, and a review of existing literature, encompassing different contemporary methods for applying machine learning in pathology. Furthermore, in Chapter 3, and 4 the key contributions of this Ph.D. research will be discussed. Finally, Chapter 5 concludes the thesis.

Chapter 2

Related Work

This chapter presents an overview of the literature related to deep models & architectures, metric learning, CBIR, tissue segmentation, and patching & WSI representation.

2.1 Deep Models & Architectures

The history of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) illustrates the remarkable progress in computer vision since 2010. CNNs, which were originally developed in the late 1980s, saw a resurgence in the 2010s with groundbreaking models like AlexNet [51], VGGNet [52], Inceptions [53], ResNet [54, 55], DenseNet [5], MobileNet [56], EfficientNet [57], RegNet [58], and ConvNeXt [7]. These architectures demonstrated the potency of deep learning for image analysis tasks and introduced innovations such as *skip connections* to tackle the challenges of training very deep networks. Meanwhile, Vision Transformers emerged as a more recent development, leveraging the disruptive Transformer architecture initially designed for natural language processing. Transformers entered the computer vision scene with models like the Vision Transformer (ViT) and have since gained momentum due to their adaptability and effectiveness in various tasks [59, 60]. Advances in both CNNs and ViTs have continued, with the introduction of efficient architectures and their application in diverse areas such as object detection, semantic segmentation, generative image synthesis, and representation learning. These advancements have been further propelled by improvements in hardware, particularly GPUs, enabling the training of increasingly complex models and reshaping the landscape of artificial intelligence and image processing.

DenseNet [5] has gained popularity as a Convolutional Neural Network (CNN) architecture known for its efficiency and effectiveness. Unlike traditional CNNs where each layer connects to the next, DenseNet introduces "dense connectivity", where every layer connects to all subsequent layers. This dense interconnection not only facilitates feature reuse but also combats the *vanishing gradient* problem [5], making it easier to train very deep networks. DenseNet has found widespread use in various computer vision applications, including image classification, object detection, and image segmentation, due to its ability to extract rich and hierarchical features from images efficiently [5, 6]. Its architectural innovations have contributed significantly to the success of deep learning in the field of Computer Vision (CV). Recent research considers the DenseNet architecture a dependable choice for representing images in the field of histopathology [61]. Riasatian et al. [6] also used DenseNet to fine-tune using the TCGA dataset for better representation in histopathology and named as *KimiaNet*.

ViTs [59, 60] have become versatile tools in a wide range of computer vision applications. Initially introduced as an adaptation of the Transformer architecture from natural language processing to image analysis tasks, ViTs have demonstrated their effectiveness in image classification, object detection, image segmentation, and more. Their ability to capture long-range dependencies in images and their scalability to handle both small and large datasets has made ViTs a popular choice for various computer vision tasks. Additionally, ViTs have shown promise in handling tasks that require understanding context and relationships within images, contributing to their broad adoption and potential for future innovations in the field of computer vision. ViTs are generally computationally more demanding than CNNs, and require more training data [50]. Caron et al. [50] and Oquab et al. [62] have shown that training ViTs on large unlabeled datasets to predict certain image transformations or context, ViTs can learn meaningful representations of visual data. Swin Transformer represents a significant advancement in the field of deep learning [49]. It introduces a hierarchical design with alternating stages of vision and shift operations, allowing for an efficient and scalable approach to handling images of varying resolutions. One of the standout features of the Swin Transformer is its ability to handle large images with minimal computational cost. By using shifted windows instead of traditional non-overlapping patches, it reduces the memory requirements while maintaining high accuracy.

Following the rapid proliferation of ViTs as a dominant paradigm in computer vision, CNN named ConvNeXt [7] have experienced a resurgence marked by substantial advancements. ViTs initially gained prominence for their remarkable ability to capture long-range dependencies and effectively process visual data. However, CNNs, which were once considered the standard, have reentered the research landscape with renewed vigor. This

resurgence can be attributed to several factors, including architectural innovations, more efficient training techniques, and the adaptability of CNNs to various vision tasks (inspired by swin transformer). Consequently, the competition and synergy between ViTs and CNNs have spurred significant progress in the field of computer vision, offering diverse and powerful tools to address complex visual recognition challenges.

2.2 Metric Learning

Metric learning is a crucial subfield within machine learning that focuses on the development of algorithms and techniques to learn effective distance metrics or similarity measures between data points [63]. In essence, it aims to teach a model how to quantify the similarity or dissimilarity between pairs of samples in a dataset. This learned distance metric is valuable for various ML tasks, including classification, clustering, and recommendation systems [63]. By optimizing the metric, machine learning models can better distinguish between similar and dissimilar instances, leading to improved performance in tasks that rely on measuring similarity or dissimilarity between data points. Metric learning finds applications in image retrieval, face recognition, recommendation engines, and other tasks, where the quality of the learned metric directly impacts the system's accuracy and effectiveness [64]. Numerous contemporary deep metric learning techniques rely on pairs of data samples. To elaborate formally, their loss functions are formulated based on the pairwise similarities observed within the embedding space. Pair-based deep metric learning approaches include contrastive loss [65], triplet loss [66], triplet-center loss [67], quadruplet loss [68], lifted structure loss [69], N-pairs loss [70], histogram loss [71], angular loss [72], distance weighted margin-based loss [73], and hierarchical triplet loss (HTL) [74]. Every dataset presents unique challenges regarding both classification and clustering tasks. Distance metrics that lack the capacity to adapt well to various problems are unlikely to yield effective outcomes in data classification. Consequently, the attainment of successful results with input data hinges upon the utilization of a robust distance metric.

A triplet network [66, 75], drawing inspiration from the Siamese network architecture [75], comprises three entities: positive, negative, and anchor samples. These triplet networks employ Euclidean space for the comparative analysis of these entities during the pattern recognition process, and this methodology is intrinsically tied to the principles of metric learning [66]. The triplet loss initially places emphasis on evaluating the likeness between pairs of samples from the same and distinct classes while leveraging shared weights. Triplet networks enhance discriminative capabilities by considering relationships within the same class as well as across different classes.

Cakir et al. [76] present an innovative deep metric learning methodology, drawing inspiration from the "learning to rank" paradigm, which is termed FastAP. This technique specifically seeks to optimize the rank-based Average Precision metric by employing an approximation rooted in distance quantization. Notably, this approach is meticulously adapted to be compatible with the nuances of stochastic gradient descent, ensuring efficient and effective learning dynamics.

Ranking loss and metric learning are interconnected in their pursuit of optimizing similarity metrics. Ranking loss defines how pairs or triplets of data should be ranked in terms of similarity, aligning the model's objective with similarity measurement. Recently, Kemertas et al. [77] present an information-theoretic loss function termed "RankMI" along with a corresponding training algorithm designed for deep representation learning in the context of image retrieval. The proposed framework involves iterative updates to a network, which estimates the divergence between distance distributions for pairs of embeddings corresponding to matching and non-matching instances. Simultaneously, they optimize an embedding network to maximize this estimate using sampled negative examples.

Recently Mazaheri et al. [78] introduced a ranking loss to overcome the image search bias in histopathology. In this research paper, two innovative approaches are presented to enhance the performance of image retrieval. Firstly, a ranking loss function is employed to steer the feature extraction process towards a focus on the matching aspects of the search. This involves training the model to rank matched outputs, thereby tailoring the representation learning specifically for image retrieval purposes rather than traditional class label learning. Secondly, the concept of "sequestering learning" is introduced, aiming to improve the generalization capabilities of the feature extraction process.

2.3 Content-Based Image Retrieval (CBIR)

Since the last decade, CBIR has been one of the most important fields in computer vision [79]. It allows a user to search for photographs that are similar to one another from a large database of images. CBIR has numerous real-world applications, but it is especially valuable for medical images, as linguistic features collected from medical reports are frequently insufficient representations of the content of the related medical images [80, 81, 82]. The enormous medical image archives have traditionally been bundled with textual annotations classified by professionals; however, this technique does not scale well with the ever-increasing demands of digital pathology. While CBIR systems for histopathology have received a lot of attention [83], image search and analysis for histopathological images has just recently become a focus of research, due to the rise of DP and DL [34, 3, 84, 85, 86, 15].

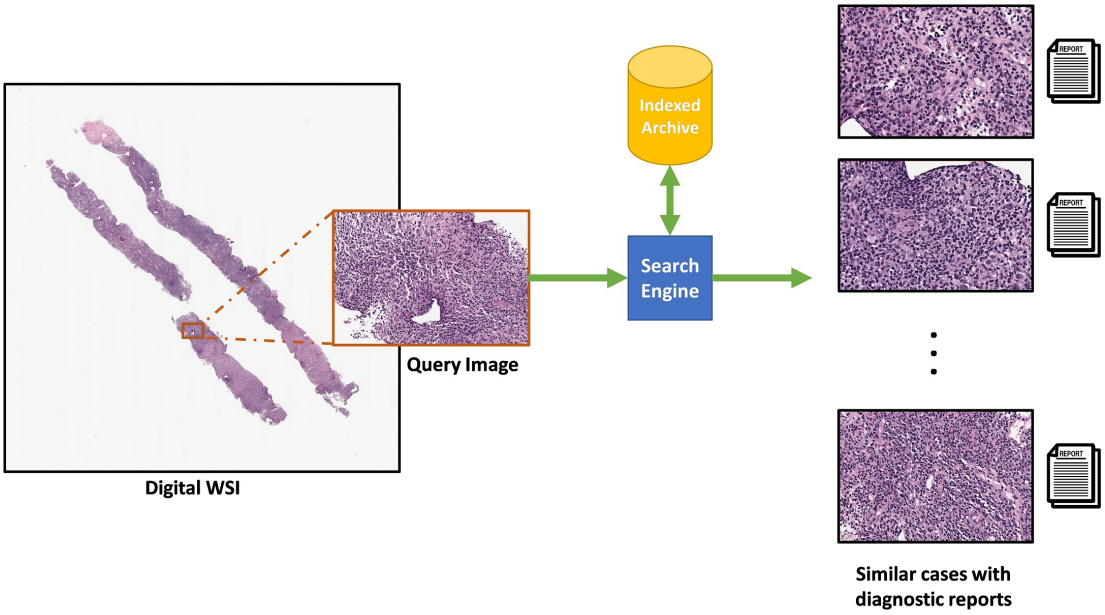

Figure 2.1: CBIR in digital pathology (image re-created using idea from [3]).

The general approach for searching in a medical image database is CBIR is shown in Fig. 2.1.

2.3.1 Opportunities

CBIR systems offer various opportunities for both pathologists and ML researchers which include virtual peer review, transfer learning, and unsupervised learning. CBIR provides pathologists with a virtual second opinion by allowing them to access information from evidently diagnosed cases from the past. Pathologists can diagnose more confidently and rapidly by retrieving similar cases from a large and well-curated database [87]. Using the knowledge contained in a big archive can reduce the overall rate of misdiagnosis, improve diagnosis efficiency by triaging, and reduce the burden on pathology labs [79, 38]. Pathologists can compare new cases with past cases regardless of geographic limitations, thus this technology can be life-saving in remote areas where pathologists are not available [38, 88, 89].

Transfer learning has become quite popular in a CBIR system. It entails applying what a deep network has learned in a field to a different one [90]. In the context of

histopathology, a deep neural network was successfully employed to extract visual information from histopathological slides even though it was trained on a million of natural images [38, 3, 91, 84]. While fine-tuning a pre-trained network improves generalization and saves computational and Graphical Processing Unit (GPU) resources [91].

There is a scarcity of labeled datasets in the pathology domain mainly due to the dependency to highly specialized pathologists and the large size of images. On the other hand, all of the most effective deep neural networks use supervised algorithms that necessitate a large amount of labeled data. Hence, unsupervised learning and clustering may be more useful for the processing of histopathology images. In the literature, Tizhoosh et al. [38] have included some important methods in computational pathology, such as hierarchical clustering. Similarly, "Yottixel" developed a CBIR system for histology archives using a series of unsupervised clustering and feature extraction approaches [3].

CBIR solutions offer substantial utility in augmenting pathologist education by providing an innovative approach to image analysis and interpretation. Leveraging advanced algorithms, CBIR facilitates comparison of medical images based on content, i.e., features [92]. Pathologists can benefit from interactive and dynamic learning experiences, as CBIR enables the exploration of diverse cases, aiding in pattern recognition and diagnostic proficiency [92]. Additionally, the integration of CBIR in educational settings promotes a comprehensive understanding of pathology, enhancing the knowledge base and decision-making skills of pathologists-in-training.

2.3.2 Challenges

Despite the efficiency and intelligence that DP and CBIR technologies bring to the pathology domain, there are a few challenges which include insufficient labeled data, complex diagnostic language, high dimensionality, computational power, demand for storage, and clinical validation.

Large datasets are required to develop a CBIR system based on ML techniques, especially DL models. Curating a large high-quality labeled dataset is not a straightforward task in the medical field, especially for histopathology. The manual annotation of a large number of histopathological images is a time-consuming task that requires domain experts, i.e., pathologists. Moreover, the majority of histopathological slides are only annotated at slide level, making pixel-by-pixel delineations necessary for supervised schemes [38]. Public data banks of labeled histopathology slides can speed up research and standardize field performance evaluation. However, there are just a few public datasets available.

ML techniques are not well suited to the highly complicated and intricate nature of histopathology's diagnosis procedure [93]. Pathologists, on the other hand, utilize highly complicated and evolved vocabulary to express a medical diagnosis, whereas ML methods normally classify images in some discrete set of classes. Therefore, it is difficult to transcribe a diagnostic language to labels that can be effectively used for training a ML model. Pathologists are occasionally unable to classify some challenging cases precisely [94]. Moreover, suspicious and complex cases with confusing patterns can also be divided into subtypes until a more advanced technique can be used to make a final diagnosis.

histopathology WSIs are large images that can reach 100,000 × 100,000 pixels or even more [95]. In fact, a single prostate biopsy can contain anywhere from 12 to 20 biopsy samples, resulting in 2.5–4 billion pixels of data per case (patient) [96, 9]. As a result, analyzing these high-resolution images necessitates resizing or down-sampling operations. Resizing images to a manageable size, on the other hand, can result in the loss of key diagnostic information [97]. Another way to avoid this problem is to *pick a few patches* (tiles) and process each one separately. While this method keeps all features, the spatial layout of the patches may not be well captured. In most cases, even a single patch is down-sampled before being fed into a deep network [38].

In terms of computational power and memory capacity, processing histopathological slides certainly presents challenges. Using such images to train a deep network will almost certainly require the use of GPU, a specialized processor designed to speed up visual operations [98, 99]. Moreover, a reliable data storage server may also be required to handle large images. The requirement of highly sophisticated GPUs, Central Processing Unit (CPU)s, and storage resources for a CBIR system in histopathology images, makes them impractical within real-life clinical settings.

The most challenging part in developing an intelligent real-world image search system is the validation phase [100]. Conducting thorough external validation with enough data (in terms of size and variety) from several hospitals is a key roadblock in the deployment of many ML systems in a real clinical context. In a recent study, ML algorithms that give diagnostic analysis utilizing medical images were evaluated [101]. According to the study, only 6% of 516 published algorithms that were suitable for validation performed well enough. In the medical profession, a broad evaluation is especially important because algorithms may have a higher specificity of a given data-set and may not have generalized well throughout.

2.3.3 CBIR in Digital Pathology

With the improvement in computational resources and DP algorithms, image search was in the spotlight of researchers during the last decade. There is a considerable literature on CBIR in digital pathology [102, 34, 103, 104, 105]. Zheng et al. [106] created an online CBIR system in 2003, in which the client sends the query image and search parameters to the server. The server then conducts similarity searches using vector dot-product as a distance metric, using feature types such as color histogram, image texture, Fourier coefficients, and wavelet coefficients. The server then delivers similar photos, together with similarity scores and feature descriptors, to the query image. On the other hand, Mehta et al. [107] proposed an offline CBIR system that uses sub-images rather than the complete digital slide. When compared to manual search, experimental results suggested that using Scale-invariant Feature Transform (SIFT) [108] to search for similar structures by indexing each sub-image yielded 80 percent accuracy for the top-5 results retrieved from a database containing 50 IHC stained pathology images (IHC) with 8 resolution levels. Akakin and Gurcan created a multi-tiered CBIR system based on WSI in 2012 that can classify and retrieve digital slides using both multi-image queries and images at the slide level [109]. Zhang et al. [110] created a scalable CBIR method to deal with WSI by employing supervised kernel hashing, which compresses a 10,000-dimensional feature vector into only ten binary bits, preserving the image's simple representation.

Most recently, a team from Google AI healthcare department introduced Similar Image Search for Histopathology (SMILY) [84]. Based on both large-scale quantitative analysis using annotated tissue regions and prospective investigations with pathologists blinded to the source of the search results, SMILY retrieves image search results with similar histologic traits, organ site, and cancer grades. An input image is condensed into a feature vector by a pre-trained network. SMILY's network is a deep-ranking network that was pre-trained on 5,000,000 natural photos from 18,000 different classifications. By computing and comparing the embeddings of input images, our network learns to extract discriminative characteristics. SMILY used a dataset manually annotated by pathologists to test the search performance in finding patches with the same histologic features. At 40×, 20×, 10×, and 5× magnification levels, top-5 scores for patch-based searches have been reported. Google employed 400 processors with 10 compute threads to build SMILY as a web-based utility. However, SMILY does not provide a "divide" approach to process WSIs. Given the rapid development of histopathological images, this level of processing cost, namely brute force patch processing of a WSI, is unsustainable for future applications, and a better level of efficiency is required.

In another recent study, Kalra et al. [3] introduced a state-of-the-art search engine

for real-time WSI retrieval in histopathology that for the first time offered a complete "divide & conquer" solution for WSI search. The authors extracted a series of images at 20× magnification from each WSI using an unsupervised color-based and spatial proximity clustering technique. *Mosaic* was the name given to a group of patches that encompassed around 5% of the tissue samples. The Mosaic was then sent to deep CNNs that had already been trained to extract deep features. The feature vectors were then barcoded, i.e., binarized, in order to speed up the indexing of WSIs. The capacity of Yottixel to represent WSI in a compact manner is its most striking feature. The most important performance enabler of Yottixel is the "Bunch of Barcodes" (BoB), a very efficient indexing technique capable of describing WSIs with a mosaic of patches that are subsequently transformed into barcodes.

In 2021, KimiaNet [6] was developed and reported applications for image representation for DP search engines. The DenseNet topology was re-trained in numerous configurations during their research. Without any pathologist annotation or hand designation of regions of interest, the training data were used from a publicly available The Cancer Genome Atlas (TCGA) dataset without any pathologist annotation or manual delineation of regions of interest. Histopathology images at 20× magnification were chosen using a clustering-based approach based on a high-cellularity score. Then, to fine-tune DenseNet, the type of malignancy linked with the WSI was used as the soft label for all extracted images from that WSI. Around 240,000 histopathology images with the size of 1000 × 1000 pixels from more than 7,000 WSIs were selected for training the DenseNet at four different stages. To improve the effectiveness of image search, feature vectors were transformed to binary codes using the *Min-Max barcoding* approach [111, 112, 113] after training the feature extractor and during the test phase [3]. For multi-organ WSI search, KimiaNet was evaluated for image search on three public histopathology datasets.

For H&E-stained histopathology images of malignant lymphoma, Hashimoto et al. [114] offer a novel case-based similar image retrieval method. They apply attention-based multiple-instance learning to compute case similarity while focusing on tumor-specific regions. Additionally, they employ contrastive distance metric learning to incorporate IHC-stain image patterns as supervised data for determining acceptable similarity between various malignant lymphoma cases. In this study, comprising 249 malignant lymphoma patients, they discovered that their proposed method had higher assessment measures than baseline case-based image retrieval systems. Furthermore, subjective examination by pathologists verified that the similarity measure based on IHC staining patterns is appropriate for expressing the similarity of H&E-stained tissue images for malignant lymphoma.

In 2022, Chen et al. [115] published a search method called self-supervised image search for histology (short SISH) based on Yottixel's [3] idea with an additional VQ-VAE-

based [116] as a feature extractor. SISH also incorporates the vEB tree-based [117] indexing and a post-search ranking algorithms. This approach aims to improve the speed and scalability of image retrieval. SISH uses the Yottixel's chain entirely: mosaic, DenseNet and barcoding. However, SISH, as a derivative of Yottixel [118], has several shortcomings: 1) it is not fast because the ranking makes it slow, 2) it is not scalable because it needs much more indexing storage than Yottixel due to additional endcoding, 3) it is not self-supervised, 4) ranking after search makes SISH a patch classifier, hence SISH cannot perform WSI-2-WSI matching. Sikaroudi et al. [118] argue that Chen et al.'s work [115] should not have been babptized as a new method with a new name since it is a mere modification of Yottixel.

In 2023, Wang et al. [119] introduced RetCCL. This approach introduces a novel network for inedxing whole-slide image. RetCCL combines clustering guidance with contrastive learning techniques to improve the retrieval accuracy of whole-slide pathology images. By leveraging cluster information, RetCCL effectively captures image representations that are semantically meaningful, leading to more precise and efficient retrieval of relevant images in digital pathology archives. One has to point to the fact that RetCCL uses Yottixel's mosaic to extract patches. However, as it employs deep features for patch selection - instead of color histogram - RetCCL is very slow. As well, RetCCL uses the post-search ranking proposed by SISH, which makes it, like SISH, a mere pacth classifier that cannot perform true WSI macthing.

2.4 Tissue Segmentation

Medical experts can easily determine the boundaries of the tissue in the WSI. However, because of the presence of color fluctuations, fatty tissues, debris and artefacts, computers may have difficulty identifying tissue regions in WSIs [120, 121]. For the automatic background removal of histopathology images two methods are mainly in focus including Otsu thresholding [122], and pre-trained U-Net [4]. Otsu and U-Net are briefly explained below.

Otsu Thresholding – The Otsu binarization method is a widely used conventional algorithm for separating pixels into foreground and background [4, 122]. For images with only two distinct histogram modes, this hand-crafted image thresholding algorithm performs well. Because an image histogram with only two different modes will only have two peaks, a good threshold will be in the middle of those two values. The Otsu method selects an ideal global threshold value in this fashion, and the same threshold value is applied to each pixel to construct the matching binary mask.

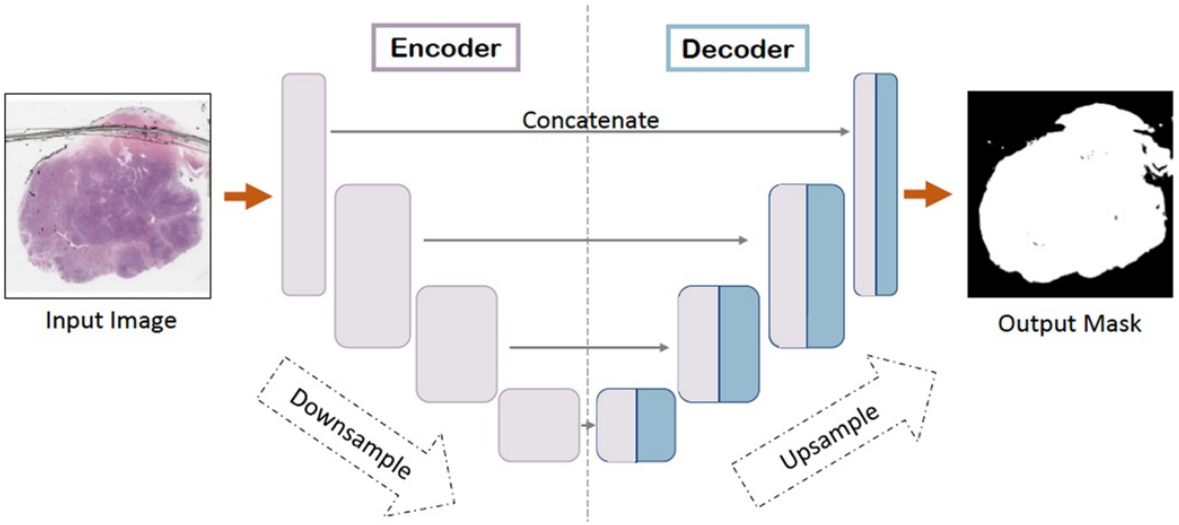

Figure 2.2: U-Net for tissue segmentation (image copied from [4]).

U-Net Segmentation – The U-Net is a fully convolutional network with a U-shape architecture that is divided into two sections, the encoder and the decoder [4, 123]. The tissue segmentation was done using the U-Net architecture with the MobileNet backbone [124]. U-Net has been trained to partition input thumbnail WSI, i.e., WSI at one of its low magnifications, into tissue and non-tissue regions using the MobileNet backbone. The blank background and artifacts such as bubbles, tissue folds, excessive stains, shattered glass, detritus, and marker traces are all non-tissue regions. Figure 2.2 demonstrates U-Net's approach to generating the tissue segmentation.

2.5 Patching & WSI Representation

Searching within gigapixel WSI archives, like many other big-data challenges, necessitates the application of a fundamental computer science approach: the "Divide & Conquer" strategy [118]. The splitting of a WSI into many small patches, i.e., sub-images, is a pivotal step in image analysis and understanding. It plays a crucial role in representing each distinct aspect, feature, or region within the WSI accurately. By breaking down the WSI into meaningful segments or patches, it becomes feasible to analyze and process. Although working with fully-annotated WSIs is desirable, having pixel-level annotations for a large number of WSIs is excessively time-consuming, if not impossible. Therefore, "patching"

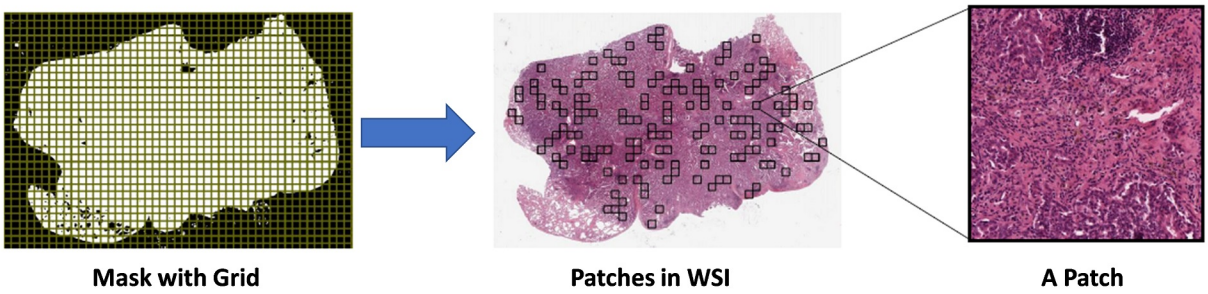

Figure 2.3: Patching from the binary mask and grid.

WSI must be unsupervised. The WSI binary mask is used to differentiate between tissue and the background, which then is used to divide the WSI into a grid to generate patches. In the literature, the Yottixel's mosaic appears to be the only unsupervised patching algorithm. With respect to its significance, it is not clear why there not any other alternative for selecting a small set of representative patches from a WSI. Figure 2.3 shows the generic example to sample patches from a given WSI.

In 2020, Kalra et al. [3] introduced an unsupervised patch extraction scheme "*mosaic*" in which patches extracted from the WSI are organized into a predetermined number of categories or classes through the utilization of a clustering technique, specifically the k-means algorithm. This clustering algorithm operates in an unsupervised manner, autonomously grouping WSI patches into clusters characterized by shared tissue patterns. Subsequently, a fractional subset, typically ranging from 5% to 20%, of patches from each cluster is uniformly selected to construct a mosaic. This mosaic effectively serves as a representative depiction of the entire tissue area within the WSI. The Yottixel's mosaic approach is being used as an unsupervised patch extractor for the WSI search and matching framework proposed by Chen et al. [115] in SISH and Wang et al. [119] in RetCCL. However, Yottixel necessitates predefined empirical parameters, a requirement that may diminish efficiency and elevate redundancy.

2.6 Summary

The review of literature for WSI search and retrieval shows two key issues: 1) we need innovation to address the patching challenge, and 2) we still need to represent the patches with more expressive representations. This research, therefore, attempted to contribute

to both issues. The thesis first introduces Selection of Distinct Morphologies (SDM) algorithm as an alternative for Yottixel's mosaic. To offer a complete solution, the last chapter also introduces a new fine-tuning to generate better deep features for image retrieval applications.

Chapter 3

Selection of Distinct Morphologies to Divide & Conquer the Whole Slide Images

3.1 Introduction

Progress in ML has demonstrated considerable potential in augmenting the efforts of healthcare practitioners [125]. Nevertheless, the adoption and confidence in clinical applications require the development of cutting-edge algorithms that exhibit high accuracy and performance [34]. The emergence of digital pathology has opened new horizons for histopathology [126]. ML algorithms are able to operate on digitized slides to assist pathologists with different tasks. Diverse repositories of digital pathology scans are progressively transitioning from conceptualization to actualization. The volume of data encompassed within these archives is both remarkable and daunting in its scale [3].

The representation of WSIs holds immense importance across a wide spectrum of applications within the fields of pathology, medicine, and beyond. WSIs are essentially high-resolution digital images that capture the entirety of a histopathology glass slide, providing a comprehensive view of tissue specimens under examination. Deep models, such as CNNs, ViTs, and other sophisticated architectures, have been instrumental in extracting meaningful and interpretable features from WSIs, leading to advanced applications. DL representation of WSIs involves the use of neural networks to automatically learn hierarchical and abstract features from the vast amount of visual information contained in these high-resolution images. These learned representations enable computers to understand and

interpret the complex structures and patterns present in histopathological slides. With the DL and meaningful representation of the WSI, the applications are diverse, ranging from automated disease diagnosis and prognosis prediction to drug discovery, telepathology consultations, and search and matching techniques in CBIR.

Second opinions (or *consultations*) in histopathology are of paramount importance as they serve as a crucial quality control measure, enhancing diagnostic accuracy and reducing the risk of misdiagnosis, especially for complex or ambiguous cases [15]. WSI search offers a valuable avenue for obtaining a virtual or computational second opinion. By leveraging advanced CBIR techniques, pathologists can compare a patient's WSI with a database of evidently diagnosed cases, aiding in the identification of similar patterns and anomalies. This approach provides a data-driven, objective perspective that complements the pathologist's evaluation, contributing to more reliable diagnoses and fostering a collaborative and evidence-based approach to pathology. The endeavor of conducting searches within extensive archives of gigapixel WSIs, akin to addressing large-scale big-data challenges, necessitates the implementation of a well-defined computational methodology characterized by the principle of "Divide and Conquer".

Despite the critical role of patch selection as an initial step in the analysis of WSI, this phase has not been extensively investigated. The predominant methods in the literature use brute force patching where the entire WSI is tiled into thousands of patches [127, 128, 129]. Leveraging the entirety of patches extracted from the WSI for retrieval tasks is computationally prohibitive for clinical utility due to the substantial processing resources required. In the literature, a search engine was introduced in 2020 that proposed a sophisticated patching technique called *mosaic* [3]. Yottixel's mosaic functions as a pivotal component during the primary "Divide" stage, effectively partitioning the formidable task of processing WSIs into discrete, manageable parts, with each part symbolized by an individual patch within the mosaic [3]. In the realm of scientific literature, subsequent to Yottixel, two additional search engines, denoted as SISH [115, 118] and RetCCL [119], were introduced in the years 2022 and 2023, respectively. However, both of these search engines used Yottixel's patching scheme to divide the WSI whereas SISH is a slightly modified version of Yottixel [118].

While Yottixel's mosaic method stands as a cutting-edge unsupervised approach for patch selection in the existing scientific literature, it does incorporate certain empirical parameters, including the utilization of 9 clusters for k-means clustering and the selection of 5% to 20% of the total patches within each of the k=9 clusters. These parameters, however, may not comprehensively encompass all the diverse facets and characteristics inherent in the complex tissue morphology of a WSI. Given the intricate nature of tissue morphology in such images, it is plausible that there exist more than nine distinct features

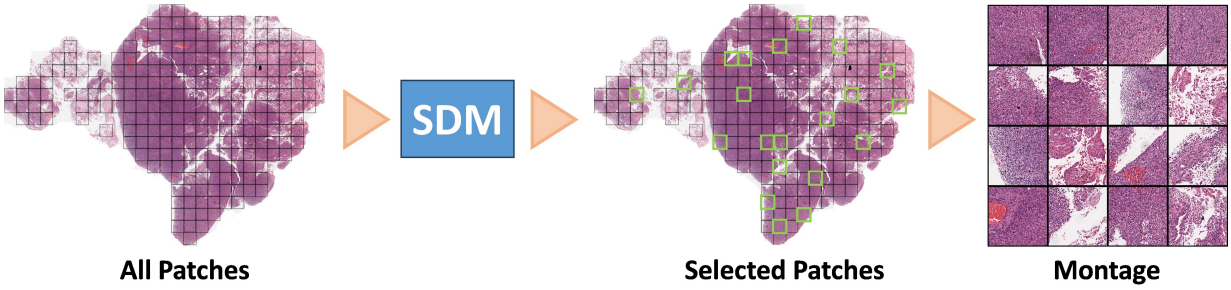

Figure 3.1: **Conceptual Overview.** The overall process to generate a montage from the WSI using SDM.

and aspects that merit consideration. As well, determining the proper level of cluster sampling may not be straightforward. All these considerations underscore the urgent need for the development of a more advanced unsupervised patch selection methodology capable of comprehensively capturing and representing all the diverse aspects and characteristics inherent in a WSI for all types of biopsy.

In this chapter, the initial contribution of this Ph.D. research is introduced — a novel unsupervised patch selection methodology that comprehensively captures the discrete attributes of a WSI without necessitating any empirical input from the user. This methodology is designated as the "Selection of Distinct Morphologies (SDM)", which is further explained in the methods Section. 3.2. Furthermore, the evaluation of the proposed method is described in Section. 3.3 followed by the discussion and conclusion Section. 3.4.

3.2 Methodology

Although it is important to have comprehensive annotations for the WSIs, manual delineations for a large number of WSIs are prohibitively time-consuming or even infeasible. Therefore, in most scenarios, the utilization of *unsupervised patching* becomes inevitable. For this reason, an unsupervised technique is introduced to represent all distinct features of a WSI using fewer patches, termed a "montage". Building such montages serves as a fundamental component crucial for facilitating numerous downstream WSI operations, image search being just one of them. Figure 3.1 shows the steps for producing a montage from a WSI using the proposed SDM method.

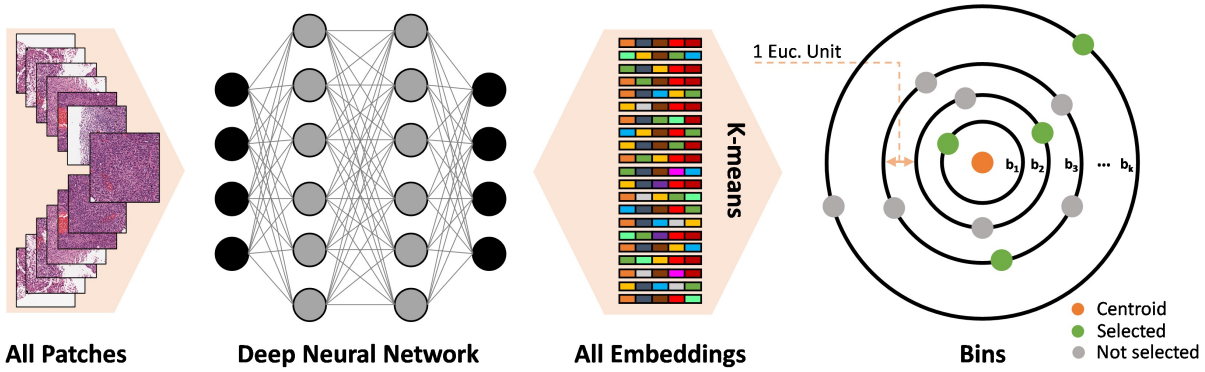

Figure 3.2: **The overall SDM process.** Commencing with the extraction of all patches from the WSI at low magnification (say at 2.5x), these patches subsequently undergo processing through a deep network (say DenseNet [5]), resulting in the generation of embeddings for each patch. After obtaining all embeddings, k-means clustering is applied around a *single centroid*, resulting in the calculation of the Euclidean distance of each patch from the centroid. Patches exhibiting similar Euclidean distances are organized into distinct Euclidean bins. Finally, one patch is selected from each bin to build the montage.

3.2.1 Selection of Distinct Morphologies (SDM)

Patch selection is a fundamental step in digital pathology for many CAD techniques, leading to enhanced diagnostic capabilities and improved patient care. To obtain representative patches that effectively capture the content of a WSI, the SDM framework is introduced in this work. This framework aims to create a "montage" comprising a rather small number of patches that exhibit diversity while maintaining their meaningfulness within the context of the WSI (see Figure 3.1). The algorithm for creating a montage using SDM is outlined in Algorithm 1, and also illustrated in Figure 3.2.

Algorithm 1 Creation of the Montage using SDM

Require: WSI Image

Ensure: Set of selected patches P_s as output

$m \leftarrow$ Set the lower magnification for patching
$s \leftarrow$ Set the patch size at low magnification
$t \leftarrow$ Set a minimum tissue threshold for each patch
$o \leftarrow$ Set the overlap percentage between each adjacent patch
Procedure
$I_m \leftarrow$ OpenWSI(m) ▷ Open the WSI at lower magnification (m)
$M_m \leftarrow$ TissueSegmentation (I_m) ▷ Extract the tissue regions
$T \leftarrow$ Patching (I_m, M_m, s, o) ▷ Perform dense patching with s size and o overlap
for each T **do**
    $G \leftarrow$ TissuePercentage (T) ▷ Calculate tissue percentage for each patch
    $P \leftarrow T$ if $G > t$ ▷ Get the patches with tissue percentage over threshold
end for
$E \leftarrow$ GetEmbeddings(P) ▷ Push the patches P_t through a deep network
$C, D \leftarrow$ k-means(E) ▷ Get the centroid and the Euclidean distances for all the patches
$D_r \leftarrow$ Roundoff(D) ▷ Round off the distances to the nearest integer
$B \leftarrow$ Binned(D_r) ▷ Generate the bin for each integer distance
$P_s \leftarrow B$ ▷ Select a patch from each bin
Return P_s ▷ Return the final selection of distinct patches
End Procedure

Initially, we process the WSI I_m at a low magnification level m, *e.g.*, $m = 2.5\times$. Tissue segmentation is performed to generate a binary tissue mask M_m, for instance using U-Net segmentation [4]. Here, it is presumed that the WSIs that are being processed have been through quality control. Hence, it can be assumed that they are free from artifacts like bubbles, tissue folds, ink markers, and similar imperfections. According to the findings in the literature, a magnification of $2.5\times$ represents the minimum level at which it remains feasible to differentiate between tissue components and artifacts while also retaining some intricate details [4]. Using the tissue mask M_m, dense patching is performed all over the tissue region to extract all the patches with patch size $s_l \times s_l$, and patch overlap o at $2.5\times$. Empirically, we use $s_l = 128$, $2.5\times$ magnification, and $o = 5\%$.

In the literature [3, 34], the patch size of 1024×1024 at $20\times$ magnification is used and thus we use the same. Once a WSI entirely tiled, a subset of patches $P = \{p_1, p_2, \ldots, p_N\}$ with tissue threshold $\geq t$ (*i.e.*, 70%) are selected (here, N is the total number of patches

in subset P). Subsequently, these selected patches P are fed into a deep neural network $f(\cdot)$ to extract the corresponding set of embeddings $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\}$. Empirically, we use DenseNet-121 [5] pre-trained on natural images of ImageNet [130].

Here, the selection of DenseNet [5] is a choice to mitigate any potential bias towards specific histological features (*i.e.*, any properly trained network can be used). The primary goal is to identify various structural elements and edges within the WSI in order to effectively distinguish and capture the multitude of intricate tissue details.

All embedding vectors in E are then used to get one centroid $\mathbf{c}$ embedding vector computed as the mean of embedding vectors $\mathbf{e}_i$, where $\mathbf{e}_i \in E$, and $i = \{1, 2, \ldots, N\}$. $\mathbf{c}$ is computed as

$$\mathbf{c} = \frac{1}{|N|} \sum_{i=0}^{N} \mathbf{e}_i. \tag{3.1}$$

Calculating the mean of the entire dataset, notably in techniques like Principal Component Analysis (PCA) [131], proves advantageous for evaluating variance within the data. Once, we calculated the centroid of the WSI, the set of Euclidean distances $D = \{d_1, d_2, \ldots, d_N\}$ from the centroid is computed for each patch in P. Euclidean distance is measured to quantify the degree of dissimilarity between patches. Individual distances d_i are computed as

$$d_i = \|\mathbf{e}_i - \mathbf{c}\|_2, \tag{3.2}$$

where $d_i \in D$, and here $i = \{1, 2, \ldots, N\}$.

To compute the centroid $\mathbf{c}$, we used the k-means algorithm with only one centroid. Subsequently, these distances D are discretized by rounding them to the nearest integer $r(d_i)$.

Discretized patches that exhibit similar Euclidean distances are grouped together in the set Euclidean bins $B = \{b_1, b_2, \ldots, b_K\}$ since their proximity in terms of Euclidean distance suggests similarity (here, K is the number of Euclidean bins which in turn represents the final number of selected patches). In this process, it is not required to manually specify the number of bins as this is the case for Yottixel's mosaic when it defines the number of clusters. By contrast, K is dynamically determined based on the variability in the Euclidean distances among the patches. This adaptability allows the proposed method to effectively capture diverse numbers of distinct tissue regions within the WSI. Finally,

a single patch is randomly chosen from each Euclidean bin, considering that all patches within the same bin are regarded as similar. Figure 3.3 shows the discrete Euclidean bins and selected patches from each bin. These selected set of patches $P_s = \{p_{s1}, p_{s2}, \ldots, p_{sK}\}$ constitute distinct patches called WSI's *montage*.

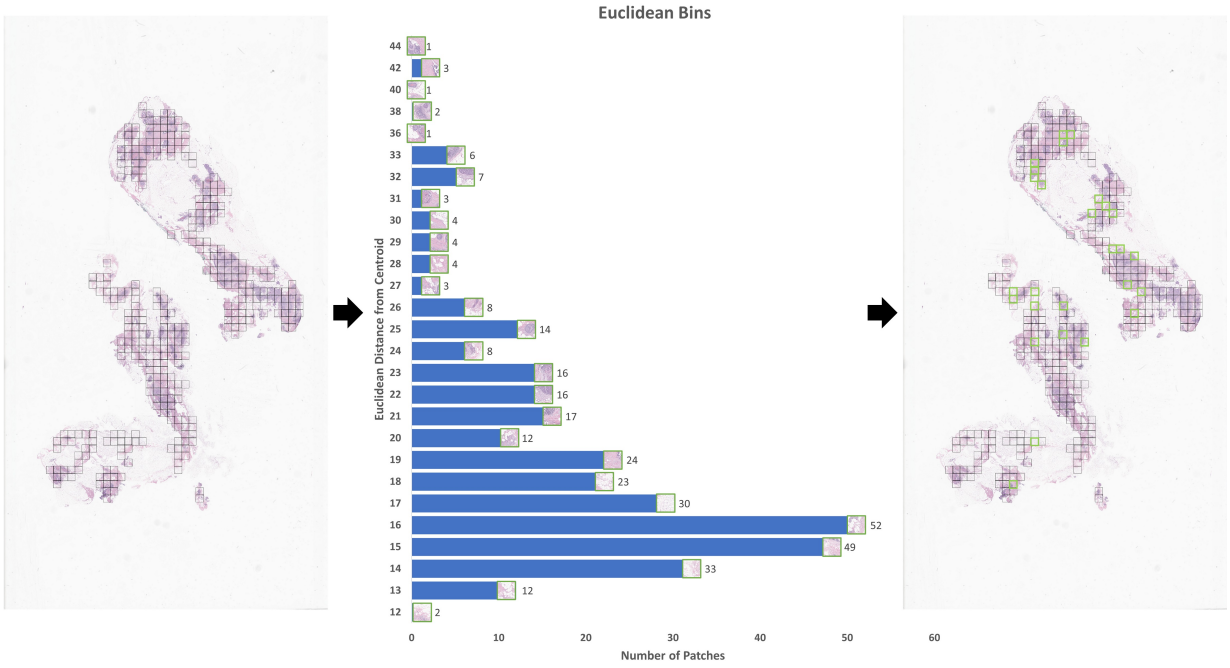

Figure 3.3: **Discrete Euclidean bins within SDM.** The bar chart visually represents the distribution of patches from the WSI across various Euclidean bins. Patches grouped within the same Euclidean bin exhibit similarity. Randomly selected patches (displayed at the top of each bin) represent the montage.

3.2.2 Atlas for WSI Matching

After identifying a unique set of patches from the WSI at a lower magnification level (say 2.5×), these patches are subsequently extracted at higher magnification (say 20×) with a patch size of 1024 × 1024 pixels. This process generates a montage that contains fewer patches than contained in WSI. This approach enhances computational efficiency and minimizes storage space requirements for subsequent processing without compromising the distinct information in the WSI. The patches in a montage are converted to a set of

barcodes using the MinMax algorithm [112, 3]. To achieve this, the patches are initially converted into feature vectors using KimiaNet [6], which is a DenseNet-121 [5] model trained on histological data from TCGA. Global average pooling is applied to the feature maps obtained from this last convolutional layer, resulting in a feature vector with a dimension of 1024. Following feature extraction, we employ the discrete differentiation of the MinMax algorithm [112, 3], to convert the feature vectors into binary representations known as a "barcode". This barcode is lightweight and enables rapid Hamming distance-based searches [3]. While it's possible to directly assess image similarity using deep features and metrics like the Euclidean distance, there is a notable concern regarding computational and storage efficiency, particularly when conducting searches within a large databases spanning various primary sites. While the utilization of Hamming distance for barcodes is acknowledged here, it is noteworthy to consider exploring alternative approaches such as data compression techniques, including but not limited to Huffman encoding or Vector Quantization [132]. Following the processing and binarization of all WSIs using the SDM method, the resulting barcodes are preserved as a reference "atlas" (structured database of patients with known outcomes). This atlas can subsequently be employed for the matching process when handling new patients, enabling efficient search applications.

Matching WSI to one another poses significant challenges due to various factors. One key challenge arises from the inherent variability in the number of patches extracted from different WSIs. Since WSIs can vary in size and complexity, the number of patches derived from them can differ substantially. Additionally, factors such as variations in tissue preparation, staining quality, and imaging conditions can introduce further complexity. All these factors make it challenging to establish WSI-to-WSI matching, requiring sophisticated computational methods to address these variations and ensure robust matching in histopathological analysis. To overcome this challenge, Kalra et al. [3] introduced a novel approach called the "*median of minimum*" distances within the search engine Yottixel. This technique aims to enhance the robustness of WSI-to-WSI matching. It does so by considering the minimum distances between patches in two WSIs and then selecting the median of these minimum values as a cumulative measure of WSI similarity (see Figure 3.4). In this study, the median-of-minimum method has been adopted to perform WSI-to-WSI matching within the atlas.

3.3 Evaluation & Results

The verification and validation of histological similarity represent formidable challenges. A comprehensive validation scenario would ideally entail the comparison of numerous pa-

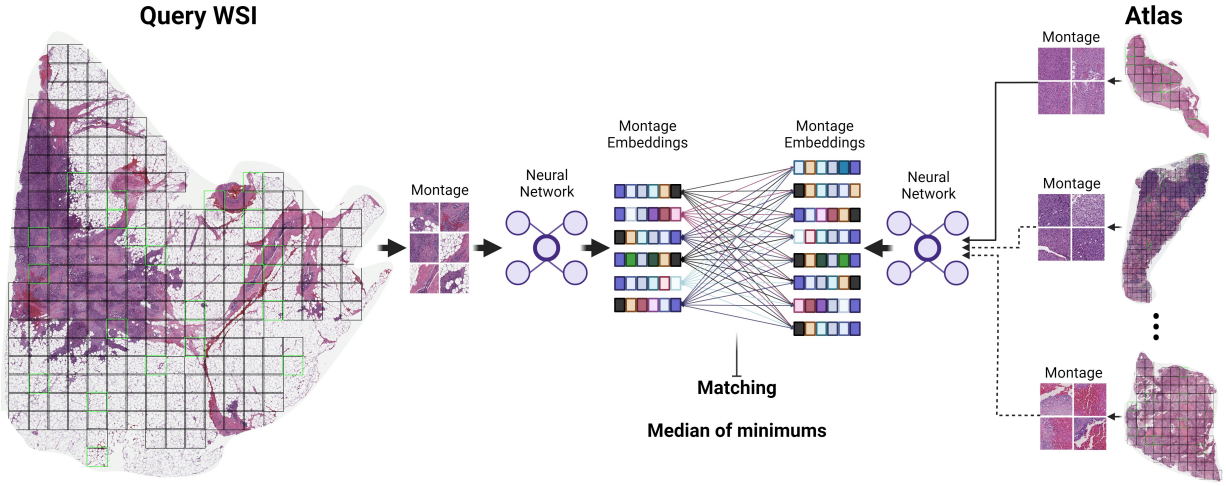

Figure 3.4: **WSI-Level Search**. The process involves matching one WSI to another using the *median of minimum distances* [3]. For each query WSI, its patch embeddings are compared with the patch embeddings of every WSI in the archive.

tients across diverse healthcare institutions, involving multiple pathologists conducting visual inspections over an extended timeframe. In this research, the performance of the search task was quantified by approaching it as a classification problem for simplification purposes. One of the primary advantages of employing classification methodologies lies in their ease of validation; each image can be categorized as either belonging to a specific class or not, a binary concept that allows for performance quantification through tallying misclassified instances. Nonetheless, it's essential to acknowledge that the notion of similarity in image search is a fundamentally continuous subject matter (in many cases, a straightforward yes/no answer may not suffice) and predominantly a matter of degree (ranging from *almost identical* to *utterly dissimilar*). Moreover, distance measures, such as Euclidean distance, which assesses dissimilarity between two feature vectors representing images, are typically used to gauge the extent of similarity (or dissimilarity) between images. The classification-based assessments we employ may tend to be overly cautious when evaluating search outcomes and hence may overlook shared anatomical traits among various tumor types.

SDM montage has been extensively evaluated on various public and private histopathology datasets using a "leave-one-out" WSI search and matching as a downstream task on each dataset and compared with the state-of-the-art Yottixel's mosaic. For public dataset evaluation, the following datasets have been used: TCGA [133], BReAst Carcinoma Subtyping (BRACS) [134], and Prostate cANcer graDe Assessment (PANDA) (PANDA) [135].

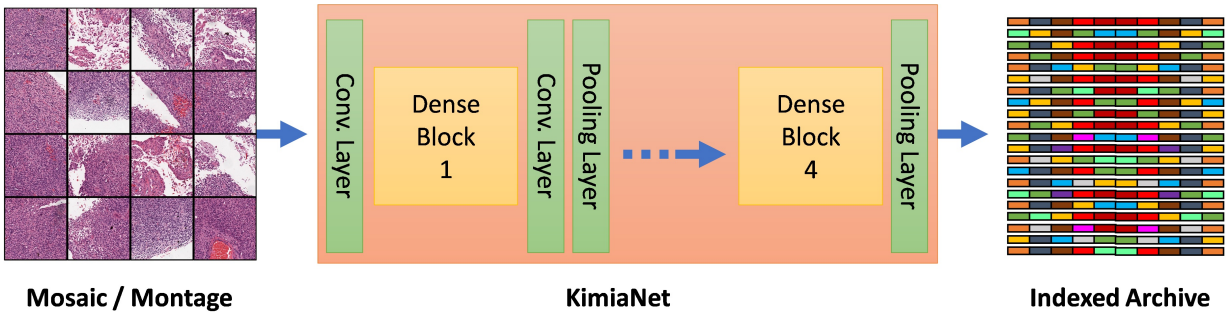

Figure 3.5: Feature extraction to generate the atlas (indexed archive) after pushing the selected patches (Yottixel's Mosaic and SDM's Montage) through the KimiaNet [6].

On the other hand, for the private dataset evaluation, we have used Breast Cancer (BC), Alcoholic Steatohepatitis (ASH) and Non-alcoholic Steatohepatitis (NASH) Liver, and Colorectal Cancer (CRC) datasets from Mayo Clinic, Rochester, USA.

All experiments have been conducted on Dell PowerEdge XE8545 with 2× AMD EPYC 7413 CPUs, 1023 GB RAM, and 4× NVIDIA A100-SXM4-80GB using TensorFlow (TF) version of DenseNet and KimiaNet. We used TF 2.12.0, python 3.9.16, CUDA 11.8, and CuDNN 8.6 on a Linux operating system. In all experimental procedures, two distinct patch selection methodologies were employed: Yottixel's mosaic and SDM's montage. Subsequently, patches were extracted at a 20× magnification, with dimensions measuring 1000 × 1000 for the mosaic and 1024 × 1024 for the montage. Here, a patch size of 1024 for montage is selected due to its favorable memory management properties, originating from its status as a power of 2 and its divisibility by 16. This particular size facilitates computational efficiency and aligns with architectural requirements, particularly for Vision Transformers (ViTs), thereby ensuring future compatibility and optimized processing. Following patching, feature extraction was executed using KimiaNet (DenseNet121 trained on TCGA data) [6]. These features were subsequently transformed into barcodes, characterized by their lightweight nature and ability to facilitate swift Hamming distance-based searches [112, 3]. Figure 3.5 depicts the comprehensive sequence of operations encompassing feature extraction and the subsequent creation of an atlas. This atlas functions as a fundamental asset, tested via a "leave-one-out" search and matching experiment, a notably rigorous method particularly suited for datasets of small to medium size, with the aim of retrieving the highest-ranking matching WSIs. The computer vision literature typically emphasizes **top-n** accuracy, where success is determined if any one of the top-n search results is accurate. In contrast, our approach relies on "**majority-n** accuracy", which we find to be a significantly more dependable validation scheme for medical imaging [3, 34].

Under this scheme, a search is deemed correct only when the majority of the top-n search results are accurate. The advantage of a search process lies in its capability to retrieve multiple top-matching results, enabling the opportunity to examine these foremost matches to explain the required decision.

Once, the top matching results (through majority voting) are compiled, then the most commonly used evaluation metrics for verifying the performance of image search and CBIR algorithms are precision, recall, and F1-score [136, 3, 6, 137]. The harmonic mean of precision and recall is the F1-score. The F1-score is considered more reliable than accuracy, especially in scenarios where there is an imbalance between the classes. Accuracy, which measures the overall correctness of predictions, may be misleading when dealing with imbalanced datasets, where one class significantly outweighs the other. The F1-score, on the other hand, combines precision and recall, providing a balanced measure that considers both false positives and false negatives. F1-score is defined as:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3.3}$$

Here, a concise overview of the performance of the proposed method is provided using results from three datasets: TCGA 3.3.1, CRC - Mayo Clinic 3.3.2, and Breast Cancaer - Mayo Clinic 3.3.3. Detailed results from the other three datasets (BRACS A.1.2, PANDA A.1.3, and Liver - Mayo Clinic A.1.5) are included in the Appendix A.

3.3.1 Public – The Cancer Genome Atlas (TCGA)

TCGA is a public and comprehensive repository in the field of cancer research. Established by the National Institutes of Health (NIH) and the National Cancer Institute (NCI), TCGA represents a collaborative effort involving numerous research institutions. Its primary mission is to analyze and catalog genomic and clinical data from a wide spectrum of cancer types. It is the largest publicly available dataset for cancer research. The dataset contains 25 anatomic sites with 32 cancer subtypes of almost 33,000 patients.

The KimiaNet [4] underwent a training process utilizing the TCGA dataset, using the ImageNet weights from DenseNet as initial values. This process involved the utilization of 7,375 diagnostic H&E slides to extract a substantial dataset of over 240,000 patches, each with dimensions measuring 1000×1000, for training KimiaNet. Additionally, a set of 1553 slides was set aside for evaluation purposes, comprising a test dataset consisting of 777 slides and a validation dataset encompassing 776 slides.

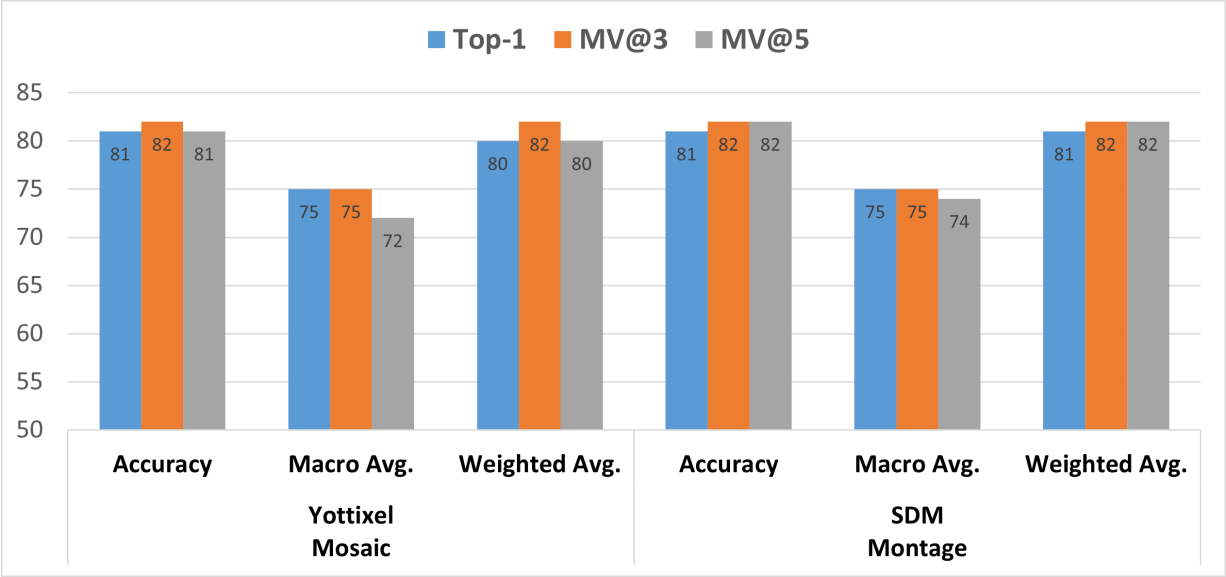

Figure 3.6: Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the TCGA dataset. The diagram shows comparable performance of SDM montage against Yottixel when comparing top-1 and MV@3 retrievals. Hoverver, SDM performs marginally better than Yottixel when comparing MV@5 retrievals.

From 1553 evaluation slides that were not involved in the fine-tuning of KimiaNet [6], 1466 were used in the evaluation of this study (see Table. 3.1 for a detailed breakdown of the dataset).

To assess how the performance of SDM montage compares to Yottixel's mosaic, a leave-one-out evaluation was conducted to retrieve the most similar cases. The evaluation involved multiple retrieval criteria, including the top-1 retrieval, the majority agreement among the top 3 retrievals (MV@3), and the majority agreement among the top 5 retrievals (MV@5). The accuracy, macro average, and weighted average at top-1, MV@3, and MV@5 are reported in Figure 3.6. Moreover, confusion matrices and chord diagrams at MV@5 are illustrated in Figure 3.7. Additional confusion matrices and chord diagrams of top-1, and MV@3 retrievals are provided in Appendix A (see Figure A.1, and A.2, respectively). A **chord diagram** serves as a graphical depiction presenting the interrelations and connections among data points arranged in a circular layout. Notably, its structure bears resemblance to that of confusion matrices, with the distinction that confusion matrices

adopt a square layout. It is particularly effective in illustrating the connections or flows between entities or categories. The diagram consists of arcs, or "chords", that connect pairs of points, with the width of each chord representing the strength or frequency of the connection. In our evaluation, chords that are matching with another subtype represents the mismatch of of the subtype. Ideally we should not have any cross connections to say that our search results match completely with the query. However, tissue similarity may create cross connections that are anatomically correct. Table 3.2 and 3.3 show the detailed results including precision, recall, and F1-score for Yottixel's mosaic and SDM's montage, respectively. In addition to conventional accuracy metrics, we also conducted a comparative analysis of the number of patches extracted per WSI by each method (see the boxplots in Figure 3.8 for the depiction of the patch distribution per WSI). To visually represent the extracted patches, t-distributed Stochastic Neighbor Embedding (t-SNE) projections of these patches are also provided in Figure 3.9.

Through this experiment, we observed that SDM exhibited comparable performance to the Yottixel mosaic concerning top-1 retrieval and the majority agreement among the top 3 retrievals. However, notably, the SDM montage demonstrated marginally superior performance by +2% in macro avg. of F1-scores, and +1% in accuracy and weighted avg. of F1-scores as compared to the Yottixel mosaic when it came to the majority agreement among the top 5 retrievals, highlighting its effectiveness in capturing relevant information in this specific retrieval context (see Figure 3.6). Another notable advantage of employing the SDM montage method becomes evident when examining Figure 3.8, which illustrates the number of patches selected. In comparison to the Yottixel mosaic, SDM proves to be more efficient by selecting a fewer number of patches while having comparable performance. This not only conserves storage space but also eliminates the redundancy & need for empirical determination of the ideal number of patches to select. Additionally, it has come to our attention that Yottixel is more prone to overlooking WSIs in comparison to SDM. Specifically, our observations reveal that Yottixel processed 1462 WSIs, whereas SDM successfully processed the entirety of 1466 WSIs. Finally, the t-SNE map, in Figure 3.9, shows that SDM have more discernible pattern than the Yottixel's extracted patches.

Primary Diagnoses	Acronym	Slides
Adrenocortical Carcinoma	ACC	11
Bladder Urothelial Carcinoma	BLCA	68
Brain Lower Grade Glioma	BLGG	79
Breast Invasive Carcinoma	BRCA	178
Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	CESC	39
Cholangiocarcinoma	CHOL	8
Colon Adenocarcinoma	COAD	62
Esophageal Carcinoma	ESCA	28
Glioblastoma Multiforme	GBM	70
Head and Neck Squamous Cell Carcinoma	HNSC	63
Kidney Chromophobe	KICH	22
Kidney Renal Clear Cell Carcinoma	KIRC	99
Kidney Renal Papillary Cell Carcinoma	KIRP	53
Liver Hepatocellular Carcinoma	LIHC	70
Lung Adenocarcinoma	LUAD	74
Lung Squamous Cell Carcinoma	LUSC	84
Mesothelioma	MESO	9
Ovarian Serous Cystadenocarcinoma	OV	20
Pancreatic Adenocarcinoma	PAAD	24
Pheochromocytoma and Paraganglioma	PCPG	30
Prostate Adenocarcinoma	PRAD	77
Rectum Adenocarcinoma	READ	21
Sarcoma	SARC	26
Skin Cutaneous Melanoma	SKCM	49
Stomach Adenocarcinoma	STAD	55
Testicular Germ Cell Tumors	TGCT	26
Thymoma	THYM	6
Thyroid Carcinoma	THCA	101
Uterine Carcinosarcoma	UCS	6
Uveal Melanoma	UVM	8

Table 3.1: Comprehensive details regarding the TCGA dataset utilized in this study, encompassing the corresponding acronyms and the number of slides attributed to each primary diagnosis.

Yottixel Mosaic

	Top-1			MV@3			MV@5			
Primary Diagnoses	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Slides
ACC	0.86	0.55	0.67	0.83	0.45	0.59	0.80	0.36	0.50	11
BLCA	0.68	0.85	0.76	0.63	0.84	0.72	0.61	0.84	0.70	68
BLGG	0.90	0.87	0.88	0.88	0.87	0.88	0.86	0.85	0.85	79
BRCA	0.91	0.94	0.93	0.92	0.93	0.92	0.88	0.96	0.92	178
CESC	0.78	0.46	0.58	0.87	0.51	0.65	0.88	0.59	0.71	39
CHOL	0.45	0.62	0.53	0.50	0.50	0.50	0.33	0.12	0.18	8
COAD	0.64	0.68	0.66	0.70	0.73	0.71	0.71	0.79	0.75	62
ESCA	0.45	0.50	0.47	0.52	0.43	0.47	0.44	0.39	0.42	28
GBM	0.84	0.88	0.86	0.84	0.86	0.85	0.80	0.83	0.81	66
HNSC	0.79	0.71	0.75	0.84	0.78	0.81	0.82	0.73	0.77	63
KICH	0.90	0.86	0.88	1.00	0.86	0.93	1.00	0.82	0.90	22
KIRC	0.87	0.90	0.89	0.89	0.95	0.92	0.85	0.95	0.90	99
KIRP	0.78	0.79	0.79	0.84	0.81	0.83	0.83	0.74	0.78	53
LIHC	0.90	0.81	0.86	0.85	0.83	0.84	0.85	0.83	0.84	70
LUAD	0.75	0.72	0.73	0.77	0.72	0.74	0.78	0.68	0.72	74
LUSC	0.72	0.76	0.74	0.73	0.86	0.79	0.70	0.85	0.77	84
MESO	0.40	0.22	0.29	0.50	0.11	0.18	0.00	0.00	0.00	9
OV	0.80	0.80	0.80	0.80	0.80	0.80	0.84	0.80	0.82	20
PAAD	0.60	0.62	0.61	0.62	0.62	0.62	0.61	0.58	0.60	24
PCPG	0.93	0.90	0.92	0.93	0.90	0.92	0.82	0.90	0.86	30
PRAD	0.92	0.95	0.94	0.94	0.95	0.94	0.94	0.95	0.94	77
READ	0.18	0.19	0.19	0.32	0.29	0.30	0.30	0.14	0.19	21
SARC	0.77	0.77	0.77	0.77	0.77	0.77	0.80	0.77	0.78	26
SKCM	0.95	0.78	0.85	0.88	0.76	0.81	0.83	0.71	0.77	49
STAD	0.66	0.71	0.68	0.68	0.78	0.73	0.69	0.78	0.74	55
TGCT	0.95	0.81	0.88	0.96	0.85	0.90	0.95	0.81	0.88	26
THYM	1.00	0.67	0.80	1.00	0.67	0.80	1.00	0.50	0.67	6
THCA	0.94	0.97	0.96	0.94	0.98	0.96	0.97	0.99	0.98	101
UCS	0.67	1.00	0.80	0.67	1.00	0.80	0.67	1.00	0.80	6
UVM	1.00	0.88	0.93	1.00	0.88	0.93	1.00	0.88	0.93	8
Total Slides										**1462**

Table 3.2: Detailed precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using the Yottixel mosaic. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the TCGA dataset.

SDM Montage

	Top-1			MV@3			MV@5			
Primary Diagnoses	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Slides
ACC	0.89	0.73	0.80	0.80	0.73	0.76	0.80	0.73	0.76	11
BLCA	0.74	0.79	0.77	0.68	0.82	0.75	0.66	0.87	0.75	68
BLGG	0.82	0.82	0.82	0.86	0.87	0.87	0.83	0.85	0.84	79
BRCA	0.89	0.96	0.92	0.88	0.98	0.93	0.84	0.99	0.91	178
CESC	0.80	0.62	0.70	0.66	0.54	0.59	0.75	0.62	0.68	39
CHOL	0.50	0.25	0.33	0.67	0.25	0.36	0.00	0.00	0.00	8
COAD	0.71	0.79	0.75	0.70	0.77	0.73	0.71	0.84	0.77	62
ESCA	0.43	0.43	0.43	0.54	0.50	0.52	0.58	0.54	0.56	28
GBM	0.80	0.81	0.81	0.86	0.84	0.85	0.81	0.80	0.81	70
HNSC	0.82	0.78	0.80	0.83	0.76	0.79	0.88	0.79	0.83	63
KICH	0.95	0.82	0.88	0.95	0.91	0.93	1.00	0.86	0.93	22
KIRC	0.91	0.90	0.90	0.93	0.94	0.93	0.88	0.95	0.91	99
KIRP	0.75	0.83	0.79	0.79	0.83	0.81	0.84	0.79	0.82	53
LIHC	0.82	0.80	0.81	0.84	0.81	0.83	0.84	0.84	0.84	70
LUAD	0.76	0.73	0.74	0.71	0.74	0.73	0.75	0.74	0.75	74
LUSC	0.72	0.75	0.74	0.77	0.76	0.77	0.78	0.77	0.78	84
MESO	0.67	0.22	0.33	1.00	0.11	0.20	1.00	0.11	0.20	9
OV	0.88	0.75	0.81	0.84	0.80	0.82	0.83	0.75	0.79	20
PAAD	0.64	0.58	0.61	0.60	0.50	0.55	0.68	0.54	0.60	24
PCPG	0.90	0.87	0.88	0.93	0.83	0.88	0.96	0.83	0.89	30
PRAD	0.93	0.96	0.94	0.95	0.96	0.95	0.94	0.96	0.95	77
READ	0.31	0.24	0.27	0.31	0.19	0.24	0.33	0.19	0.24	21
SARC	0.83	0.73	0.78	0.86	0.69	0.77	0.90	0.73	0.81	26
SKCM	0.80	0.82	0.81	0.86	0.76	0.80	0.87	0.67	0.76	49
STAD	0.74	0.84	0.79	0.72	0.84	0.77	0.74	0.82	0.78	55
TGCT	0.81	0.81	0.81	0.85	0.85	0.85	0.88	0.85	0.86	26
THYM	0.80	0.67	0.73	1.00	0.67	0.80	1.00	0.33	0.50	6
THCA	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	101
UCS	0.75	1.00	0.86	0.75	1.00	0.86	0.75	1.00	0.86	6
UVM	1.00	0.88	0.93	1.00	0.88	0.93	1.00	0.88	0.93	8
Total Slides										**1466**

Table 3.3: Detailed precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using the SDM Montage. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA dataset.

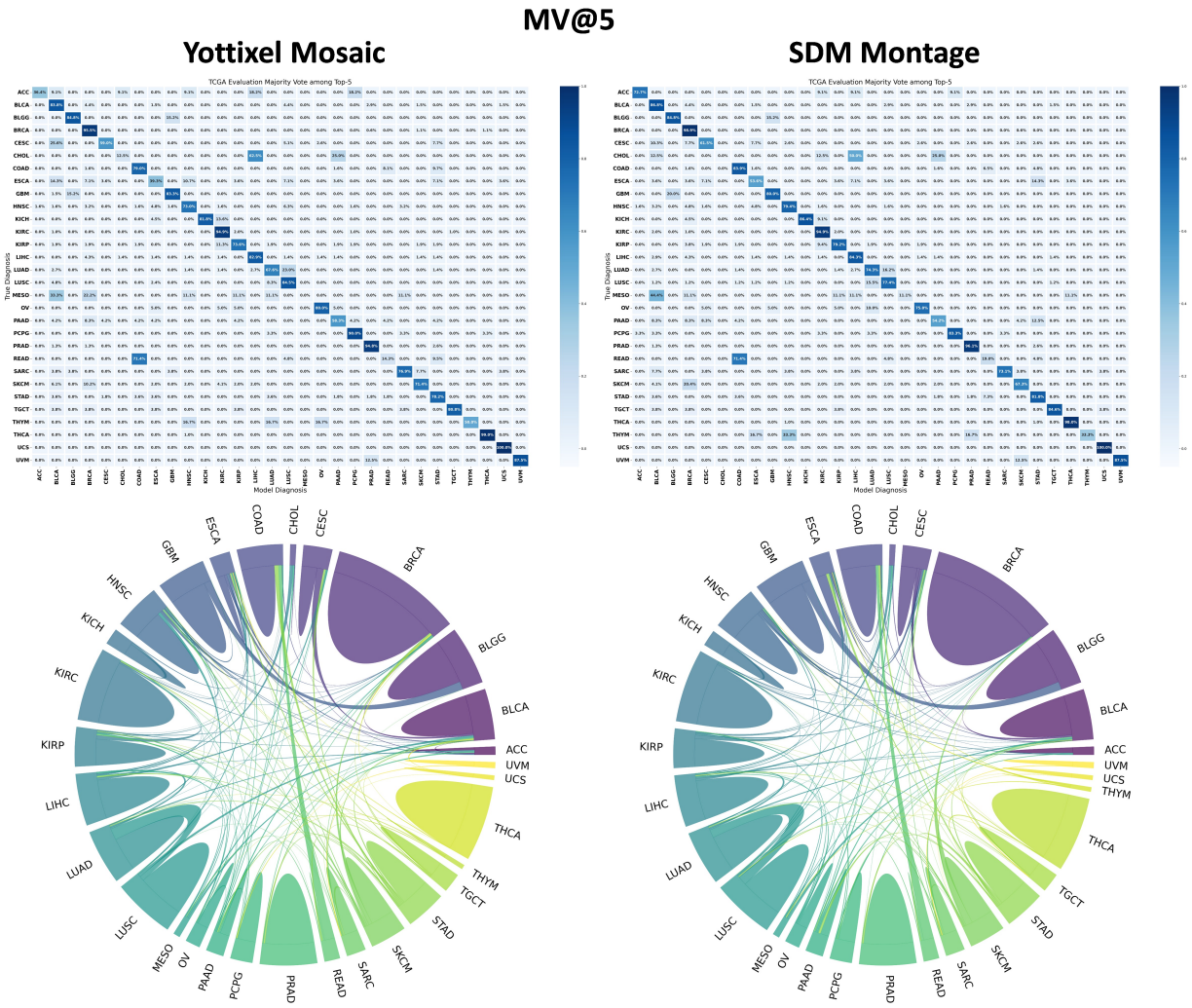

Figure 3.7: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the TCGA dataset.

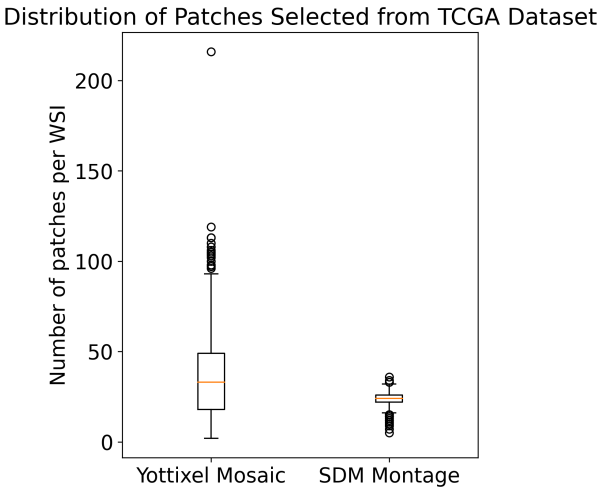

Figure 3.8: The boxplot illustrates the distribution of patches selected for each WSI in the TCGA dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 33 ± 21. Conversely, for the SDM Montage, the median number of selected patches is 24 ± 4. Here, SDM selects significantly fewer patches than Yottixel.

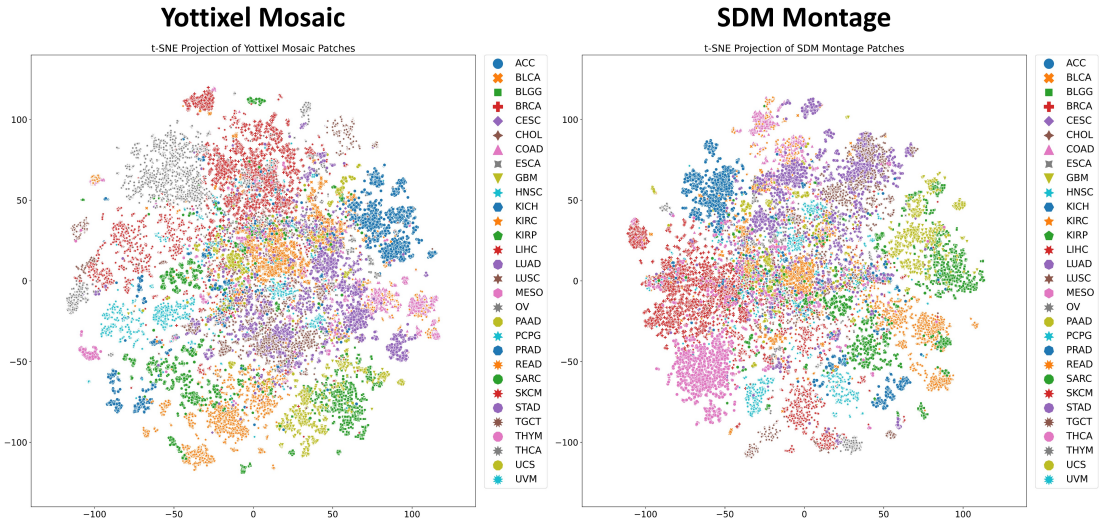

Figure 3.9: The t-SNE projection displays the embeddings of all patches extracted from the TCGA dataset using Yottixel's mosaic (left) and SDM's montage (right).

3.3.2 Private – Colorectal Cancer (CRC)

The Colorectal Cancer (CRC) dataset, sourced from Mayo Clinic, Rochester, USA, encompasses a collection of 209 WSIs, with a primary focus on colorectal histopathology. This dataset is categorized into three distinct groups, specifically Cancer Adjacent polyps (CAP), Non-recurrent polyps (POP-NR), and Recurrent polyps (POP-R), all of which pertain to colorectal pathology. Importantly, all the slides in this dataset were subjected to scanning at a magnification level of 40x (see Table 3.4 for more details).

Primary Diagnoses	Acronyms	Slides
Cancer Adjacent Polyps	CAP	63
Non-recurrent Polyps	POP-NR	63
Recurrent Polyps	POP-R	83

Table 3.4: Comprehensive dataset particulars pertaining to the Colorectal Cancer dataset utilized in this experiment, encompassing relevant acronyms and the number of slides attributed to each primary diagnosis.

To assess the effectiveness of the SDM montage in comparison to Yottixel's mosaic, we conducted a leave-one-out evaluation to retrieve the most similar cases using the CRC dataset. The evaluation criteria encompass multiple retrieval scenarios, including the top-1 retrieval, the majority consensus among the top 3 retrievals (MV@3), and the majority consensus among the top 5 retrievals (MV@5). The results, including accuracy, macro average, and weighted average scores at the top-1, MV@3, and MV@5 levels, are presented in Figure 3.10. Table 3.5 shows the detailed statistical results including precision, recall, and F1-score. Moreover, confusion matrices and chord diagrams at MV@5 are shown in Figure 3.11. Additional confusion matrices and chord diagrams of Top-1, and MV@3 retrievals are provided in Appendix A (see Figure A.15, and A.16 respectively). In addition to the traditional accuracy metrics, we conducted a comparative examination of the number of patches extracted per WSI by each individual method. For a visual depiction of this distribution across the complete dataset, we refer to the boxplots provided in Figure 3.12. To visually illustrate the extracted patches, we used t-SNE projections, as demonstrated in Figure 3.13.

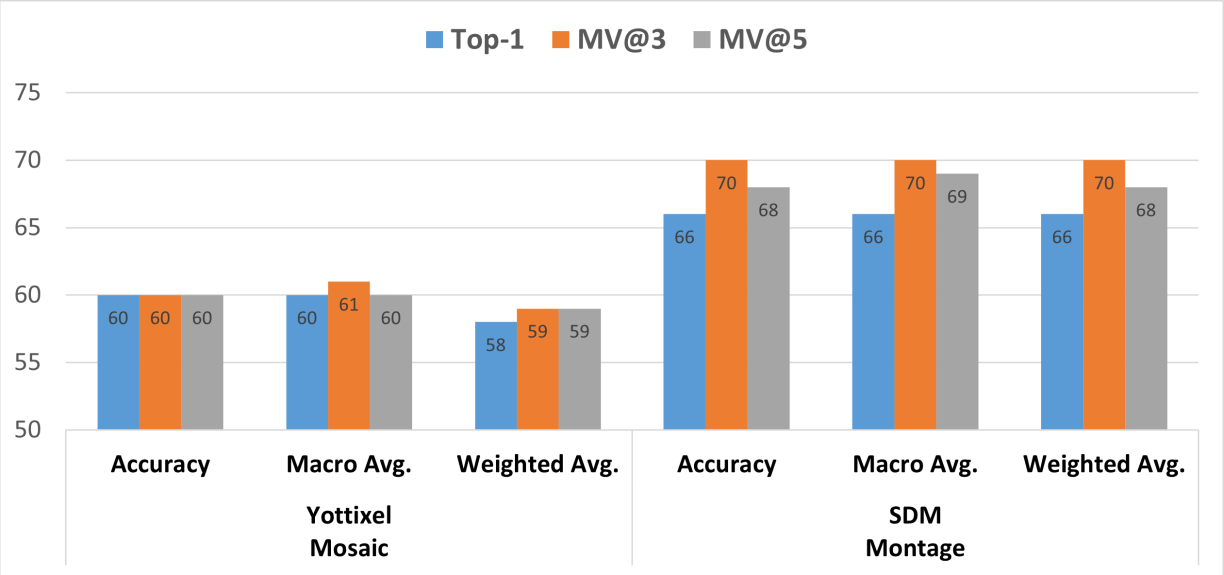

Figure 3.10: Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the CRC dataset. The diagram shows that SDM montage significantly outperforms Yottixel's mosaic.

		Top-1			**MV@3**			**MV@5**			
	Primary Diagnoses	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Slides
Yottixel Mosaic	CAP	0.75	0.71	0.73	0.72	0.70	0.71	0.77	0.79	0.78	63
	POP-NR	0.52	0.81	0.63	0.53	0.78	0.63	0.51	0.76	0.61	63
	POP-R	0.58	0.35	0.44	0.59	0.40	0.47	0.56	0.34	0.42	83
Total Slides											**209**
SDM Montage	CAP	0.77	0.70	0.73	0.81	0.79	0.80	0.81	0.79	0.80	63
	POP-NR	0.64	0.68	0.66	0.67	0.67	0.67	0.67	0.65	0.66	63
	POP-R	0.59	0.60	0.60	0.64	0.65	0.65	0.60	0.63	0.62	83
Total Slides											**209**

Table 3.5: Precision, recall, F1-score, and the number of slides processed for each sub-type are shown in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the CRC dataset.

During our experimentation, the SDM montage manifested a marked performance su-

periority over the Yottixel mosaic. Specifically, we observed enhancements in the macro-average of F1-scores by +6%, +9%, and +9% for top-1 retrieval (top-1), majority consensus within the top 3 retrievals (MV@3), and majority consensus within the top 5 retrievals (MV@5), respectively. From an accuracy perspective, the SDM method demonstrated increments of +6%, +10%, and +8% for the top-1, MV@3, and MV@5 retrievals, respectively. These results emphasize the SDM method's adeptness in assimilating and representing critical data effectively within the retrieval paradigm, as delineated in the referenced Figure 3.10. Furthermore, an additional noteworthy benefit of implementing the SDM montage method comes to the forefront when examining Figure 3.12, which depicts the number of selected patches. In contrast to the Yottixel mosaic, SDM proves to be more resource-efficient by opting for a smaller patch selection. This not only leads to storage conservation but also eliminates the redundancy and the necessity for an empirical determination of the optimal patch count to select.

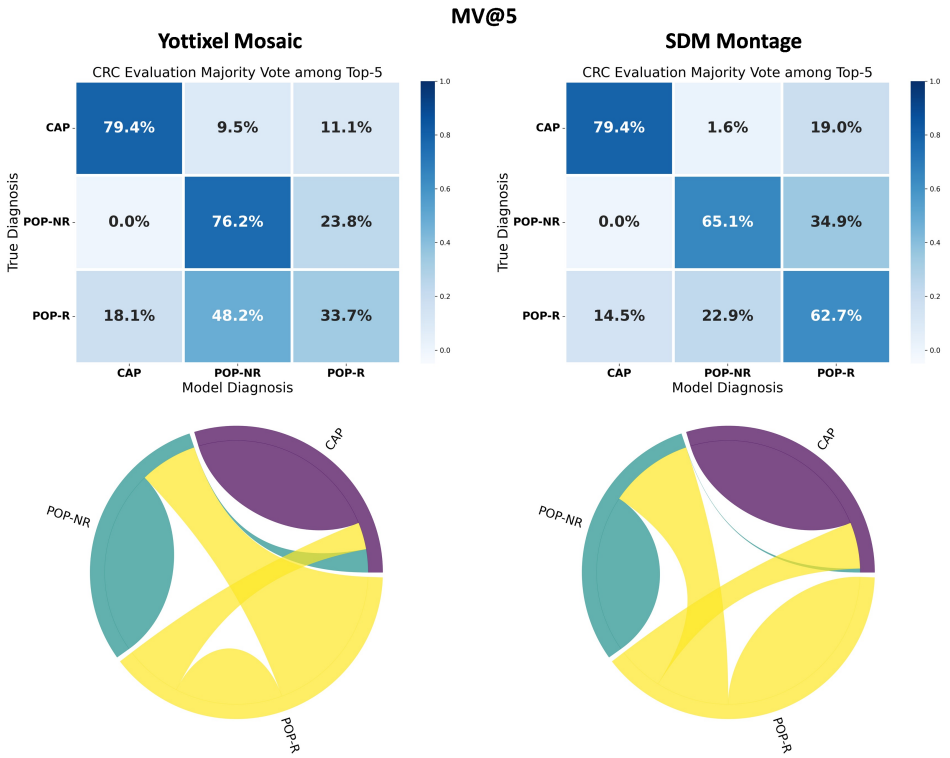

Figure 3.11: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the CRC dataset.

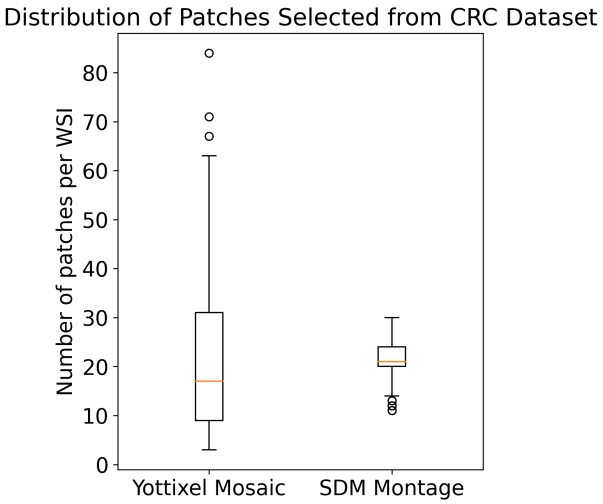

Figure 3.12: The boxplot illustrates the distribution of patches selected for each WSI in the CRC dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 17 ± 15. On the other hand, for the SDM Montage, the median number of selected patches is 21 ± 4. Here, SDM selects significantly fewer patches than Yottixel.

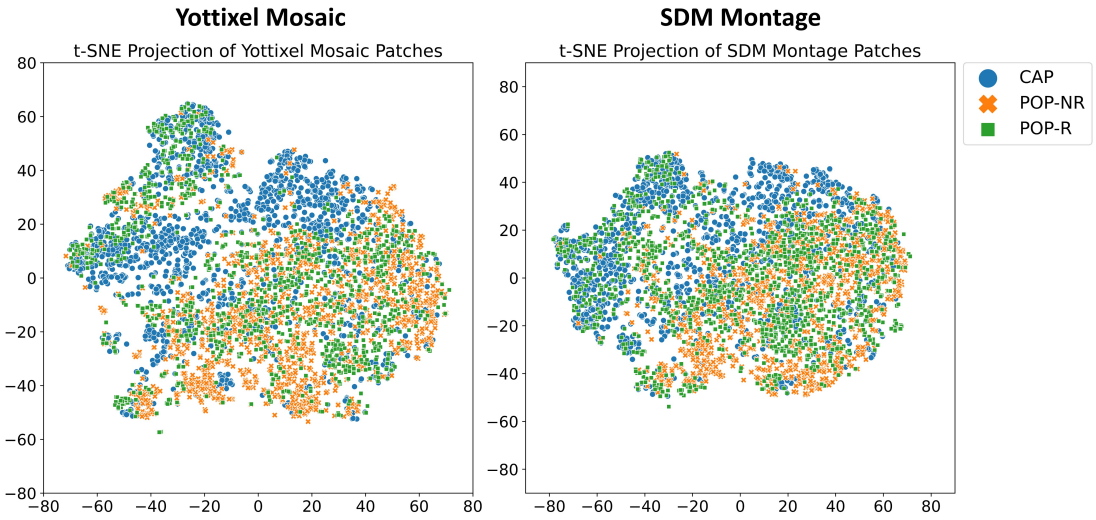

Figure 3.13: The t-SNE projection displays the embeddings of all patches extracted from the CRC dataset using Yottixel's mosaic (left) and SDM's montage (right).

3.3.3 Private – Breast Cancer (BC)

Breast tumor slides were acquired from patients at Mayo Clinic, Rochester, USA. There are 16 different subtypes of breast tumors were employed in this experiment. All of the biopsy slides were digitized at 40× magnification and linked to their respective diagnoses at the WSI level (see Table 3.6 for more details).

Primary Diagnoses	Acronyms	Slides
Adenoid Cystic Carcinoma	ACC	3
Adenomyoepthelioma	AME	4
Ductal Carcinoma In Situ	DCIS	10
Ductal Carcinoma In Situ, - Columnar Cell Lesions Including - Flat Epithelial Atypia, - Atypical Ductal Hyperplasia	DCIS, CCLIFEA, ADH	3
Intraductal Papilloma, Columnar Cell Lesions	IP, CCL	3
Invasive Breast Carcinoma of No Special Type	IBC NST	3
Invasive Lobular Carcinoma	ILC	3
Lobular Carcinoma In Situ + Atypical Lobular Hyperplasia	LCIS + ALH	2
Lobular Carcinoma In Situ, - Flat Epithelial Atypia, - Atypical Lobular Hyperplasia	LCIS, FEA, ALH	2
Malignant Adenomyoepithelioma	MAE	4
Metaplastic Carcinoma	MC	5
Microglandular Adenosis	MGA	2
Microinvasive Carcinoma	MIC	2
Mucinous Cystadenocarcinoma	MCC	5
Normal Breast	Normal	21
Radial Scar Complex Sclerosing Lesion	RSCSL	2

Table 3.6: Detailed information related to the BC dataset, inclusive of the respective acronyms and the number of slides associated with each primary diagnosis.

To assess the performance of the SDM's montage against Yottixel's mosaic, we conducted a leave-one-out evaluation to retrieve the most similar cases using the BC dataset. The evaluation criteria encompass the top-1 retrieval. The results, including accuracy, macro average, and weighted average scores at the top-1 are presented in Figure 3.14.

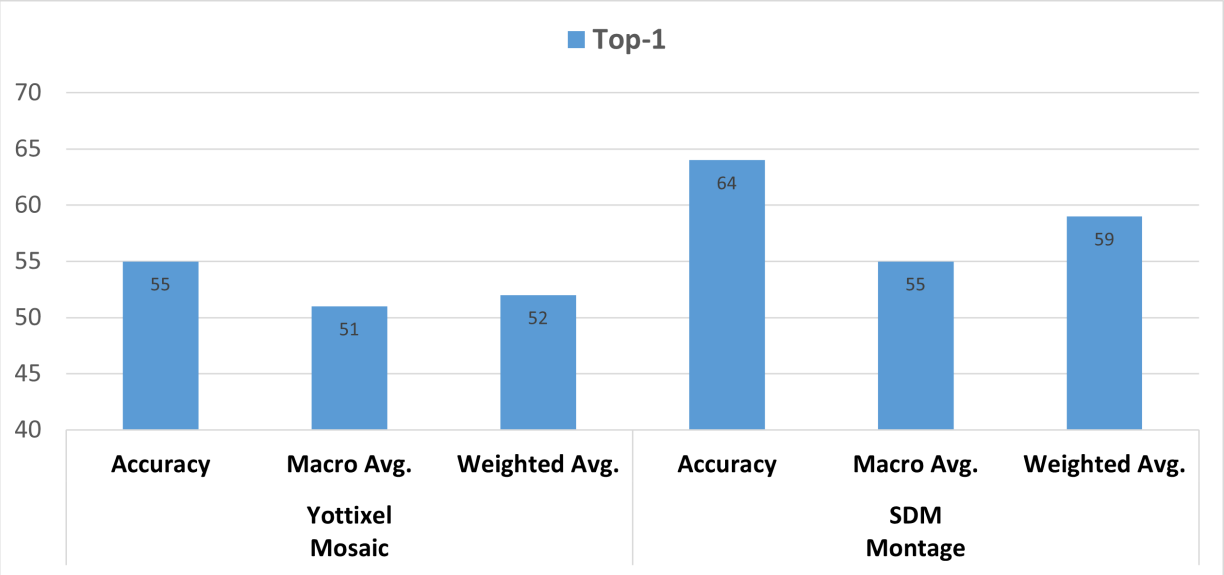

Figure 3.14: Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval in the Breast Cancer dataset. The diagram shows that SDM montage significantly outperforms Yottixel's mosaic.

Table 3.7 shows the detailed statistical results including precision, recall, and F1-score. Moreover, Confusion matrices and chord diagram at top-1 are shown in Figure 3.15. In addition to these accuracy metrics, a comparative analysis of the number of patches extracted per WSI by each respective method is also presented in Figure 3.16 for a visual representation of the distribution over the entire dataset. To visually illustrate the extracted patches, we used t-SNE projections, as demonstrated in Figure 3.17.

Our experimental findings showcased the superior performance of SDM, particularly evident in the top-1 retrieval result by +9% in accuracy, +4% in macro avg. of F1-scores, and +7% in weighted average as illustrated in Figure 3.14. Furthermore, our observations shed light on an intriguing aspect of Yottixel's behavior in comparison to SDM. Specifically, it has come to our attention that Yottixel displays a proclivity for overlooking certain WSIs within the dataset. To elaborate, our analysis reveals that Yottixel processed a total of 73 WSIs, whereas SDM demonstrated a more comprehensive approach by successfully processing all 74 WSIs. This observation underscores the robustness and completeness of the SDM method in handling the entire dataset, further emphasizing its merits in WSI analysis and retrieval applications.

	Yottixel Mosaic				SDM Montage			
	Top-1				Top-1			
Primary Diagnoses	Precision	Recall	F1-score	Slides	Precision	Recall	F1-score	Slides
ACC	0.00	0.00	0.00	3	0.00	0.00	0.00	3
AME	0.50	0.25	0.33	4	0.33	0.25	0.29	4
DCIS	0.67	0.20	0.31	10	0.50	0.60	0.55	10
DCIS, CCLIFEA, ADH	0.75	1.00	0.86	3	0.75	1.00	0.86	3
IP, CCL	1.00	0.33	0.50	3	1.00	1.00	1.00	3
IBC NST	0.00	0.00	0.00	3	0.00	0.00	0.00	3
ILC	0.38	1.00	0.55	3	0.00	0.00	0.00	3
LCIS + ALH	0.50	1.00	0.67	2	0.67	1.00	0.80	2
LCIS, FEA, ALH	0.67	1.00	0.80	2	1.00	1.00	1.00	2
MAE	0.80	1.00	0.89	4	1.00	1.00	1.00	4
MC	0.75	0.75	0.75	4	1.00	0.80	0.89	5
MGA	0.33	1.00	0.50	2	0.00	0.00	0.00	2
MIC	0.00	0.00	0.00	5	0.00	0.00	0.00	5
MCC	1.00	1.00	1.00	2	1.00	1.00	1.00	2
Normal	0.74	0.67	0.70	21	0.66	0.90	0.76	21
RSCSL	0.20	0.50	0.29	2	1.00	0.50	0.67	2
Total Slides				73				74

Table 3.7: Precision, recall, F1-score, and the number of slides processed for each sub-type are shown in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval in the Breast Cancer dataset.

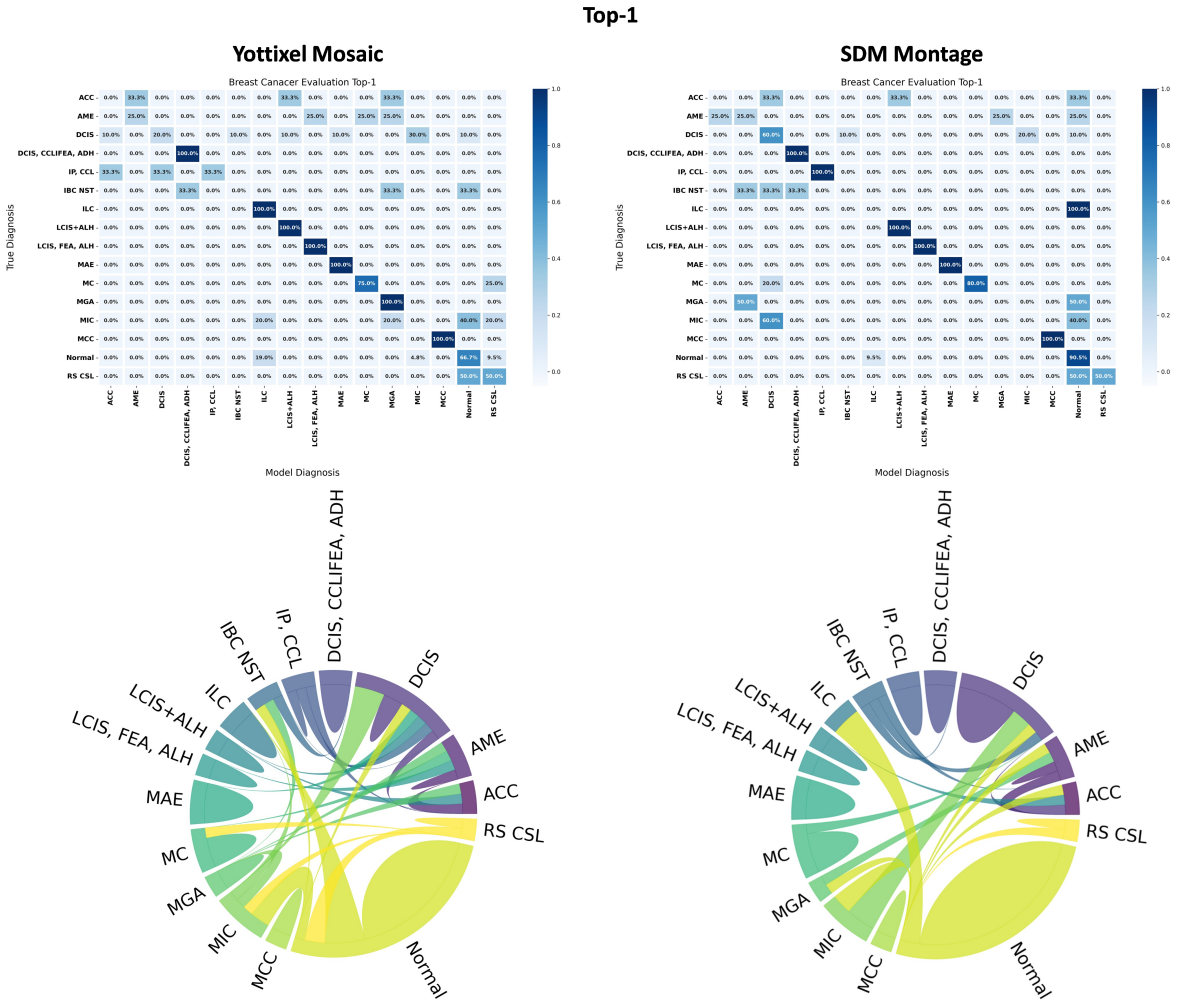

Figure 3.15: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval from the BC dataset.

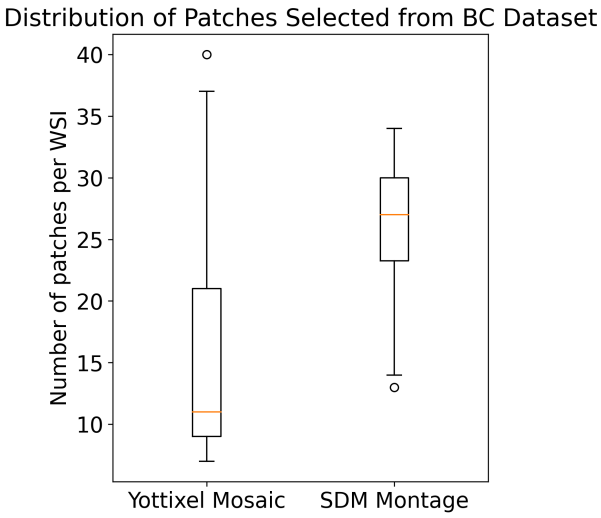

Figure 3.16: The boxplot illustrates the distribution of patches selected for each WSI in the Breast Cancer (BC) dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 11 ± 9. Conversely, for the SDM Montage, the median number of selected patches is 27 ± 5. Here, SDM selects slightly more patches than Yottixel's mosaic.

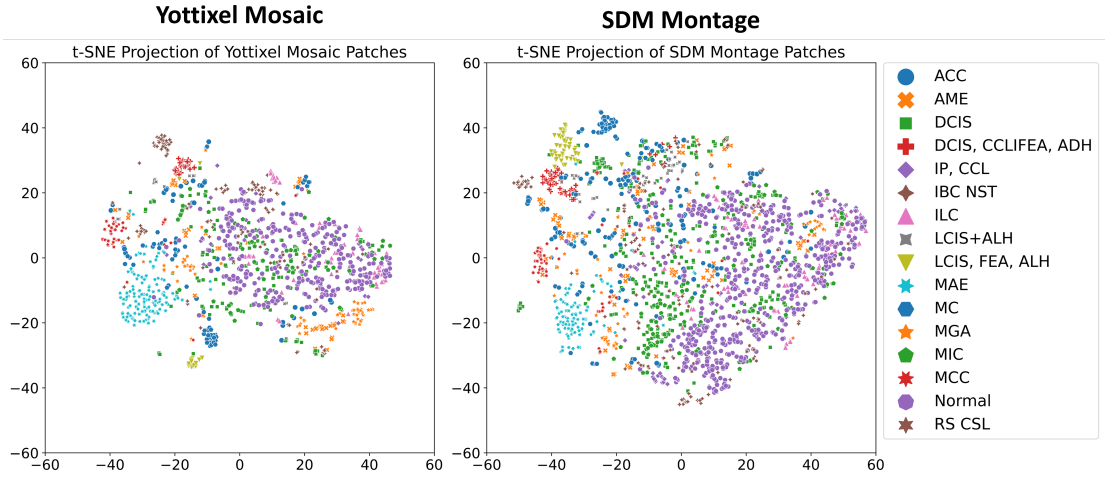

Figure 3.17: The t-SNE projection displays the embeddings of all patches extracted from the BC dataset using Yottixel's mosaic (left) and SDM's montage (right).

3.4 Discussion & Conclusion

Unsupervised WSI-to-WSI search holds significant importance, particularly when searching through extensive archives of medical images. It offers the invaluable capability of generating a virtual second opinion based on previously established and confidently diagnosed cases. By leveraging unsupervised search techniques, medical practitioners can efficiently compare a new WSI to a repository of historical cases without requiring prelabeled data. This not only aids in confirming diagnoses but also enhances the potential for discovering similar cases, patterns, or treatment insights, ultimately advancing the quality and accuracy of medical decision-making in histopathology and other fields of medicine. To execute WSI-to-WSI search effectively, it is imperative to employ a sophisticated divide-and-conquer strategy. WSIs are typically gigapixel and intricate images that are impractical to process in their entirety due to their size and complexity. Therefore, the divide-and-conquer approach involves breaking down the WSI into smaller, more manageable patches. These sections can be systematically analyzed and compared to other WSIs or reference images. Relying on a small number of meaningful patches is a crucial aspect of WSI-to-WSI matching. Incorporating a diverse range of patches from Whole WSIs is critical for capturing the rich and varied information contained within tissue samples. This diversity not only helps mitigate bias and enhances the robustness of matching algorithms but also ensures a comprehensive assessment of tissue features and anomalies. By accommodating the inherent diversity within WSIs, utilizing a varied set of patches can boost diagnostic accuracy. This approach not only refines the quality of research insights but also strengthens the ability to generalize findings across a wider array of cases.

For the specific objective at hand, we have introduced a methodology referred to as "SDM", which stands for Selection of Distinct Morphologies (presented in Section 3.2). The primary aim of SDM is to systematically choose a set of patches from a larger pool, with the intention of encompassing all diverse and unique morphological characteristics present within a given WSI. These meticulously selected patches collectively constitute what we term a "montage." The proposed methodology has undergone rigorous testing across six distinct datasets, comprising three publicly available datasets and three privately acquired datasets. In the evaluation process, we conducted a comprehensive comparative analysis with the Yottixel mosaic [3], which is the sole existing patch selection method documented in the literature. This extensive testing thoroughly assesses the effectiveness and performance of our approach in relation to the established benchmark provided by Yottixel mosaic [3]. In Figure 3.18, a systematic ranking methodology is presented to assess the efficacy of two distinct methods: Yottixel mosaic and SDM montage, across multiple datasets, employing a range of evaluation metrics. The criteria employed to evaluate

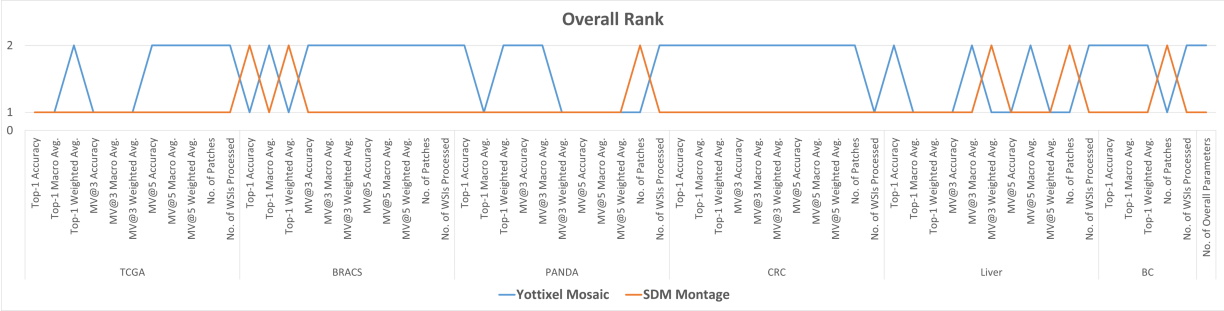

Figure 3.18: A comprehensive ranking scheme was devised to evaluate the performance of the two methods: Yottixel mosaic and SDM montage, across multiple datasets using various metrics. In this scheme, a rank of '1' signifies superior performance of a method relative to the other, a rank of '2' indicates inferior performance, and identical ranks of '1' for both methods denote comparable performance. After aggregating the results across all metrics, Yottixel mosaic achieved an average rank of 1.64, while SDM montage recorded a more favorable score of 1.09.

and rank the algorithms encompass various metrics, including accuracies, macro averages, weighted averages, the number of WSIs successfully processed per dataset, the number of patches extracted for each dataset, and the cumulative number of parameters essential for the algorithm's operation. Within this ranking paradigm, a designation of '1' denotes that a method exhibits a performance edge over its counterpart, while a '2' suggests subpar performance. Receiving identical rankings of '1' for both methods suggests they exhibit parity in their performance outcomes. Notably, the method is considered superior even when its performance exceeds the other method by a mere 1%. Upon consolidating the rankings overall metrics, Yottixel mosaic registered an average ranking of 1.64, in contrast to the SDM montage which secured a more commendable average of 1.09. An inspection of the figure clearly illustrates the SDM montage consistently achieving a '1' rank more often than the Yottixel mosaic. The proximity of its average rank to '1' further accentuates that, in an overarching assessment, SDM montage markedly outperforms Yottixel mosaic. Additionally, Figure 3.19 shows an overall comparison of accuracy, macro average, and weighted average at top-1, MV@3, and MV@5 using both Yottixel mosaic and SDM montage methods across all datasets used in this experiment.

The investigation underscores the paramount significance of an adept patch selection strategy in the context of WSI search and matching applications. The robustness and precision of such applications hinge on the ability to meticulously curate informative patches from the vast and intricate WSIs. In this regard, our proposed approach, SDM has demon-

strated remarkable efficacy through extensive experimentation on diverse datasets, including both publicly available and privately acquired ones. Throughout our evaluations, it has been consistently discerned that the proposed methodology outperforms the prevailing state-of-the-art patch selection technique, as epitomized by the Yottixel mosaic. The Yottixel approach necessitates the specification of certain empirical parameters, such as the percentage of patch selection and the number of color clusters, which poses challenges in determining optimal values across diverse datasets given the non-universal applicability of any single configuration. In contrast, the SDM approach obviates the need for such empirical parameterizations, inherently optimizing the selection to capture the distinct morphological features present in the WSI. Taken together, our findings affirm that a robust patch selection strategy is indispensable for enhancing the effectiveness of WSI search and matching applications, with our proposed method showcasing substantial advancements in this critical domain.

For future work, subsequent to a thorough evaluation across diverse datasets, one should assess the proposed methodology by engaging pathologists. This evaluation aims to solicit visual assessments from the pathologists, specifically focusing on the search and matching performance of the proposed algorithm.

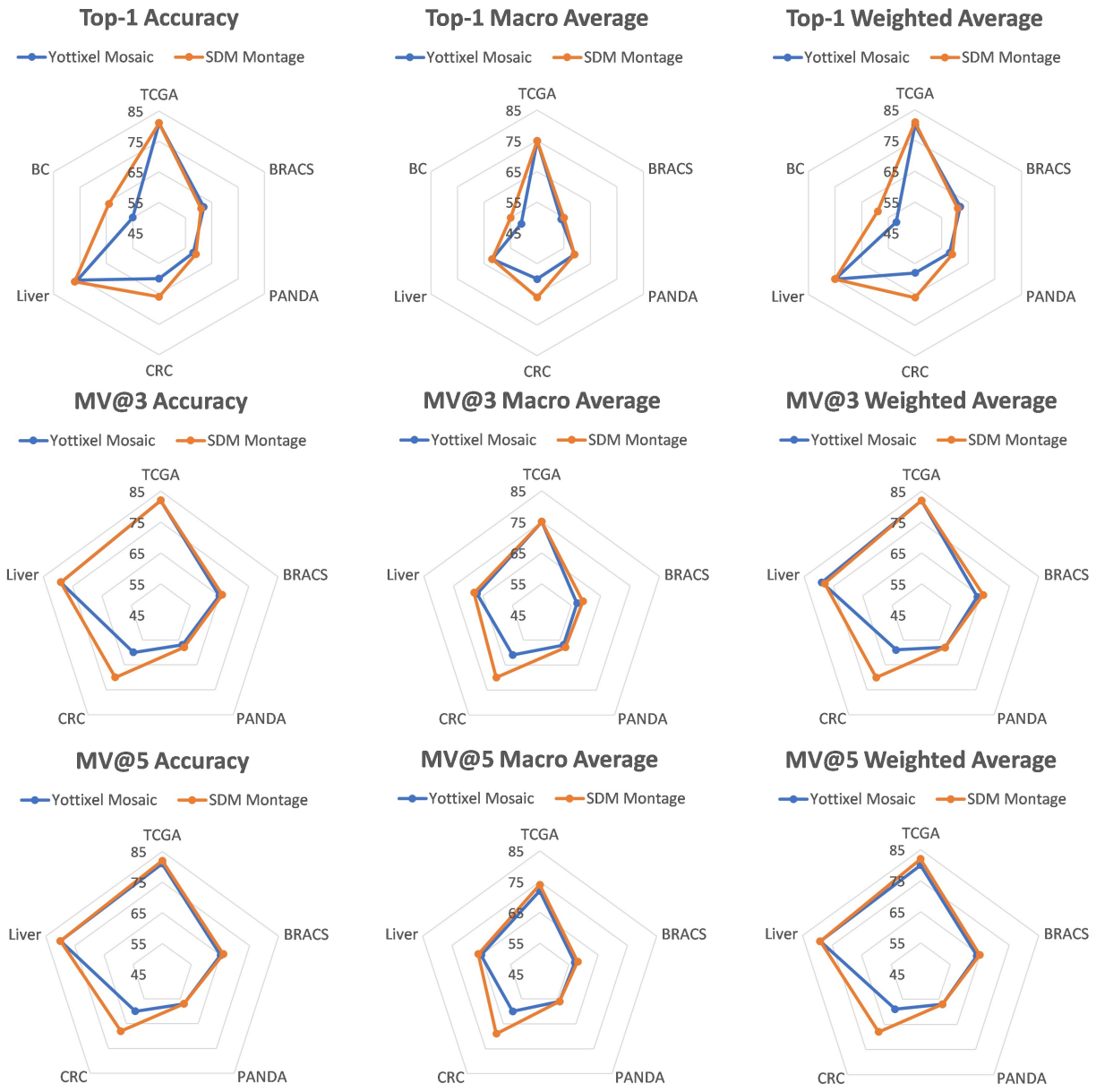

Figure 3.19: The collective accuracy, both macro and weighted averages, at top-1, MV@3, and MV@5 using both Yottixel mosaic and SDM montage methods across all datasets employed for evaluation.

Chapter 4

NeXtPath – Representation Learning for Image Search

In this chapter, the second contribution of this thesis is introduced — a novel representation learning methodology that can distinguish between different tissue types specialized for image search in the latent space. This methodology is further explained in the methods Section 4.1. Furthermore, the evaluation of the proposed method is described in Section 4.2 followed by the discussion and conclusion in Section 4.3.

4.1 Methodology

In DL, representation learning is often achieved through deep neural networks that learn to encode input data into a series of layers of increasing abstraction. Learning discriminative representations of raw data is crucial for the efficacy of image retrieval systems. In the domain of image retrieval, the objective is to transform raw image pixels into a feature space where the pertinent characteristics that differentiate one image from another are emphasized. To enhance the discriminative power of these representations, techniques such as metric learning can be employed. For this reason, the second major contribution of this doctoral research pertains to learning discriminative representations by utilizing metric learning approaches. This method works by optimizing the feature space to ensure that similar images are closer together while dissimilar images are further apart, thereby streamlining the retrieval process.

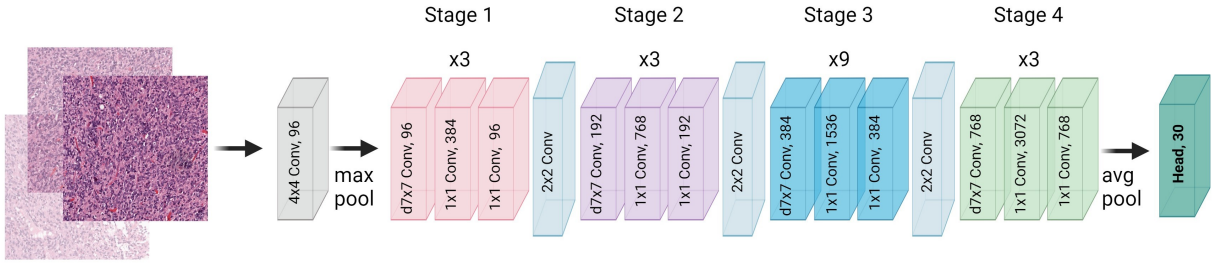

Figure 4.1: ConvNeXt-tiny [7] architecture for NeXtPath.

In this chapter, a new **ranking loss** function is introduced, specifically designed to enhance image retrieval tasks. This function aids the model in learning distinctive representations. Additionally, this chapter details the implementation of cutting-edge CNN architectures and validates that the process of learning discriminative representations can be effectively refined using various network structures.

4.1.1 NeXtPath: Fine-Tuned ConvNeXt on TCGA FFPE Slides

Recently, the emergence of ConvNeXt [7] has demonstrated competitive performance and scalability with state-of-the-art vision transformers. Notably, ConvNeXt, being a CNN, possesses the distinct advantage of requiring relatively smaller amounts of data for fine-tuning while still achieving commendable results. This attribute sets it apart from ViTs, as it enables ConvNeXt to perform effectively in scenarios where data availability may be limited, further underscoring its practical utility in various applications.

Riasatian et al. [6] used the TCGA dataset, the largest publicly available repository for diagnostic slides, to fine-tune a DenseNet [5]. The data (a total of 8,611 Formalin-fixed Paraffin-embedded Tissue (FFPE) WSIs, 7,126 training slides, 741 validation slides, and 744 test slides) were processed based on cellularity, and the processed data is publicly available. FFPE tissue samples represent a gold standard in both clinical and research settings for the preservation and preparation of biopsy materials. The patches were extracted from 30 primary diagnoses at 20× magnification with a patch size of 1000 × 1000 pixels.

For the fine-tuning of NeXtPath (ConvNeXt-tiny [7] fine-tuned on the diagnostic slides of TCGA, see Figure 4.1 for the architecture), a publicly available high-quality dataset from Riasatian et al. [6] was used. Based on the findings in the literature, the optimal results are achieved when fine-tuning the entire network. Consequently, we conducted fine-tuning in a single configuration using the PyTorch, wherein the entire network was fine-tuned using a

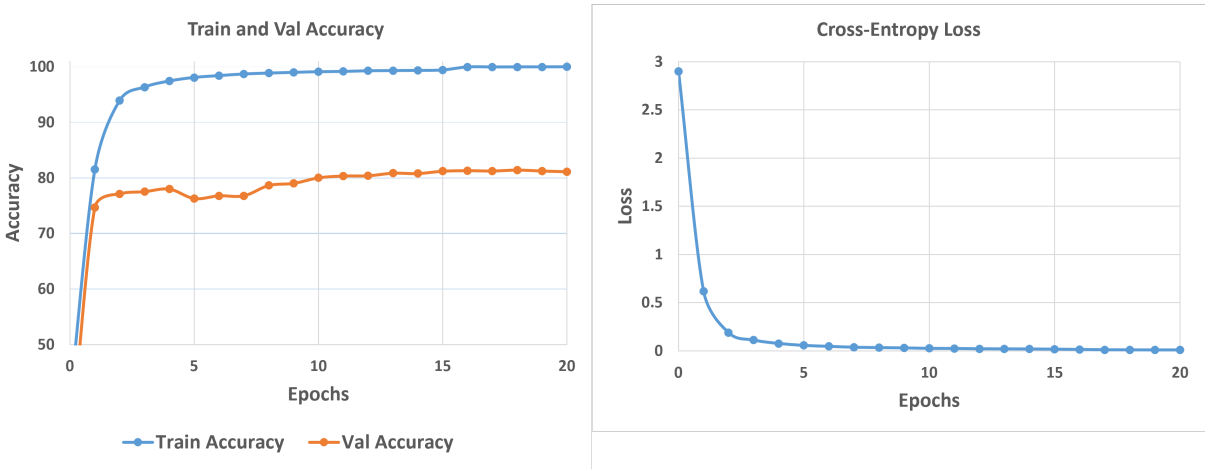

Figure 4.2: (left) Training and validation accuracy for the NeXtPath (ConvNeXt fine-tuned with TCGA diagnostic slides over a span of 20 epochs), exhibited a peak training accuracy of 99.99% and a maximum validation accuracy of 81.45%. (right) Training loss over a span of 20 epochs using cross-entropy loss function.

dataset comprising 240,527 high-cellular patches for training and 24,492 validation patches originating from 30 primary diagnostic categories. The NeXtPath was fine-tuned for 20 epochs with AdamW optimizer with 0.0001 as the initial learning rate with 0.01 weight decay. Cross-entropy was selected as the loss function to gauge the classification model's performance. For initializing the model weights, we employed a pre-trained model, trained on 21,841[1] classes from the ImageNet dataset. For data augmentation, random rotation (90, 270, and -90 degrees only) was used with 50% probability, random verticle flip with 50% probability, random horizontal flip with 50% probability, random crop (with sizes 224, 384, 512, and 786) with 20% probability and then resized back to 1000 $\times$ 1000, and random color jitter (brightness 0.2, contrast 0.2, saturation 0.1 and hue 0.05) with 20% probability. During the training process, the weights were saved for the highest validation accuracy. Figure 4.2 shows the convergence behavior of NeXtPath with training accuracy and loss curve using cross-entropy. The highest training and validation accuracy is 99.99% and 81.45%, respectively.

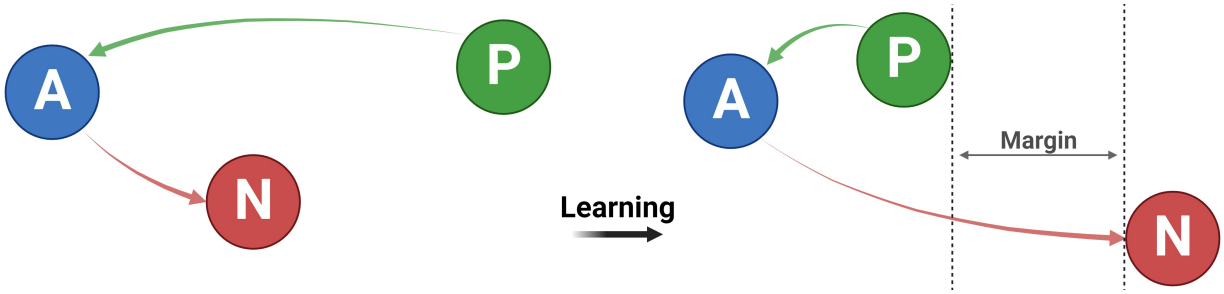

Figure 4.3: Triplet-loss is a loss function used primarily in metric learning. The goal of this loss function is to ensure that an anchor (query) pulls the positive (similar) sample closer and pushes away the negative (dissimilar) sample by some margin.

4.1.2 Metric Learning for Image Search

Triplet loss is a powerful loss function commonly used in the domain of metric learning. By focusing on relative distances between sets of three data points anchor, positive, and negative, it provides a framework to learn embeddings or representations where similar items are pulled closer and dissimilar items are pushed apart in the embedded space (see Figure 4.3 and Equation 4.1) [138]. This inherently tailors the learned metric to emphasize relationships between data instances, ensuring that the metric reflects the underlying structure and similarities within the data. As such, triplet loss provides a direct and effective mechanism to drive metric learning, optimizing the representation of data in a way that meaningful distances are preserved. Typically, triplet loss is employed by utilizing three neural networks that share identical weights [66], processing three distinct inputs: an anchor, a positive sample, and a negative sample. This shared-weight architecture ensures consistent feature transformation across all three inputs. The objective is to refine the embedded feature space such that the anchor's representation is closer to the positive sample than to the negative sample. This tailored feature space optimization enhances tasks like similarity search and clustering by emphasizing discriminative characteristics inherent to the data.

$$L(a, p, n) = \max\left(0, d(f(a) - f(p)) - d(f(a) - f(n)) + \alpha\right) \tag{4.1}$$

where L is the triplet loss with anchor a, positive p, and negative n as inputs to the loss function. $f(a)$, $f(p)$, and $f(n)$ are the embeddings from the network when using a, p,

[1]An extended version of the ImageNet-1K dataset, which originally has 1,000 classes, now includes approximately 14 million images spread across 21,841 distinct classes.

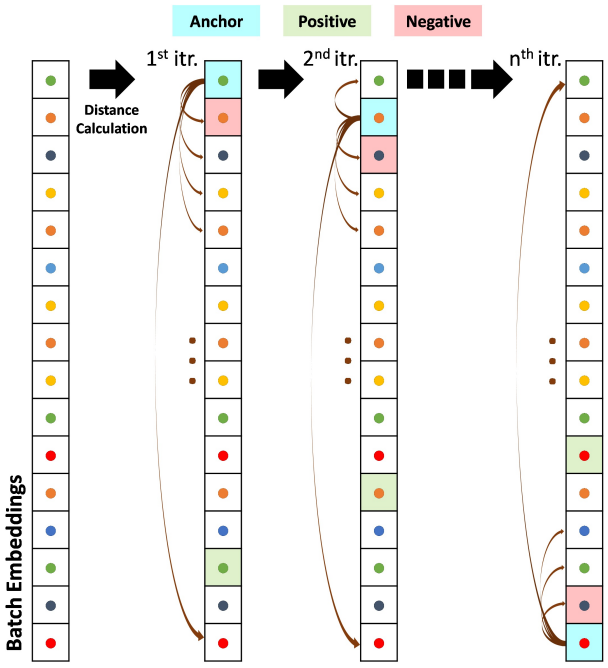

Figure 4.4: Ranking loss for image search algorithm is illustrated in this image. It is an iterative process with batch embeddings as an input. Within each iteration, a point from the batch is designated as the anchor, and distances to all other points are computed and subsequently arranged in ascending order. From this order, the first dissimilar embedding from the anchor is chosen as a negative embedding, and the last similar embedding as the anchor is chosen as a positive embedding. Subsequently, generating anchor-positive, and anchor-negative pairs. This procedure progresses iteratively until the final anchor point is paired with both its positive and negative counterparts.

and n as inputs. $d(f(a), f(p))$ and $d(f(a), f(n))$ denote the Euclidean distances between the anchor and the positive vectors, and between the anchor and the negative vectors, respectively. α is the margin, and the $\max(0, \cdot)$ function ensures that the loss is non-negative.

The absence of deep learning models tailored specifically for the image search is a real challenge. To address this gap, this thesis introduces an advanced learning guidance strategy termed "ranking loss for image search". This novel mechanism facilitates training a DNN that is optimized for the nuanced demands of image search in histopathology.

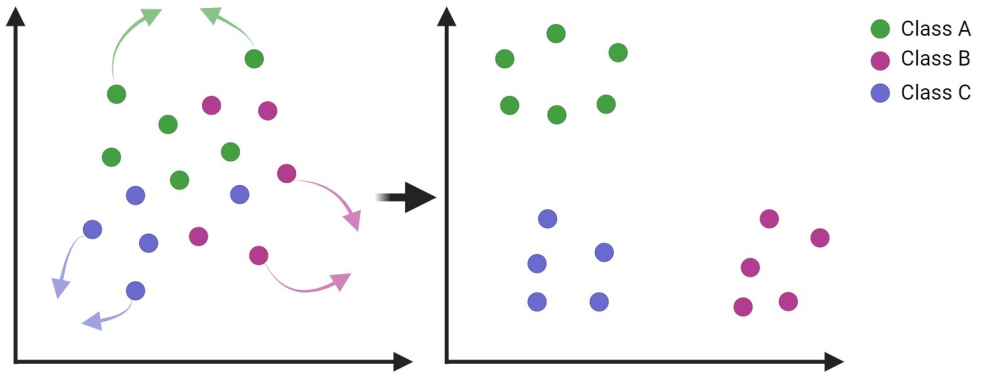

Figure 4.5: Multi-class learning process using "ranking loss for image search" to pull the similar class embeddings closer and push the dissimilar embeddings away.

Ranking loss for image search is a modified triplet loss function designed specifically for image search and matching. The fundamental concept behind ranking loss is to increase the number of accurately retrieved instances, emphasizing the importance of prioritizing relevant items during the retrieval process. The algorithm to find the anchor-positive and anchor-negative pairs is outlined in Algorithm 2 and depicted in Figure 4.4. This algorithm is designed for a single network instead of a triple network setup to pull similar class instances together and push the dissimilar instances far apart with some margin. The search mechanism hinges on distance-based matching, ranking entities by proximity. Drawing inspiration from this, a batch of embeddings, E_b, was probed to identify positive (p) and negative (n) pairs corresponding to a given anchor (a), which acts as the query. Initially, inputs are quantized based on instances per class (C), serving both as an accuracy metric and a filter. Classes with fewer than two instances are excluded from subsequent steps. At the same time, those with more than two are eligible for pair matching, given that a minimum of two instances is requisite for effective querying. For a given anchor, each point within the batch is compared based on their distances, denoted as (D), utilizing the Hamming distance in this study as the embeddings are converted into the barcodes using Min-Max algorithm [112, 113, 3] to speed up the process. Subsequently, the batch undergoes an ascending sort operation, S, anchored on the derived distances D. Navigating through sorted batch S, the initial embedding from a disparate class is selected as the negative pair, while the most distant corresponding class embedding is chosen as the positive pair. The rationale behind this approach is to distance the **near-dissimilar** class from the anchor while drawing the **far-similar** class embeddings closer, ensuring the proximity of similar entities. During the training phase, the network contracts the distance

Algorithm 2 Ranking Loss for Image Search

Require: Input Batch Embeddings
Require: Input Batch size must be at least 3× number of classes
Ensure: a Set of anchor-positive and anchor-negative pairs as output

$E_b \leftarrow$ get the batch embeddings as input
Procedure
$C \leftarrow$ quantification(E_b) ▷ count the number of instances per class
for each $E_b > 2$ **do**
    $a \leftarrow$ getAnchor (E_b) ▷ iteratively set as anchor each embedding in E_b
    $D \leftarrow$ DistanceCalculations(a, E_b) ▷ distance calculation from anchor a to every point in E_b
    $S \leftarrow$ SortDistances(D) ▷ sort the distances in ascending order
    $p \leftarrow$ findPositive(a, S) ▷ find the last instance of similar class as anchor a in S
    $n \leftarrow$ findNegative(a, S) ▷ find the first instance of dissimilar class as anchor a in S
    $A \leftarrow$ calculateAccuracy(a, S, C) ▷ how many instances of same class as anchor a are in top S out of total number of instances C
end for
$(a_b, p_b, n_b) \leftarrow$ makePairs(a, p, n) ▷ make anchor-positive and anchor-negative pairs for whole batch
$A_b \leftarrow$ mean(A) ▷ calculate mean accuracy of the batch
Return (a_b, p_b, n_b), A_b ▷ return the final pairs and mean accuracy for the input batch
End Procedure

between embeddings of identical classes and expand the separation between those of dissimilar classes. This promotes an optimized representational learning specifically tailored for image retrieval (see Figure 4.5 for pictorial representation of this process). Additionally, class-specific accuracy A is computed based on the ratio of correct nearest matches to the total occurrences of that specific class. For instance, if an input batch contains ten samples of a given class and, post-matching and sorting, 5 samples from that class appear within the top nine matches, the resulting accuracy for that class is determined to be 55.56%. Conclusively, the average accuracy across all classes is computed and presented, representing the aggregated accuracy A_b for the entire batch.

4.2 Evaluation & Results

In evaluating the representations procured from the neural networks, the search was treated as a classifier to streamline the evaluation process. The inherent benefit of leveraging classification approaches is their straightforward validation mechanics. Specifically, each image is delineated into a categorical context, affirming its affiliation to a predetermined class or negating it. Such a binary demarcation facilitates performance metric extraction by enumerating instances of misclassification. Complementing this, metrics grounded in distance calculations, like the Hamming distance, play an instrumental role. This measure quantifies the divergence between two feature vectors representing images. This dual-pronged approach, combining classification and distance metrics, provides a comprehensive framework for evaluating the efficacy of the neural networks in extracting and representing salient features from the histopathology patches.

All experiments have been conducted on a Dell PowerEdge XE8545 with 2× AMD EPYC 7413 CPUs, 1023 GB RAM, and 4× NVIDIA A100-SXM4-80GB using PyTorch deep learning framework. We used PyTorch 2.0.0, Python 3.9.16, and CUDA 11.7 on a Linux operating system. Two state-of-the-art CNNs were used, namely KimiaNet (a fine-tuned version of DenseNet121 tailored for histological applications) [6] and NeXtPath (a contemporary CNN developed for the 2020s [7], fine-tuned for histopathology). For training and validation, publicly accessible datasets, specifically TCGA and BRACS [134], were utilized. This chapter evaluates the representations obtained from the afore-mentioned two neural networks. These networks were initially trained using the prevalent cross-entropy loss function, followed by the application of a proposed ranking loss, specifically designed for representation learning in image search applications. The evaluation criterion uses the Hamming distance. We conducted an evaluation for the patch retrievals. In this methodology, a query patch representation is compared against the entire reference atlas to identify the closest histological match. A distinct advantage of the search and matching is the ability to retrieve multiple closely matching representations, enabling the formation of a consensus based on the top n retrievals (here $n = 1$, 3, and 5).

4.2.1 The Cancer Genome Atlas (TCGA)

TCGA repository covers 25 anatomical sites, featuring 32 cancer subtypes and data from nearly 33,000 WSIs [6, 133]. For retrieval evaluation, we constructed a digital atlas comprising 110,032 high-cellular test patches, representing an extensive spectrum of over 30 distinct tumor types. This atlas was derived from the 744 TCGA Test WSIs which were not used in

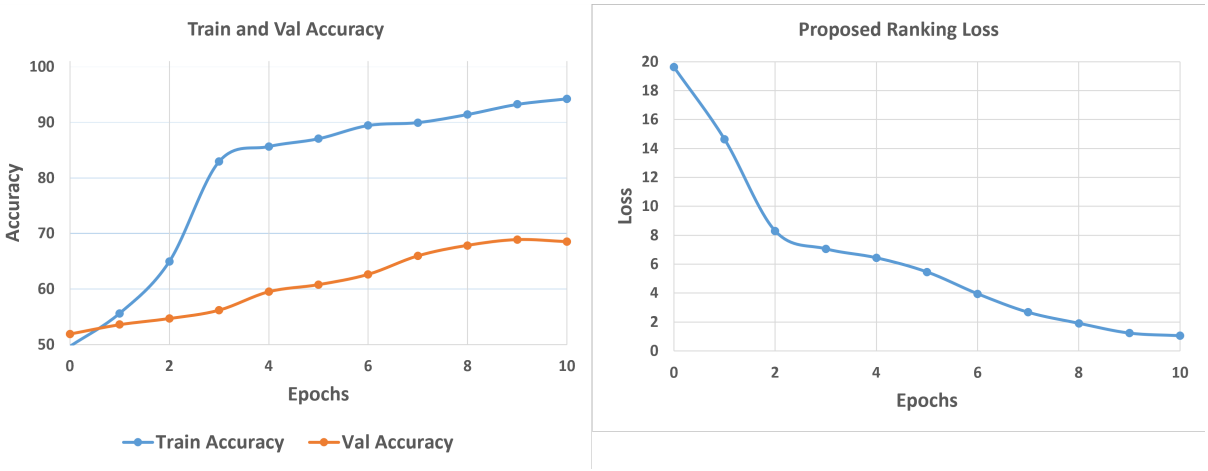

Figure 4.6: (left) Training and validation accuracy for the KimiaNet + Ranking (fine-tuned with the proposed ranking loss), exhibited a peak training accuracy of 94.25% and a maximum validation accuracy of 68.88%. (right) Training loss over a span of 10 epochs using the proposed ranking loss function tailored specifically for search and matching.

the training process. In our evaluation phase, a total of 24,492 high-cellular patches from 722 TCGA validation WSIs, served as the query set for histological comparisons against the reference atlas. The overarching objective was to facilitate precise histological matching by leveraging the comprehensive representations encapsulated within our digital atlas. Feature vectors for the atlas and query cohorts were derived using two specific models: **KimiaNet** and **NeXtPath**. Both models were trained initially with cross-entropy loss and subsequently with the proposed ranking loss for comparison. The adoption of these two networks aimed to enable a systematic comparison of their effectiveness in feature representation learning.

KimiaNet

KimiaNet [6] represents a leading-edge DNN specifically tailored for histopathological analyses and is publicly available. Distinguishingly, it is among the few networks trained on an extensive field of view, accommodating patch sizes of 1000×1000. The network underwent training on over 240,000 histological patches of this dimension, sourced from publicly available diagnostic WSIs in the TCGA dataset. For the purposes of this evaluation, we employed the official version of KimiaNet, which was trained using the PyTorch framework, ensuring an equitable comparison.

To assess the efficacy of representations produced by the baseline KimiaNet (previously fine-tuned with the cross-entropy loss function), I subsequently fine-tuned the same network employing the novel Ranking loss, a method I introduced. This loss is specially designed to optimize the learning of representations for histological matching tasks. In this study, a uniform dataset was employed for both training and validation stages, encompassing over 240,000 training patches and exceeding 24,000 validation patches. Building on the foundational architecture of DenseNet-121, the terminal three dense blocks were subjected to a fine-tuning process spanning 10 epochs. This was conducted with a predetermined learning rate of 0.00001, accompanied by a decay parameter of 0.01. The optimization procedure was orchestrated using the AdamW optimizer, leveraging the weights from KimiaNet as the initial weight configuration. To enhance model robustness and generalizability, a comprehensive set of data augmentation techniques were integrated. These encompassed random rotations restricted to angles of 90, 270, and -90 degrees, applied with a 50% probability; random vertical and horizontal flips, each with a probability of 50%; random cropping, with potential sizes being 224, 384, 512, and 786, and a 20% probability, subsequently resized to the dimensions of 1000×1000. Additionally, the dataset underwent random color jittering, characterized by variations in brightness (0.2), contrast (0.2), saturation (0.1), and hue (0.05), with each adjustment bearing a 20% probability of application. Figure 4.6 depicts the convergence trajectory of KimiaNet over the training period, illustrating the training & validation accuracy and the associated loss curve when using the proposed ranking loss function. Throughout the training progression, the peak training accuracy achieved was 94.25%, while the maximum validation accuracy recorded was 68.88%. To ensure optimal generalization capabilities, the model's weights were preserved based on the peak validation performance, given that the validation set remains external to the training process. In this context, the most favorable model weights corresponded to a validation accuracy of 68.88%. In subsequent sections of this analysis, the KimiaNet model, which was fine-tuned using the proposed ranking loss, will be referred to as "KimiaNet + Ranking". Figure 4.7 presents a visualization of the latent space representations of the training, validation, and testing datasets derived from KimiaNet and KimiaNet + Ranking.

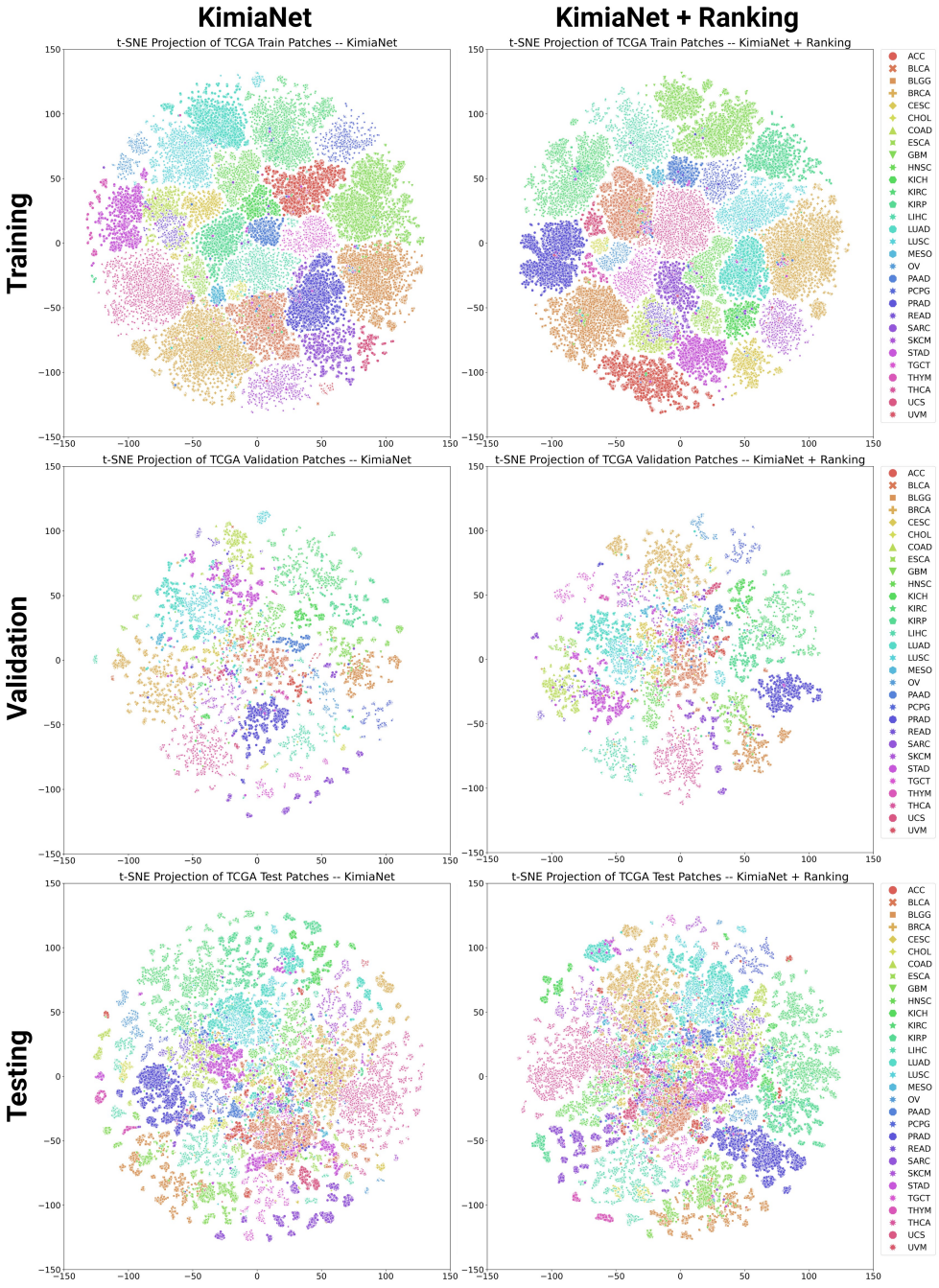

Figure 4.7: t-SNE projections for all the embeddings of 240,527 training patches, 24,492 validation patches, and 110,032 test patches from KimiaNet and KimiaNet + Ranking.

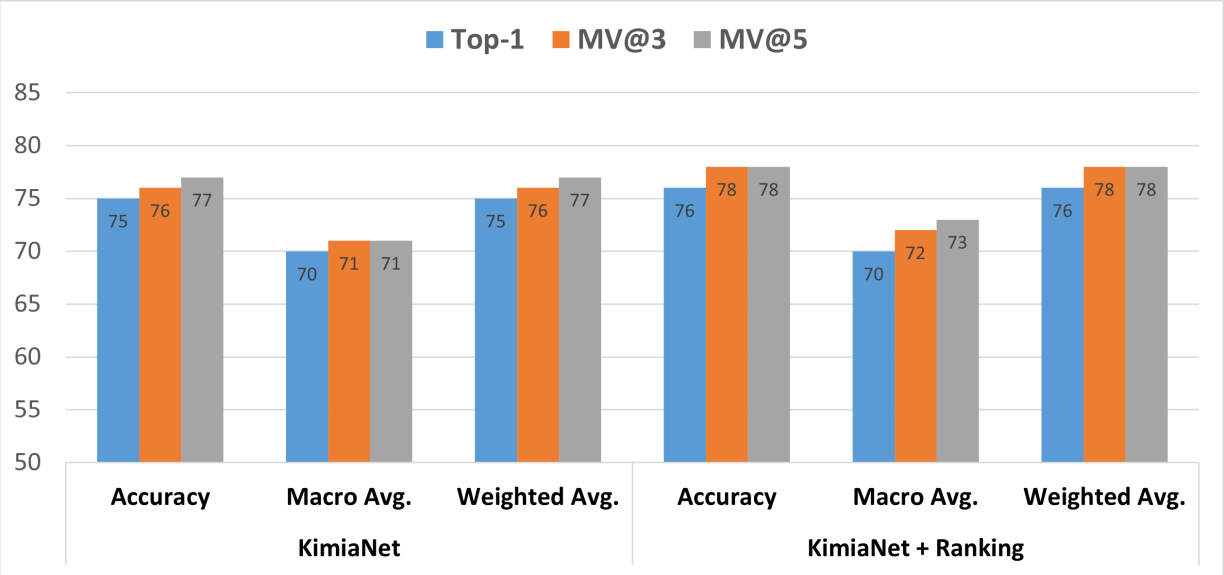

Figure 4.8: Accuracy, macro average of f1-scores, and weighted average of F1-scores are shown for the patch matching when using features from KimiaNet (fine-tuned using cross-entropy loss) [6], and KimiaNet + Ranking (fine-tuned using the proposed ranking loss). The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA validation patches used as query and test patches as a reference atlas. KimiaNet trained with the proposed ranking loss performs slightly better than the KimiaNet trained with cross-entropy loss.

A retrieval-based analysis was executed to evaluate the efficacy of KimiaNet (fine-tuned using cross-entropy loss) in comparison with KimiaNet + Ranking (fine-tuned using the proposed ranking loss). Leveraging both query and atlas cohorts, the objective was to identify and retrieve the cases that bore the highest similarity, thereby gauging the relative performance of the model in such a matching paradigm. The evaluation involved multiple retrieval criteria, including the top-1 retrieval (Top-1), the majority agreement among the top 3 retrievals (MV@3), and the majority agreement among the top 5 retrievals (MV@5). The accuracy, macro average, and weighted average at Top-1, MV@3, and MV@5 are reported in Figure 4.8. Moreover, confusion matrices and chord diagrams at MV@5 are illustrated in Figure 4.9. Additional confusion matrices and chord diagrams of Top-1, and MV@3 retrievals are provided in Appendix B (see Figure B.1, and B.2, respectively). Table 4.1 and 4.2 show the detailed results including precision, recall, and f1-score for the representations generated using KimiaNet and KimiaNet + Ranking, respectively. Fur-

thermore, the t-SNE visualizations of the feature vectors from KimiaNet and KimiaNet + Ranking are shown in Figure 4.7.

In our research, we observed that the performance of representations (also referred to as embeddings) obtained from KimiaNet + Ranking surpassed that of the conventional state-of-the-art KimiaNet. In terms of model accuracy, KimiaNet + Ranking demonstrated consistently higher performance. For Top-1, MV@3, and MV@5 retrievals, we observed performance enhancements of +1%, +2%, and +1%, respectively. However, in terms of the macro-average F1-scores, KimiaNet + Ranking exhibited superior performance, surpassing the baseline by +1% and +2% for MV@3 and MV@5 retrievals, respectively. This performance was consistent during Top-1 retrieval. Additionally, for the weighted average of f1-scores, similar to the accuracy, KimiaNet + Ranking shows the performance enhancement by +1%, +2%, and +1% when retrieving Top-1, MV@3, and MV@5 retrievals, respectively (see Figure. 4.8).

KimiaNet

	Top-1			MV@3			MV@5			
Primary Diagnoses	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score	Patches
ACC	0.64	0.81	0.71	0.63	0.82	0.71	0.65	0.85	0.74	264
BLCA	0.66	0.71	0.68	0.67	0.72	0.70	0.67	0.73	0.70	1383
BLGG	0.90	0.81	0.85	0.91	0.82	0.86	0.91	0.83	0.87	1030
BRCA	0.84	0.89	0.87	0.84	0.91	0.88	0.84	0.92	0.88	2113
CESC	0.45	0.32	0.37	0.46	0.29	0.36	0.51	0.30	0.38	457
CHOL	0.38	0.66	0.48	0.38	0.56	0.46	0.40	0.58	0.48	200
COAD	0.50	0.50	0.50	0.51	0.50	0.50	0.50	0.50	0.50	925
ESCA	0.39	0.32	0.35	0.41	0.33	0.36	0.42	0.32	0.36	552
GBM	0.77	0.81	0.79	0.77	0.82	0.80	0.78	0.83	0.80	800
HNSC	0.67	0.75	0.71	0.68	0.77	0.72	0.68	0.78	0.73	809
KICH	0.96	0.82	0.88	0.96	0.83	0.89	0.96	0.83	0.89	489
KIRC	0.89	0.87	0.88	0.89	0.88	0.89	0.90	0.88	0.89	2105
KIRP	0.70	0.85	0.77	0.70	0.84	0.77	0.72	0.85	0.78	821
LIHC	0.80	0.72	0.76	0.81	0.75	0.78	0.82	0.75	0.78	1312
LUAD	0.72	0.55	0.62	0.75	0.57	0.65	0.75	0.57	0.64	1028
LUSC	0.73	0.69	0.71	0.74	0.71	0.73	0.74	0.71	0.72	1510
MESO	0.63	0.25	0.36	0.66	0.26	0.38	0.74	0.28	0.41	163
OV	0.83	0.86	0.85	0.84	0.87	0.85	0.85	0.87	0.86	447
PAAD	0.62	0.67	0.64	0.62	0.66	0.64	0.62	0.67	0.64	358
PCPG	0.89	0.92	0.90	0.90	0.93	0.91	0.90	0.93	0.92	630
PRAD	0.95	0.90	0.92	0.95	0.91	0.93	0.95	0.91	0.93	1347
READ	0.12	0.18	0.14	0.10	0.14	0.12	0.10	0.14	0.11	324
SARC	0.85	0.76	0.80	0.85	0.77	0.80	0.86	0.77	0.81	703
SKCM	0.77	0.70	0.73	0.80	0.73	0.77	0.81	0.75	0.78	799
STAD	0.59	0.67	0.63	0.61	0.70	0.65	0.61	0.70	0.65	1031
TGCT	0.91	0.93	0.92	0.92	0.93	0.92	0.93	0.93	0.93	671
THCA	0.88	0.93	0.90	0.89	0.93	0.91	0.89	0.94	0.91	1859
THYM	0.93	0.58	0.71	0.95	0.60	0.73	0.95	0.60	0.73	131
UCS	0.64	0.72	0.68	0.66	0.74	0.70	0.67	0.74	0.70	141
UVM	0.90	0.94	0.92	0.89	0.92	0.91	0.89	0.93	0.91	90
Total Patches										**24492**

Table 4.1: Detailed precision, recall, f1-score, and the number of patches processed for each subtype are shown in this table using the validation patches when matched against the test patches using KimiaNet for feature extraction. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA Patch dataset.

KimiaNet + Ranking

	Top-1			MV@3			MV@5			
Primary Diagnoses	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score	Patches
ACC	0.59	0.60	0.59	0.63	0.62	0.62	0.62	0.67	0.64	264
BLCA	0.66	0.65	0.66	0.67	0.67	0.67	0.69	0.67	0.68	1383
BLGG	0.94	0.80	0.86	0.94	0.84	0.88	0.91	0.86	0.88	1030
BRCA	0.84	0.90	0.87	0.85	0.91	0.88	0.88	0.91	0.90	2113
CESC	0.42	0.34	0.38	0.45	0.35	0.39	0.53	0.40	0.46	457
CHOL	0.37	0.65	0.47	0.39	0.66	0.49	0.42	0.74	0.54	200
COAD	0.49	0.54	0.51	0.51	0.55	0.53	0.54	0.61	0.58	925
ESCA	0.51	0.35	0.42	0.51	0.36	0.42	0.48	0.46	0.47	552
GBM	0.74	0.83	0.78	0.77	0.84	0.80	0.78	0.74	0.76	800
HNSC	0.64	0.81	0.71	0.66	0.82	0.73	0.66	0.86	0.75	809
KICH	0.94	0.74	0.83	0.94	0.76	0.84	0.91	0.79	0.84	489
KIRC	0.92	0.89	0.90	0.92	0.90	0.91	0.93	0.89	0.91	2105
KIRP	0.73	0.85	0.78	0.75	0.85	0.79	0.73	0.85	0.78	821
LIHC	0.80	0.73	0.77	0.81	0.74	0.77	0.83	0.72	0.77	1312
LUAD	0.72	0.63	0.67	0.73	0.65	0.69	0.73	0.69	0.71	1028
LUSC	0.72	0.70	0.71	0.73	0.70	0.72	0.76	0.73	0.74	1510
MESO	0.49	0.30	0.37	0.57	0.31	0.40	0.40	0.40	0.40	163
OV	0.80	0.82	0.81	0.81	0.82	0.81	0.81	0.82	0.82	447
PAAD	0.73	0.72	0.73	0.76	0.74	0.75	0.79	0.77	0.78	358
PCPG	0.88	0.94	0.91	0.89	0.95	0.92	0.91	0.95	0.93	630
PRAD	0.97	0.90	0.93	0.97	0.90	0.93	0.97	0.88	0.92	1347
READ	0.14	0.21	0.17	0.16	0.24	0.19	0.16	0.16	0.16	324
SARC	0.88	0.80	0.84	0.88	0.81	0.84	0.87	0.78	0.82	703
SKCM	0.76	0.69	0.72	0.75	0.73	0.74	0.78	0.73	0.75	799
STAD	0.69	0.67	0.68	0.70	0.69	0.70	0.70	0.64	0.67	1031
TGCT	0.87	0.90	0.89	0.88	0.91	0.90	0.80	0.93	0.86	671
THCA	0.92	0.95	0.93	0.92	0.95	0.94	0.91	0.96	0.93	1859
THYM	0.90	0.62	0.73	0.90	0.62	0.73	0.88	0.69	0.77	131
UCS	0.52	0.78	0.63	0.53	0.80	0.64	0.53	0.83	0.65	141
UVM	0.82	0.92	0.87	0.89	0.92	0.91	0.89	0.91	0.90	90
Total Patches										**24492**

Table 4.2: Detailed precision, recall, f1-score, and the number of patches processed for each subtype are shown in this table using the validation patches when matched against the test patches using KimiaNet (trained with the proposed ranking loss) for feature extraction. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA Patch dataset.

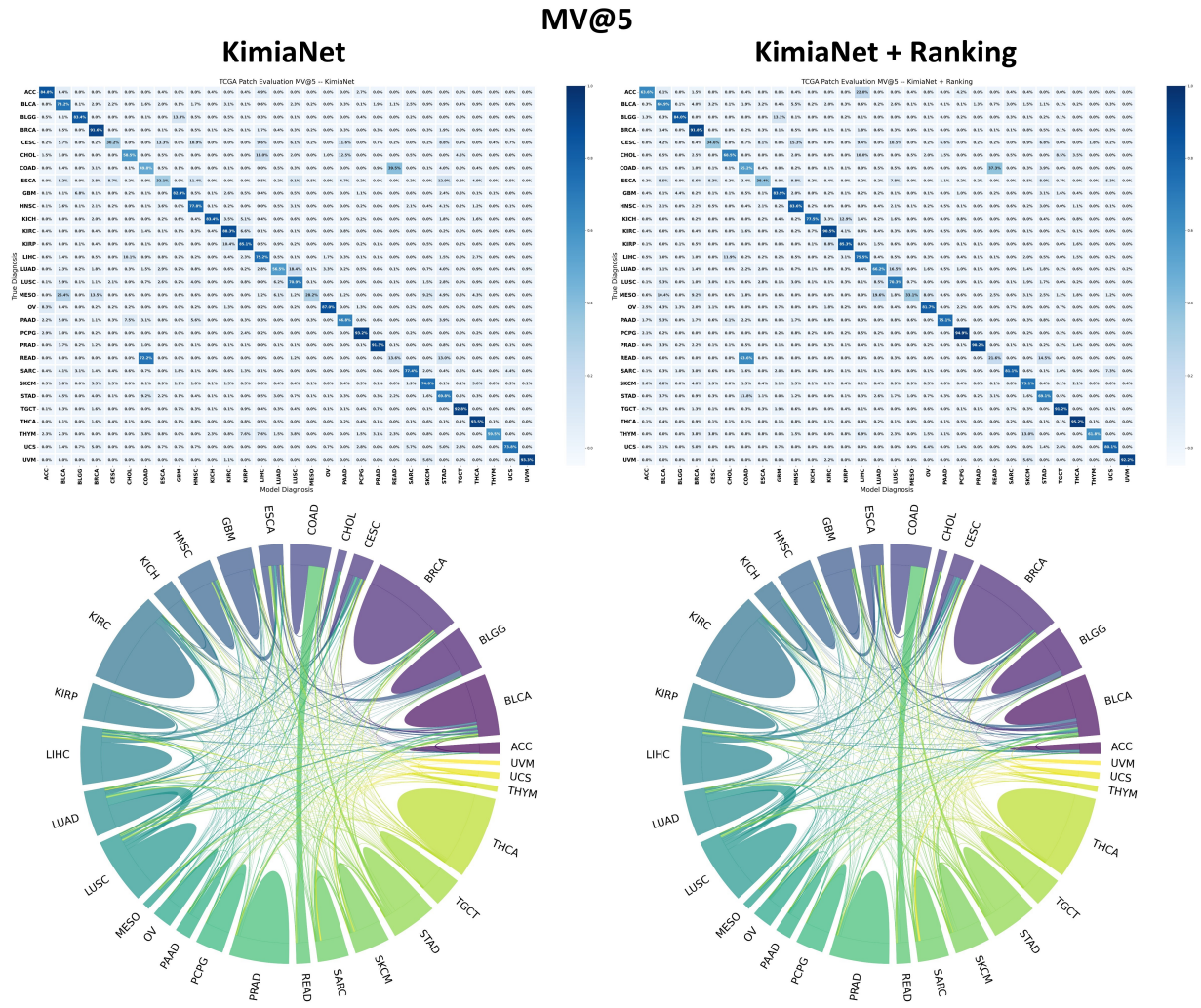

Figure 4.9: Confusion matrices and chord diagrams from KimiaNet (left column), and KimiaNet + Ranking (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the TCGA Patch dataset.

NeXtPath

Relative to KimiaNet [6], NeXtPath was also trained using the cross-entropy loss function on more than 240,000 high-cellularity patches (prepared by Riasatin et al. [6]) of dimensions 1000x1000. This suggests that NeXtPath, like KimiaNet, operates with a broader

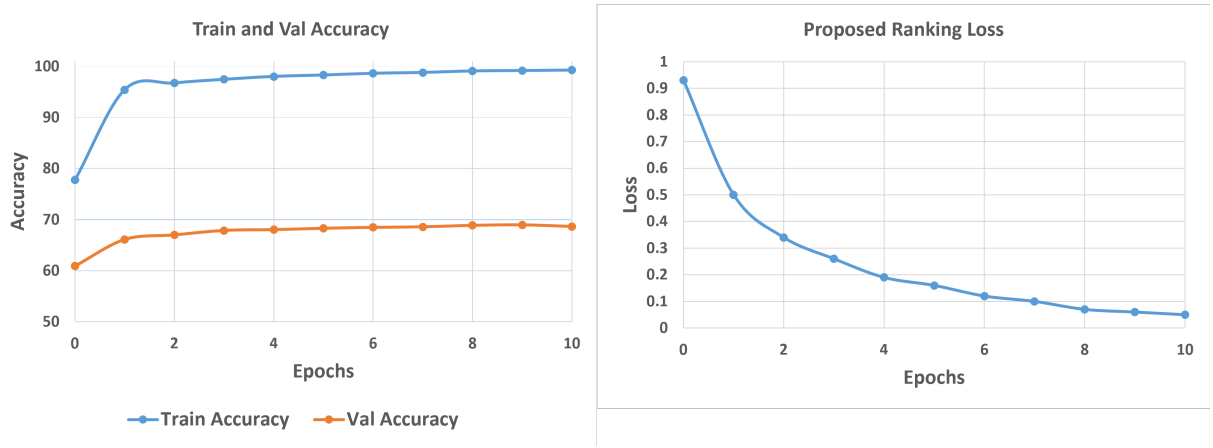

Figure 4.10: (left) Training and validation accuracy for the NeXtPath + Ranking (fine-tuned with the proposed ranking loss), exhibited a peak training accuracy of 99.25% and a maximum validation accuracy of 68.95%. (right) Training loss over a span of 10 epochs using the proposed ranking loss function tailored specifically for search and matching.

field of view. The initial training process with cross-entropy loss has been explained in Section 4.1.1. In the context of top 1 retrieval (Top-1) accuracy, NeXtPath demonstrates a modest advantage of +1% over KimiaNet. However, the performances converge when evaluating the majority accuracy among the top 3 retrievals (MV@3) and the top 5 retrievals (MV@5). In a deeper dive into the macro averages of F1-scores, KimiaNet consistently outpaces NeXtPath by +1% across Top-1, MV@3, and MV@5 retrievals. In contrast, while assessing the weighted average of F1-scores, both models showcase comparable results for Top-1 and MV@3 retrievals. Nevertheless, in the MV@5 retrieval, KimiaNet secures a lead of +1% over NeXtPath. The results can be compared in Figure 4.8 and Figure 4.12.

In this research, we endeavored to assess the quality of representations generated by the baseline NeXtPath, previously adapted using the cross-entropy loss function. To this end, I further refined the same network with the introduction of a newly proposed Ranking loss. This loss mechanism, a novel contribution, has been meticulously crafted to optimize representation learning tailored to histological matching tasks. For the training and validation phases, a consistent dataset was utilized, comprising over 240,000 training samples and more than 24,000 validation samples. Utilizing ConvNeXt-Tiny [7] core architecture, the last two stages underwent a ten-epoch fine-tuning regimen. The process was orchestrated with an initial learning rate of 0.00001, complemented by a decay rate of 0.01. The AdamW optimizer drove the optimization, with the NeXtPath's weights serving as the foundational weight setup. In an effort to foster model resilience and broader applicability, an array of

data augmentation techniques was deployed. Specifically, these involved random rotations limited to 90, 270, and -90 degrees with a 50% probability, vertical and horizontal flips (each with a 50% probability), and random cropping with potential sizes being 224, 384, 512, and 786, and a 20% probability, subsequently resized to the dimensions of 1000×1000. The dataset also underwent color variations, characterized by variations in brightness (0.2), contrast (0.2), saturation (0.1), and hue (0.05), with each adjustment bearing a 20% probability. Figure 4.10 visualizes NeXtPath's convergence patterns throughout the training duration, delineating both the training and validation accuracy curves, alongside the corresponding loss curve when leveraging the proposed ranking loss. Within this training spectrum, the apex of training accuracy was registered at 99.25%, with the validation accuracy peaking at 68.95%. To ascertain maximum generalizability, model weights were retained based on the zenith of validation performance, emphasizing the separation of validation data from the training phase. This best model coincided with a validation accuracy of 68.95%. In the following sections of this analysis, the version of the NeXtPath model that has been fine-tuned using the proposed ranking loss will be denoted as "NeXtPath + Ranking". Figure 4.11 provides a visual representation of the latent space mappings for the training, validation, and testing datasets as procured from both the original NeXtPath and its NeXtPath + Ranking counterpart.

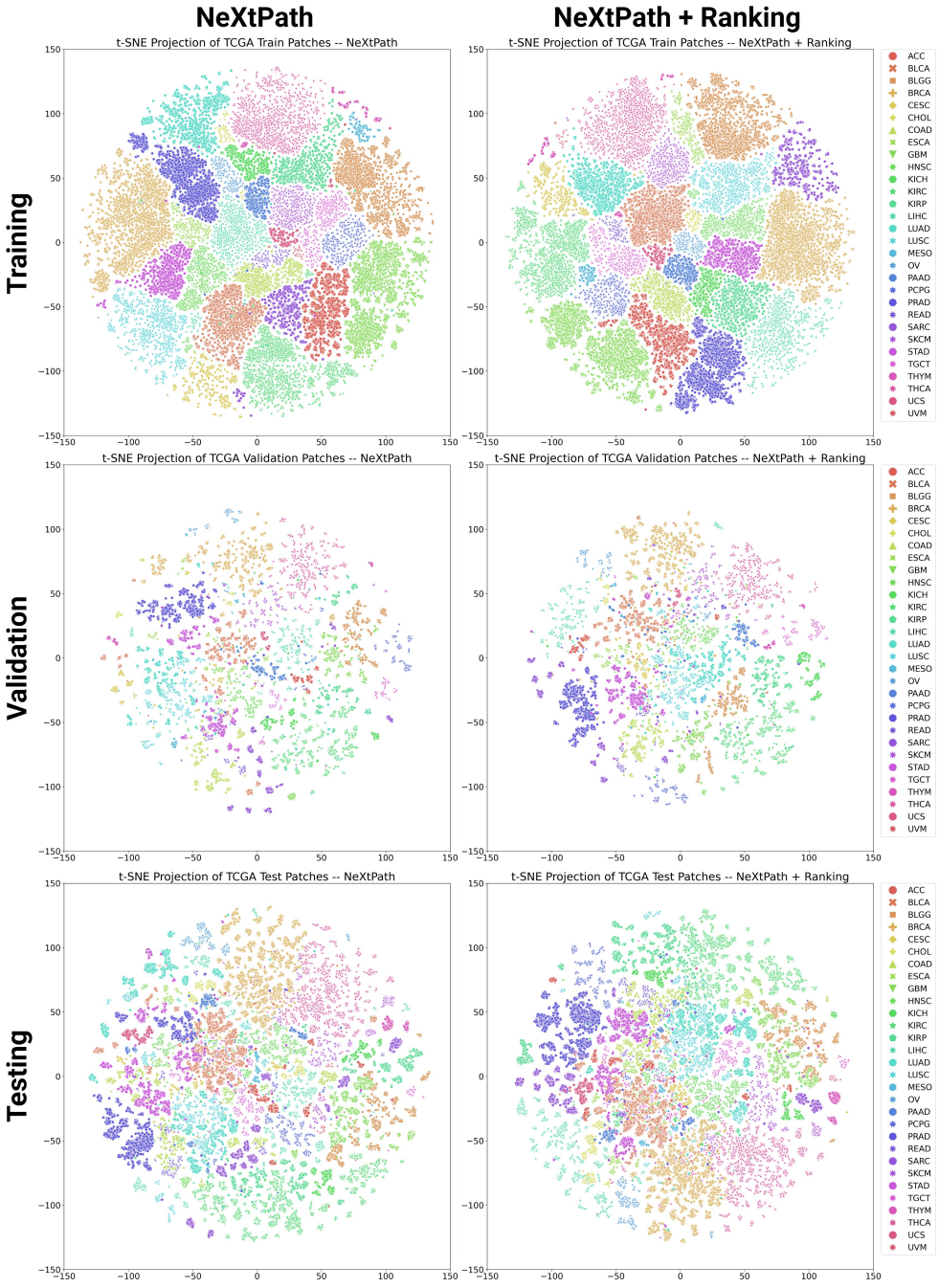

Figure 4.11: t-SNE projections for all the embeddings of 240,527 training patches, 24,492 validation patches, and 110,032 test patches from NeXtPath and NeXtPath + Ranking.

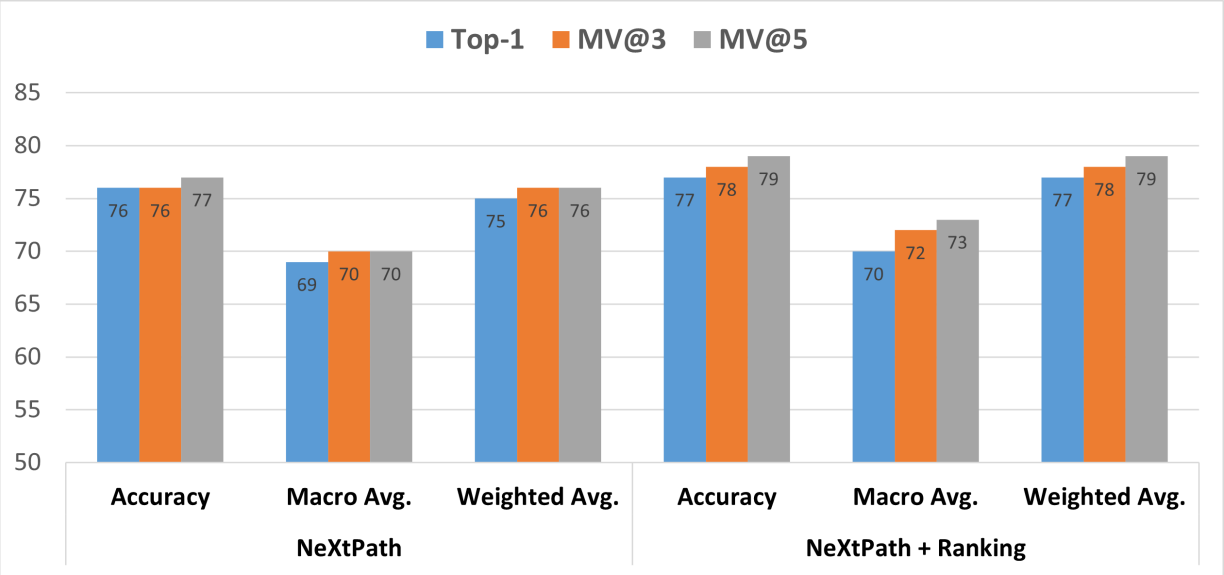

Figure 4.12: Accuracy, macro average of f1-scores, and weighted average of f1-scores are shown for the patch matching when using features from NeXtPath (fine-tuned using cross-entropy loss), and NeXtPath + Ranking (fine-tuned with the proposed ranking loss). The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA test and validation patches combined when used as query and training patches as a reference atlas. NeXtPath trained with the proposed ranking loss performs slightly better than the NeXtPath trained with cross-entropy loss.

A retrieval-based analysis was executed to evaluate the efficacy of NeXtPath (fine-tuned using cross-entropy loss) in comparison with NeXtPath + Ranking (fine-tuned using the proposed ranking loss). Leveraging both query and atlas cohorts, the objective was to identify and retrieve the cases that bore the highest similarity, thereby gauging the relative performance of the model in such a matching paradigm. The evaluation involved multiple retrieval criteria, including the top-1 retrieval (Top-1), the majority agreement among the top 3 retrievals (MV@3), and the majority agreement among the top 5 retrievals (MV@5). The accuracy, macro average, and weighted average at Top-1, MV@3, and MV@5 are reported in Figure 4.12. Moreover, confusion matrices and chord diagrams at MV@5 are illustrated in Figure 4.13. Additional confusion matrices and chord diagrams of Top-1, and MV@3 retrievals are provided in Appendix B (see Figure B.3, and B.4, respectively). Table 4.3 and 4.4 show the detailed results including precision, recall, and f1-score for the representations generated using NeXtPath and NeXtPath + Ranking, respectively.

Furthermore, the t-SNE visualizations of the feature vectors from NeXtPath and NeXtPath + Ranking are shown in Figure 4.11.

In our analysis, we observed that the performance of representations (also referred to as embeddings) obtained from NeXtPath + Ranking surpassed that of the conventional state-of-the-art NeXtPath. In terms of model accuracy, NeXtPath + Ranking demonstrated consistently higher performance. For Top-1, MV@3, and MV@5 retrievals, we observed performance enhancements of +1%, +2%, and +2%, respectively. However, in terms of the macro-average F1-scores, NeXtPath + Ranking exhibited superior performance, surpassing the baseline by +1%, +2%, and +3% for Top-1, MV@3 and MV@5 retrievals, respectively. Additionally, for the weighted average of f1-scores, NeXtPath + Ranking shows the performance enhancement by +2%, +2%, and +3% when retrieving Top-1, MV@3, and MV@5 retrievals, respectively (see Figure. 4.12).

NeXtPath

	Top-1			MV@3			MV@5			
Primary Diagnoses	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score	Patches
ACC	0.65	0.72	0.68	0.67	0.73	0.70	0.65	0.73	0.69	264
BLCA	0.64	0.67	0.66	0.65	0.68	0.66	0.65	0.69	0.67	1383
BLGG	0.91	0.81	0.86	0.92	0.83	0.87	0.92	0.83	0.87	1030
BRCA	0.84	0.92	0.88	0.85	0.92	0.89	0.85	0.93	0.89	2113
CESC	0.44	0.25	0.32	0.48	0.25	0.33	0.47	0.24	0.32	457
CHOL	0.31	0.50	0.38	0.36	0.56	0.43	0.38	0.58	0.46	200
COAD	0.52	0.61	0.56	0.53	0.63	0.58	0.53	0.63	0.58	925
ESCA	0.38	0.31	0.34	0.38	0.29	0.33	0.37	0.29	0.32	552
GBM	0.74	0.80	0.77	0.75	0.81	0.78	0.76	0.82	0.79	800
HNSC	0.65	0.76	0.70	0.65	0.78	0.71	0.65	0.78	0.71	809
KICH	0.95	0.70	0.80	0.94	0.70	0.80	0.94	0.69	0.80	489
KIRC	0.90	0.91	0.91	0.90	0.92	0.91	0.90	0.92	0.91	2105
KIRP	0.81	0.84	0.83	0.83	0.85	0.84	0.83	0.84	0.84	821
LIHC	0.79	0.68	0.73	0.82	0.70	0.75	0.83	0.70	0.76	1312
LUAD	0.68	0.63	0.65	0.69	0.64	0.66	0.69	0.64	0.66	1028
LUSC	0.71	0.66	0.68	0.71	0.67	0.69	0.72	0.67	0.70	1510
MESO	0.58	0.18	0.27	0.54	0.17	0.25	0.60	0.16	0.25	163
OV	0.72	0.81	0.76	0.74	0.82	0.77	0.73	0.82	0.77	447
PAAD	0.67	0.72	0.69	0.67	0.72	0.69	0.70	0.72	0.71	358
PCPG	0.89	0.95	0.92	0.89	0.95	0.92	0.88	0.94	0.91	630
PRAD	0.95	0.90	0.92	0.95	0.90	0.93	0.95	0.91	0.93	1347
READ	0.10	0.11	0.11	0.10	0.11	0.11	0.10	0.10	0.10	324
SARC	0.82	0.79	0.81	0.83	0.81	0.82	0.83	0.81	0.82	703
SKCM	0.71	0.70	0.70	0.72	0.72	0.72	0.73	0.72	0.72	799
STAD	0.60	0.63	0.62	0.61	0.65	0.63	0.62	0.66	0.64	1031
TGCT	0.95	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.96	671
THCA	0.90	0.96	0.93	0.90	0.96	0.93	0.90	0.96	0.93	1859
THYM	0.93	0.53	0.68	0.93	0.53	0.68	0.90	0.55	0.68	131
UCS	0.70	0.67	0.69	0.69	0.75	0.72	0.64	0.74	0.69	141
UVM	0.83	0.97	0.89	0.89	0.94	0.92	0.91	0.93	0.92	90
Total Patches										**24492**

Table 4.3: Detailed precision, recall, f1-score, and the number of patches processed for each subtype are shown in this table using the validation patches when matched against the test patches using NeXtPath for feature extraction. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA Patch dataset.

NeXtPath + Ranking

	Top-1			MV@3			MV@5			
Primary Diagnoses	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score	Patches
ACC	0.63	0.73	0.68	0.64	0.74	0.68	0.65	0.75	0.70	264
BLCA	0.69	0.69	0.69	0.69	0.72	0.70	0.69	0.73	0.71	1383
BLGG	0.90	0.85	0.87	0.90	0.86	0.88	0.91	0.87	0.89	1030
BRCA	0.87	0.93	0.90	0.87	0.93	0.90	0.87	0.94	0.90	2113
CESC	0.54	0.35	0.43	0.60	0.40	0.48	0.63	0.44	0.52	457
CHOL	0.37	0.57	0.45	0.42	0.62	0.50	0.42	0.64	0.51	200
COAD	0.54	0.59	0.56	0.56	0.60	0.58	0.57	0.61	0.59	925
ESCA	0.51	0.37	0.43	0.54	0.37	0.44	0.57	0.41	0.48	552
GBM	0.76	0.75	0.76	0.78	0.77	0.77	0.78	0.78	0.78	800
HNSC	0.67	0.78	0.72	0.69	0.79	0.74	0.70	0.79	0.74	809
KICH	0.92	0.69	0.79	0.93	0.70	0.80	0.93	0.72	0.81	489
KIRC	0.92	0.91	0.91	0.92	0.91	0.92	0.92	0.92	0.92	2105
KIRP	0.74	0.85	0.79	0.76	0.85	0.80	0.77	0.85	0.81	821
LIHC	0.83	0.72	0.77	0.85	0.75	0.79	0.86	0.75	0.80	1312
LUAD	0.70	0.66	0.68	0.71	0.67	0.69	0.72	0.68	0.70	1028
LUSC	0.70	0.69	0.69	0.73	0.69	0.71	0.74	0.70	0.72	1510
MESO	0.48	0.14	0.22	0.60	0.18	0.28	0.63	0.20	0.30	163
OV	0.72	0.80	0.76	0.72	0.81	0.76	0.72	0.80	0.76	447
PAAD	0.74	0.74	0.74	0.78	0.74	0.76	0.81	0.73	0.77	358
PCPG	0.90	0.95	0.93	0.89	0.95	0.92	0.91	0.96	0.93	630
PRAD	0.97	0.92	0.95	0.97	0.93	0.95	0.97	0.93	0.95	1347
READ	0.07	0.08	0.07	0.09	0.10	0.09	0.09	0.09	0.09	324
SARC	0.86	0.82	0.84	0.85	0.83	0.84	0.85	0.84	0.84	703
SKCM	0.71	0.71	0.71	0.71	0.73	0.72	0.72	0.73	0.73	799
STAD	0.62	0.67	0.64	0.63	0.69	0.66	0.64	0.70	0.67	1031
TGCT	0.88	0.96	0.92	0.91	0.96	0.94	0.92	0.97	0.94	671
THCA	0.90	0.96	0.93	0.91	0.96	0.93	0.91	0.96	0.93	1859
THYM	0.84	0.55	0.66	0.91	0.56	0.70	0.89	0.56	0.69	131
UCS	0.67	0.79	0.73	0.69	0.82	0.75	0.70	0.83	0.76	141
UVM	0.94	0.92	0.93	0.95	0.92	0.94	0.95	0.92	0.94	90
Total Patches										**24492**

Table 4.4: Detailed precision, recall, f1-score, and the number of patches processed for each subtype are shown in this table using the validation patches when matched against the test patches using NeXtPath (trained with the proposed ranking loss) for feature extraction. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the TCGA Patch dataset.

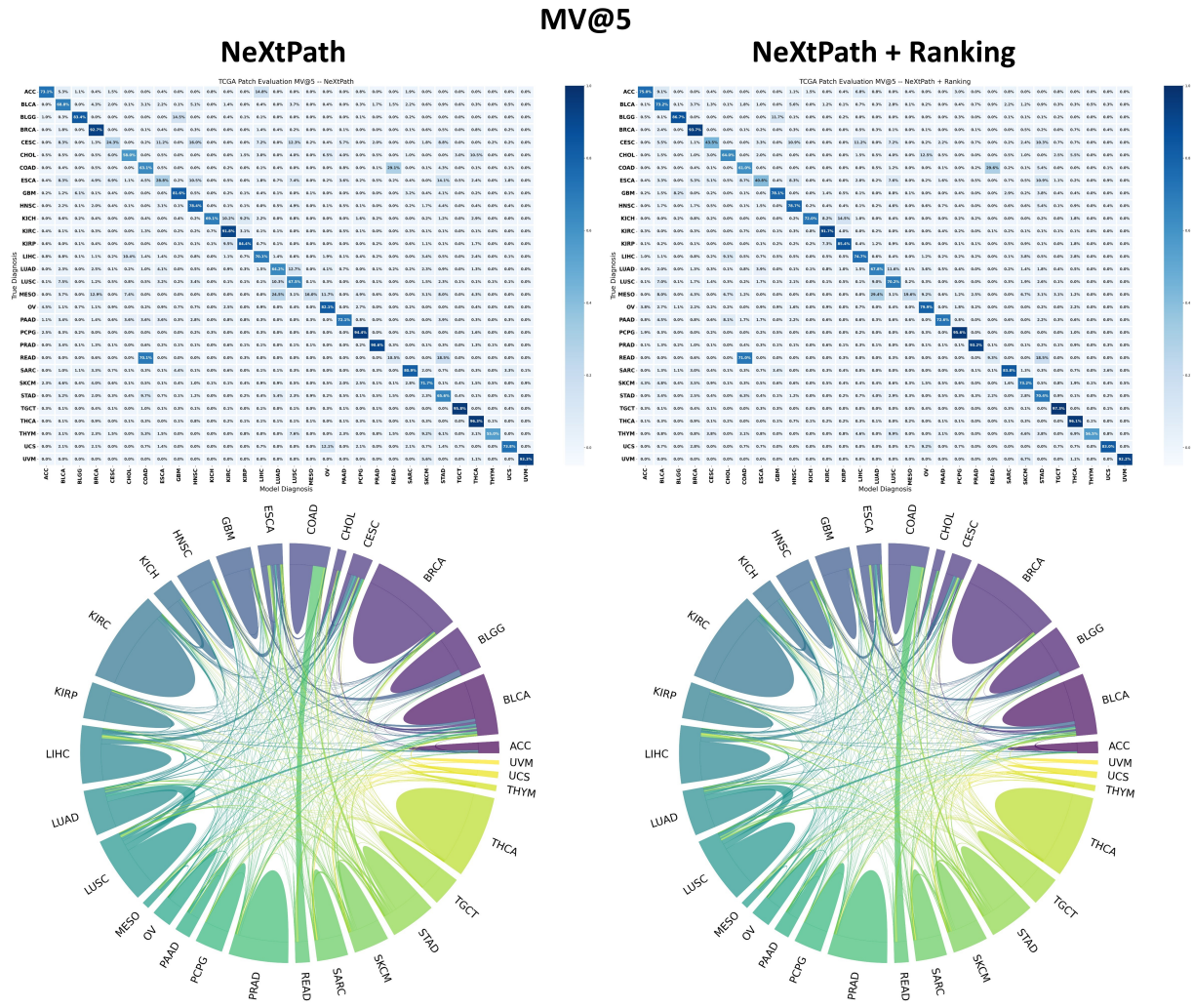

Figure 4.13: Confusion matrices and chord diagrams from KimiaNet (left column), and NeXtPath (right column) trained with the proposed ranking loss. The evaluations are based on the majority of the top 5 retrievals when evaluating the TCGA Patch dataset.

4.2.2 BReAst Carcinoma Subtyping (BRACS)

Another dataset, BRACS, was also used to evaluate the representation learning to differentiate between different subtypes of the tumors. The BRACS dataset comprises a total of 547 WSIs derived from 189 distinct patients [134]. Brancati et al. [134] provided

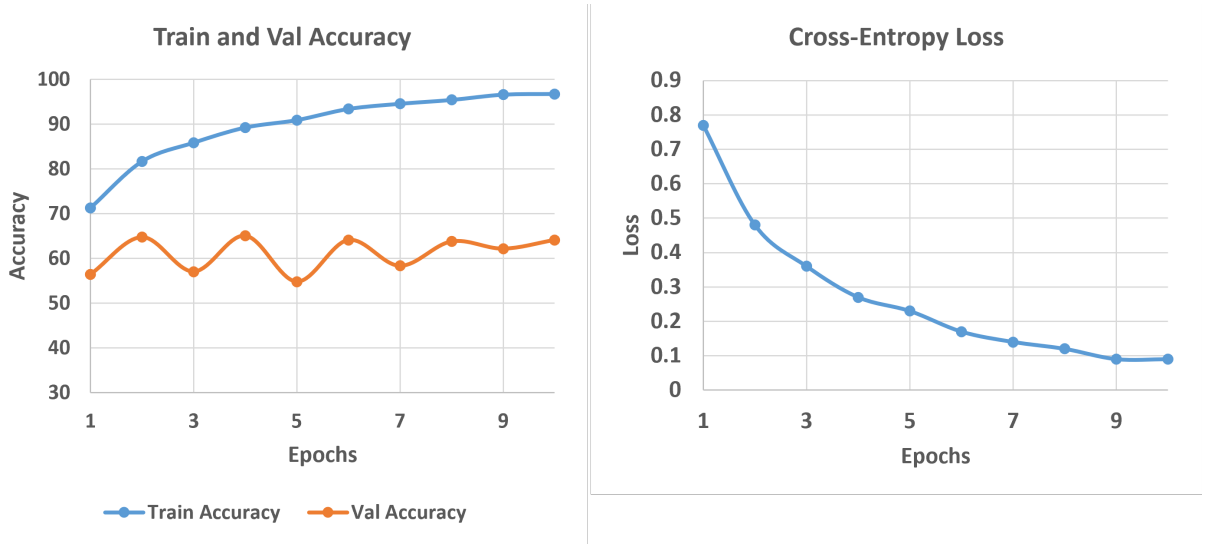

Figure 4.14: (left) Training and validation accuracy for the NeXtPath fine-tuned with BRACS ROI images over a span of 10 epochs, exhibited a peak training accuracy of 96.71% and a maximum validation accuracy of 65.06%. (right) Training loss over a span of 10 epochs using cross-entropy loss function.

the images of seven different subtypes of breast tumors extracted from these 547 WSIs. Notably, all slides have been scanned utilizing an Aperio AT2 scanner, with a resolution of 0.25 μm per pixel and a magnification factor of 40×. The dataset is categorized into two main subsets: WSI and Region of Interest (ROI). Within the WSI subset, there are three primary tumor groups [134], whereas, the ROI subset is divided into seven distinct tumor types [134]. For this study, since we are conducting a WSI-to-WSI matching, we utilized the WSI subset to perform histological matching. Table 4.5 shows more details about the data used in this experiment. In the conducted experiment, ROIs extracted from seven distinct subtypes were employed for the purpose of downstream representation learning. The NeXtPath model was harnessed for training, with fine-tuning restricted to the last stage (stage 4) of the architecture (stages can be seen in Figure. 4.1 for reference). It is noteworthy that the publicly available ROIs possess varied dimensions, a decision justified by the original authors to encompass the complete morphological patterns characteristic of each specific tumor type. Consequently, the training was executed with a batch size of one, utilizing the cross-entropy loss function. For the training, a learning rate of 0.0001, with a decay parameter of 0.01 is used with AdamW optimizer. The weights from NeXtPath (ConvNeXt-tiny fine-tuned using TCGA dataset) were used as the initial

Primary Diagnoses	Acronyms	Slides
Atypical Ductal Hyperplasia	ADH	48
Flat Epithelial Atypia	FEA	41
Normal	N	44
Pathological Benign	PB	147
Usual Ductal Hyperplasia	UDH	74
Ductal Carcinoma in Situ	DCIS	61
Invasive Carcinoma	IC	132

Table 4.5: Information concerning the BRACS dataset employed in this experiment, inclusive of the respective acronyms and the number of slides associated with each primary diagnosis.

weights. To enhance model robustness and generalizability, a comprehensive set of data augmentation techniques were integrated. These encompassed random rotations restricted to angles of 90, 270, and -90 degrees, applied with a 50% probability. Furthermore, vertical and horizontal flips (each with a 50% probability) were also used. Additionally, the dataset underwent random color jittering, characterized by variations in brightness (0.2), contrast (0.2), saturation (0.1), and hue (0.05), with each adjustment bearing a 20% probability of application. Figure 4.14 shows the convergence behavior of NeXtPath with training accuracy and loss curve using cross-entropy. The highest training and validation accuracy is 96.71% and 65.06%, respectively recorded during the training process.

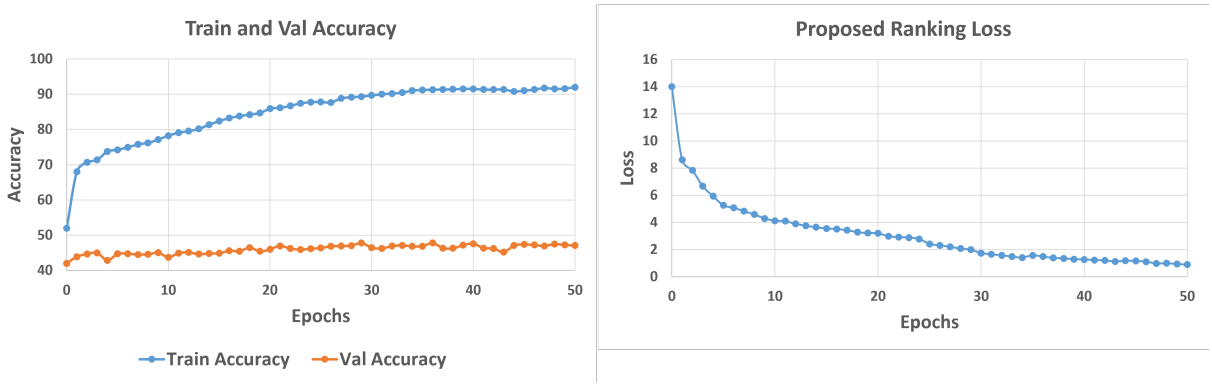

Figure 4.15: (left) Training and validation accuracy for the autoencoder fine-tuned with BRACS ROI embeddings from NeXtPath over a span of 50 epochs, exhibited a peak training accuracy of 92.00% and a maximum validation accuracy of 47.81%. (right) Training loss over a span of 50 epochs using the proposed ranking loss function.

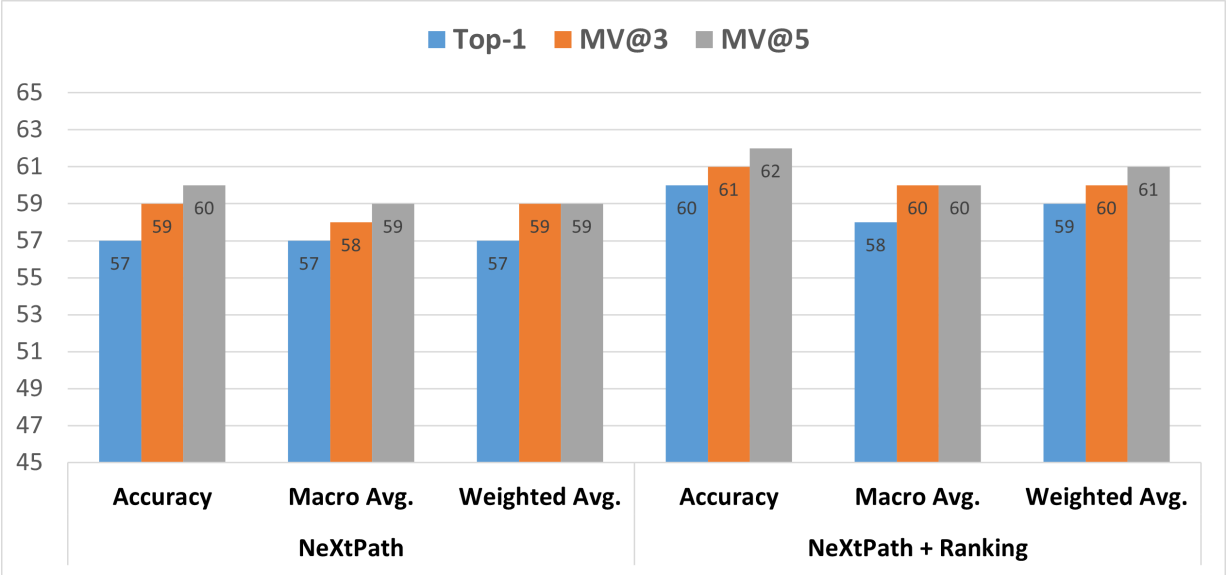

Figure 4.16: Accuracy, macro average of f1-scores, and weighted average of f1-scores are shown for the patch matching when using features from NeXtPath, and NeXtPath trained with the proposed ranking loss. The evaluations are based on the top 1 retrieval (Top-1), the majority among the top 3 retrievals (MV@3), and the majority among the top 5 retrievals (MV@5) using the BRACS test and validation patches combined when used as query and training patches as a reference atlas. NeXtPath trained with the proposed ranking loss performs slightly better than the NeXtPath trained with cross-entropy loss.

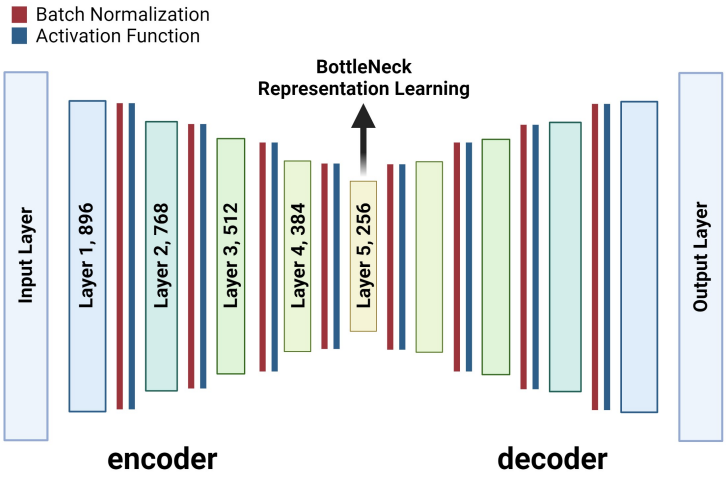

Figure 4.17: The architecture of the autoencoder used to learn distinct representations. In this study, RELU is used as the activation function

Following the preliminary training phase, feature representations were extracted using the NeXtPath model trained on BRACS ROIs. These extracted features were subsequently employed for the histological matching evaluation. A total of 3,657 training ROIs served as our reference (or atlas), while a combined set of 882 test and validation ROIs were designated as the query dataset. As a benchmark, the accuracy values obtained were 57%, 59%, and 60% for the Top-1, MV@3, and MV@5 retrieval metrics, respectively (see Figure 4.16). To optimize representation learning through the proposed ranking loss function, it necessitated training with a significantly larger batch size, ideally thrice the count of distinct subtypes. However, this was rendered impractical owing to the non-uniform patch sizes present within the dataset. Consequently, we opted to employ an autoencoder (see Figure 4.17 for autoencoder architecture), leveraging the feature representations derived from the NeXtPath model. The autoencoder's structural design incorporates linear layers, complemented by batch normalization processes and ReLU activation functions. These same representations were previously used to establish the initial histological matching baseline. Autoencoding is a well-established technique for representation learning. It employs an encoder that translates an input into a latent space, and a decoder that subsequently reconstructs the original input from this latent representation [139].

For the training, the proposed ranking loss used feature embeddings obtained from the output layer of the autoencoder for metric learning. Notably, the most generalized and compact representation from an autoencoder is typically procured from the bottleneck region of the autoencoder (output of the encoder). As an initial step, the autoencoder

was trained to replicate its input embeddings as output, utilizing the mean squared error loss function. This was facilitated through the AdamW optimizer, set at a learning rate of 0.0001 and accompanied by a decay rate of 0.01. Over a duration of 500 epochs, the autoencoder was trained to accurately reconstruct its input with a batch size of 256. Subsequent to this phase, the autoencoder underwent further refinement using our proposed loss function, with the objective of discerning and distinguishing the representations of varied subtypes. The autoencoder was trained for 50 epochs with a batch size of 256 and retained the AdamW optimizer, albeit at a reduced learning rate of 0.00001 and a decay of 0.01. Figure 4.15 shows the convergence behavior of the autoencoder with training accuracy and loss curve using the proposed ranking loss. The highest training and validation accuracy is 92.00% and 47.81%, respectively recorded during the training process. The weights of the autoencoder were saved at the highest validation accuracy for better generalization, in our case, the weights of the autoencoder were saved at 47.81% validation accuracy. Figure 4.18 provides a visual representation of the latent space mappings for the training set, and test & validation set combined as procured from both the NeXtPath and NeXtPath + Ranking (bottleneck of the autoencoder).

A retrieval-based analysis was executed to evaluate the efficacy of NeXtPath (fine-tuned using cross-entropy loss) in comparison with NeXtPath + Ranking (fine-tuned autoencoder using the proposed ranking loss). Leveraging both query and atlas cohorts, the objective was to identify and retrieve the cases that bore the highest similarity, thereby gauging the relative performance of the model in such a matching paradigm. The evaluation involved multiple retrieval criteria, including the top-1 retrieval (Top-1), the majority agreement among the top 3 retrievals (MV@3), and the majority agreement among the top 5 retrievals (MV@5). The accuracy, macro average, and weighted average of F1-scores at Top-1, MV@3, and MV@5 are reported in Figure 4.16. Moreover, confusion matrices and chord diagrams at MV@5 are illustrated in Figure 4.19. Additional confusion matrices and chord diagrams of Top-1, and MV@3 retrievals are provided in Appendix B (see Figure B.5, and B.6, respectively). Table 4.6 shows the detailed results including precision, recall, and f1-score for the representations generated using NeXtPath and NeXtPath + Ranking, respectively. Furthermore, the t-SNE visualizations of the feature vectors from NeXtPath and NeXtPath + Ranking are shown in Figure 4.18.

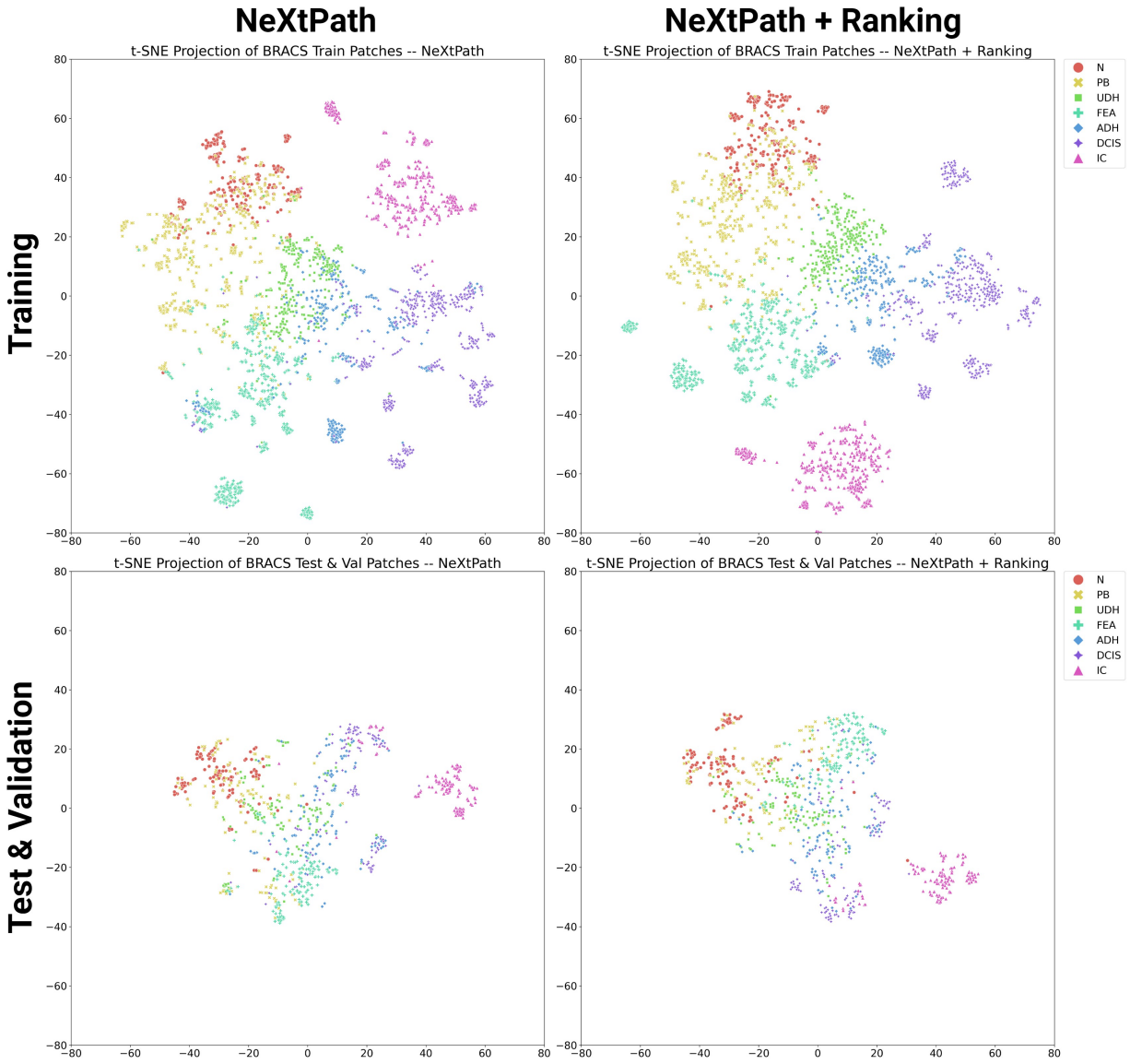

Figure 4.18: t-SNE projections for all the embeddings of 3657 training patches, and 882 test & validation patches combined from NeXtPath and NeXtPath + Ranking (bottleneck of autoencoder trained using the proposed ranking loss).

In our analysis, we observed that the performance of embeddings obtained from NeXtPath + Ranking (bottleneck of the autoencoder) surpassed that of the conventional state-of-the-art NeXtPath. In terms of model accuracy, NeXtPath + Ranking demonstrated consis-

tently higher performance. For Top-1, MV@3, and MV@5 retrievals, we observed performance enhancements of +3%, +2%, and +2%, respectively. However, in terms of the macro-average F1-scores, NeXtPath + Ranking exhibited superior performance, surpassing the baseline by +1%, +2%, and +1% for Top-1, MV@3 and MV@5 retrievals, respectively. Additionally, for the weighted average of f1-scores, NeXtPath + Ranking shows the performance enhancement by +2%, +1%, and +2% when retrieving Top-1, MV@3, and MV@5 retrievals, respectively (see Figure. 4.16).

		Top-1			**MV@3**			**MV@5**			
	Primary Diagnoses	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score	ROIs
NeXtPath	N	0.61	0.65	0.63	0.64	0.72	0.68	0.64	0.71	0.67	127
	PB	0.43	0.49	0.46	0.43	0.49	0.46	0.43	0.50	0.46	122
	UDH	0.40	0.34	0.37	0.42	0.34	0.37	0.44	0.35	0.39	128
	FEA	0.72	0.79	0.75	0.70	0.80	0.75	0.70	0.80	0.74	132
	ADH	0.38	0.33	0.35	0.41	0.33	0.37	0.43	0.33	0.37	120
	DCIS	0.53	0.55	0.54	0.56	0.59	0.58	0.54	0.59	0.56	125
	IC	0.89	0.86	0.88	0.91	0.87	0.89	0.93	0.87	0.90	128
Total ROIs											**882**
NeXtPath + Ranking	N	0.66	0.72	0.69	0.67	0.75	0.71	0.70	0.80	0.74	127
	PB	0.42	0.52	0.46	0.44	0.52	0.48	0.47	0.50	0.49	122
	UDH	0.46	0.34	0.39	0.47	0.34	0.40	0.47	0.37	0.41	128
	FEA	0.67	0.83	0.74	0.67	0.86	0.75	0.65	0.85	0.74	132
	ADH	0.40	0.31	0.35	0.41	0.30	0.35	0.42	0.31	0.36	120
	DCIS	0.58	0.57	0.57	0.58	0.60	0.59	0.59	0.62	0.60	125
	IC	0.91	0.86	0.88	0.93	0.87	0.90	0.93	0.86	0.89	128
Total ROIs											**882**

Table 4.6: Precision, recall, F1-score, and the number of slides processed for each subtype are reported in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the BRACS dataset.

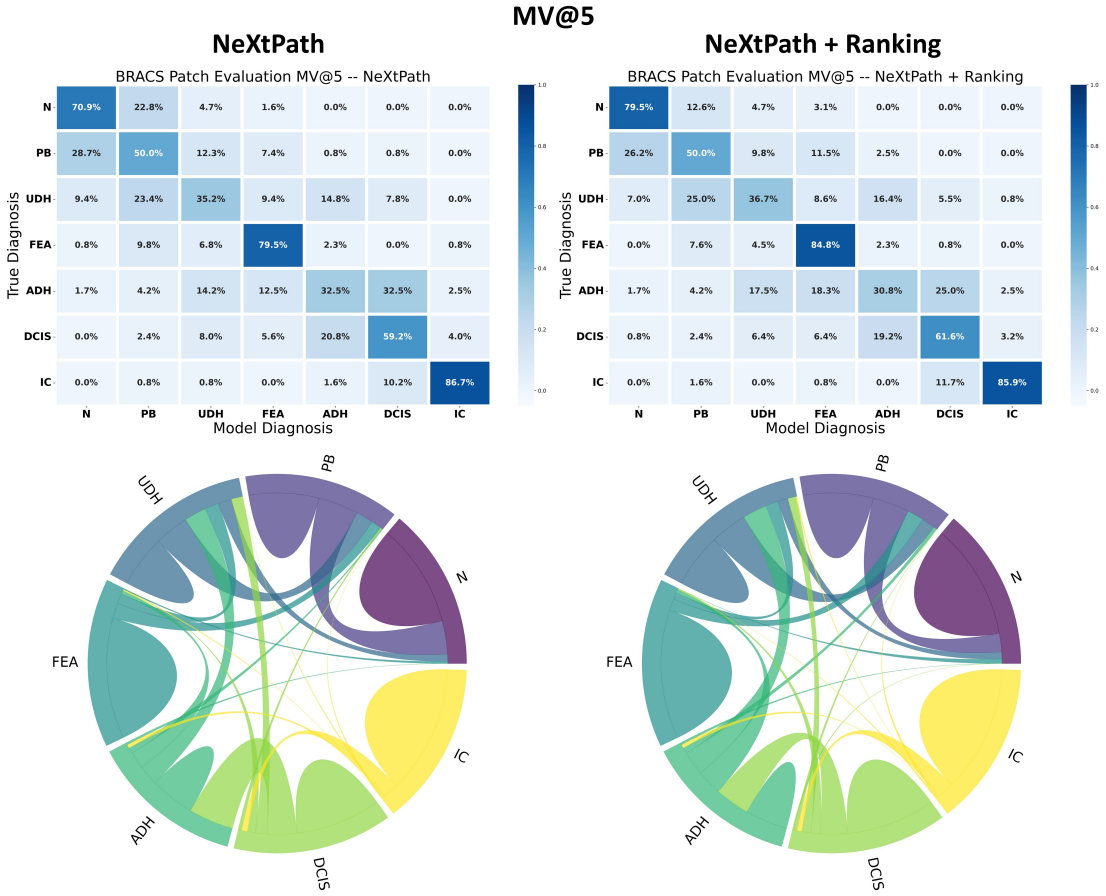

Figure 4.19: Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) autoencoder trained with the proposed ranking loss. The evaluations are based on the majority of the top 5 retrievals when evaluating the BRACS ROI dataset.

4.3 Discussion & Conclusion

Precise image retrieval is crucial in the context of histopathology, where the subtleties of tissue morphology can be indicative of diverse pathologies. Discriminative representation learning ensures that the nuanced patterns within the histological images are captured and emphasized, allowing for a nuanced and sophisticated search through the archives. Consequently, this facilitates a more informed and reliable virtual second opinion, grounding it

in a vast database of historical antecedents.

To address the targeted objective, this study presented a modified ranking loss function. This tailored loss function is specifically optimized for image retrieval tasks within the realm of histopathology. Histopathological image analysis presents a greater level of complexity compared to the analysis of natural images, due to the intricate cellular patterns and the subtle variations in tissue structure that are characteristic of such medical imagery. The representation learning methodology proposed in this study was rigorously tested using diagnostic slides from TCGA, encompassing 30 distinct types of tumors. This evaluation was conducted using two separate architectural models (KimiaNet and NeXtPath). Across both principal experimental conditions, the networks fine-tuned with the novel proposed ranking loss demonstrated superior performance when compared to their counterparts trained with the widely used cross-entropy loss. Subsequent experimentation was conducted to corroborate the results obtained from the TCGA dataset analysis. For this purpose, the BRACS [134] dataset, which includes seven distinct breast tumor categories, was utilized. The heterogeneity of image dimensions within this dataset presented a significant challenge for batch processing during training. As a result, training sessions had to be conducted with a batch size of one while fine-tuning with the cross-entropy loss. Conversely, when fine-tuning with the proposed ranking loss, an autoencoder was implemented to facilitate training on feature vectors. These vectors were initially extracted using the NeXtPath model, which had been previously fine-tuned with Regions of Interest (ROIs) from the BRACS [134] dataset. The outcomes of this additional experiment reinforced the initial findings; the representations derived post-training with the proposed ranking loss consistently outperformed those obtained through the traditional cross-entropy loss. This underscores the ranking loss's efficacy in yielding more discriminative feature representations, thereby enhancing the system's capability for precise histological matching within the complex milieu of breast tumor histopathology.

Comprehending both spatial and temporal components within the feature vector is crucial for a thorough analysis of the representations produced by the neural network. In this study, two distinct CNN frameworks have been used, specifically DenseNet-121 [5], under the adaptation KimiaNet [6] tailored for histopathological analysis, and ConvNeXt-tiny [7], modified into NeXtPath, also specialized for histological applications. Following the training regimen of NeXtpath with cross-entropy loss (similar to the KimiaNet [6] for fair comparison), we proceeded to an analytical assessment of its derived feature representations. These were benchmarked against KimiaNet's representations, given that both models were trained using a congruent dataset. To delve deeper into the intricacies of the representation vectors of the histopathological images, principal component analysis was employed. Further insight was achieved by visualizing these vectors using the t-SNE

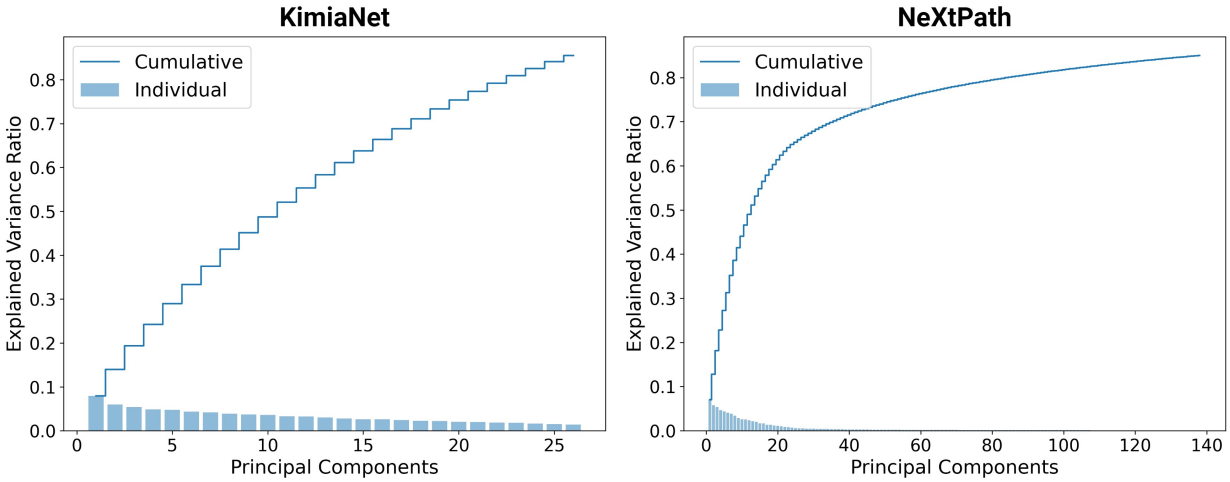

Figure 4.20: 85% of the variance explained by the top principle components (PCs). Left: The top 26 PCs from the KimiaNet embeddings. Right: The top 138 PCs from the NeXtPath embeddings contributed 85% of the variance.

embedding projections. Figure 4.20 shows the variance of the top principle components (PCs) of the feature vectors from KimiaNet and NextPath (fine-tuned using prevalWent cross-entropy loss), respectively. Only 26 top PCs out of 1024, which is approx. 2.5% of the total vector, explains 85% of the variance, whereas, on the other hand, 138 top PCs out of 786, which is approx. 18% of the total vector, explains 85% of the total variance of the representations acquired from the KimiaNet and NeXtPath, respectively. This variance from the KimiaNet shows that fewer number of PC contribute to the representation whereas more PCs contribute to the representation of NeXtPath. Furthermore, the training and testing with validation patch embeddings from both the networks, KimiaNet and NeXtPath, are visualized after training using t-SNE in Figure 4.7 and 4.11.

In conclusion, the research presented in this thesis highlights the critical importance of employing a specialized loss function to enable models to learn representations that are particularly discriminative for image retrieval applications. The ranking loss devised for image search purposes within this study has shown exceptional effectiveness, as evidenced by thorough experimental validation across two distinct publicly accessible datasets. The enhancement in performance relative to cross-entropy loss, conventionally used for classification endeavors, suggests the capability of the ranking loss to refine the accuracy of image retrieval frameworks within the intricate landscape of multi-class medical imaging. This augmentation is instrumental in supporting virtual second opinions, an essential component in the diagnostic process of intricate cancer cases.

In the context of future endeavors, subsequent to thorough assessments on two downstream tasks, one has to aim to subject the proposed method to evaluation by pathologists. This evaluation will involve a visual assessment of the search and matching performance of the algorithm, seeking additional insights and validation from domain experts to further enhance the robustness and applicability of the proposed approach.

Chapter 5

Summary and Conclusions

Advances in ML have shown great promise for assisting healthcare professionals. However, innovative algorithms with accurate performance are still necessary to gain trust and adoption in clinical settings. Histopathology remains the definitive modality for cancer diagnosis. The rapid incorporation of digital pathology into clinical practice underscores the potential intersections between histopathology and ML advancements. In the diagnostic process, particularly among novice pathologists or when confronting complex cases, it is not uncommon for pathologists to seek to characterize specific regions or patches within a slide. Historically, this would entail consulting experienced colleagues or engaging in a meticulous examination of specialized reference materials, in the quest for comparable visual representations. In the realm of computer vision, such challenges are addressed through CBIR. Within the medical context, and provided semantic equivalence between human and computerized assessments, this approach essentially serves as a mechanism to obtain a "virtual second opinion", offering a virtual perspective to support diagnostic decisions. In the realm of histopathology and cancer diagnosis, obtaining a second opinion stands as a critical component to reduce variability. Given the intricate and multifaceted nature of histopathological slides, interpretations can sometimes be subjective and vary among pathologists. Discrepancies in diagnoses, particularly in borderline or complex cases, can significantly impact treatment decisions and patient prognoses. A 'computational' second opinion acts as an additional safeguard, mitigating the risks of oversight or misinterpretation. It not only augments the confidence in the diagnostic process but also underscores the commitment to patient-centered care, ensuring that therapeutic decisions are grounded in a consensus of expert evaluations.

WSI search presents a pivotal mechanism for procuring a virtual second opinion. Through the integration of sophisticated CBIR methodologies, pathologists have the capability to

compare a patient's WSI against an atlas (repository) of previously diagnosed cases (empirical evidence). This facilitates the discernment of congruent histological patterns and irregularities. Such a methodology introduces an objective, data-informed lens that synergizes with the pathologist's clinical judgment, fortifying diagnostic accuracy and fostering an evidence-backed, collaborative paradigm in pathology. The task of executing searches amidst small or large repositories of gigapixel WSIs, mirroring the complexities inherent to large-scale data analytics, mandates the adoption of a structured computational strategy, epitomized by the "Divide & Conquer" principle.

In the context of the "divide" procedure, this thesis has put forward a novel unsupervised technique termed the **Selection of Distinct Morphology** (SDM). The core intent of SDM is to meticulously discern and collate distinct patches from a WSI, culminating in what we designate as a "*montage*". This constructed montage, based on one-class clustering, not only encapsulates the heterogeneity inherent within the WSI but also serves as a foundational component, instrumental for facilitating an array of advanced applications, prominently inclusive of image search. The overarching aim of SDM is to engineer a montage characterized by a reduced quantity of patches. Yet, it's vital that these selected patches encapsulate a broad morphological spectrum, ensuring they are both representative and meaningful in the broader context of the WSI. Through this approach, the montage becomes a condensed yet comprehensive visual summary, like Yottixel's mosaic, providing a streamlined perspective that retains the essential diagnostic information of the WSI.

Expounding on the "conquer" part, this thesis elucidated a cutting-edge methodology tailored for representation learning, aimed at discerning and differentiating varied morphological nuances via unique representations, leveraging ranking loss as the guiding metric. Training with the proposed **ranking loss** showcases proficiency in distinguishing these characteristics from those of alternate classes as it is tailored for the histological search and matching application in the latent space. Such a refined model not only fosters the in-depth comprehension of intricate histological patterns within a class but also ensures robust discriminative power across multiple classes, augmenting the accuracy and reliability of subsequent analyses and applications in the histopathological domain.

In conclusion, the collective research endeavors undertaken throughout the course of the Ph.D. program have coalesced into an exhaustive and application-oriented architectural framework. Explicitly tailored, this framework seeks to optimize the extraction of semantically rich representations from WSI within the domain of DP. A salient thrust of this structure is its emphasis on facilitating applications predominantly centered around image retrieval and histological matching search. This holistic approach not only advances the current methodologies in DP but also sets the stage for potential future innovations, paving the way for enhanced diagnostic accuracy and efficiency in histopathological evaluations.

References

[1] Preparation microscopic images. https://www.shutterstock.com/search/preparation+microscope. Accessed: 2022-02-24.

[2] Diagnosis. https://www.hl-info.ch/diagnose/. Accessed: 2022-02-24.

[3] 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.

[4] 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.

[5] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[6] Abtin Riasatian, Morteza Babaie, Danial Maleki, Shivam Kalra, Mojtaba Valipour, Sobhan Hemati, Manit Zaveri, Amir Safarpoor, Sobhan Shafiei, Mehdi Afshari, et al. Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides. *Medical Image Analysis*, 70:102032, 2021.

[7] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022.

[8] Talha Qaiser, Yee-Wah Tsang, Daiki Taniyama, Naoya Sakamoto, Kazuaki Nakane, David Epstein, and Nasir Rajpoot. Fast and accurate tumor segmentation of his-

tology images using persistent homology and deep convolutional features. *Medical image analysis*, 55:1–14, 2019.

[9] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2:147–171, 2009.

[10] Juan Rosai. Why microscopy will remain a cornerstone of surgical pathology. *Laboratory investigation*, 87(5):403–408, 2007.

[11] Mike May. A better lens on disease. *Scientific American*, 302(5):74–77, 2010.

[12] Kun-Hsing Yu, Andrew L Beam, and Isaac S Kohane. Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10):719–731, 2018.

[13] 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.

[14] Abubakr Shafique, Morteza Babaie, Mahjabin Sajadi, Adrian Batten, Soma Skdar, and HR Tizhoosh. Automatic multi-stain registration of whole slide images in histopathology. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3622–3625. IEEE, 2021.

[15] Abubakr Shafique, Morteza Babaie, Ricardo Gonzalez, and H.R. Tizhoosh. Immunohistochemistry biomarkers-guided image search for histopathology. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–5, 2023.

[16] Yuri Tolkach, Tilmann Dohmgörgen, Marieta Toma, and Glen Kristiansen. High-accuracy prostate cancer pathology using deep learning. *Nature Machine Intelligence*, 2(7):411–418, 2020.

[17] JD Pallua, A Brunner, B Zelger, M Schirmer, and J Haybaeck. The future of pathology is digital. *Pathology-Research and Practice*, 216(9):153040, 2020.

[18] Joann G Elmore, Gary M Longton, Patricia A Carney, Berta M Geller, Tracy Onega, Anna NA Tosteson, Heidi D Nelson, Margaret S Pepe, Kimberly H Allison, Stuart J Schnitt, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *Jama*, 313(11):1122–1132, 2015.

[19] Berta M Geller, Heidi D Nelson, Donald L Weaver, Paul D Frederick, Kimberly H Allison, Tracy Onega, Patricia A Carney, Anna NA Tosteson, and Joann G Elmore. Characteristics associated with requests by pathologists for second opinions on breast biopsies. *Journal of clinical pathology*, 70(11):947–953, 2017.

[20] Jeroen Van der Laak, Geert Litjens, and Francesco Ciompi. Deep learning in histopathology: the path to the clinic. *Nature medicine*, 27(5):775–784, 2021.

[21] Histopathology. https://en.wikipedia.org/wiki/Histopathology. Accessed: 2023-09-05.

[22] Pathology. https://en.wikipedia.org/wiki/Pathology. Accessed: 2022-02-25.

[23] Histology. https://en.wikipedia.org/wiki/Histology. Accessed: 2022-02-25.

[24] Kenneth A Fleming, Mahendra Naidoo, Michael Wilson, John Flanigan, Susan Horton, Modupe Kuti, Lai Meng Looi, Christopher P Price, Kun Ru, Abdul Ghafur, et al. High-quality diagnosis: an essential pathology package. 2018.

[25] Reed T Sutton, David Pincock, Daniel C Baumgart, Daniel C Sadowski, Richard N Fedorak, and Karen I Kroeker. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine*, 3(1):1–10, 2020.

[26] Andreas Kårsnäs. *Image analysis methods and tools for digital histopathology applications relevant to breast cancer diagnosis*. PhD thesis, Acta Universitatis Upsaliensis, 2014.

[27] Tatyana S Gurina and Lary Simms. Histology, staining. 2020.

[28] Hani A Alturkistani, Faris M Tashkandi, and Zuhair M Mohammedsaleh. Histological stains: a literature review and case study. *Global journal of health science*, 8(3):72, 2016.

[29] Thaína A Azevedo Tosta, Paulo Rogério de Faria, Leandro Alves Neves, and Marcelo Zanchetta do Nascimento. Computational normalization of h&e-stained histological images: Progress, challenges and future potential. *Artificial intelligence in medicine*, 95:118–132, 2019.

[30] Shaimaa Al-Janabi, André Huisman, and Paul J Van Diest. Digital pathology: current status and future perspectives. *Histopathology*, 61(1):1–9, 2012.

[31] Jan G Van den Tweel and Clive R Taylor. A brief history of pathology. *Virchows Archiv*, 457(1):3–10, 2010.

[32] Liron Pantanowitz. Digital images and the future of digital pathology. *Journal of pathology informatics*, 1, 2010.

[33] Humayun Irshad, Antoine Veillard, Ludovic Roux, and Daniel Racoceanu. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. *IEEE reviews in biomedical engineering*, 7:97–114, 2013.

[34] 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):1–15, 2020.

[35] Abubakr Shafique, Morteza Babaie, Ricardo Gonzalez, Adrian Batten, Soma Sikdar, and H.R. Tizhoosh. Composite biomarker image for advanced visualization in histopathology. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–4, 2023.

[36] S. Maryam Hosseini, Abubakr Shafique, Morteza Babaie, and H. R. Tizhoosh. Class-imbalanced unsupervised and semi-supervised domain adaptation for histopathology images. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–7, 2023.

[37] Takumi Ishikawa, Junko Takahashi, Mai Kasai, Takayuki Shiina, Yuka Iijima, Hiroshi Takemura, Hiroshi Mizoguchi, and Takeshi Kuwata. Support system for pathologists and researchers. *Journal of pathology informatics*, 6, 2015.

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

[39] Juan C Caicedo, Fabio A González, and Eduardo Romero. Content-based histopathology image retrieval using a kernel-based semantic annotation framework. *Journal of biomedical informatics*, 44(4):519–528, 2011.

[40] Kirsten L Weind, Cynthia F Maier, Brian K Rutt, and Madeleine Moussa. Invasive carcinomas and fibroadenomas of the breast: comparison of microvessel distributions–implications for imaging modalities. *Radiology*, 208(2):477–483, 1998.

[41] PH Bartels, D Thompson, M Bibbo, and JE Weber. Bayesian belief networks in quantitative histopathology. *Analytical and quantitative cytology and histology*, 14(6):459–473, 1992.

[42] PW Hamilton, N Anderson, PH Bartels, and D Thompson. Expert system support using bayesian belief networks in the diagnosis of fine needle aspiration biopsy specimens of the breast. *Journal of clinical pathology*, 47(4):329–336, 1994.

[43] Linda T Kohn, JM Corrigan, MS Donaldson, et al. Institute of medicine. to err is human: building a safer health system, 2000.

[44] Kevin M Boehm, Pegah Khosravi, Rami Vanguri, Jianjiong Gao, and Sohrab P Shah. Harnessing multimodal data integration to advance precision oncology. *Nature Reviews Cancer*, pages 1–13, 2021.

[45] Sudeep Gaudi, J Manuel Zarandona, Stephen S Raab, Joseph C English III, and Drazen M Jukic. Discrepancies in dermatopathology diagnoses: the role of second review policies and dermatopathology fellowship training. *Journal of the American Academy of Dermatology*, 68(1):119–128, 2013.

[46] MCRF Van Dijk, KKH Aben, F Van Hees, A Klaasen, WAM Blokx, LALM Kiemeney, and DJ Ruiter. Expert review remains important in the histopathological diagnosis of cutaneous melanocytic lesions. *Histopathology*, 52(2):139–146, 2008.

[47] Berta M Geller, Paul D Frederick, Stevan R Knezevich, Jason P Lott, Heidi D Nelson, Linda J Titus, Patricia A Carney, Anna NA Tosteson, Tracy L Onega, Raymond L Barnhill, et al. Pathologists' use of second opinions in interpretation of melanocytic cutaneous lesions: policies, practices, and perceptions. *Dermatologic surgery: official publication for American Society for Dermatologic Surgery [et al.]*, 44(2):177, 2018.

[48] Chetan L Srinidhi, Ozan Ciga, and Anne L Martel. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, 67:101813, 2021.

[49] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[50] 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.

[51] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[52] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems*, 27, 2014.

[53] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[54] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[55] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

[56] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[57] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[58] Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10428–10436, 2020.

[59] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[60] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold,

Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[61] 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.

[62] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

[63] Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R Scott. Multi-similarity loss with general pair weighting for deep metric learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5022–5030, 2019.

[64] Sungyeon Kim, Dongwon Kim, Minsu Cho, and Suha Kwak. Proxy anchor loss for deep metric learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3238–3247, 2020.

[65] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE, 2006.

[66] Elad Hoffer and Nir Ailon. Deep metric learning using triplet network. In *Similarity-Based Pattern Recognition: Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015. Proceedings 3*, pages 84–92. Springer, 2015.

[67] Xinwei He, Yang Zhou, Zhichao Zhou, Song Bai, and Xiang Bai. Triplet-center loss for multi-view 3d object retrieval. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1945–1954, 2018.

[68] Marc T Law, Nicolas Thome, and Matthieu Cord. Quadruplet-wise image similarity learning. In *Proceedings of the IEEE international conference on computer vision*, pages 249–256, 2013.

[69] Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4004–4012, 2016.

[70] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. *Advances in neural information processing systems*, 29, 2016.

[71] Evgeniya Ustinova and Victor Lempitsky. Learning deep embeddings with histogram loss. *Advances in neural information processing systems*, 29, 2016.

[72] Jian Wang, Feng Zhou, Shilei Wen, Xiao Liu, and Yuanqing Lin. Deep metric learning with angular loss. In *Proceedings of the IEEE international conference on computer vision*, pages 2593–2601, 2017.

[73] Chao-Yuan Wu, R Manmatha, Alexander J Smola, and Philipp Krahenbuhl. Sampling matters in deep embedding learning. In *Proceedings of the IEEE international conference on computer vision*, pages 2840–2848, 2017.

[74] Weifeng Ge. Deep metric learning with hierarchical triplet loss. In *Proceedings of the European conference on computer vision (ECCV)*, pages 269–285, 2018.

[75] Xingping Dong and Jianbing Shen. Triplet loss in siamese network for object tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pages 459–474, 2018.

[76] Fatih Cakir, Kun He, Xide Xia, Brian Kulis, and Stan Sclaroff. Deep metric learning to rank. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1861–1870, 2019.

[77] Mete Kemertas, Leila Pishdad, Konstantinos G Derpanis, and Afsaneh Fazly. Rankmi: A mutual information maximizing ranking loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14362–14371, 2020.

[78] Pooria Mazaheri, Azam Asilian Bidgoli, Shahryar Rahnamayan, and Hamid Reza Tizhoosh. Ranking loss and sequestering learning for reducing image search bias in histopathology. *Applied Soft Computing*, 142:110346, 2023.

[79] Shivam Kalra. Content-based image retrieval of gigapixel histopathology scans: a comparative study of convolution neural network, local binary pattern, and bag of visual words. Master's thesis, University of Waterloo, 2018.

[80] Scott Doyle, Mark Hwang, Shivang Naik, Michael Feldman, John Tomaszeweski, and Anant Madabhushi. Using manifold learning for content-based image retrieval of prostate histopathology. In *MICCAI 2007 Workshop on Content-based Image*

Retrieval for Biomedical Image Archives: Achievements, Problems, and Prospects, pages 53–62. Citeseer, 2007.

[81] Jennifer G. Dy, Carla E. Brodley, Avi Kak, Lynn S. Broderick, and Alex M. Aisen. Unsupervised feature selection applied to content-based retrieval of lung images. *IEEE transactions on pattern analysis and machine intelligence*, 25(3):373–378, 2003.

[82] Akshay Sridhar. *Content-based image retrieval of digitized histopathology via boosted spectral embedding (BoSE)*. Rutgers The State University of New Jersey-New Brunswick and University of Medicine and Dentistry of New Jersey, 2012.

[83] Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International journal of medical informatics*, 73(1):1–23, 2004.

[84] 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.

[85] Xiaoshuang Shi, Fuyong Xing, KaiDi Xu, Yuanpu Xie, Hai Su, and Lin Yang. Supervised graph hashing for histopathology image retrieval and classification. *Medical image analysis*, 42:117–128, 2017.

[86] Abubakr Shafique, Ricardo Gonzalez, Liron Pantanowitz, Puay Hoon Tan, Alberto Machado, Ian A Cree, and Hamid R Tizhoosh. A preliminary investigation into search and matching for tumor discrimination in world health organization breast taxonomy using deep networks. *Modern Pathology*, 37(2):100381, 2024.

[87] 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.

[88] David E Malarkey, Gabrielle A Willson, Cynthia J Willson, E Terence Adams, Greg R Olson, William M Witt, Susan A Elmore, Jerry F Hardisty, Michael C Boyle, Torrie A Crabbs, et al. Utilizing whole slide images for pathology peer review and working groups. *Toxicologic pathology*, 43(8):1149–1157, 2015.

[89] Yingci Liu and Liron Pantanowitz. Digital pathology: review of current opportunities and challenges for oral pathologists. *Journal of Oral Pathology & Medicine*, 48(4):263–269, 2019.

[90] Romain Mormont, Pierre Geurts, and Raphaël Marée. Comparison of deep transfer learning strategies for digital pathology. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2262–2271, 2018.

[91] Brady Kieffer, Morteza Babaie, Shivam Kalra, and Hamid R Tizhoosh. Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2017.

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

[93] Gil Patrus Pena and Joséde Souza Andrade-Filho. How does a pathologist make a diagnosis? *Archives of pathology & laboratory medicine*, 133(1):124–132, 2009.

[94] Camille Laurent, Marine Baron, Nadia Amara, Corinne Haioun, Mylène Dandoit, Marc Maynadié, Marie Parrens, Beatrice Vergier, Christiane Copie-Bergman, Bettina Fabiani, et al. Impact of expert pathologic review of lymphoma diagnosis: study of patients from the french lymphopath network. *Journal of Clinical Oncology*, 35(18):2008–2017, 2017.

[95] Fusheng Wang, Tae W Oh, Cristobal Vergara-Niedermayr, Tahsin Kurc, and Joel Saltz. Managing and querying whole slide images. In *Medical Imaging 2012: Advanced PACS-Based Imaging Informatics and Therapeutic Applications*, volume 8319, page 83190J. International Society for Optics and Photonics, 2012.

[96] Jason D Hipp, Anna Fernandez, Carolyn C Compton, and Ulysses J Balis. Why a pathology image should not be considered as a radiology image. *Journal of pathology informatics*, 2, 2011.

[97] Fuyong Xing, Yuanpu Xie, Hai Su, Fujun Liu, and Lin Yang. Deep learning in microscopy image analysis: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 29(10):4550–4568, 2017.

[98] Guilherme Andrade, Renato Ferreira, George Teodoro, Leonardo Rocha, Joel H Saltz, and Tahsin Kurc. Efficient execution of microscopy image analysis on cpu, gpu, and mic equipped cluster systems. In *2014 IEEE 26th International Symposium on Computer Architecture and High Performance Computing*, pages 89–96. IEEE, 2014.

[99] Victor Campos Camunez, Francesc Sastre, Maurici Yagües, Míriam Bellver, Xavier Giró Nieto, and Jordi Torres Viñals. Distributed training strategies for a computer vision deep learning algorithm on a distributed gpu cluster. In *Procedia Computer Science*, pages 315–324. Elsevier, 2017.

[100] Christopher J Kelly, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, 17(1):195, 2019.

[101] Dong Wook Kim, Hye Young Jang, Kyung Won Kim, Youngbin Shin, and Seong Ho Park. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. *Korean journal of radiology*, 20(3):405–410, 2019.

[102] Morteza Babaie, Shivam Kalra, Aditya Sriram, Christopher Mitcheltree, Shujin Zhu, Amin Khatami, Shahryar Rahnamayan, and Hamid R Tizhoosh. Classification and retrieval of digital pathology scans: A new dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 8–16, 2017.

[103] 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*, 2021.

[104] Daisuke Komura, Keisuke Fukuta, Ken Tominaga, Akihiro Kawabe, Hirotomo Koda, Ryohei Suzuki, Hiroki Konishi, Toshikazu Umezaki, Tatsuya Harada, and Shumpei Ishikawa. Luigi: Large-scale histopathological image retrieval system using deep texture representations. *biorxiv*, page 345785, 2018.

[105] Akshay Sridhar, Scott Doyle, and Anant Madabhushi. Content-based image retrieval of digitized histopathology in boosted spectrally embedded spaces. *Journal of pathology informatics*, 6, 2015.

[106] Lei Zheng, Arthur W Wetzel, John Gilbertson, and Michael J Becich. Design and analysis of a content-based pathology image retrieval system. *IEEE transactions on information technology in biomedicine*, 7(4):249–255, 2003.

[107] Neville Mehta, Alomari Raja'S, and Vipin Chaudhary. Content based sub-image retrieval system for high resolution pathology images using salient interest points. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3719–3722. IEEE, 2009.

[108] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999.

[109] Hatice Cinar Akakin and Metin N Gurcan. Content-based microscopic image retrieval system for multi-image queries. *IEEE transactions on information technology in biomedicine*, 16(4):758–769, 2012.

[110] Xiaofan Zhang, Wei Liu, Murat Dundar, Sunil Badve, and Shaoting Zhang. Towards large-scale histopathological image analysis: Hashing-based image retrieval. *IEEE Transactions on Medical Imaging*, 34(2):496–506, 2014.

[111] 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.

[112] Hamid R Tizhoosh, Shujin Zhu, Hanson Lo, Varun Chaudhari, and Tahmid Mehdi. Minmax radon barcodes for medical image retrieval. In *Advances in Visual Computing: 12th International Symposium, ISVC 2016, Las Vegas, NV, USA, December 12-14, 2016, Proceedings, Part I 12*, pages 617–627. Springer, 2016.

[113] 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.

[114] Noriaki Hashimoto, Yusuke Takagi, Hiroki Masuda, Hiroaki Miyoshi, Kei Kohno, Miharu Nagaishi, Kensaku Sato, Mai Takeuchi, Takuya Furuta, Keisuke Kawamoto, et al. Case-based similar image retrieval for weakly annotated large histopathological images of malignant lymphoma using deep metric learning. *Medical Image Analysis*, 85:102752, 2023.

[115] Chengkuan Chen, Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Andrew J Schaumberg, and Faisal Mahmood. Fast and scalable search of whole-slide images via self-supervised deep learning. *Nature Biomedical Engineering*, 6(12):1420–1434, 2022.

[116] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

[117] Peter van Emde Boas. Preserving order in a forest in less than logarithmic time. In *16th Annual Symposium on Foundations of Computer Science (sfcs 1975)*, pages 75–84. IEEE, 1975.

[118] Milad Sikaroudi, Mehdi Afshari, Abubakr Shafique, Shivam Kalra, and HR Tizhoosh. Comments on 'fast and scalable search of whole-slide images via self-supervised deep learning'. *arXiv preprint arXiv:2304.08297*, 2023.

[119] 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.

[120] Péter Bándi, Rob van de Loo, Milad Intezar, Daan Geijs, Francesco Ciompi, Bram van Ginneken, Jeroen van der Laak, and Geert Litjens. Comparison of different methods for tissue segmentation in histopathological whole-slide images. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 591–595, 2017.

[121] Péter Bándi, Maschenka Balkenhol, Bram van Ginneken, Jeroen van der Laak, and Geert Litjens. Resolution-agnostic tissue segmentation in whole-slide histopathology images with convolutional neural networks. *PeerJ*, 7:e8242, 2019.

[122] Takayuki Otsu and Masatoshi Yoshida. Role of initiator-transfer agent-terminator (iniferter) in radical polymerizations: Polymer design by organic disulfides as iniferters. *Die Makromolekulare Chemie, Rapid Communications*, 3(2):127–132, 1982.

[123] 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.

[124] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[125] 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.

[126] Vipul Baxi, Robin Edwards, Michael Montalto, and Saurabh Saha. Digital pathology and artificial intelligence in translational medicine and clinical practice. *Modern Pathology*, 35(1):23–32, 2022.

[127] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.

[128] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147, 2021.

[129] 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):56, 2019.

[130] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[131] George H Dunteman. *Principal components analysis*, volume 69. Sage, 1989.

[132] David A Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101, 1952.

[133] Katarzyna Tomczak, Patrycja Czerwińska, and Maciej Wiznerowicz. Review the cancer genome atlas (tcga): an immeasurable source of knowledge. *Contemporary Oncology/Współczesna Onkologia*, 2015(1):68–77, 2015.

[134] Nadia Brancati, Anna Maria Anniciello, Pushpak Pati, Daniel Riccio, Giosuè Scognamiglio, Guillaume Jaume, Giuseppe De Pietro, Maurizio Di Bonito, Antonio Foncubierta, Gerardo Botti, et al. Bracs: A dataset for breast carcinoma subtyping in h&e histology images. *Database*, 2022:baac093, 2022.

[135] Wouter Bulten, Kimmo Kartasalo, Po-Hsuan Cameron Chen, Peter Ström, Hans Pinckaers, Kunal Nagpal, Yuannan Cai, David F Steiner, Hester van Boven, Robert Vink, et al. Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. *Nature medicine*, 28(1):154–163, 2022.

[136] Shiv Ram Dubey. A decade survey of content based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[137] Rasoolijaberi, Maral. Multi-magnification search in digital pathology. Master's thesis, 2021.

[138] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[139] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

[140] Feature. https://en.wikipedia.org/wiki/Feature_(machine_learning)#:~:text=In%20pattern%20recognition%20and%20machine,facilitate%20processing%20and%20statistical%20analysis. Accessed: 2022-06-09.

[141] Jonathan I Epstein, William C Allsbrook Jr, Mahul B Amin, Lars L Egevad, ISUP Grading Committee, et al. The 2005 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *The American journal of surgical pathology*, 29(9):1228–1242, 2005.

[142] Manuel Salto-Tellez and Ian A Cree. Cancer taxonomy: pathology beyond pathology. *European Journal of Cancer*, 115:57–60, 2019.

[143] Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International journal of medical informatics*, 73(1):1–23, 2004.

[144] Xiaoshuang Shi, Fuyong Xing, KaiDi Xu, Yuanpu Xie, Hai Su, and Lin Yang. Supervised graph hashing for histopathology image retrieval and classification. *Medical image analysis*, 42:117–128, 2017.

APPENDICES

Appendix A

Additional Content for Chapter 3

A.1 Extended Results for the Proposed SDM Framework

A.1.1 Public – The Cancer Genome Atlas (TCGA)

Additional confusion matrices and chord diagrams of Top-1, and MV@3 retrievals are shown in fig. A.1, and A.2 when evaluating the TCGA dataset.

Figure A.1: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the TCGA dataset.

Figure A.2: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the TCGA dataset.

A.1.2 Public – BReAst Carcinoma Subtyping (BRACS)

The BRACS dataset comprises a total of 547 WSIs derived from 189 distinct patients [134]. In the context of the leave-one-out search and matching experiment, all 547 WSIs were employed from the dataset. Notably, all slides have been scanned utilizing an Aperio AT2 scanner, with a resolution of 0.25 μm per pixel and a magnification factor of 40×. The dataset is categorized into two main subsets: WSI and Region of Interest (ROI). Within the WSI subset, there are three primary tumor Groups [134]. Whereas, the ROI subset is divided into seven distinct tumor types [134]. For this study, since we are conducting a WSI-to-WSI matching, we utilized the WSI subset to perform histological matching. Table A.1 shows more details about the data used in this experiment.

Primary Diagnoses	Acronyms	Slides	Group	Group Acronyms	Slides
Atypical Ductal Hyperplasia	ADH	48	Atypical Tumours	AT	89
Flat Epithelial Atypia	FEA	41			
Normal	N	44	Benign Tumours	BT	265
Pathological Benign	PB	147			
Usual Ductal Hyperplasia	UDH	74			
Ductal Carcinoma in Situ	DCIS	61	Malignant Tumours	MT	193
Invasive Carcinoma	IC	132			

Table A.1: Information concerning the BRACS dataset employed in this experiment, inclusive of the respective acronyms and the number of slides associated with each primary diagnosis and group.

To evaluate the performance of the SDM montage against Yottixel's mosaic, we retrieved the top similar cases using leave-one-out evaluation. The assessments rely on several retrieval criteria, including the top-1 retrieval, the majority agreement among the top 3 retrievals (MV@3), and the majority agreement among the top 5 retrievals (MV@5). The accuracy, macro average, and weighted average at top-1, MV@3, and MV@5 are shown in Figure A.3. Table A.2 shows the detailed results including precision, recall, and F1-score. Moreover, confusion matrices and chord diagrams at Top-1, MV@3, and MV@5 are shown in Figure A.4, A.5, and A.6, respectively. In addition to these accuracy metrics, a comparative analysis of the number of patches extracted per WSI by each respective method is also presented in Figure A.7 for a visual representation of the distribution over the entire dataset. To visually illustrate the extracted patches, we used t-SNE projections, as demonstrated in Figure A.8.

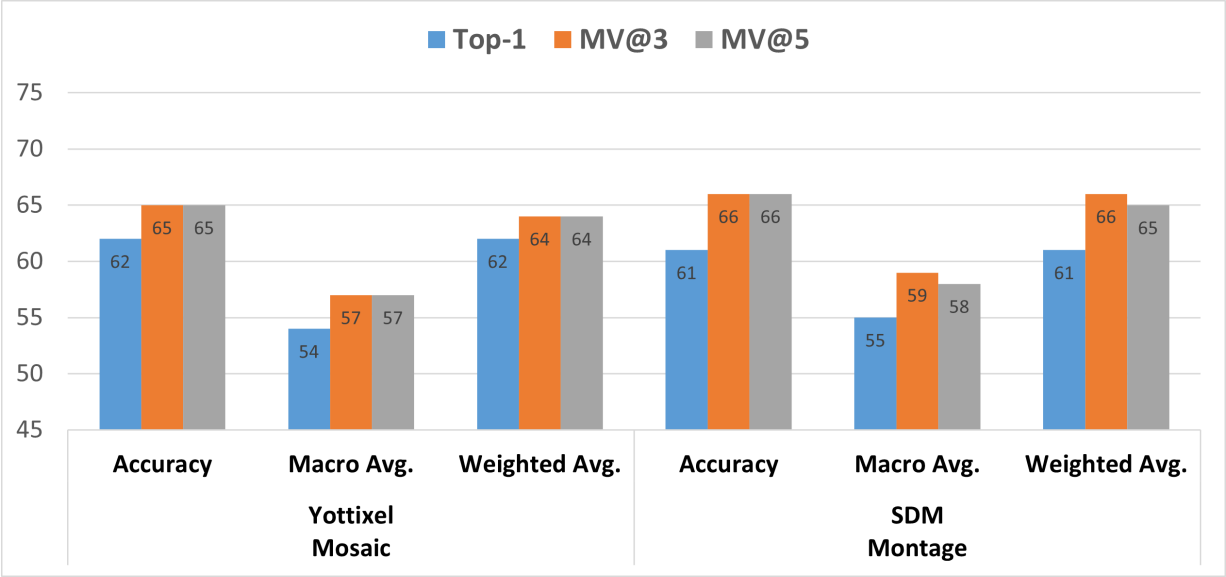

Figure A.3: Accuracy, macro average of F1-scores, and weighted average of F1-scores are reported from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the BRACS dataset.

		Top-1			**MV@3**			**MV@5**			
	Groups	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Slides
Yottixel Mosaic	AT	0.26	0.26	0.26	0.32	0.27	0.29	0.36	0.27	0.31	86
	BT	0.66	0.74	0.69	0.66	0.80	0.72	0.65	0.81	0.72	248
	MT	0.74	0.62	0.68	0.79	0.63	0.70	0.76	0.62	0.69	193
Total Slides											**527**
SDM Montage	AT	0.30	0.35	0.32	0.34	0.34	0.34	0.34	0.31	0.33	89
	BT	0.70	0.69	0.69	0.72	0.79	0.75	0.70	0.83	0.76	265
	MT	0.65	0.62	0.64	0.73	0.63	0.68	0.76	0.59	0.66	193
Total Slides											**547**

Table A.2: Precision, recall, F1-score, and the number of slides processed for each subtype are reported in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals using the BRACS dataset.

In the course of this experiment, the SDM montage demonstrated the performance advantage over the Yottixel mosaic. Notably, it exhibited improvements of +1%, +2%, and +1% in the macro average of F1-scores concerning top-1 retrieval, majority agreement

among the top 3 retrievals, and majority agreement among the top 5 retrievals, respectively. In terms of accuracy, SDM underperforms at top-1 retrieval by one percent whereas it outperforms at MV@3 and MV@5 retrievals by one percent. These findings underscore the method's effectiveness in capturing relevant information within the specific context of retrieval, as visualized in Figure A.3. Furthermore, our analysis unveiled an important aspect of Yottixel's behavior in comparison to SDM. Specifically, our investigation revealed that Yottixel failed to process some WSIs and it processed a total of 527 WSIs, whereas SDM demonstrated a more comprehensive approach by successfully processing all 547 WSIs as shown in Table A.2. This observation highlights the robustness and completeness of the SDM method in managing the entire dataset, further emphasizing its advantages in applications related to the analysis and retrieval of WSIs. In contrast to the Yottixel mosaic, SDM exhibits reduced variability in the number of patches per WSI as seen in Figure A.7. This is attributed to the absence of an empirical parameter dictating patch selection, as opposed to Yottixel's approach of utilizing 5% of the total patches. Such a methodological shift not only optimizes storage utilization but also curtails redundancy and obviates the necessity for empirical determination of an optimal patch count.

Figure A.4: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the BRACS dataset.

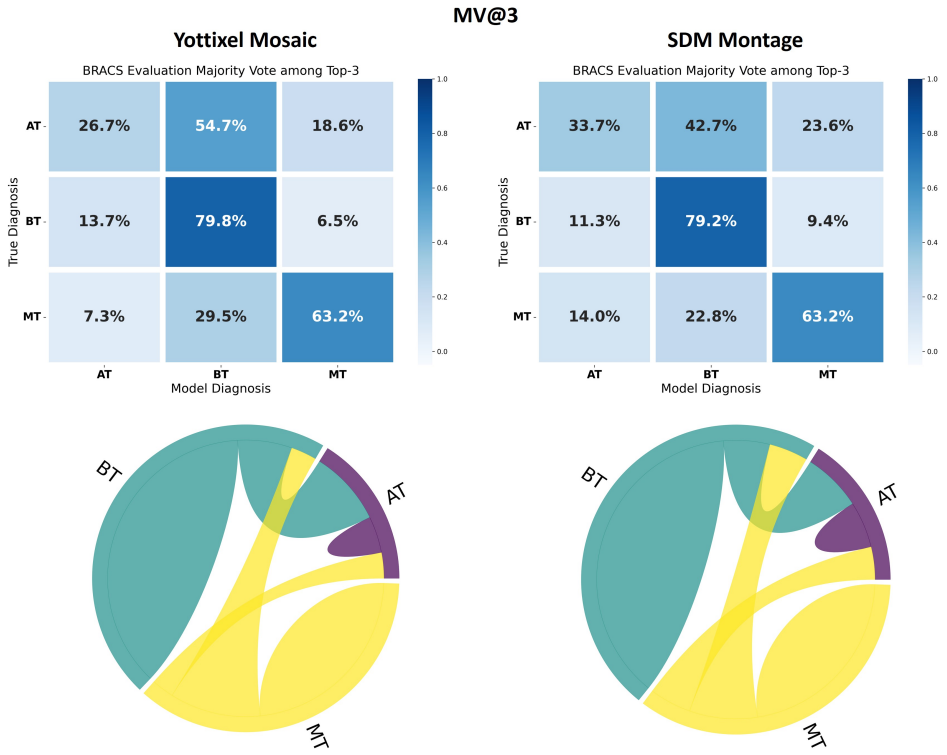

Figure A.5: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the BRACS dataset.

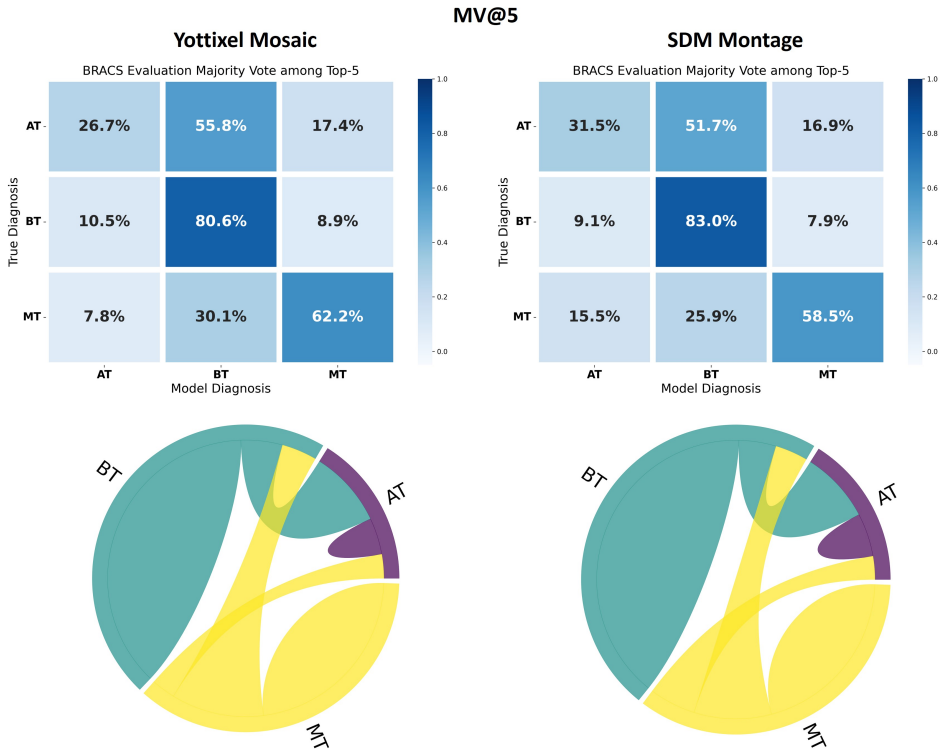

Figure A.6: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the BRACS dataset.

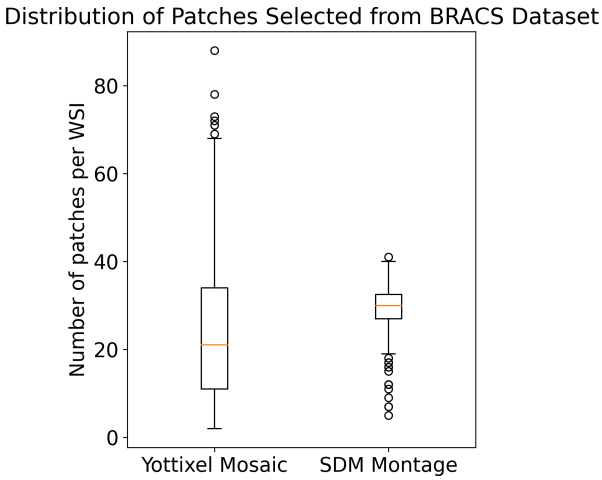

Figure A.7: The boxplot illustrates the distribution of patches selected for each WSI in the BRACS dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 21 ± 16. On the other hand, for the SDM Montage, the median number of selected patches is 30 ± 5.

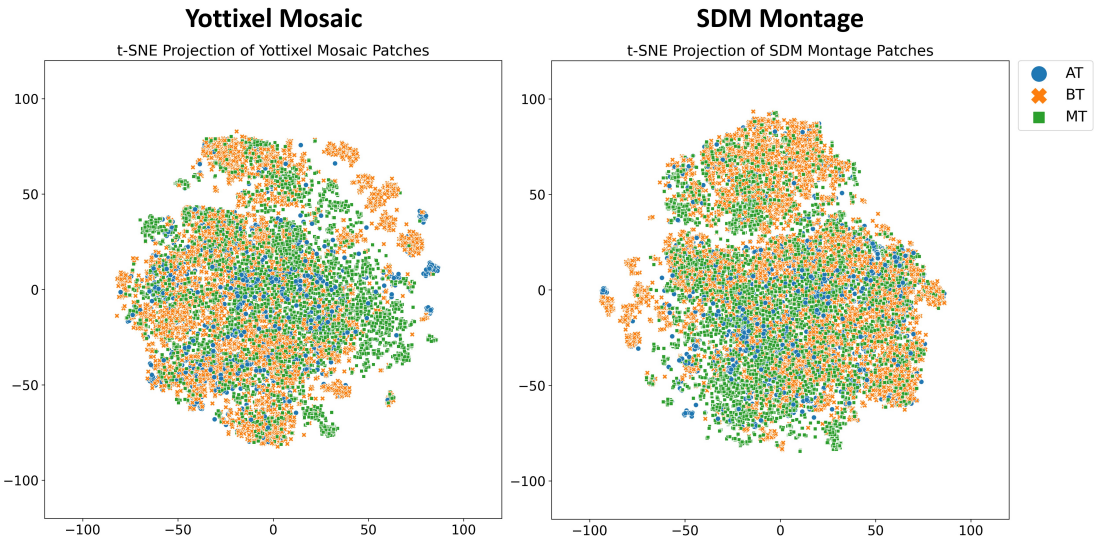

Figure A.8: The t-SNE projection displays the embeddings of all patches extracted from the BRACS dataset using Yottixel's mosaic (left) and SDM's montage (right).

A.1.3 Public – Prostate cANcer graDe Assessment (PANDA)

PANDA is the largest publicly available dataset of prostate biopsies, put together for a global AI competition [135]. The data is provided by Karolinska Institute, Solna, Sweden, and Radboud University Medical Center (RUMC), Nijmegen, Netherlands. All slides from RUMC were scanned at 20× using a 3DHistech Pannoramic Flash II 250 scan. On the other hand, all the WSIs from Karolinska Institute were digitized at 20× using a Hamamatsu C9600-12 scanner, and an Aperio ScanScope AT2 scanner. In entirety, a dataset comprising 12,625 whole slide images (WSIs) of prostate biopsies was amassed and partitioned into 10,616 WSIs for training and 2,009 WSIs for evaluation purposes. In our experiment, we used the publicly available training cohort of 10,616 WSIs with their International Society of Urological Pathology (ISUP) scores for an extensive leave-one-out search and matching experiment (see Table. A.3 for more details).

In recent years, there have been significant advancements in both the diagnosis and treatment of prostate cancer. As we entered the new millennium, there was a significant effort to update and modernize the Gleason system. In 2005, the ISUP organized a consensus conference. The gathering attempted to provide a clearer understanding of the patterns that make up different Gleason grades. It also established practical guidelines for how to apply these patterns and introduced what is now known as the ISUP score from zero to five based on the severity of the cancer [141, 135].

ISUP Grade	Slides
0	2889
1	2665
2	1343
3	1242
4	1246
5	1223

Table A.3: Comprehensive dataset particulars pertaining to the Prostate cANcer graDe Assessment (PANDA) dataset, encompassing relevant ISUP grade and the number of slides attributed to each grade.

To assess the performance of the SDM montage in comparison to Yottixel's mosaic, we conducted a leave-one-out evaluation to retrieve the most similar cases. This evaluation involves multiple criteria for retrieval assessment, including the top-1 retrieval, as well as evaluating the majority consensus among the top 3 retrievals (MV@3), and the top 5

retrievals (MV@5). The results include accuracy, macro average, and weighted average scores for each of these criteria, as depicted in Figure A.9. Table A.4 shows the detailed statistical results including precision, recall, and F1-score. Moreover, confusion matrices and chord diagrams at Top-1, MV@3, and MV@5 are shown in Figure A.10, A.11, and A.12, respectively. In addition to these accuracy metrics, a comparative analysis of the number of patches extracted per WSI by each respective method is also presented in Figure A.13 for a visual representation of the distribution over the entire dataset. To visually illustrate the extracted patches, we used t-SNE projections, as demonstrated in Figure A.14.

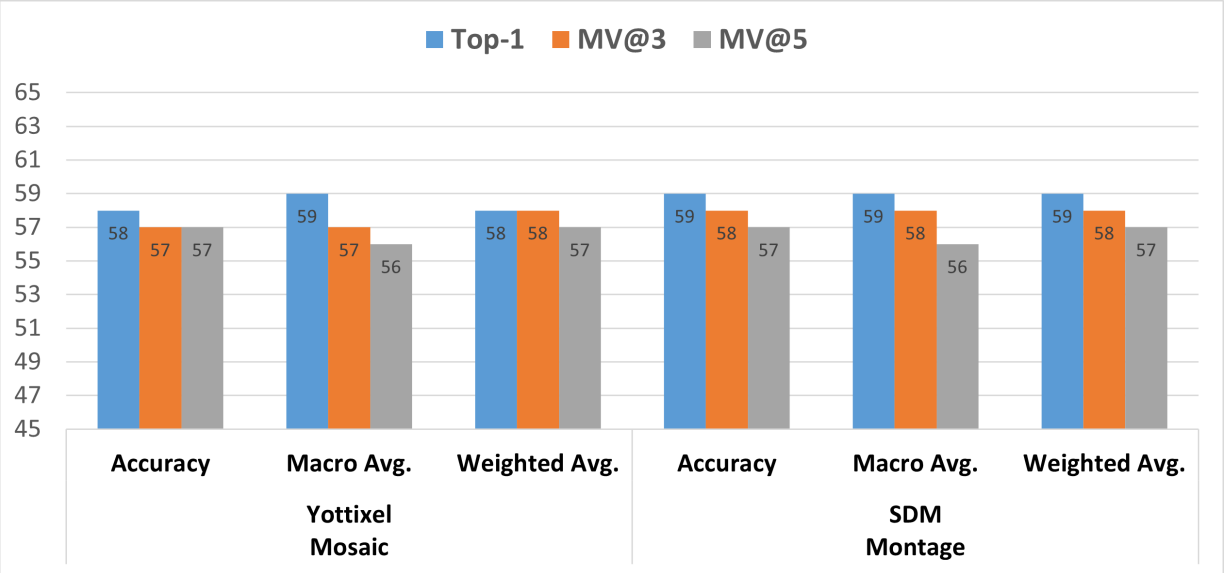

Figure A.9: Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals in the PANDA dataset.

		Top-1			MV@3			MV@5			
	ISUP Grade	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Slides
Yottixel Mosaic	0	0.60	0.60	0.60	0.60	0.64	0.62	0.60	0.68	0.63	2853
	1	0.50	0.57	0.53	0.49	0.60	0.54	0.48	0.62	0.54	2655
	2	0.50	0.48	0.49	0.49	0.42	0.46	0.51	0.37	0.43	1332
	3	0.62	0.58	0.60	0.62	0.54	0.57	0.60	0.50	0.55	1230
	4	0.61	0.54	0.57	0.62	0.49	0.55	0.62	0.43	0.51	1225
	5	0.77	0.68	0.72	0.77	0.65	0.71	0.80	0.61	0.69	1201
Total Slides											**10496**
SDM Montage	0	0.63	0.63	0.63	0.62	0.67	0.64	0.61	0.70	0.65	2889
	1	0.51	0.57	0.54	0.50	0.58	0.53	0.48	0.60	0.53	2665
	2	0.48	0.47	0.48	0.50	0.43	0.46	0.49	0.38	0.43	1343
	3	0.64	0.59	0.62	0.62	0.55	0.58	0.60	0.49	0.54	1242
	4	0.60	0.58	0.59	0.60	0.52	0.56	0.60	0.45	0.52	1246
	5	0.77	0.68	0.72	0.77	0.64	0.70	0.78	0.61	0.68	1223
Total Slides											**10608**

Table A.4: Precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals in the PANDA dataset.

PANDA is one of the most extensive publicly available datasets for prostate cancer analysis. In this research, our empirical findings shed light on the comparative efficacy of our proposed method when compared to the Yottixel mosaic. Specifically, our findings indicate that SDM exhibited comparable performance to the Yottixel mosaic concerning accuracy with majority agreement among the top 5 retrievals. However, a noteworthy distinction emerged when considering accuracy at top-1 and the majority agreement among the top 3 retrievals. Regarding the macro-averaged F1-scores, both top-1 and MV@5 exhibit analogous outcomes. However, for MV@3, the SDM method demonstrates a 1% enhancement, as depicted in the Figure A.9. This highlights the proficiency of the SDM method in assimilating pertinent information for retrieval tasks without the reliance on empirical parameters, a contrast to the Yottixel approach. Specifically, Yottixel necessitates predefined settings for both cluster count and patch selection percentage. Moreover, our analysis revealed an intriguing facet of Yottixel's behavior in comparison to SDM. Specifically, it has come to our attention that Yottixel exhibits a tendency to overlook certain WSIs within the dataset. Our observations indicate that Yottixel processed a total of 10,496 WSIs, while SDM demonstrated a more comprehensive approach, successfully processing 10,608 WSIs out of the 10,616 WSIs as shown in Table A.4. This observation underscores the robustness and completeness of the SDM method in managing the entire dataset, further emphasizing its advantages in applications related to the analysis and retrieval of WSIs in the context of prostate cancer research. A notable inference from the box plot depicted in

Figure A.13reveals that for fine needle biopsies (which constitute a significant portion of the PANDA dataset), the Yottixel 5% methodology selects a reduced number of patches in comparison to the SDM approach.

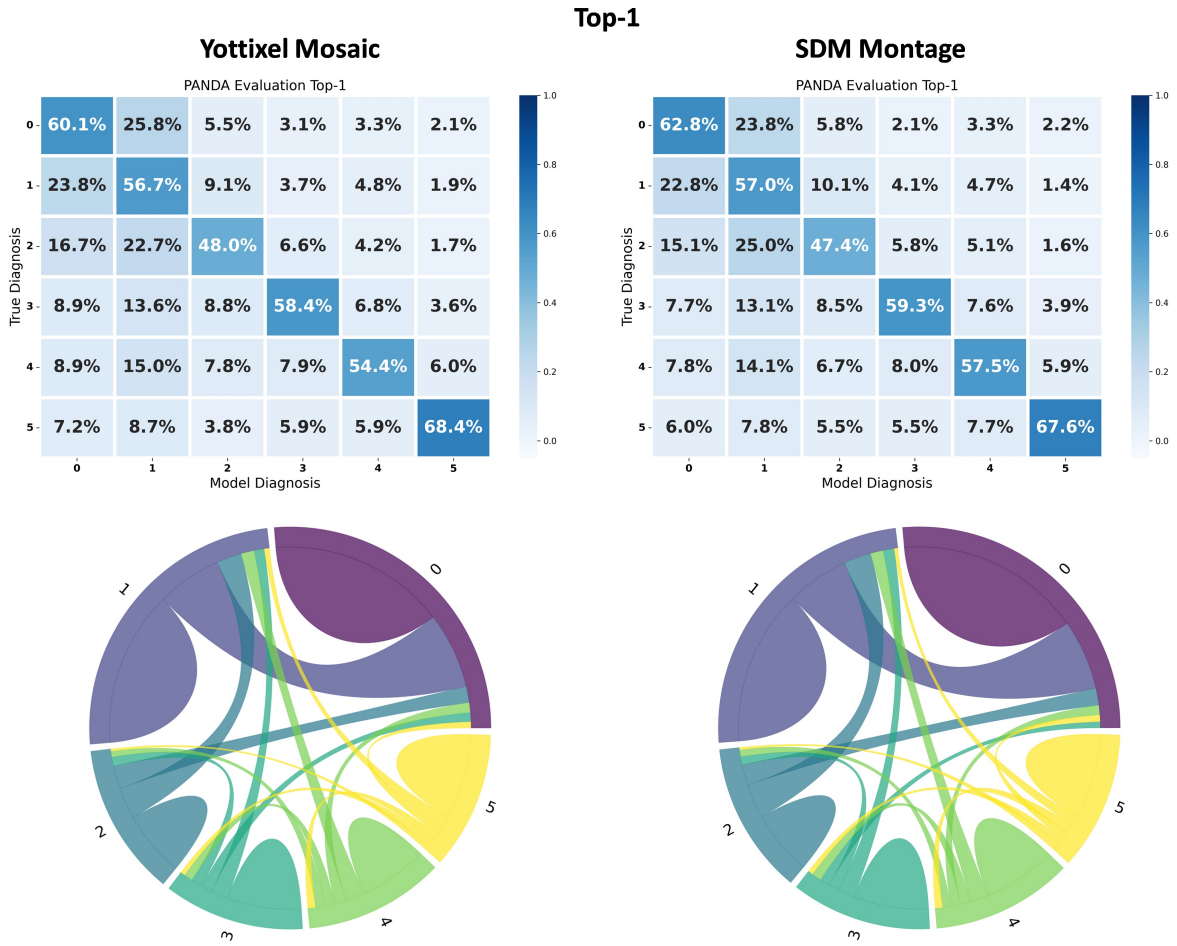

Figure A.10: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the PANDA dataset.

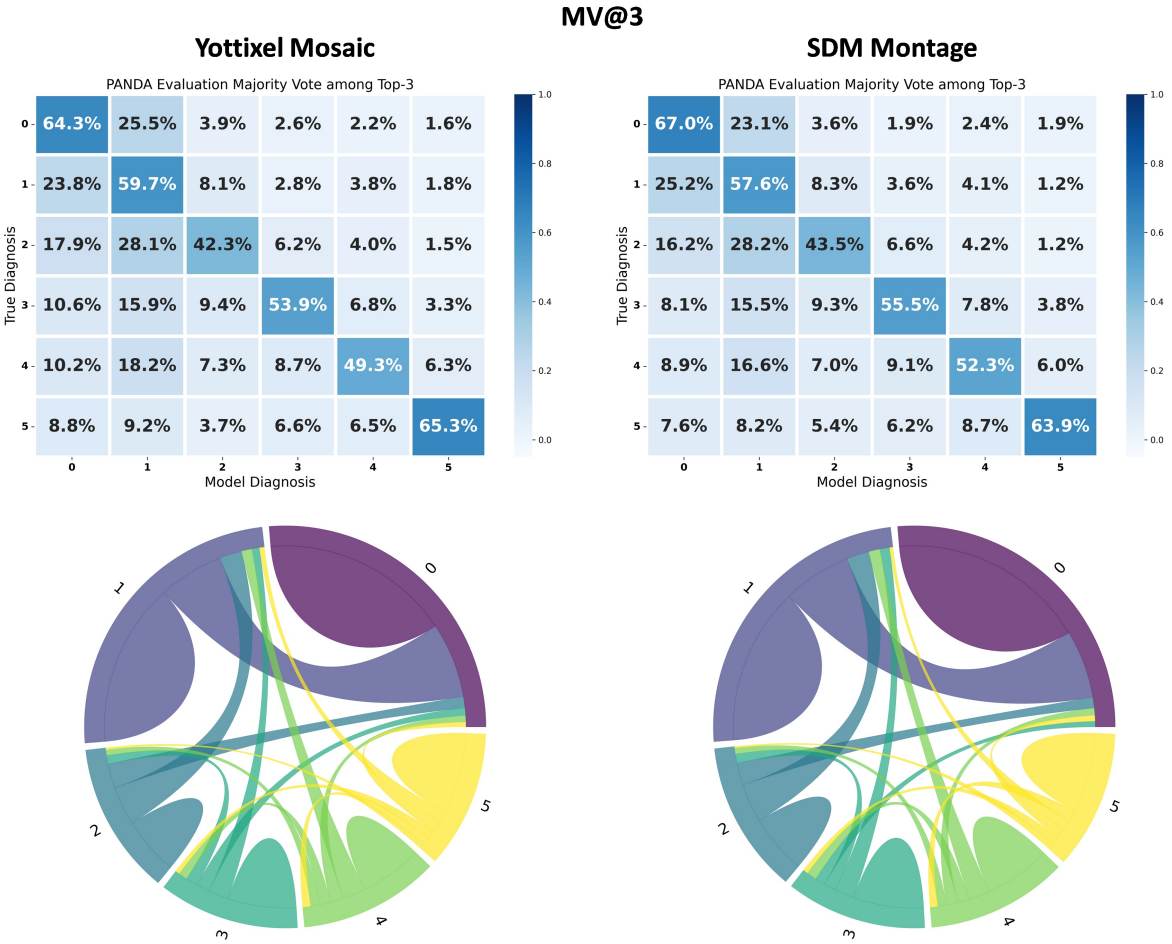

Figure A.11: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the PANDA dataset.

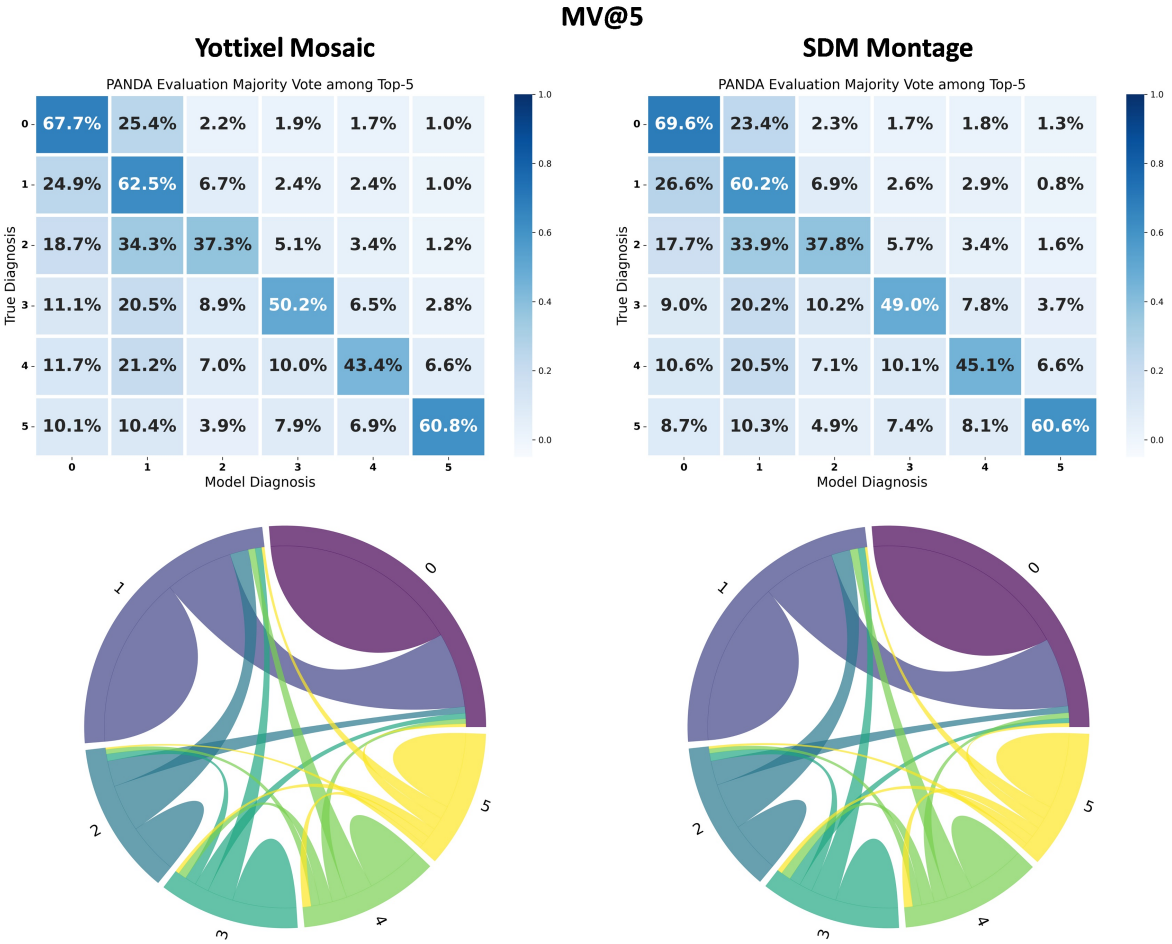

Figure A.12: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the PANDA dataset.

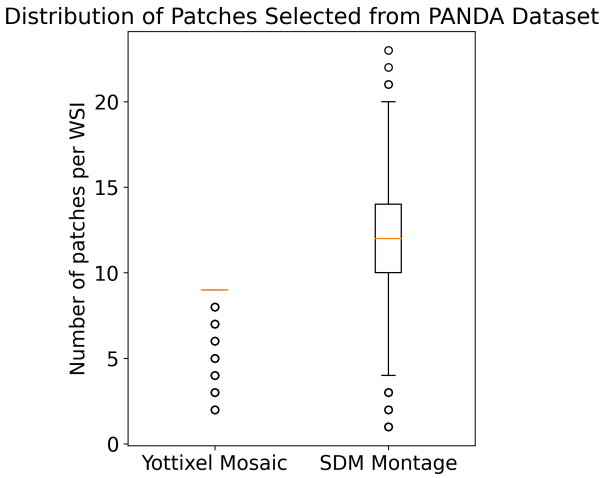

Figure A.13: The boxplot illustrates the distribution of patches selected for each WSI in the PANDA dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 9 ± 2. On the other hand, for the SDM Montage, the median number of selected patches is 12 ± 3.

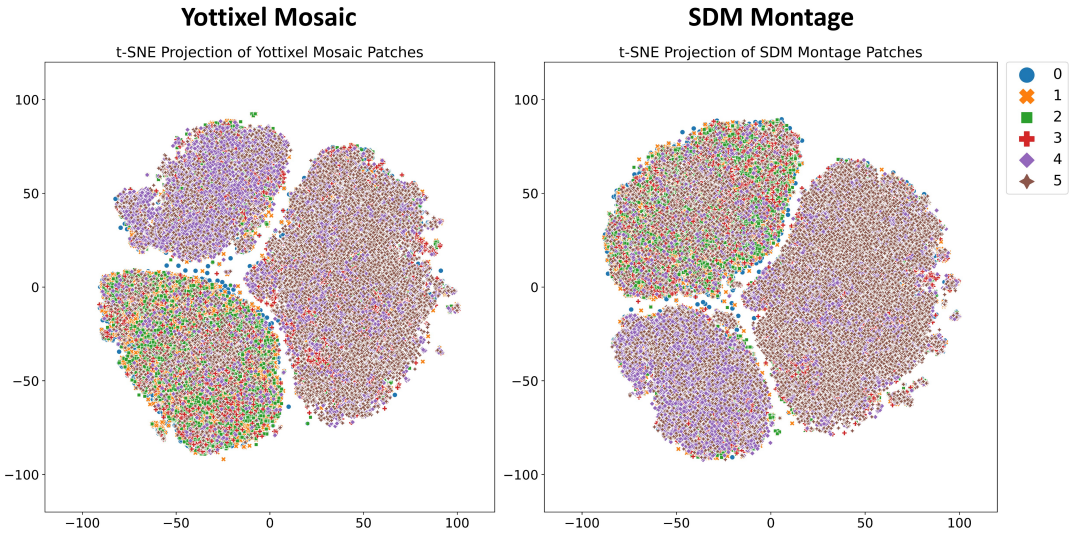

Figure A.14: The t-SNE projection displays the embeddings of all patches extracted from the PANDA dataset using Yottixel's mosaic (left) and SDM's montage (right).

A.1.4 Private – Colorectal Cancer (CRC)

Additional confusion matrices and chord diagrams of Top-1, and MV@3 retrievals are shown in fig. A.15, and A.16 when evaluating the CRC dataset.

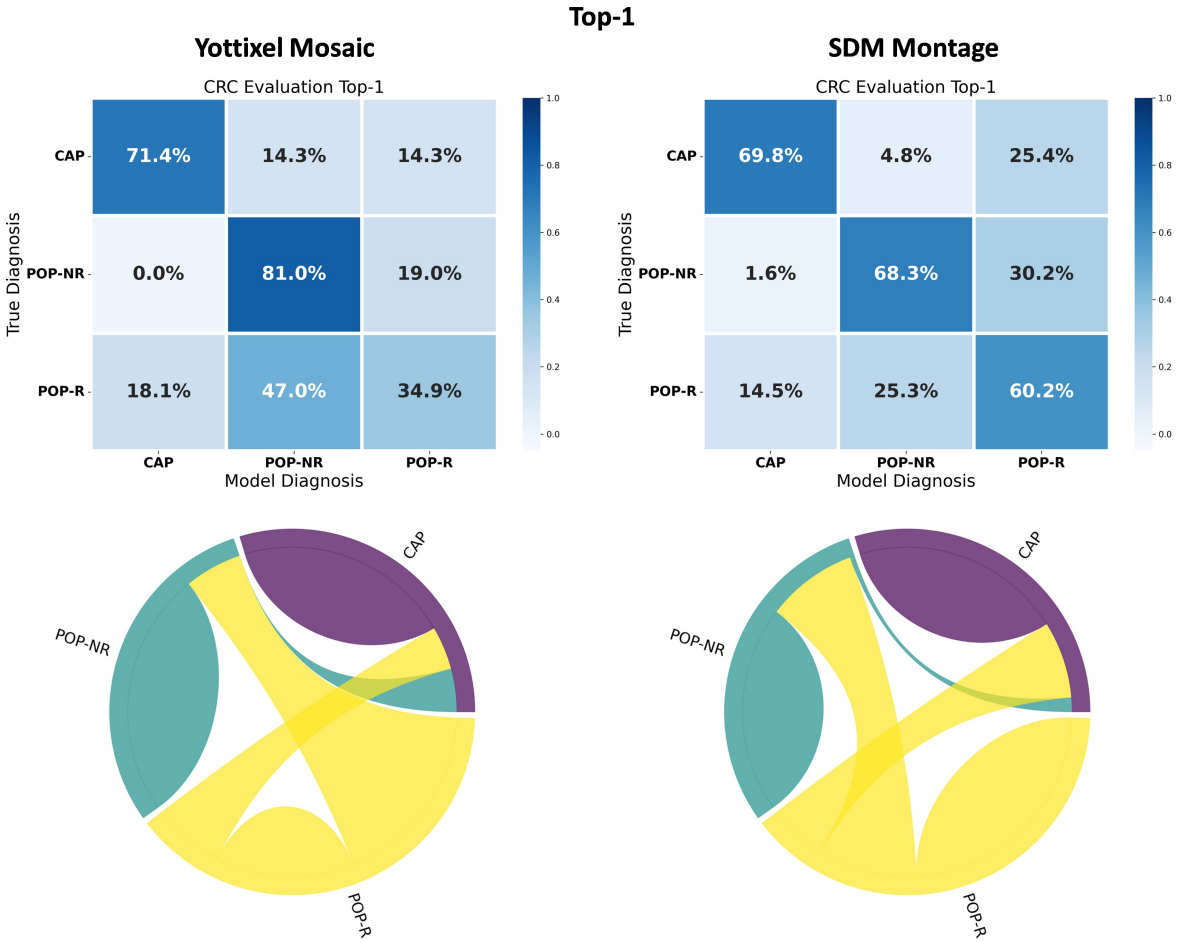

Figure A.15: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the CRC dataset.

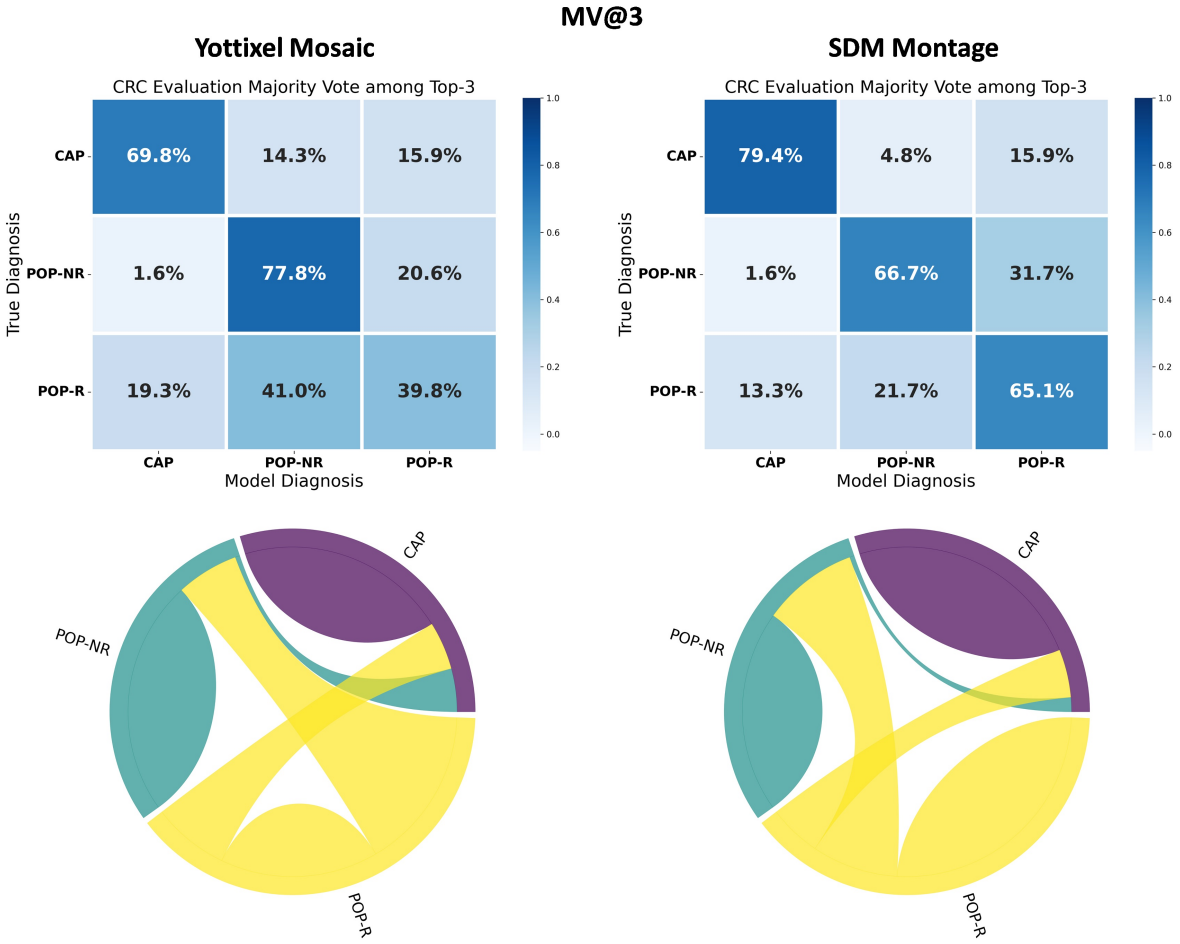

Figure A.16: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the CRC dataset.

A.1.5 Private – Liver ASH vs. NASH

Liver biopsy slides were acquired from patients who had been diagnosed with either Alcoholic Steatohepatitis (ASH) or Non-Alcoholic Steatohepatitis (NASH) at Mayo Clinic, Rochester, USA. The ASH diagnosis was established through a comprehensive review of patient records and expert assessments that considered medical history, clinical presentation, and laboratory findings. For the NASH group, liver biopsies were selected from a cohort of morbidly obese patients undergoing bariatric surgery. All of the biopsy slides were digitized at 40× magnification and linked to their respective diagnoses at the WSI level (see Table A.5 for more details).

Primary Diagnoses	Acronyms	Slides
Alcoholic Steatohepatitis	ASH	150
Non-alcoholic Steatohepatitis	NASH	158
Normal	Normal	18

Table A.5: Information related to the Liver dataset, inclusive of the respective acronyms and the number of slides associated with each primary diagnosis.

To assess the effectiveness of the SDM montage in comparison to Yottixel's mosaic, we conducted a leave-one-out evaluation to retrieve the most similar cases using the Liver dataset. The evaluation criteria encompass multiple retrieval scenarios, including the top-1 retrieval, the majority consensus among the top 3 retrievals (MV@3), and the majority consensus among the top 5 retrievals (MV@5). The results, including accuracy, macro average, and weighted average scores at the top-1, MV@3, and MV@5 levels, are presented in Figure A.17. Table A.6 shows the detailed statistical results including precision, recall, and F1-score. Moreover, Confusion matrices and chord diagrams at Top-1, MV@3, and MV@5 are shown in Figure A.18, A.19and A.20, respectively. In addition to these accuracy metrics, a comparative analysis of the number of patches extracted per WSI by each respective method is also presented in Figure A.21 for a visual representation of the distribution over the entire dataset. To visually illustrate the extracted patches, we used t-SNE projections, as demonstrated in Figure A.22.

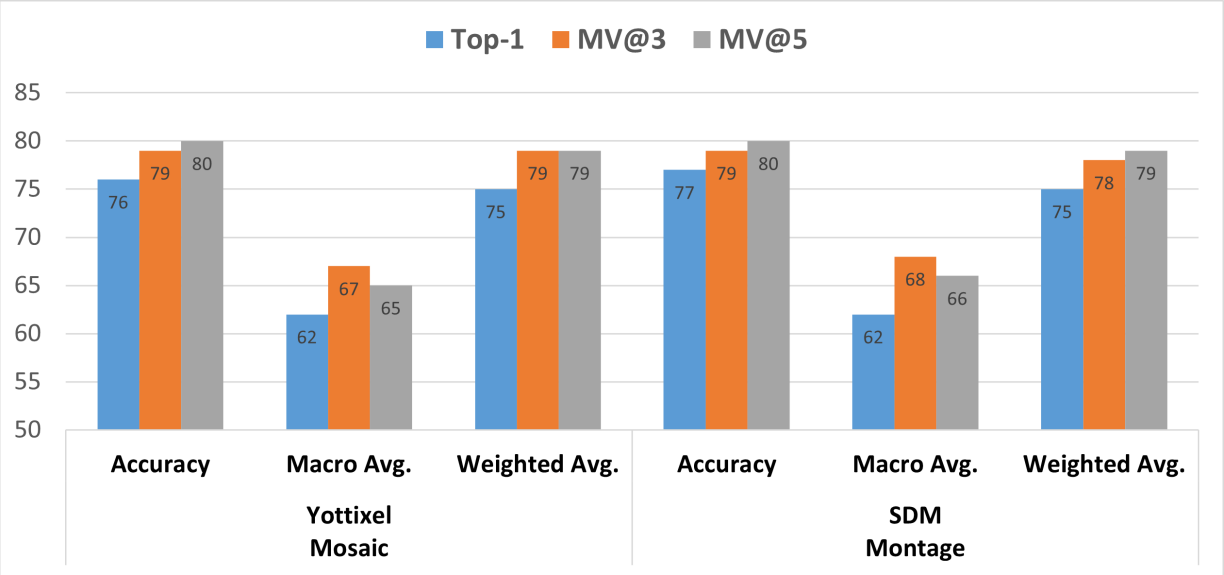

Figure A.17: Accuracy, macro average of F1-scores, and weighted average of F1-scores are shown from Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals in the Liver dataset.

		Top-1			**MV@3**			**MV@5**			
	Primary Diagnoses	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Slides
Yottixel Mosaic	Ash	0.81	0.73	0.76	0.87	0.73	0.80	0.89	0.73	0.81	150
	Nash	0.72	0.85	0.78	0.74	0.91	0.81	0.74	0.93	0.83	158
	Normal	0.75	0.19	0.30	1.00	0.25	0.40	1.00	0.19	0.32	16
Total Slides											**324**
SDM Montage	Ash	0.84	0.72	0.78	0.87	0.73	0.79	0.87	0.76	0.81	150
	Nash	0.71	0.88	0.79	0.73	0.90	0.81	0.75	0.91	0.82	158
	Normal	1.00	0.17	0.29	1.00	0.28	0.43	1.00	0.22	0.36	18
Total Slides											**326**

Table A.6: Precision, recall, F1-score, and the number of slides processed for each subtype are shown in this table using Yottixel mosaic, and SDM montage. The evaluations are based on the top 1 retrieval, the majority among the top 3 retrievals, and the majority among the top 5 retrievals in the Liver dataset.

In our empirical assessments, the SDM approach displayed performance metrics closely aligned with the Yottixel mosaic. This similarity in performance was especially pronounced in the MV@-3 and MV@-5 retrieval outcomes. Notably, there was an enhancement of +1%

in the macro-average of F1-scores when employing the SDM technique. The nuanced differences and advantages of the SDM approach over the Yottixel mosaic in specific retrieval scenarios are further elucidated in the referenced Figure A.17. From an accuracy standpoint, the SDM method exhibited a marginal improvement of one percentage point for top-1 retrieval. Nonetheless, its performance remained largely analogous to that of the Yottixel mosaic when evaluated at MV@3 and MV@5 retrieval metrics as seen in Figure A.17. Moreover, our observations have unveiled an intriguing aspect of Yottixel's behavior in contrast to SDM. It shows that Yottixel processed a total of 324 WSIs, while SDM successfully processed all 326 WSIs. From a detailed examination of the box plot presented in Figure A.21, it becomes evident that for fine needle biopsies — a predominant category within the Liver dataset — the Yottixel 5% strategy tends to opt for fewer patches relative to the SDM method.

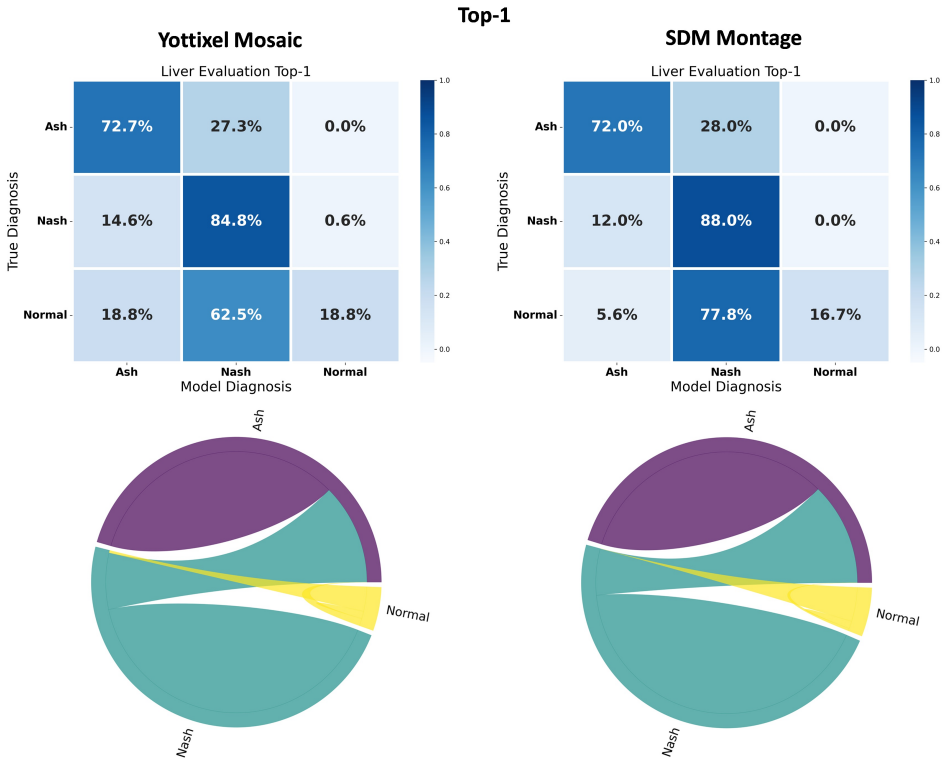

Figure A.18: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the top 1 retrieval when evaluating the Liver dataset.

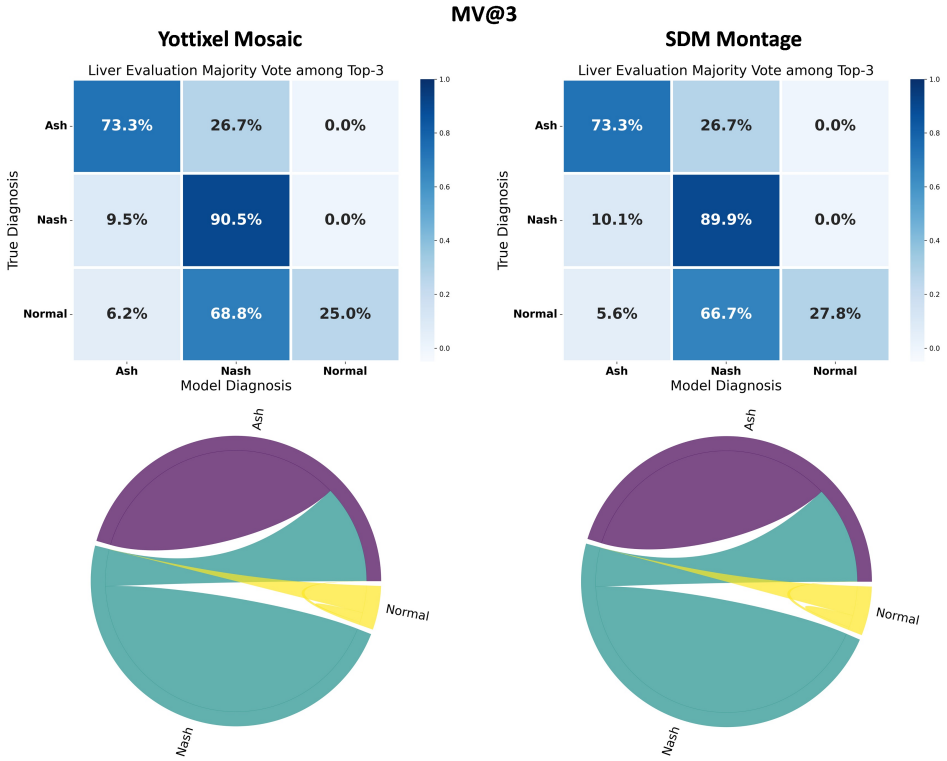

Figure A.19: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the Liver dataset.

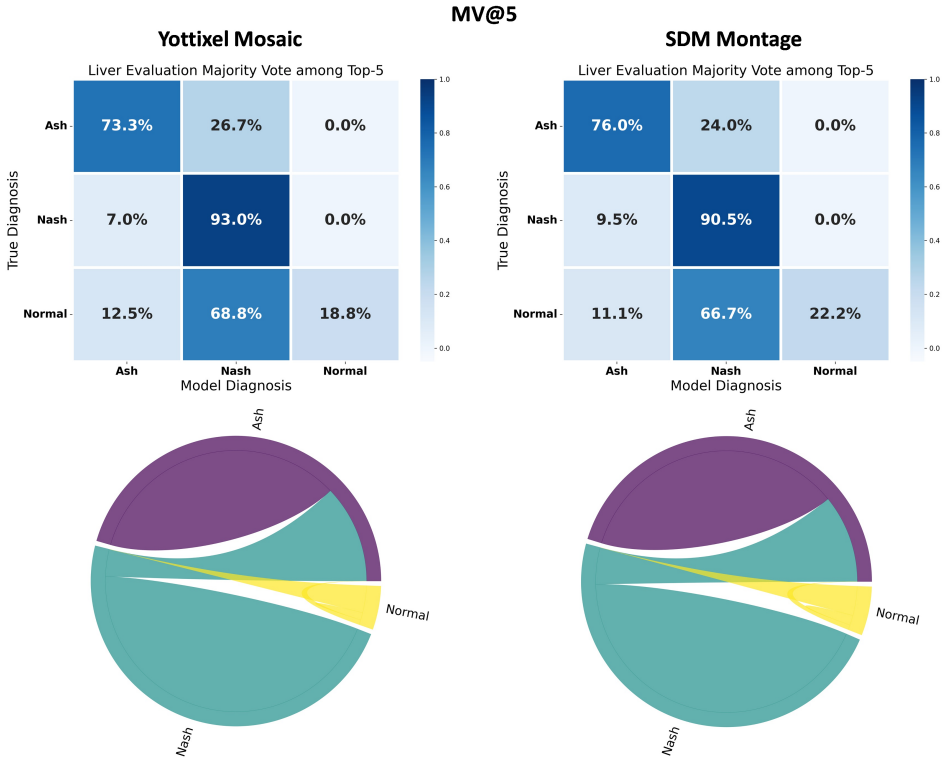

Figure A.20: Confusion matrices and chord diagrams from Yottixel mosaic (left column), and SDM montage (right column). The evaluations are based on the majority of the top 5 retrievals when evaluating the Liver dataset.

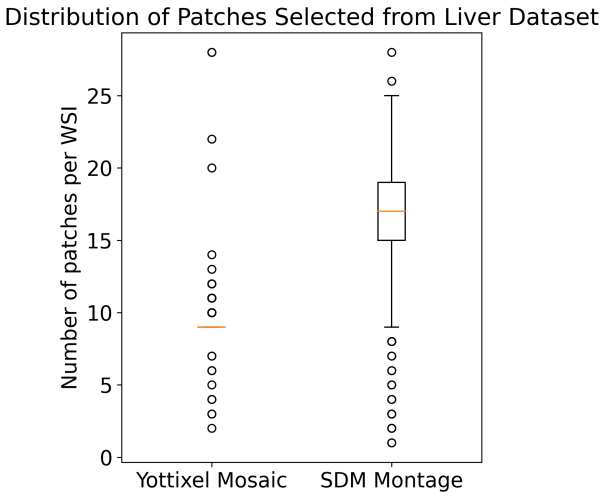

Figure A.21: The boxplot illustrates the distribution of patches selected for each WSI in the Liver dataset from both the Yottixel Mosaic and SDM Montage. Additionally, it provides statistical measures for these distributions. Specifically, for the Yottixel Mosaic, the median number of selected patches is 9 ± 3. Conversely, for the SDM Montage, the median number of selected patches is 17 ± 4.

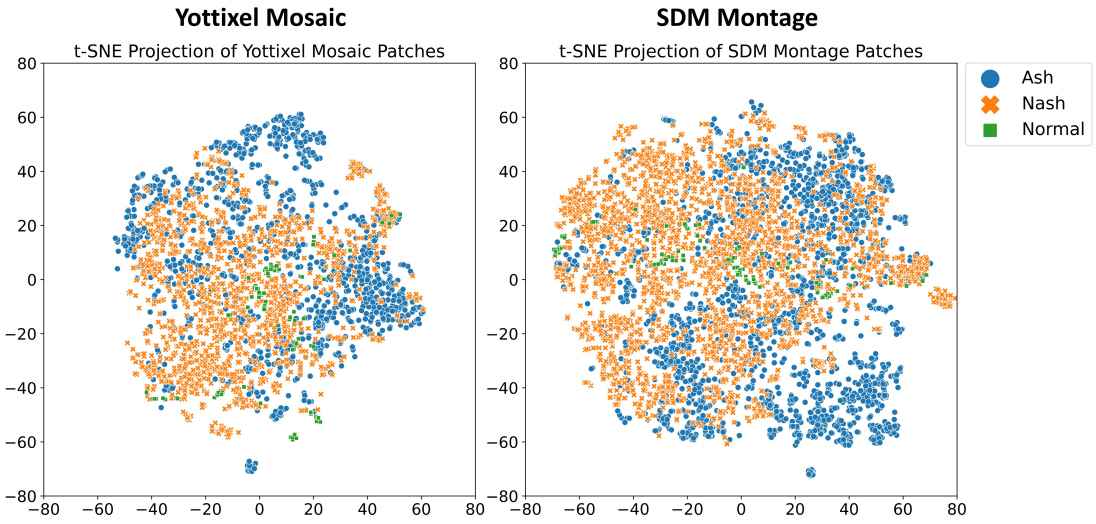

Figure A.22: The t-SNE projection displays the embeddings of all patches extracted from the Liver dataset using Yottixel's mosaic (left) and SDM's montage (right).

Appendix B

Additional Content for Chapter 4

B.1 Extended Results for TCGA Retrieval Evaluation

B.1.1 KimiaNet

Additional confusion matrices and chord diagrams of Top-1, and MV@3 retrievals are shown in fig. B.1, and B.2 when evaluating the TCGA Patch-Level dataset.

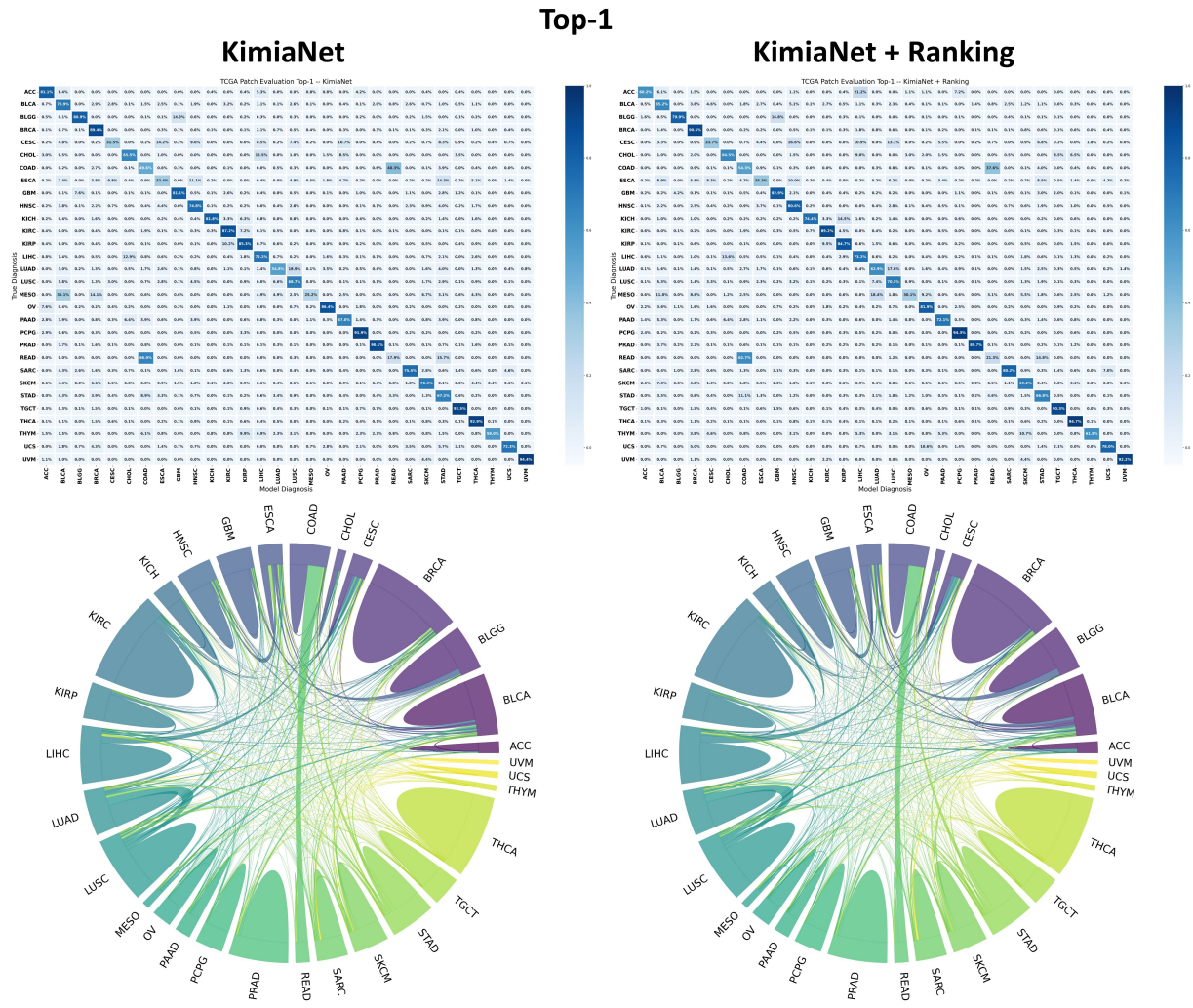

Figure B.1: Confusion matrices and chord diagrams from KimiaNet (left column), and KimiaNet + Ranking (right column). The evaluations are based on the top 1 retrieval when evaluating the TCGA Patch dataset.

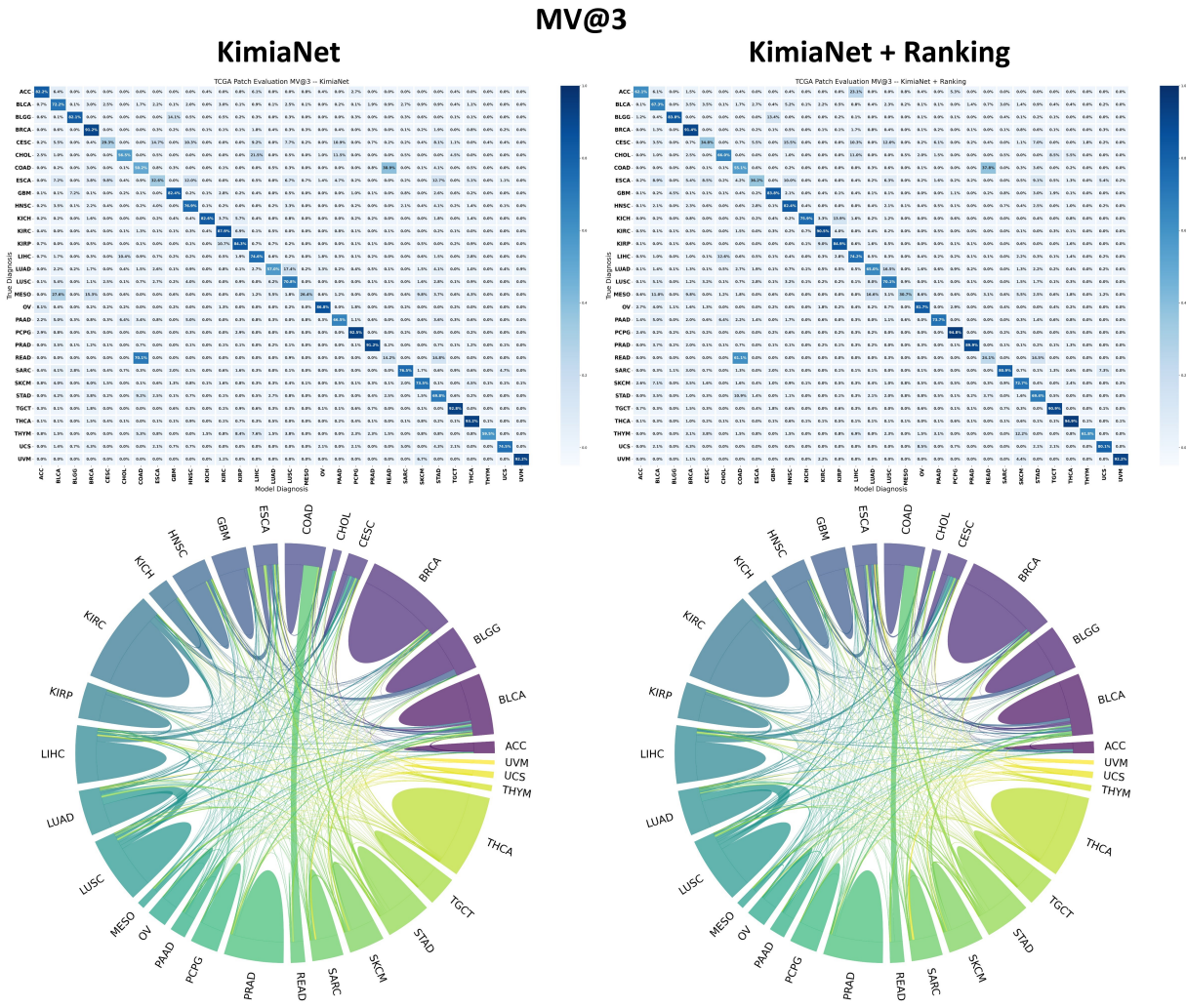

Figure B.2: Confusion matrices and chord diagrams from KimiaNet (left column), and KimiaNet + Ranking (right column). The evaluations are based on the majority of the top 3 retrievals when evaluating the TCGA Patch dataset.

B.1.2 NeXtPath

Additional confusion matrices and chord diagrams of Top-1, and MV@3 retrievals are shown in fig. B.3, and B.4 when evaluating the TCGA Patch-Level dataset from the models trained using the proposed ranking loss.

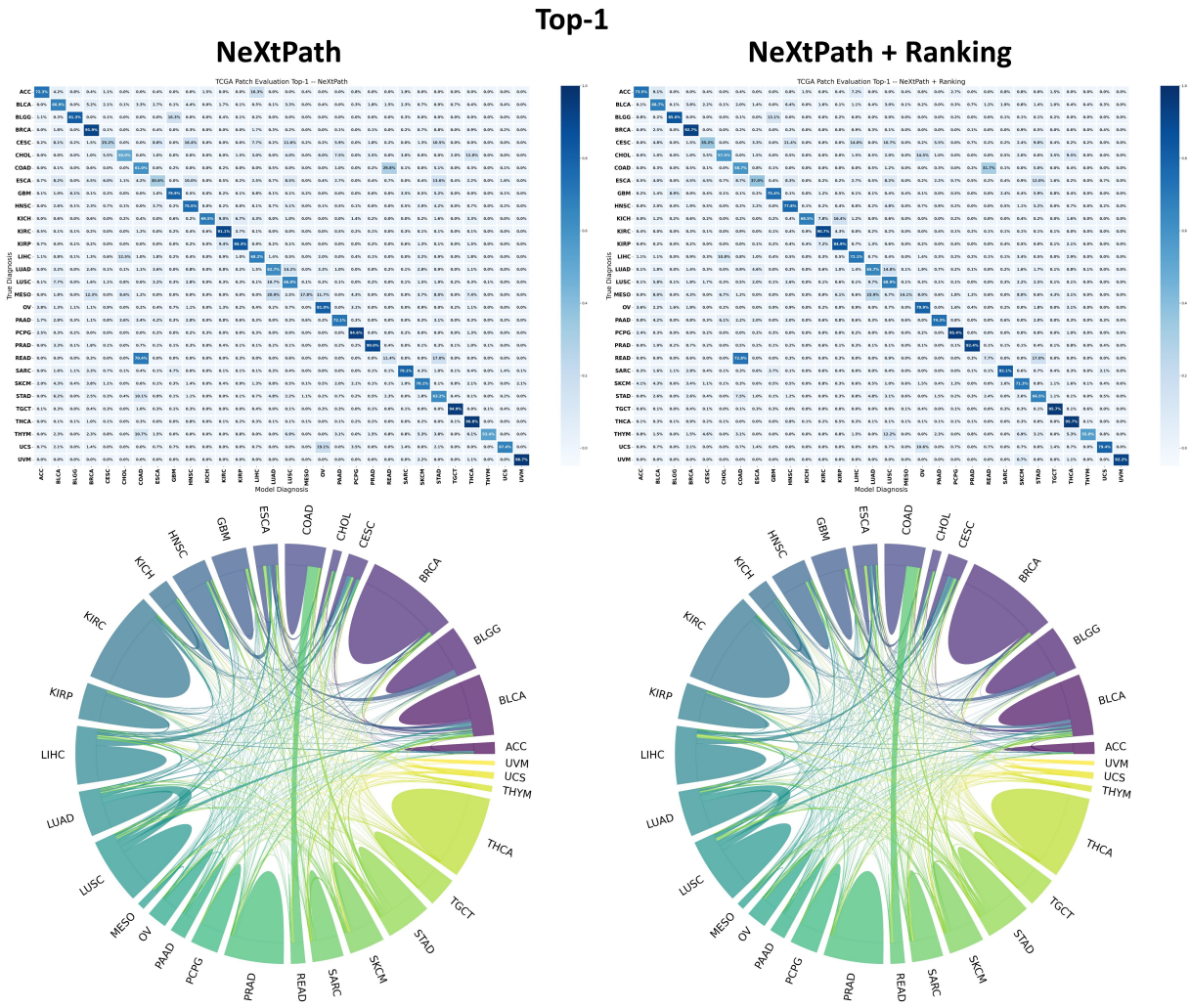

Figure B.3: Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) trained with the proposed ranking loss. The evaluations are based on the top 1 retrieval when evaluating the TCGA Patch dataset.

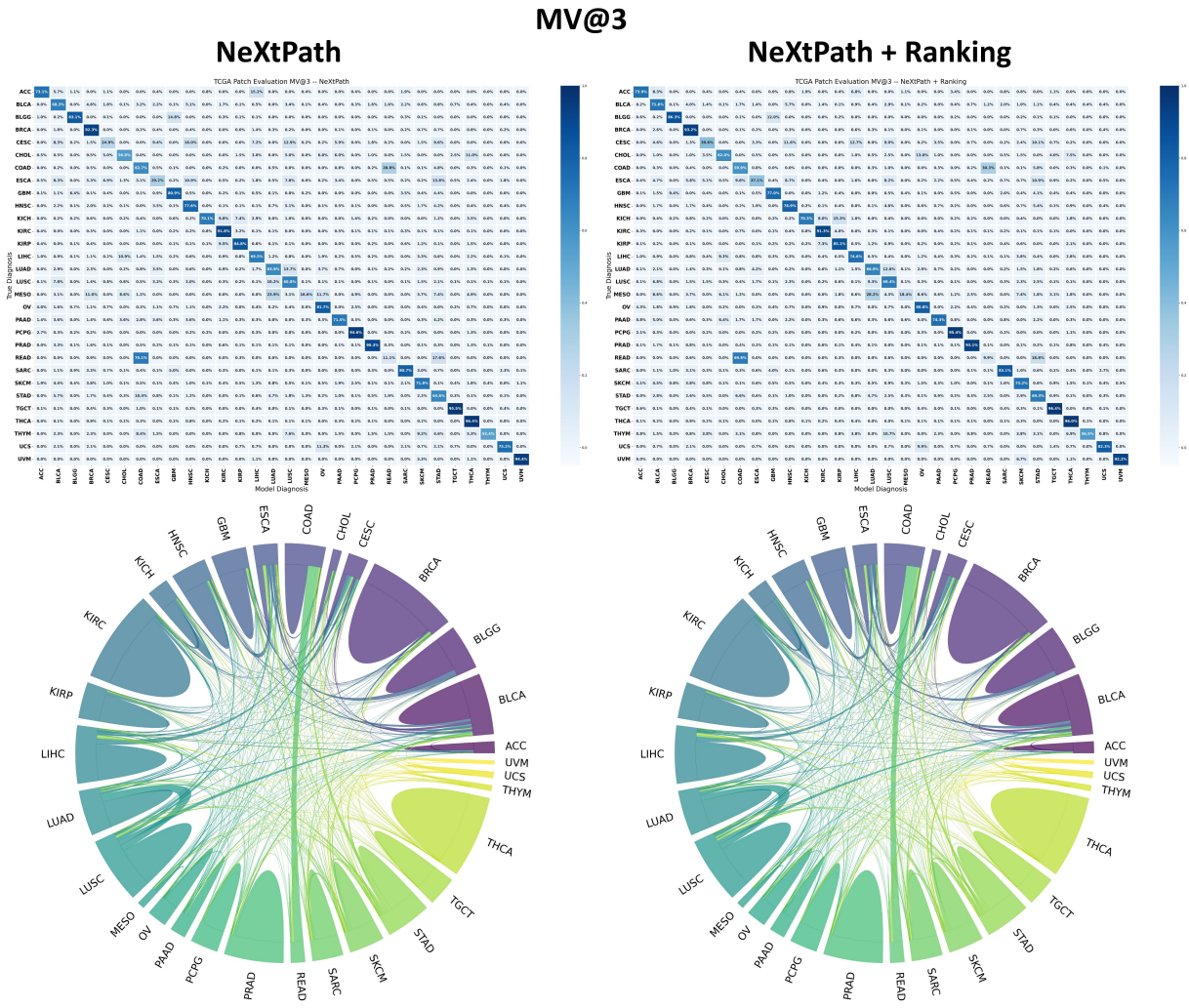

Figure B.4: Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) trained with the proposed ranking loss. The evaluations are based on the majority of the top 3 retrievals when evaluating the TCGA Patch dataset.

B.2 Extended Results for BRACS Retrieval Evaluation

Additional confusion matrices and chord diagrams of Top-1, and MV@3 retrievals are shown in fig. B.5, and B.6 when evaluating the BRACS Patch-Level dataset from the model trained using the cross-entropy loss and the proposed ranking loss.

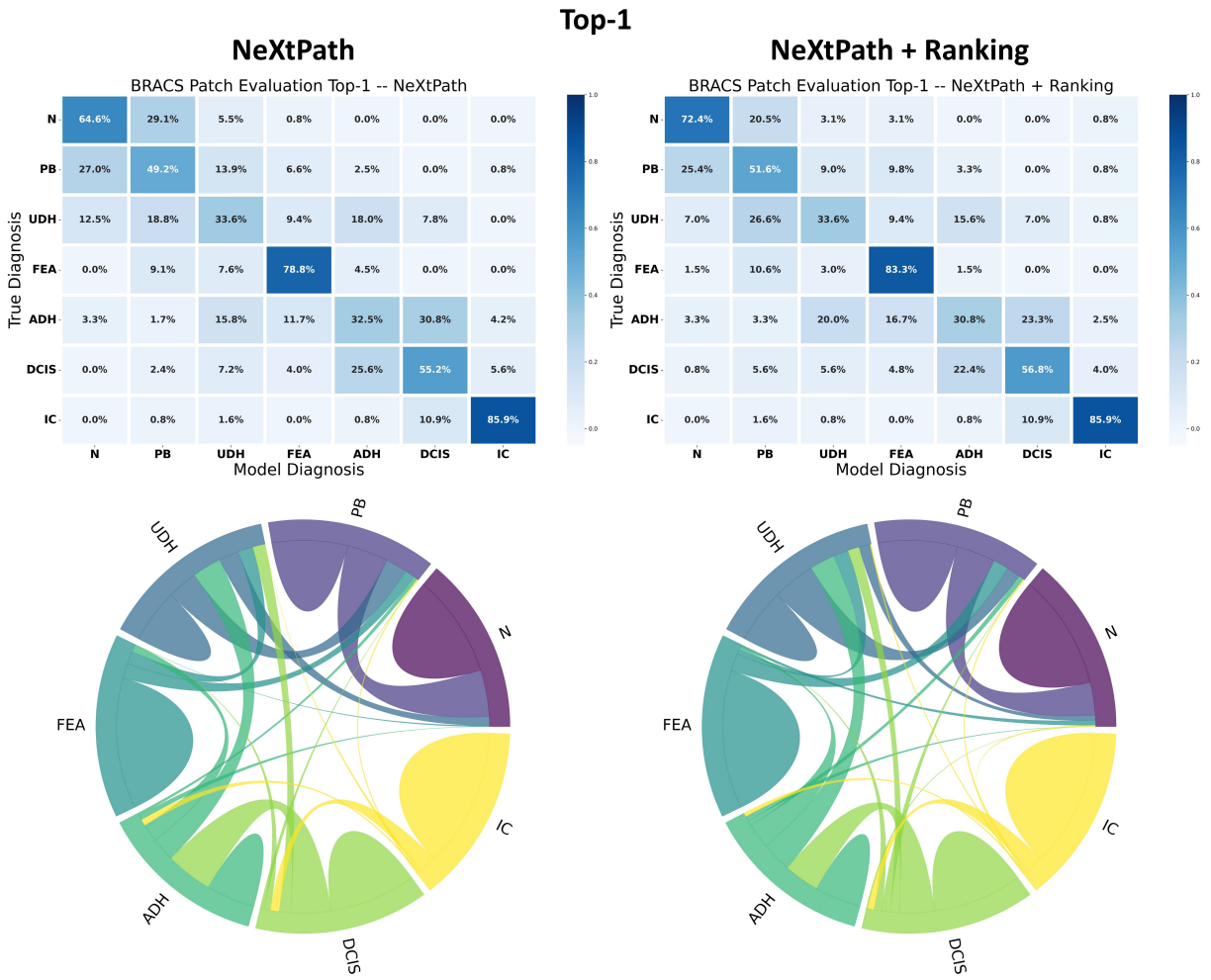

Figure B.5: Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) trained with the proposed ranking loss. The evaluations are based on the top 1 retrieval when evaluating the BRACS ROI dataset.

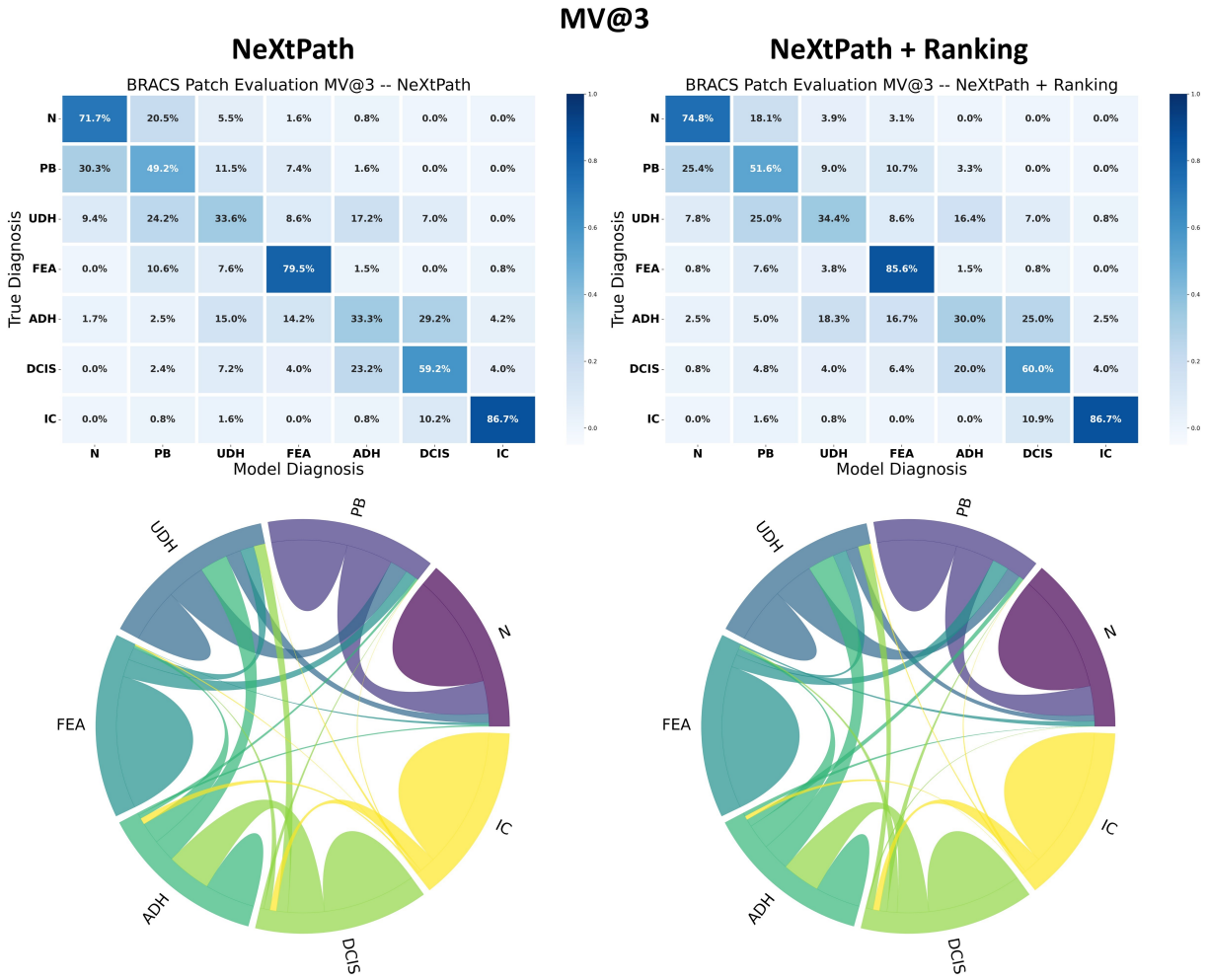

Figure B.6: Confusion matrices and chord diagrams from NeXtPath (left column), and NeXtPath + Ranking (right column) trained with the proposed ranking loss. The evaluations are based on the majority of the top 3 retrievals when evaluating the BRACS ROI dataset.

Glossary

histopathology Histopathology is the diagnosis and study of diseases of the tissues, and involves examining tissues and/or cells under a microscope. 2–4, 11, 13, 14, 16, 17

pathology Pathology is a branch of medical science that involves the study and diagnosis of disease through the examination of surgically removed organs, tissues (biopsy samples), bodily fluids, and in some cases the whole body (autopsy). 1–4