

June 2024

# From Graph Theory for Robust Deep Networks to Graph Learning for Multimodal Cancer Analysis

Asim Waqas  
*University of South Florida*

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>



Part of the [Artificial Intelligence and Robotics Commons](#), and the [Oncology Commons](#)

---

## Scholar Commons Citation

Waqas, Asim, "From Graph Theory for Robust Deep Networks to Graph Learning for Multimodal Cancer Analysis" (2024). *USF Tampa Graduate Theses and Dissertations*.  
<https://digitalcommons.usf.edu/etd/10575>

This Dissertation is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact [digitalcommons@usf.edu](mailto:digitalcommons@usf.edu).

From Graph Theory for Robust Deep Networks to Graph Learning for Multimodal  
Cancer Analysis

by

Asim Waqas

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Electrical Engineering  
College of Engineering  
University of South Florida

Co-Major Professor: Ghulam Rasool, Ph.D.  
Co-Major Professor: Mia Naeini, Ph.D.  
Nidhal Carla Bouaynaya, Ph.D.  
Ismail Uysal, Ph.D.  
Paul Stewart, Ph.D.

Date of Approval:  
June 20, 2024

Keywords: Deep Learning, Neural Architecture Search, Pan-cancer, Foundation Models

Copyright © 2024, Asim Waqas

## **Dedication**

This work is the culmination of invaluable support and guidance from many individuals, some of whom are acknowledged below. Foremost, I extend my deepest gratitude to my advisor, Professor Ghulam Rasool. It has been a profound honor to work under his mentorship. The avid discussions throughout my time with Dr. Rasool, his perspective of science in particular and life, in general, has been the most valuable part of learning for me. I am deeply grateful for having Dr. Rasool as my advisor and mentor.

I also thank my co-advisor, Professor Mia Naeini, for accepting me to her mentorship as the on-campus advisor, and for her time in discussing and evaluating my research. I extend my gratitude to Professor Nidhal Bouaynaya for her invaluable guidance during my time at Rowan University and for accepting to be on my committee. She has been a source of inspiration for many including myself and my wife. I thank Dr. Paul Stewart for the countless discussions in the last two years and for teaching me the world of omics. He has been very supportive in the cancer-related research presented here. I also want to thank Professor Ismail Uysal for his feedback and help in compiling this work. I also thank Dr. Issam El Naqa for accepting to chair my committee and for his supportive feedback in the research discussions. I am thankful to all the past and present members of our research group at Rowan and at Moffitt Cancer Center who have helped me immensely to grow academically and professionally. A special thanks to my friend and colleague Aakash Tripathi. You have been very supportive in all of my work at Moffitt and this would not have been possible without your help and guidance. I also want to thank my friends, Dr. Jehandad, Dr. Hikmat, Salman Saleh, Zaheer Arif, Ijaz Haider, Mohid, Hasan, Husain, and others for always supporting me and making me feel at home in the U.S.

I would especially like to thank my friend and mentor, Dr. Hamza Farooq, who not only encouraged me to pursue the academic path but also continually supported me in all my endeavors.

My first publication is attributed to our discussions with you, and I owe a debt of gratitude to you and your family for everything.

Lastly, I would like to thank my beloved wife Sabeen, and my beautiful daughters Hajra and Khadija, who have been with me through thick and thin. I thank my parents Naik Muhammad and Zahida, my uncle Tausif, my siblings Shaiq and Zoha, my parents-in-law Dr. Munir and Itrat, my siblings-in-law, and my entire family for your love and unconditional support. Special thanks to Professor Rasool's family, Muneera Rasool, and others. I could not have reached where I am today without any of you, and I cherish your support always.

## **Acknowledgments**

I am grateful to Professor Ghulam Rasool for his financial support throughout my Ph.D. through grants supported in part by funds from the National Science Foundation (NSF: # 2008690 and # 2234468), and Moffitt Cancer Center.

## Table of Contents

List of Tables .....	viii
List of Figures .....	x
Abstract .....	xiv
Chapter 1: Introduction .....	1
Chapter 2: Exploring Robust Architectures for Deep Artificial Neural Networks .....	5
2.1 Note to Reader .....	5
2.2 Introduction .....	5
2.3 Results .....	10
2.3.1 Graph Design Space .....	10
2.3.2 Performance Trends of DANNs .....	11
2.3.3 Robustness Analysis .....	14
2.3.4 Effect of Task and Model Complexity .....	18
2.4 Discussion .....	20
2.5 Methods .....	21
2.5.1 Generating Random Graphs .....	21
2.5.2 Graph-Theoretic Measures .....	22
2.5.3 From Graphs to DANNs .....	25
2.5.4 Datasets .....	26
2.5.5 Robustness Analysis .....	26
2.5.6 Statistical Analysis .....	28
2.6 Data and Code Availability .....	29
2.7 Graph to DANN Transformation .....	29
2.8 Further Results and Limitations of the Study .....	29
2.9 Curvature vs. Test Accuracy .....	32
2.10 Average Degree vs. Test Accuracy .....	33
2.11 Global Efficiency vs. Test Accuracy .....	34
2.12 Frameworks and Hyperparameters .....	35
2.13 Compute Resources and Wall-clock Times .....	37
2.14 Sample Images .....	38
2.15 Algorithm for Robust Model Selection .....	39
Chapter 3: Brain Tumor Segmentation and Surveillance with Deep Artificial Neural Networks .....	41
3.1 Note to Reader .....	41
3.2 Introduction .....	41

3.3 Theoretical Background.....	42
3.3.1 Image Segmentation.....	42
3.3.2 Brain Tumor Segmentation.....	44
3.3.3 Tumor Surveillance.....	45
3.3.4 Deep Learning Segmentation Task.....	46
3.3.5 Motivation.....	47
3.3.6 Challenges.....	49
3.4 Brain Tumor Segmentation Using Deep Artificial Neural Networks.....	49
3.4.1 Image Segmentation in Computer Vision Realm .....	50
3.4.2 Deep Artificial Neural Networks and Image Segmentation .....	52
3.4.3 DL-based Image Segmentation Architectures .....	53
3.4.3.1 Convolutional Neural Networks .....	53
3.4.3.2 Fully Convolutional Networks.....	53
3.4.3.3 Encoder-Decoder Based Models.....	54
3.4.3.4 Other DL Models for Image Segmentation .....	55
3.4.4 Brain Tumor Segmentation Task Challenge.....	55
3.5 Inception Modules in Brain Tumor Segmentation.....	56
3.5.1 Segmentation Using Inception/ Dilated Inception Modules.....	57
3.5.2 BraTS Dataset and Pre-processing.....	58
3.5.3 Deep Artificial Neural Network Architectures .....	58
3.5.3.1 Inception Module .....	59
3.5.3.2 Dilated Inception U-Net.....	60
3.5.3.3 Modified DSC as Objective/Loss Function .....	61
3.5.4 Experimental Setup and Results .....	62
3.5.4.1 Results - Inception Modules .....	63
3.5.4.2 Results - DIU-Net .....	63
3.6 Uncertainty Estimation in Brain Tumor Segmentation .....	64
3.6.1 Variational Learning .....	65
3.6.2 Variational Density Propagation.....	65
3.6.3 Extended Variational Density Propagation.....	66
3.6.3.1 First Convolutional Layer .....	67
3.6.3.2 Non-linear Activation Function .....	67
3.6.3.3 Max-pooling Layer .....	68
3.6.3.4 Flattening Operation .....	68
3.6.3.5 Fully-connected Layer .....	68
3.6.3.6 Softmax Function.....	69
3.6.3.7 Objective Function.....	69
3.6.3.8 Back-propagation.....	70
3.6.4 Application to Brain Tumor Segmentation in MRI Images.....	70
3.7 Tumor Surveillance.....	71
3.7.1 Rationale for Tumor Surveillance.....	72
3.7.2 Surveillance Techniques .....	73
3.7.3 Community-level Active Surveillance.....	75
3.7.4 Surveillance of Brain Tumor .....	76
3.7.5 An Example of Surveillance Study.....	77
3.8 Conclusion .....	79

Chapter 4: Revolutionizing Digital Pathology with the Power of Generative Artificial Intelligence and Foundation Model .....	81
4.1 Note to Reader .....	81
4.2 Introduction.....	81
4.3 AI in Digital Pathology.....	84
4.3.1 Digital Pathology .....	86
4.3.2 Computational Pathology.....	86
4.3.3 Classical Machine Learning in Digital Pathology .....	88
4.3.4 Task-specific AI in Digital Pathology .....	88
4.3.4.1 Convolutional Neural Networks (CNNs).....	89
4.3.4.2 Recurrent Neural Networks (RNNs).....	89
4.3.4.3 Graph Neural Networks (GNNs) .....	89
4.3.4.4 Transformers .....	90
4.3.5 AI-based Algorithms Used in Pathology .....	90
4.3.6 Limitations of Task-specific AI.....	92
4.4 Foundation Models and Generative AI.....	94
4.4.1 Foundation Models .....	95
4.4.2 Characteristics of Foundation Models .....	96
4.4.3 Types of Foundation Models .....	98
4.4.3.1 Large Language AI Models .....	98
4.4.3.2 Vision-Language AI Models.....	99
4.4.4 Training Foundation Models and Generative AI.....	100
4.4.5 Challenges and Limitations of Foundation Models.....	101
4.5 Transformers, Foundation Models, and Digital Pathology.....	102
4.5.1 Qualitative Image Analysis.....	104
4.5.2 Image Synthesis, De-noising, and Virtual Staining.....	106
4.5.3 Detecting Zebras (New Disease Identification).....	108
4.5.4 Patient Engagement .....	108
4.5.5 Education and Training.....	108
4.5.6 AI-Driven Standardization in Digital Pathology Workflow .....	109
4.5.7 Trustworthiness and Acceptability of AI by Pathologists .....	110
4.6 Conclusion .....	112
 Chapter 5: Digital Pathology and Multimodal Learning on Oncology Data .....	 114
5.1 Note to Reader .....	114
5.2 Introduction.....	114
5.3 From Conventional to Digital Pathology (DP).....	117
5.4 Computational Pathology (CPATH) – AI in DP .....	118
5.5 DP, CPATH, and Multimodal AI .....	120
5.5.1 2019.....	121
5.5.2 2020.....	122
5.5.3 2021.....	123
5.5.4 2022.....	125
5.5.5 2023.....	128
5.6 Discussion.....	129
5.7 Opportunities.....	131



5.7.1 Multimodal Patient Stratification.....	131
5.7.2 Non-Invasive Alternatives, Personalized Medicine.....	131
5.7.3 Patient Empowerment.....	132
5.7.4 Improving Outcomes .....	132
5.7.5 Morphological Associations, Biomarker Discovery.....	132
5.7.6 Integrative Diagnostics .....	133
5.7.7 Correlational Studies.....	133
5.7.8 Medical Report Integration .....	133
5.8 Challenges.....	133
5.8.1 Data Availability, Integration, Missingness.....	133
5.8.2 Interoperability.....	134
5.8.3 Data Modeling and Analysis.....	134
5.8.4 Transparency and Clinical Adoption .....	134
5.8.5 Personnel Training.....	135
5.8.6 Explainability, Trust, and Reproducibility.....	135
5.9 Conclusion .....	135

## Chapter 6: Multimodal Data Integration for Oncology in the Era of Deep Neural

Networks: A Review.....	136
6.1 Note to Reader .....	136
6.2 Introduction.....	136
6.3 Fundamentals of Multimodal Learning (MML) .....	140
6.3.1 Data Modalities in Oncology.....	140
6.3.1.1 Molecular Data.....	141
6.3.1.2 Imaging Data.....	141
6.3.1.3 Clinical Data .....	141
6.3.2 Taxonomy of MML .....	142
6.3.2.1 Pre-processing.....	142
6.3.2.2 Feature Extraction.....	142
6.3.2.3 Data Fusion .....	143
6.3.2.4 Primary Learner .....	143
6.3.2.5 Final Classifier.....	143
6.3.3 Data Fusion Strategies .....	144
6.3.3.1 Early Fusion.....	144
6.3.3.2 Intermediate Fusion .....	144
6.3.3.3 Late Fusion.....	145
6.3.4 MML for Oncology Datasets.....	145
6.4 Graph Neural Networks (GNNs) in Multimodal Learning.....	146
6.4.1 The Graph Data.....	147
6.4.1.1 Graph Data .....	147
6.4.1.2 Tasks for Graph Data .....	147
6.4.2 ML for Graph Data .....	148
6.4.2.1 Traditional (Shallow) Methods.....	149
6.4.2.2 DNN-based Methods - GNNs.....	149
6.4.3 GNNs and ML Using Unimodal Oncology Datasets.....	153
6.4.3.1 Pathology Datasets.....	153

6.4.3.2 Radiology Datasets .....	154
6.4.3.3 Molecular Datasets.....	155
6.4.4 MML – Data Fusion at Pre-Learning Stage.....	155
6.4.5 MML – Data Fusion Using Cross-Modality Learning .....	156
6.4.6 MML – Data Fusion in Post-Learning Regime .....	157
6.5 Transformers in MML .....	159
6.5.1 Model Architecture .....	160
6.5.2 Multimodal Transformers .....	160
6.5.2.1 Early Fusion.....	162
6.5.2.2 Cross-Attention Fusion .....	162
6.5.2.3 Hierarchical Fusion.....	162
6.5.2.4 Late Fusion.....	163
6.5.3 Transformers for Processing Oncology Datasets.....	164
6.6 MML - Challenges and Opportunities .....	164
6.6.1 Large Amounts of High-quality Data .....	164
6.6.2 Data Registration and Alignment.....	165
6.6.3 Pan-Cancer Generalization and Transference.....	165
6.6.4 Missing Data Samples and Modalities.....	166
6.6.5 Imbalanced Data .....	166
6.6.6 Explainability and Trustworthiness .....	166
6.6.7 Over-smoothing in GNNs .....	166
6.6.8 Modality Collapse.....	167
6.6.9 Dynamic and Temporal Data .....	167
6.6.10 Data Privacy and Federated Learning.....	167
6.6.11 Other Challenges.....	167
6.7 Conclusion .....	168

## Chapter 7: Building Flexible, Scalable, and Machine Learning-Ready Multimodal

OncologyDatasets .....	169
7.1 Note to Reader .....	169
7.2 Introduction.....	169
7.2.1 Contributions of MINDS .....	171
7.3 Background and Literature Review .....	173
7.3.1 Data Characterization Pipeline .....	174
7.3.2 Traditional Data Management - BioBanks .....	175
7.3.3 Data Commons.....	176
7.3.4 The Big Data Approach .....	178
7.3.5 Summary of Gaps in Existing Methods.....	180
7.4 Methodology .....	181
7.4.1 Requirements of Data Management System.....	181
7.4.2 MINDS Architecture.....	182
7.4.2.1 Stage-1: Data Acquisition .....	184
7.4.2.2 Stage-2: Data Processing .....	186
7.4.2.3 Stage-3: Data Serving .....	189
7.4.3 Cloud Deployment.....	190
7.4.3.1 MINDS Infrastructure on AWS .....	192

7.4.3.2 Benefits of Cloud as a PaaS Platform .....	196
7.4.3.3 Scalability Across Different Platforms .....	198
7.4.4 On-Premise Deployment.....	200
7.4.5 User Application .....	200
7.5 Results and Discussion .....	201
7.5.1 Multimodal Data Consolidation.....	201
7.5.2 Storage Optimization .....	203
7.5.3 Horizontal Scalability .....	204
7.5.4 Cohort Building .....	205
7.5.5 Query Responsiveness .....	206
7.5.6 Data Tracking and Reproducibility.....	208
7.5.7 Integrated Analytics .....	208
7.5.8 Limitations and Future Improvements.....	209
7.6 Conclusions.....	210
Chapter 8: SeNMo: A Self-Normalizing Deep Learning Model for Enhanced Multi- Omics Data Analysis in Oncology.....	212
8.1 Introduction.....	212
8.1.1 Multimodal and Multiomics Data.....	212
8.1.2 Pan-cancer Perspective .....	213
8.1.3 Existing Landscape of Pan-cancer Multi-omics Analysis .....	213
8.1.4 Challenges and Opportunities .....	216
8.2 Materials and Methods.....	220
8.2.1 Data Acquisition .....	220
8.2.2 Data Modalities.....	221
8.2.3 Pre-processing.....	224
8.2.4 Features Integration .....	227
8.2.5 Clinical End-points .....	229
8.2.5.1 Overall Survival (OS) .....	230
8.2.5.2 Primary Cancer Type .....	232
8.2.6 SeNMo Deep Learning Model.....	232
8.2.7 Training and Evaluation.....	235
8.2.7.1 Loss Function.....	236
8.2.7.2 Concordance Index (C-Index).....	238
8.2.7.3 Cox Log-rank Function.....	239
8.2.7.4 Hyperparameters Search .....	240
8.2.7.5 Frameworks, Compute Resources, Wall-clock Times.....	241
8.3 Results.....	242
8.3.1 Pan-Cancer Multimodal Analysis - Overall Survival .....	242
8.3.2 Cancer-wise Multimodal Analysis - Overall Survival.....	245
8.3.3 Out-of-distribution Evaluation and Fine-tuning .....	247
8.3.4 Patient Stratification.....	247
8.3.5 Primary Cancer Type Prediction.....	248
8.4 Discussion and Conclusion .....	249
8.5 Hyperparameters Search - Pan-cancer Training .....	253
8.6 Hyperparameters Search - Finetuning CPTAC-LSCC .....	253

8.7 Hyperparameters Search - Finetuning Moffitt-SCC .....	254
Chapter 9: Embedding-based Multimodal Learning on Pan-Squamous Cell Carcinomas for Improved Survival Outcomes.....	255
9.1 Introduction.....	255
9.1.1 Multiscale, Heterogeneous Data in Cancer.....	256
9.1.2 Squamous Cell Carcinoma.....	257
9.1.3 Multimodal Learning (MML).....	257
9.1.4 State-of-the-Art and Challenges in Oncology MML.....	258
9.2 Methods.....	259
9.2.1 PARADIGM Architecture .....	259
9.2.1.1 Modality-Specific Embeddings .....	261
9.2.1.2 Concatenation and Aggregation.....	261
9.2.1.3 Patient-Specific Graphs .....	262
9.2.1.4 Joint Graph Learning .....	262
9.2.2 Learning Strategy.....	262
9.2.2.1 Modality-Specific Learning.....	263
9.2.2.2 Concatenation and Aggregation of Embeddings .....	264
9.2.2.3 Patient Graphs and Joint Graph Learning.....	265
9.2.3 Model Evaluation.....	266
9.3 Results.....	266
9.3.1 Multimodal Integration of Oncology Database System.....	266
9.3.2 Modality-Specific Models, Fine-tuning, and Transfer Learning.....	268
9.3.3 Prediction of OS in PanSCC Cohort.....	268
9.3.3.1 Four-modality Analysis .....	269
9.3.3.2 Three-modality Analysis.....	270
9.3.3.3 Two-modality Analysis.....	272
9.3.3.4 Internal Data Analysis.....	274
9.4 Discussion.....	274
9.4.1 Embedding-based Approach.....	277
9.4.2 Graph Structure and Contextual Learning.....	277
9.4.3 Effect of Graph Sparsity .....	277
9.4.4 Handling Missing Modalities.....	278
9.4.5 Hierarchical Learning .....	278
Chapter 10: Conclusion and Future Work .....	280
References.....	282
Appendix A: Copyright Permissions .....	391
About the Author .....	End Page

## List of Tables

Table 2.1	Curvature vs. test accuracy .....	32
Table 2.2	Average degree vs. test accuracy .....	33
Table 2.3	Global efficiency vs. test accuracy .....	34
Table 2.4	Frameworks and packages used in our codebase.....	35
Table 2.5	Training & eval hyperparameters for our DANN experiments .....	36
Table 3.1	Architecture of the two models, i.e., exVDP and deterministic CNN .....	72
Table 3.2	Segmentation results using DSC for BraTS test dataset.....	73
Table 4.1	Definitions of key terminologies .....	85
Table 4.2	Summary of AI models used in digital pathology .....	91
Table 5.1	Key pathology terminologies.....	116
Table 5.2	Key AI terminologies.....	119
Table 5.3	Summary of multimodal AI models for DP in this review.....	130
Table 6.1	Oncology data tokenization and embeddings selection techniques.....	161
Table 7.1	Definitions of key cloud components. ....	184
Table 7.2	Comparison of the storage size. ....	202
Table 7.3	Number of cases by programs from GDC open cases present in MINDS.....	203
Table 8.1	Feature reduction summary of pan-cancer data. ....	229
Table 8.2	Summary of patient characteristics for pan-cancer data in this study. ....	230
Table 8.3	Hyperparameters for training and finetuning.....	240

Table 8.4	Frameworks and packages used in our codebase.....	242
Table 8.5	C-Index for test and ensemble inference across cancer types.....	246

## List of Figures

Figure 2.1	Exploring robustness of DANNs with graph-theoretic measures.....	9
Figure 2.2	The graph design space for generating random graphs .....	10
Figure 2.3	Test accuracies of different Deep Artificial Neural Networks .....	12
Figure 2.4	Test accuracy vs. entropy for ResNet-18 on ImageNet .....	15
Figure 2.5	Test accuracy vs. entropy for ResNet-18 on Tiny ImageNet.....	16
Figure 2.6	Accuracy vs. entropy for CNNs.....	17
Figure 2.7	Effect of task and model complexity on entropy-robustness relationship.....	19
Figure 2.8	Schematic for graph to DANN transformation.....	29
Figure 2.9	Test accuracy vs. entropy for MLPs trained on CIFAR-10 dataset.....	30
Figure 2.10	Additional results for ResNet-18 on ImageNet dataset .....	30
Figure 2.11	Results for ResNet-18 trained and evaluated on Tiny ImageNet dataset .....	31
Figure 2.12	Results for robustness evaluation of CNN on CIFAR-100.....	31
Figure 2.13	Correlation of entropy and curvature with other graph measures.....	31
Figure 2.14	Comparison of clean images from ImageNet dataset with adversarial examples .....	38
Figure 2.15	Comparison of clean images from ImageNet dataset with noisy images .....	39
Figure 3.1	Use of image segmentation in Marvel movies and medical imaging .....	43
Figure 3.2	Summary of medical imaging file formats .....	45
Figure 3.3	Important computer vision tasks.....	47

Figure 3.4	Few image segmentation techniques in computer vision .....	50
Figure 3.5	Artificial neuron model and ANN model. ....	52
Figure 3.6	Basic architecture of CNN and FCN .....	54
Figure 3.7	Image patches with annotated tumor (glioma) sub-regions.....	57
Figure 3.8	DIU-Net, Inception module, and Dilated Inception module architectures .....	61
Figure 3.9	A schematic layout of the proposed variational density propagation CNN .....	66
Figure 3.10	Segmentation results of exVDP and deterministic CNN.....	74
Figure 3.11	Volume growth curves of grade 2 gliomas of two patients diagnosed with oligos.....	79
Figure 4.1	Pathology, digital pathology, and computational pathology .....	84
Figure 4.2	A schematic layout of various ML algorithms and AI models in DP.....	87
Figure 4.3	The evolution of machine learning models used in digital pathology .....	94
Figure 4.4	A prospective schematic layout using foundation model and GenAI.....	105
Figure 4.5	Surgery and biopsy reports generated by ChatGPT.....	107
Figure 5.1	A template of the future road map for DP and CPATH .....	117
Figure 5.2	A schematic layout of multimodal learning in cancer care.....	122
Figure 6.1	Publications involving DL, GNNs, in medical domain, and MML.....	137
Figure 6.2	An overview of data collected from population to a tissue .....	139
Figure 6.3	Detailed look into data modalities acquired for cancer care.....	140
Figure 6.4	Taxonomy, stages, and techniques of multimodal data fusion .....	145
Figure 6.5	Common graphs, graph learning tasks, and types of representation learning. ....	148
Figure 6.6	Convolution operation for graphs vs. image data .....	151
Figure 6.7	Data processing pipeline for medical images using GNNs .....	152



Figure 6.8	Graph data processing pipeline for molecular and textual data.....	156
Figure 6.9	The original Transformer architecture is presented.....	161
Figure 6.10	Four different fusion strategies using multimodal Transformers.....	163
Figure 7.1	Comparison of developing ML model with/ without using MINDS.....	172
Figure 7.2	Genome Characterization Pipeline as an example.....	174
Figure 7.3	The 7 Vs of Big Data.....	179
Figure 7.4	MINDS architecture implements a 3-stage pipeline.....	183
Figure 7.5	Overview of the MINDS architecture implemented on AWS.....	185
Figure 7.6	Quicksight analytics and visualization generated using clinical data from MINDS.....	191
Figure 7.7	The AWS Glue crawler automates ETL in MINDS through a 5-step workflow.....	195
Figure 7.8	Demonstrating the feasibility of deploying MINDS across cloud platforms.....	199
Figure 7.9	Overview of the workflow in MINDS.....	206
Figure 8.1	Overview of the SeNMo framework.....	220
Figure 8.2	Features processing pipeline for pan-cancer data.....	229
Figure 8.3	Summary of the number of cases in the pan-cancer data.....	233
Figure 8.4	Architecture of the SeNMo encoder network.....	236
Figure 8.5	Pan-cancer C-Index results for OS prediction.....	243
Figure 8.6	Cancer-wise C-Index results for OS prediction.....	245
Figure 8.7	KM comparative analysis of using SeNMo in stratifying patient outcomes.....	248
Figure 8.8	Pan-cancer primary cancer type prediction.....	250
Figure 8.9	Hyperparameters search for training on Pan-cancer multiomics data.....	253
Figure 8.10	Hyperparameters search for finetuning on CPTAC-LSCC multiomics data.....	253

Figure 8.11	Hyperparameters search for fine-tuning on Moffitt’s Lung SCC multiomics data.....	254
Figure 9.1	The schematic layout of multi-modal data integration framework.....	260
Figure 9.2	A schematic layout of MINDS multimodal data for PARADIGM .....	267
Figure 9.3	C-Index on clinical, pathology reports, WSIs, molecular data .....	269
Figure 9.4	C-Index for OS predictions on clinical, pathology reports, and WSIs .....	270
Figure 9.5	C-Index for OS predictions on clinical, molecular, and WSIs SCC data. ....	271
Figure 9.6	C-Index for OS predictions on clinical, pathology reports, and molecular data .....	272
Figure 9.7	Performance comparison of OS prediction on clinical and WSIs SCC data .....	273
Figure 9.8	Performance comparison of OS prediction on clinical and pathology reports .....	275
Figure 9.9	Performance comparison of OS prediction on clinical and molecular SCC data .....	276
Figure 9.10	C-indices from different models trained on combinations of modalities .....	276
Figure 9.11	Impact of input graph sparsity on the predictive performance of GNNs .....	278

## **Abstract**

This dissertation explores the intersection of graph theory and deep learning, focusing on enhancing the robustness of deep neural networks (DNNs) and applying these advancements to complex problems like cancer diagnosis and treatment. We investigate the structural properties of graphs and their influence on neural network performance, particularly in multimodal learning. The work delves into the design space of DNN architectures using graph-theoretic measures, transforming graphs into DNN architectures for various tasks, and examining their robustness against noise and adversarial attacks. The study extends to medical imaging, highlighting advanced DNN architectures like U-Net for brain tumor segmentation. It addresses the evolution of digital pathology, the challenges of task-specific AI/ML models, and the transformative potential of foundation models and generative AI. The integration of multimodal oncology data through Graph Neural Networks (GNNs) and Transformers is explored, showcasing their potential in improving diagnostic and prognostic models. The development of the Multimodal Integration of Oncology Data System (MINDS) and SeNMo, a deep learning model for multi-omics data, underscores the significance of harmonizing diverse data types for personalized cancer care. We also proposed a GNN-based hierarchical relational model, PARADIGM, that enhances survival predictions by integrating multimodal datasets. The compilation of articles is structured into eight chapters, each focusing on different aspects of learning, from theoretical foundations to practical applications, offering a comprehensive overview of the field and its implications for future research and clinical practice in oncology and computational pathology.

## Chapter 1: Introduction

The intersection of graph theory and deep learning represents a rapidly evolving frontier in artificial intelligence (AI) research. This compilation of different articles aims to explore how graph-theoretic principles can enhance the robustness of deep neural networks (DNNs) and how graphs can be applied to complex, real-world problems, such as cancer diagnosis and treatment. By investigating the structural properties of graphs and their influence on neural network performance, especially in the realm of multimodal learning, we provide a comprehensive literature that bridges theoretical insights with practical application of graphs and graph learning in oncology.

The architectures of deep artificial neural networks (DANNs) are a subject of extensive research, driven by the quest to enhance their predictive performance. However, the robustness of these architectures—particularly their resilience to noise and adversarial attacks—remains less thoroughly explored, especially in computer vision applications. We present a detailed investigation into the relationship between the robustness of DANNs and their underlying graph architectures. Our exploration begins by delving into the design space of DANN architectures using graph-theoretic robustness measures. We transform these graphs into DANN architectures, applying them to various image classification tasks. This approach allows us to investigate how the robustness of trained DANNs against noise and adversarial attacks correlates with their underlying graph structures. Notably, we demonstrate that the robustness of DANNs can be quantified prior to training using graph structural properties such as topological entropy and Olivier-Ricci curvature. These measures show significant reliability, especially for complex tasks and larger DANNs, suggesting that structural properties can be predictive of network robustness.

Next we transition to the medical domain and study machine learning and its applications on medical data. In the realm of medical imaging, brain tumor segmentation stands out as a

critical application. This process involves the pixel-level delineation of brain tumor structures in medical images, such as Magnetic Resonance Imaging (MRI). Accurate brain segmentation is essential for radiotherapy treatment planning and improving tumor surveillance. Despite the challenges posed by the complex topology of anatomical structures, noise from image acquisition, and the heterogeneity of tumor signals, machine learning (ML) techniques, including DNNs, have significantly advanced classification and segmentation tasks. In this context, supervised learning models, particularly architectures like U-Net and its variants, have shown remarkable performance in brain tumor segmentation.

Digital pathology has revolutionized the traditional practice of analyzing tissue under a microscope by transforming it into a computer vision workflow. Whole slide imaging enables pathologists to view and analyze microscopic images on a computer monitor, facilitating computational pathology. AI and ML have propelled computational pathology to new heights, achieving near or above-human performance in many image processing tasks. However, the performance of task-specific AI/ML models is often constrained by the availability of annotated training datasets and their generalizability to new datasets or unseen variations.

Cancer research, with its intricate complexity of biological, clinical, and molecular characteristics, demands sophisticated data integration techniques. Traditional analytical methods often fall short of capturing the comprehensive landscape of cancer's multifaceted characteristics. The integration of multimodal data—ranging from clinical records to digitized histopathology slides and molecular data—presents opportunities for enhancing the accuracy and reliability of cancer diagnosis and treatment. Graph Neural Networks (GNNs) and Transformers have emerged as powerful tools for multimodal learning, capable of synthesizing information from diverse sources to provide a nuanced understanding of cancer. The advent of foundation models and generative AI marks a significant shift in AI/ML development. Foundation models are large AI models trained on extensive datasets and later fine-tuned for specific tasks using modest amounts of annotated data. These models offer in-context learning, self-correction, and prompt adaptation to user feedback, presenting transformative potential for digital pathology and beyond.

However, the problem is that large scale, ML-ready data is not readily available for research. We present a multimodal data integration systems, called Multimodal Integration of Oncology Data System (MINDS), that consolidates disparate data into an interconnected, patient-centric framework. MINDS exemplifies a flexible, scalable, and cost-effective metadata framework for fusing public cancer data into a comprehensive system, enabling researchers to uncover diagnostic and prognostic insights through harmonized multimodal data. We also extended MINDs to an embedding or vector database system called HoneyBee that provides the feature vectors from embedding models hosted on the open-source Hugging Face platform.

Having embedding models to generate unimodal embeddings is a pre-requisite in multimodal learning. However, in the realm of multi-omics research, there is no deep learning model that has been trained across a wide breadth of data heterogeneity. We build a deep neural network called SeNMo (Self-normalizing Network for Multi-omics) that has demonstrated robust performance in handling high-dimensional multi-omics data. SeNMo's ability to predict patient outcomes using diverse molecular data types across multiple cancer types showcases the transformative potential of multi-omics approaches in oncology.

Finally, we present a GNN-based hierarchical relational model for predicting clinical outcomes in cancer. This model integrates multimodal, heterogeneous datasets to generate relational embeddings and improve survival predictions. By converging individual data modalities into a unified view, our solution offers a comprehensive framework for understanding the genetic, physiological, and psychosocial aspects of cancer.

In summary, this work explores the intersection of deep learning, graph theory, and multimodal data integration, highlighting the potential of advanced AI/ML models to transform cancer diagnosis, treatment, and research. The subsequent chapters delve deeper into these topics, presenting empirical findings and discussing their implications for the future of oncology and computational pathology. This compilation is structured into eight chapters, each presenting a different aspect of learning. Chapter 1 introduces the fundamental concepts of graph structures and how graph-theoretic measures are related to the robustness of DNNs. Chapter 2 delves into brain tumor

segmentation using advanced DNN architectures. Chapter 3 discusses the evolution and challenges of digital pathology and introduces foundation models. Chapter 4 explores the integration of multimodal oncology data and its implications for cancer care. Chapter 5 reviews the advancements in GNNs and Transformers for multimodal learning in oncology. Chapter 6 presents the MINDS framework for integrating diverse oncology data. Chapter 7 introduces SeNMo, a multi-omics model for predicting cancer outcomes. Finally, Chapter 8 proposes a GNN-based model for clinical outcome prediction, emphasizing the convergence of heterogeneous data.

## **Chapter 2: Exploring Robust Architectures for Deep Artificial Neural Networks**

### **2.1 Note to Reader**

This chapter has been previously published in Nature Communications Engineering as: Waqas, A., Farooq, H., Bouaynaya, N.C. et al. Exploring Robust Architectures for Deep Artificial Neural Networks. Nat. Commun Eng 1, 46 (2022), and has been reproduced with permission from Nature Publishing [735].

### **2.2 Introduction**

The architecture or structure of a deep artificial neural network (DANN) is defined by the connectivity patterns among its constituent artificial neurons. The mere presence or absence of a connection between two neurons or a set of neurons may provide a useful prior and improve the predictive performance of a DANN. A range of architectures has been developed over years to tackle various machine learning tasks in computer vision, natural language processing, and reinforcement learning [283, 360, 361, 624, 614, 395, 364, 386, 287, 671, 362, 355, 356, 358, 353, 363, 624, 355]. In general, the process of the development of DANN architectures is manual, iterative and time consuming. AutoML and neural architecture search (NAS) attempt to use machine learning and search the design space of DANNs for architectures that may yield maximum test data accuracy. After the selection of a suitable DANN architecture for the given task, the optimal values of the connections (parameters or weights) are found using the training dataset and the well-known gradient descent or one of its variant algorithms [433, 189]. Recently, considerable research efforts have been focused on automating the laborious task of DANN architecture design and development using techniques of autoML and NAS [757, 778, 797, 183, 456]. However, all such efforts are



primarily focused on improving the test accuracy of the DANN on the given task. In the real world, DANNs face the challenging problem of maintaining their predictive performance in the face of uncertainties and noise in the input data [175, 11, 356, 359]. The noise can be in the attributes of the input samples (attribute noise), and it can be in the associated class label (label noise). DANNs exhibit intrinsic robustness to label noise, and the test accuracy of DANNs drop only marginally against the tens of percentage increase in the label noise [223, 354, 721, 66, 356, 359, 353, 360, 364]. So, the term “noise” in this work refers to the more hostile sample noise, called the attribute noise. The challenge of noise in the attributes of data is further exacerbated for mission-critical application areas, such as clinical diagnosis, autonomous driving, financial decision-making, cyberspace, and defense. Although we have used computer vision for hypothesis testing in this work, we believe that the proposed concepts are equally applicable to other fields mentioned above. Ideally, a real world deployment-ready DANN should be robust to or equivalently maintain its predictive performance against two different types of noise, natural and malicious. The natural noise is related to the out-of-distribution generalization. Such noise is caused by the day-to-day changes in input data, e.g., changes in the hardware or software configurations used for processing input data. The malicious or adversarial noise is generated by an adversary for fooling the DANN into producing an erroneous decision [672, 354]. The adversarial attacks can be at training time (poisoning attack) or at inference time (evasion attack) [50, 359, 364]. The attacks themselves can be targeted in the feature-space [672] or in the problem space [556]. The attacker can have a perfect knowledge (white-box attacks), limited knowledge (gray-box attacks), or zero knowledge (black-box attacks) [556]. Here, the knowledge  $\theta = (D, X, f, w)$  is a set that may contain training data  $D$ , features  $X$ , model  $f$  and its parameters  $w$ . Black-box attacks are strict subset of white-box attacks, and white-box attacks perform better than other attacks against a DANN [115]. This means that gradient-based attacks outperform gradient-free attacks [115]. Following these facts and the Kerckhoffs’ principle [351], in this work we have employed white-box attacks assuming attacker’s perfect knowledge at inference time. Moreover, this work focuses on evaluating the inherent robustness of DANNs to identify the architectures that have a natural relative immunity to

adversaries and insults. Mechanisms on improving the robustness of DANNs is not covered in this work.

It has been shown with the help of Percolation theory that the architecture or structure underlying a network of any real-world system may play a key role in defining its robustness to various insults and attacks [67, 356, 359]. Graph-theoretic measures, such as network topological entropy and Ollivier-Ricci curvature, successfully quantify the functional robustness of various networks [678]. Examples include studying the behavior of cancer cells, analyzing the fragility of financial networks, studying the robustness of brain networks, tracking changes attributable to age and Autism Spectrum Disorder (ASD), and explaining cognitive impairment in Multiple Sclerosis (MS) patients [606, 607, 211, 212]. Recently, the relationship between the architectures of DANNs (quantified by various graph-theoretic measures before training) and their predictive accuracy (available after training) has been established [757, 778]. Various graph-theoretic measures (e.g., path length and clustering coefficient) calculated from the architectures of DANNs are quantitatively linked to their accuracy on various image classification tasks. However, the relationship between the graph-theoretic measures related to the robustness (e.g., entropy and Ollivier-Ricci curvature) of the architecture of DANNs and their performance against natural and adversarial noise has never been explored. Establishing such a relationship will allow the autoML and NAS research community to design and develop robust DANNs without training and testing these architectures.

Graph-theoretic measures that are related to the vulnerability and robustness of networks can be categorized into graph connectivity measures, adjacency spectrum measures, and Laplacian spectrum measures [227]. Based on the graph properties such as Ollivier-Ricci curvature, betweenness centrality, and shortest path length between nodes, more advanced network measures have been recently proposed. For example, graph and node-based multifractal analysis [755, 588], and fitness-based network efficiency for heterogeneous nodes [583] quantify the topology and robustness of complex networks. In this work, we study graph-theoretic properties of architectures of DANNs to quantify their test-time robustness. Specifically, we use the graph measures of topological entropy and curvature of the architecture of DANNs as robustness metrics. We have

considered all three aforementioned categories of graph-robustness measures in our experiments. Our choice of reporting the curvature and entropy as the robustness measures of DANNs is based on empirical evidence presented in this paper. We make two distinct research contributions to the robustness analysis of DANNs: (1) We establish a quantitative relationship between the graph-theoretic robustness measures of entropy and curvature of DANNs (available before training) and the robustness of these DANNs to natural and adversarial noise (evaluated after training DANNs). Previous studies explored graph measures that relate to the performance of DANNs, but robustness of DANNs through graph-robustness measures has never been studied. We show that graph entropy and curvature are related to DANNs' robustness and these structural measures can identify robust architectures of DANNs even before training for the given task. (2) We show that relationship between the graph robustness measured using entropy and Ollivier-Ricci curvature and the robustness performance of DANN against noise and adversarial attacks becomes significantly stronger for complex tasks, larger datasets, and bigger DANNs. Given that the sizes of DANNs and the complexity of tasks/datasets are growing significantly for many real-world applications, the strong entropy-robustness relationship assumes greater importance. The autoML/NAS design problems where robustness of DANNs is vital, our analysis can help identify robust architectures without the need to train and test these DANNs under various noisy conditions.

In Fig. 2.1, we provide an overview of the proposed approach. Fig. 2.1(a) illustrates how graph-theoretic measures are often applied in Network Science (NetSci) to study various real-world networks. The illustrated examples include biological systems such as brain networks, economic systems such as financial networks, and social systems such as social networks. Path length, graph connectivity, efficiency, degree measures, clustering coefficient, centrality, spectral measures (curvature, entropy), and fractal analysis are the graph-theoretic measures that researchers have employed for studying real-world networks [606, 211, 212, 607, 635, 707].

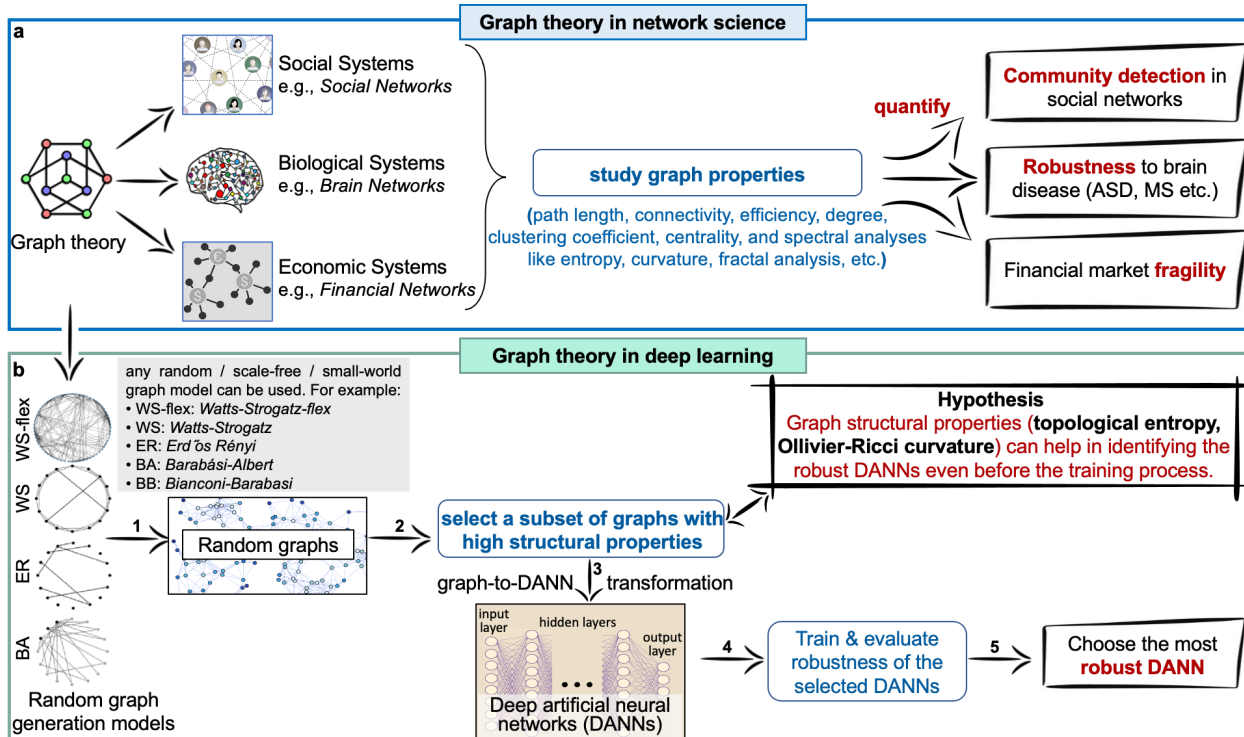


Figure 2.1 Exploring robustness of DANNs with graph-theoretic measures.

Fig. 2.1(b) illustrates our proposed methodology. Our approach consists of five steps: (1) build random, scale-free, or small-world networks or graphs using classical families of graphs, (2) calculate graph-theoretic measures of these random graphs in the graph domain and select a small subset from the entire design space for further analysis, (3) convert selected random graphs into architectures of DANNs (e.g., MLPs, CNNs, ResNets, EfficientNets), (4) train, validate and test these DANNs under different natural noise and adversarial conditions, and (5) analyze and link the robustness of architectures (measured with graph-theoretic properties) to the performance of trained DANNs against natural noise and adversarial attacks. We hypothesize that the graph-theoretic measures that quantify the robustness of networks/graphs in the NetSci domain will also provide insight into the robustness of DANNs in the deep learning domain. We provide empirical evidence to support our hypothesis. We use the term DANN for deep artificial neural networks,

graphs for unweighted directed acyclic graphs, and network for various networks as used in the network science (NetSci) domain.

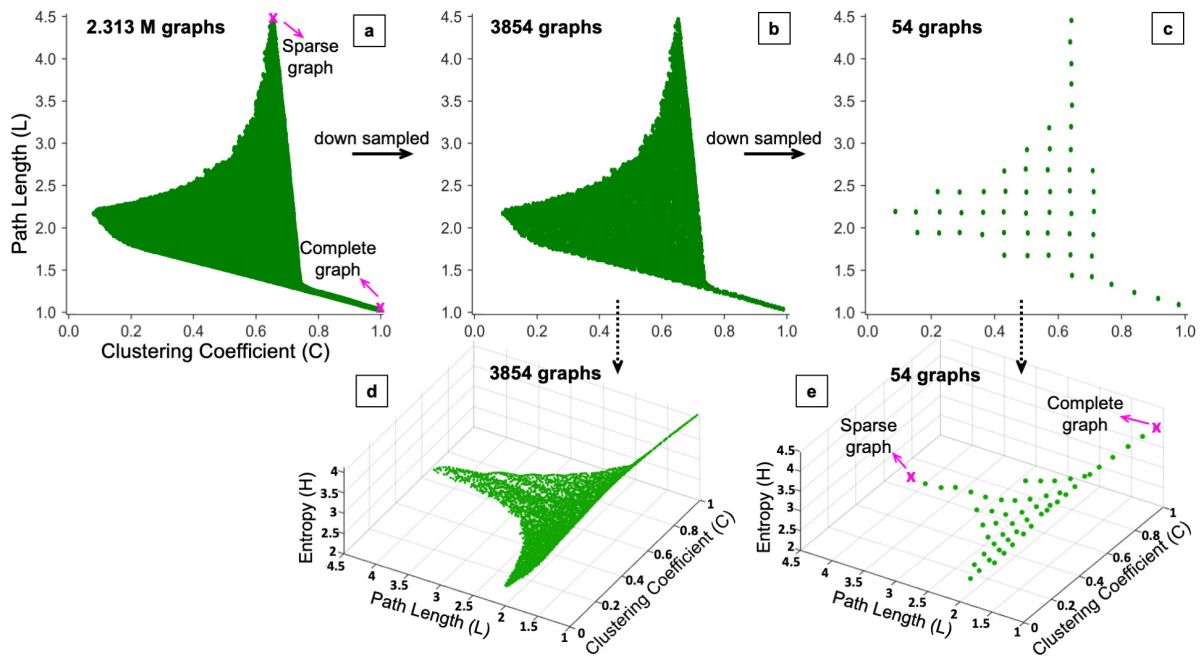


Figure 2.2 The graph design space for generating random graphs.

## 2.3 Results

### 2.3.1 Graph Design Space

Being a sub-field of autoML, Neural Architecture Search (NAS) is the process for searching suitable architectures of neural networks for a given task [207]. Design space is a component of NAS and is composed of a set of architectures of neural networks [569]. We use two graph measures, average path length ( $L$ ) and clustering coefficient ( $C$ ), for exploring the graph design space. Extensively used in prior works [742, 654, 71], these measures smoothly span the whole design space of the random graphs as illustrated in Fig. 2.2. We generate 2.313 Million (M) candidate random graphs using Watts-Strogatz flex (WS-flex) graph generator for a range of  $C$  and  $L$  values as illustrated in Fig. 2.2(a). We chose WS-flex because its graphs are superset of graphs generated

by three classical methods including, Watts-Strogatz (WS), Erdős Rényi (ER), and Barabási-Albert (BA) [742, 208, 23]. We downsample 2.313 M candidate WS-flex graphs into coarser bins of 3854 and 54 graphs (Fig. 2.2(b)&2.2(c)), where each bin has at least one representative graph. We visualize our candidate graphs using their average path length( $L$ ), clustering coefficient ( $C$ ), and entropy ( $H$ ). Entropy is a graph-theoretic measure for robustness and we visualize our design space spanned by  $(C, L, H)$ , as depicted in Fig. 2.2(d)&2.2(e). Fig. 2.2(a)&2.2(e) also depict the extreme cases of complete and sparse graphs. For a complete graph, we have  $(C, L, H) = (1.0, 1.0, 4.1)$ .

We transform the downsampled 54 graphs into DANNs using the technique of relational graphs proposed by You *et al.* [778]. The same 54 graphs are transformed into multiple types of DANNs including, multilayer perceptrons (MLPs), convolutional neural networks (CNNs), and residual neural networks (ResNets). Because of the flexibility and generalizability of the relational graphs, our graphs to DANNs transformation framework can be translated into diverse architectures, including MLPs, CNNs, ResNets, EfficientNets, etc. We use four image classification datasets of varying complexity to train and evaluate DANNs built using 54 different graph structures. These datasets include Canadian Institute For Advanced Research for ten classes (CIFAR-10), hundred classes (CIFAR-100), Tiny ImageNet, and ImageNet [384, 341, 598]. The robustness of trained DANNs is quantified by subjecting these models to various levels and types of natural and malicious noise. We used three types of additive noise, Gaussian, Speckle, and Salt&Pepper. For malicious noise, we employ three different adversarial attacks with varying severity levels. These include Fast Gradient Sign Method (FGSM) [254], Projected Gradient Descent (PGD) [466], and Carlini Wagner (CW) [116].

### 2.3.2 Performance Trends of DANNs

Fig. 2.3 presents predictive performance of different MLPs, CNNs, and ResNets built using 54 selected graphs and trained on four different image classification datasets. Performance evaluation of the trained DANNs is done using randomly selected 30 different sets of clean, adversarial, and noisy images. The test accuracy numbers presented in Fig. 2.3 are average values across all tests.

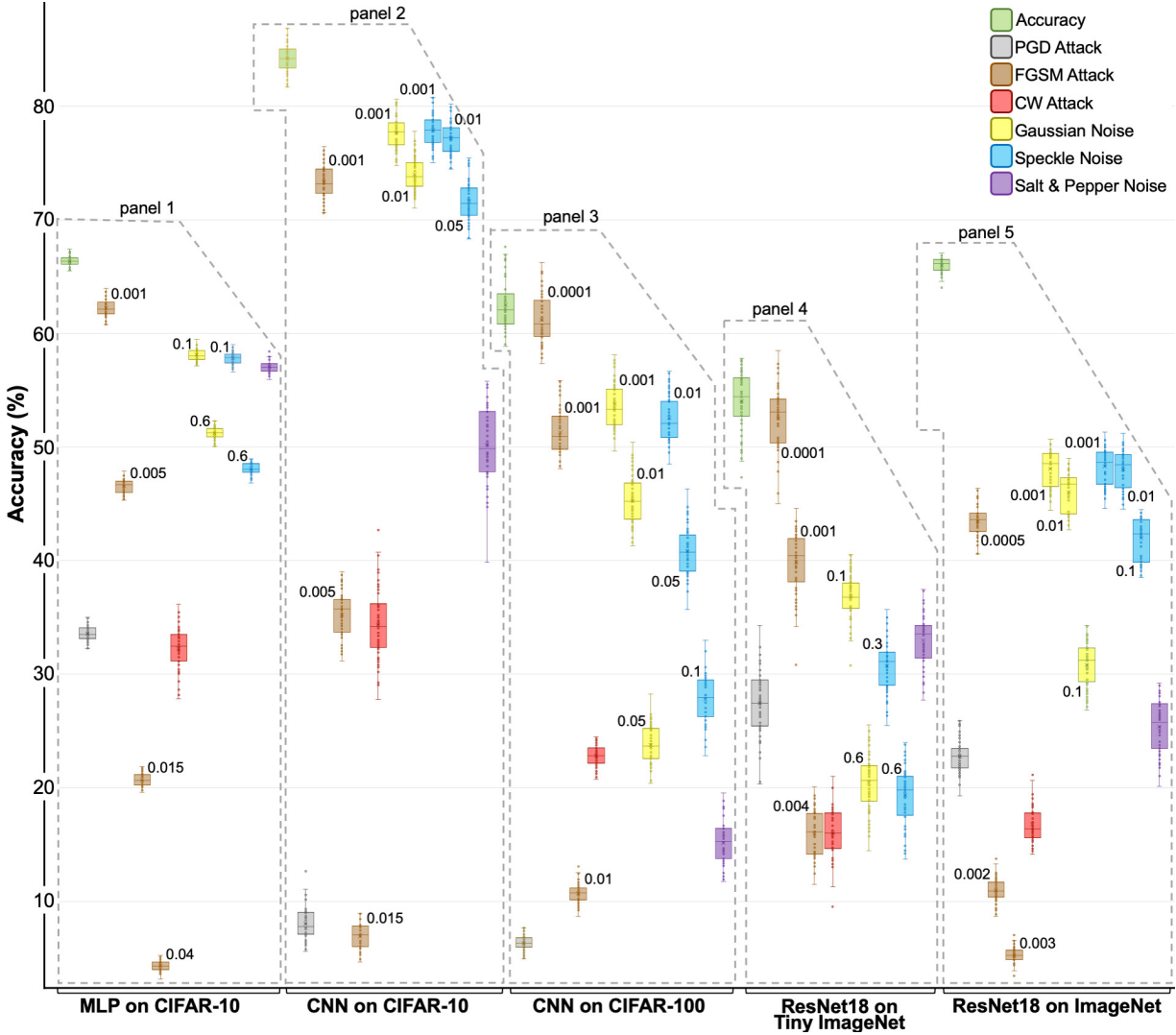


Figure 2.3 Test accuracies of different Deep Artificial Neural Networks.

MLPs on CIFAR-10 shown by panel 1 of Fig. 2.3 presents test accuracies of 54 MLPs under different conditions. The average clean test accuracy is  $66.3 \pm 0.46\%$ , which drops to  $33.6 \pm 0.64\%$  under PGD attack and to  $32.3 \pm 1.9\%$  for the CW attack. With FGSM attack levels of  $\epsilon = [0.001, 0.005, 0.015, 0.04]$ , the test accuracy drops to  $[62.2 \pm 0.72\%, 46.6 \pm 0.6\%, 20.7 \pm 0.54\%, 4.34 \pm 0.44\%]$ . For low noise level of natural or additive noise ( $\sigma_{\text{noise}}^2 = 0.1$ ), test accuracy under Gaussian noise is  $58.1 \pm 0.50\%$  and under speckle noise  $57.8 \pm 0.52\%$ . For high noise level ( $\sigma_{\text{noise}}^2 = 0.6$ ), the test

accuracy under Gaussian noise is  $51.2 \pm 0.55\%$  and under speckle noise  $48.1 \pm 0.55\%$ . Under Salt&Pepper noise (*salt vs. pepper=0.5*), the test accuracy is  $57.06 \pm 0.5\%$ .

CNNs on CIFAR-10 shown by panel 2 of Fig. 2.3 shows the average test accuracies of 8-layer CNNs built from the same 54 candidate graphs. We observe that the average clean test accuracy for CNNs is  $84.19 \pm 1.26\%$ , dropping to  $8.07 \pm 1.43\%$  under PGD attack, and to  $34.35 \pm 3.12\%$  under CW attack. We noticed similar trends for various levels of FGSM attacks, as well as for the Gaussian, speckle, salt&pepper noise.

CNNs on CIFAR-100 shown by panel 3 of Fig. 2.3, presents CNNs trained on CIFAR-100 dataset. The average test accuracy is  $62.49 \pm 2.24\%$  for clean test dataset, which reduces to  $6.35 \pm 0.64\%$  for the PGD attack, and  $22.83 \pm 0.94\%$  for the CW attack. With FGSM attack levels of  $\epsilon=[0.0001, 0.001, 0.01]$ , the test accuracy is  $[61.19 \pm 2.30\%, 51.32 \pm 2.06\%, 10.71 \pm 0.83\%]$ . For Gaussian noise levels of  $\sigma_{\text{noise}}^2=[0.001, 0.01, 0.05]$ , the test accuracy of CNNs is  $[53.75 \pm 2.01\%, 45.29 \pm 2.12\%, 23.76 \pm 1.66\%]$ . For speckle noise levels of  $\sigma_{\text{noise}}^2=[0.01, 0.05, 0.1]$ , the test accuracy is  $[52.54 \pm 2.00\%, 40.82 \pm 2.11\%, 27.79 \pm 2.10\%]$ . For salt&pepper noise, the test accuracy is  $15.14 \pm 1.81\%$ . The drop in test accuracy for all cases is significantly more than that of CIFAR-10 dataset.

ResNet-18 on Tiny ImageNet given by panel 4 of Fig. 2.3 shows 54 different ResNets trained on Tiny ImageNet. The average clean test accuracy is  $54.08 \pm 2.54\%$ ,  $27.45 \pm 2.90\%$  under PGD attack, and  $16.00 \pm 2.10\%$  under CW attack. For the FGSM attack levels of  $\epsilon=[0.0001, 0.001, 0.004]$ , the accuracy is  $[52.53 \pm 2.85\%, 39.86 \pm 2.78\%, 15.97 \pm 2.04\%]$ . For Gaussian noise levels of  $\sigma_{\text{noise}}^2=[0.1, 0.6]$ , the test accuracy is  $[36.81 \pm 1.96\%, 20.44 \pm 2.47\%]$ . For speckle noise levels of  $\sigma_{\text{noise}}^2=[0.3, 0.6]$ , the test accuracy is  $[30.73 \pm 2.31\%, 19.34 \pm 2.53\%]$ . For salt&pepper noise, the test accuracy is  $33.00 \pm 2.24\%$ .

ResNet-18 on ImageNet shown by panel 5 of Fig. 2.3 presents ResNets trained using ImageNet. Due to the large number of images available for training, the average clean test accuracy of all 54 ResNets-18 was  $66.0 \pm 0.62\%$ , a significant improvement over Tiny ImageNet experiments ( $54.08 \pm 2.54\%$ ). Under PGD attack, the test accuracy drops to  $22.75 \pm 1.39\%$ , and to  $16.78 \pm 1.54\%$



under CW attack. For the FGSM attack levels of  $\epsilon=[0.0005, 0.002, 0.003]$ , the accuracies are  $[43.47 \pm 1.40\%, 10.96 \pm 1.09\%, 5.27 \pm 0.65\%]$ . Similar trends are observed for the additive Gaussian and speckle noise under the  $\sigma_{\text{noise}}^2=[0.001, 0.01, 0.1]$ . For salt&pepper noise, the test accuracy drops to  $25.39 \pm 2.35\%$ .

For the comparison of MLPs vs. CNNs on CIFAR-10, we observed that CNNs achieve higher accuracy on the clean test data as compared to MLPs on CIFAR-10 dataset. However, under adversarial conditions (FGSM, PGD, and CW attacks), the drop in the performance of CNNs is significantly higher than MLPs as shown in panels 1 and 2 of Fig. 2.3. The test accuracy drop is  $\sim 76\%$  for CNNs compared to  $\sim 33\%$  for MLPs under PGD attack. For the CW attack, the accuracy drop for CNNs is  $\sim 50\%$  compared to  $\sim 34\%$  for MLPs. The same trend was observed for all severity levels of the FGSM attack. Generally, as expected CNNs outperform MLPs under clean test conditions; however, MLPs are more robust to adversarial perturbations as compared to CNNs. We argue that the observed fragility of CNNs is linked to their weight sharing and shift-invariant characteristics, which was previously noted by Zhang et al. [796].

### 2.3.3 Robustness Analysis

Our work is a cross-pollination between graph theory and deep learning. We attempt to link the robustness of graphs underlying the architectures of DANNs to their performance against noise and adversarial attacks. On the graph theory side, we use entropy and Ollivier-Ricci curvature to quantify the robustness of graphs. These graphs, in turn, are used to build architectures of DANNs. On the deep learning side, we train these DANNs and quantify their robustness using test accuracy against various types of noise and adversarial attacks. Entropy and Ollivier-Ricci curvature have been extensively studied in the NetSci. These measures have been shown to capture the robustness of cancer networks [678, 606], track changes in brain networks caused by age and Autism Spectrum Disorder [211], explain cognitive impairment in patients with Multiple Sclerosis [212], identify financial market’s fragility [607], and detect communities in complex social networks [635]. We study the robustness of DANNs and establish the statistical correlation of the observed robustness

with entropy and curvature. The correlation results for entropy of graphs and robustness of DANNs for different datasets are given in Fig. 2.4, 2.5, and 2.6. The correlation results between the robustness of DANNs and graph measures such as curvature, average degree, and global efficiency are provided in Supplementary Notes 2.9, 2.10, and 2.11, respectively.

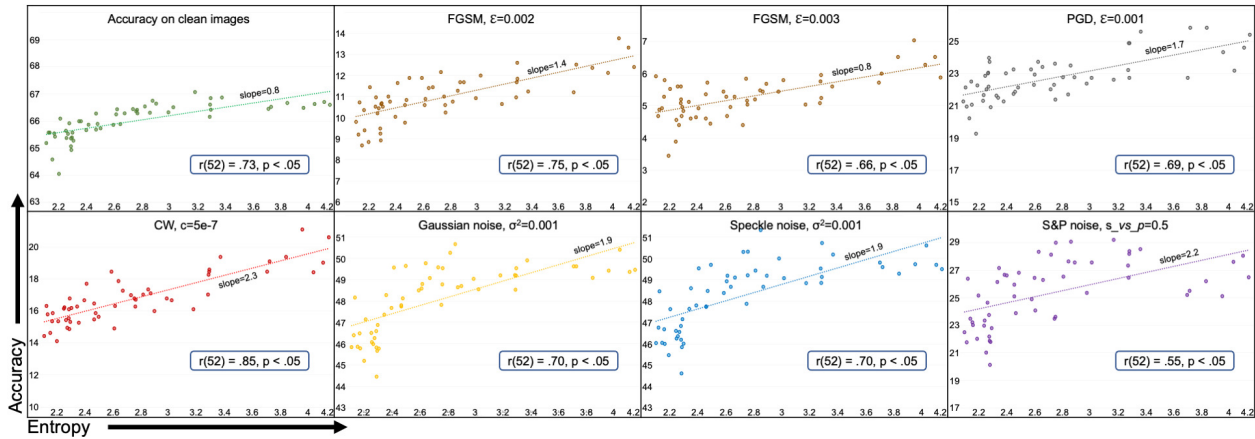


Figure 2.4 Test accuracy vs. entropy for ResNet-18 on ImageNet.

ResNet-18 on ImageNet and Tiny ImageNet shown by Fig. 2.4 presents 54 ResNet-18 DANNs trained on ImageNet and tested on clean images, adversarial examples generated with FGSM, PGD, and CW attacks, and images with additive Gaussian, speckle, and salt&pepper noise. Each sub-plot shows entropy ( $H$ ) of the underlying graph structure and the test accuracy of corresponding ResNet-18 under various conditions. The Pearson product-moment correlation coefficient values between entropy and accuracy along with  $p$  values are shown on each sub-plot. There was a very strong positive correlation between the two variables,  $r = 0.73$ ,  $n = 54$ ,  $p < 0.05$  for the clean test dataset. We note similar behavior under PGD and CW attacks, that is, a strong correlation between entropy and accuracy exists,  $r = 0.69$  for PGD and  $r = 0.85$  for CW,  $p < 0.05$  for both. Similarly, strong positive correlation trends exist for various severity levels of FGSM attack, Gaussian, speckle, and salt&pepper noise. The correlations in these experiments are all statistically significant ( $p < 0.05$ ). In general, across all types of adversarial attacks and noises, the DANNs corresponding to graphs

with higher entropy showed stronger robustness and vice versa. Additional results are provided in Supplementary Figs. 2.10 and 2.11.

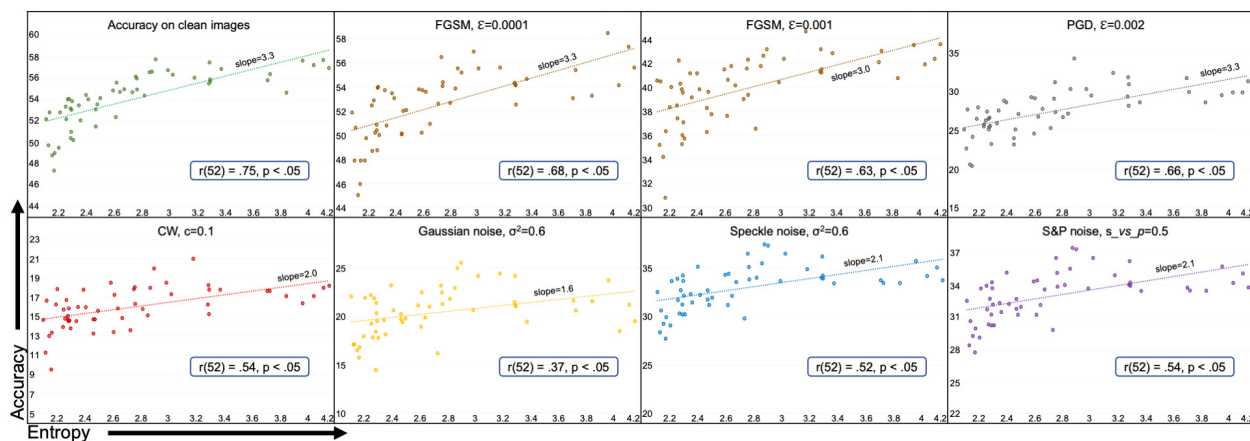


Figure 2.5 Test accuracy vs. entropy for ResNet-18 on Tiny ImageNet.

Fig. 2.5 presents test accuracy vs. entropy plots for 54 ResNet-18 models trained using Tiny ImageNet and tested under various noisy conditions. We observe a strong positive correlation between entropy and predictive performance under all noise conditions, except Gaussian noise where correlation is moderate. However, there is a notable decrease in the Pearson product-moment correlation coefficient values in all noise categories compared to the same DANNs when trained and tested on ImageNet. As Tiny ImageNet is a subset of ImageNet with only 200 distinct classes instead of 1,000, the observed decrease in the correlation may be linked to the reduction in complexity of the task, i.e., 200 classes instead of 1,000.

In Fig. 2.6(a)&(b), we present accuracy vs. entropy plots for the 54 8-layer CNNs trained on CIFAR-100 and CIFAR-10 datasets and tested under various noisy conditions. For the CIFAR-100 experiments, we observe very strong correlation between entropy and predictive performance except for CW ( $r = 0.40, p < .05$ ), PGD ( $r = 0.39, p < .05$ ) adversarial attacks, and salt&pepper noise ( $r = 0.47, p < .05$ ). For CIFAR-10 dataset, there is a strong correlation between entropy and predictive performance except for the PGD, CW attacks and salt&pepper noise which were not statistically significant.

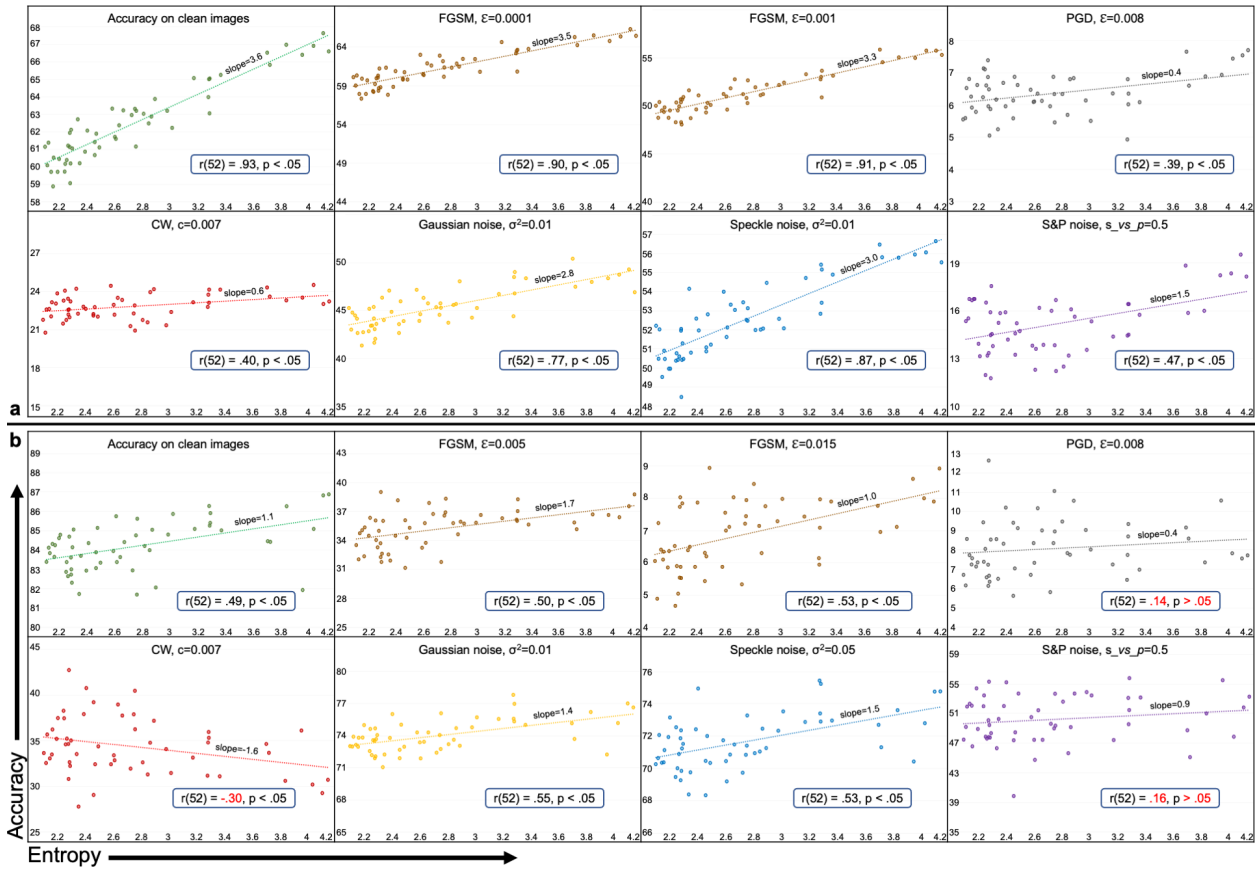


Figure 2.6 Accuracy vs. entropy for CNNs.

We opine that the weak correlation between graph entropy and DANNs’ performance under PGD and CW attacks is due to the strong nature of PGD and CW attacks on relatively simple classification tasks of CIFAR compared to Tiny ImageNet and ImageNet. This opinion was strengthened from the evaluation results of the CNNs on a more straightforward classification task of CIFAR-10. We observe that the correlation of entropy with the predictive performance of CNNs reduces for all categories. Moreover, the entropy’s correlation with accuracy under CW attack becomes negative. Under PGD attack and salt&pepper noise, it becomes insignificant with  $p > 0.05$  as highlighted by the red text in respective subplots of Fig. 2.6. Additional results are provided in Fig. 2.12.

### 2.3.4 Effect of Task and Model Complexity

We observed that DANNs’ robustness, evaluated under noisy conditions, and the robustness of underlying graph structures, quantified using entropy, are strongly correlated. Moreover, this correlation has a strong dependence on the complexity of the model and/or the dataset. In our settings, the model complexity refers to the number of parameters in the model and the task complexity refers to the number of classes in the dataset. As the complexity of the task and/or model increases, the correlation between robust performance and entropy of DANNs increases, as shown in Fig. 2.7.

In Fig. 2.7(a), models evaluated for the 10-class and 100-class tasks are 8-layer CNNs. For 200-class and 1000-class tasks, ResNet-18 models were evaluated. We note that for the same 8-layer CNNs, increasing the complexity of the task (from 10 classes of CIFAR-10 to 100 classes of CIFAR-100) results in increase in the correlation values as noted by the Student’s t-test ( $t = -2.31, n = 34, p < .05$ ). The same holds true for increasing the task complexity from 200 classes of Tiny ImageNet to 1000 classes of ImageNet and using the same ResNet-18 models ( $t = -4.66, n = 23, p < .05$ ). Comparing two different models evaluated on separate tasks, we notice insignificant increase in the correlation values (i.e.,  $p > 0.05$ ) as in the case of 8-layer CNN models evaluated on CIFAR-100 dataset versus ResNet-18 models evaluated on Tiny ImageNet dataset. While Student’s t-test is used to compare two related samples, we used the F-test to see the equality of the two unrelated populations (CNNs vs. ResNet-18s) for performance variance on separate tasks. We observe that there is significance difference between variances of the two analyzed sets of experiments,  $F(14, 14) = 2.61, p < 0.05$ . In Fig. 2.7(b), we present the effect of increasing the model complexity measured by the number of parameters against the entropy-robustness correlation. We observe that for the same CIFAR-100 dataset, as the model complexity increases from  $\sim 0.3$  M parameters in ResNet-29 to  $\sim 1.3$  M in CNN, the entropy-robustness correlation increases significantly ( $t = -6.8, n = 23, p < .05$ ). Similarly, the entropy-robustness correlation increases significantly ( $t = -5.85, n = 27, p < .05$ ) when model complexity increases from  $\sim 0.5$  M parameters in ResNet-41 to  $\sim 1.3$  M in CNN for CIFAR-100 task. We also note that

this increase in significance is large when difference between the number of parameters between models is large. Our analysis of DANNs' robustness show that a correlation exists between graph entropy and robustness of DANNs, and this correlation has a strong dependence on the complexity of task and model.

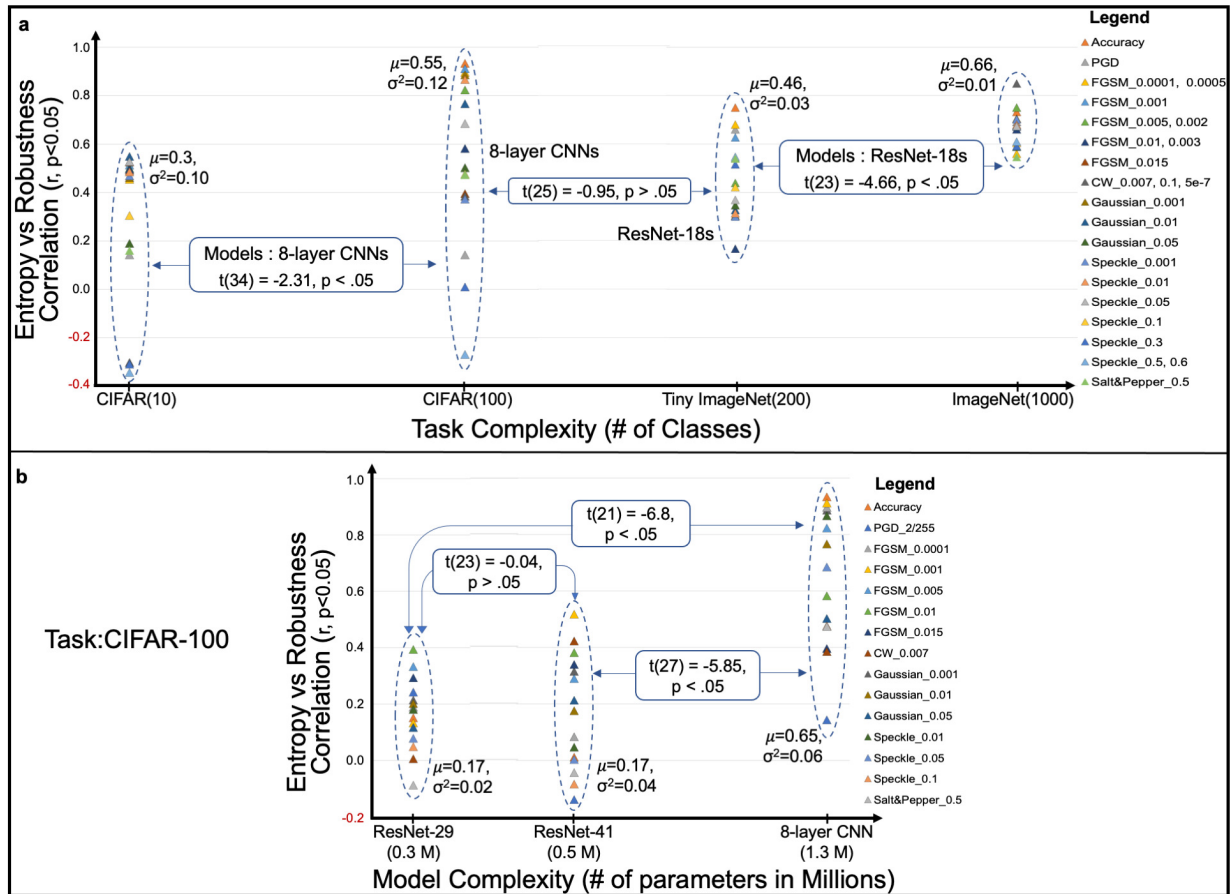


Figure 2.7 Effect of task and model complexity on entropy-robustness relationship.

We summarize our results and robustness analysis here. The graph structural entropy of complex DANNs, like ResNets, is a strong indicator of their robustness against all of the analyzed additive noise and adversarial attacks on complex tasks such as 1000-class ImageNet. As we reduced the task complexity to 200-class Tiny ImageNet, the entropy-robustness relationship decreased for all categories of noise and adversarial attacks. Similarly, graph entropy of relatively simple DANNs, like 8-layer CNNs, is a strong indicator of robustness against relatively simple adversarial attacks,

like FGSM, and all types of additive noise on the CIFAR-100 task. But under strong adversarial attacks, such as PGD and CW, the entropy-robustness relationship reduces for 8-layer CNNs. Similar trends were observed when testing the same 8-layer CNNs on the CIFAR-10 task. The only limitation we encountered in our analysis was that of MLPs trained and evaluated over CIFAR-10 dataset, as given in Section 2.8 and Fig. 2.9.

## 2.4 Discussion

In this work, we have shown that graph structural properties such as entropy and curvature can quantify the robustness of DANNs before training. We calculated entropy and curvature of a set of random graphs, which were later transformed into architectures of different types of DANNs. The DANNs were trained and their robustness was evaluated using different types of natural and adversarial noise. We noted that the robustness of trained DANNs was highly correlated with their graph measures of entropy and curvature. We also noted that the said correlations were even stronger for relatively large models and complex tasks.

Currently various autoML and NAS techniques are being developed to search for accurate model architectures for the given datasets and/or tasks [183, 797, 456, 58, 814]. We argue that for many mission-critical applications, the robustness of these models is equally or in some cases more important than accuracy. However, as there are currently no assured ways of estimating the robustness of DANNs in the graph design space except training and testing the candidate DANNs in the deep learning domain. We suggest that the users of autoML/NAS techniques should incorporate entropy and Ollivier-Ricci curvature information into their search framework. Using the graph representation of DANNs within the defined search space, the structure and associated topological properties of DANNs, such as entropy and curvature can be studied. Integrating these structural measures with the existing performance criterion shall enable the autoML/NAS algorithms to quantitatively and qualitatively select robust DANNs instead of exhaustively searching through all possible candidates. The users and autoML/NAS algorithms can directly identify and choose the most robust model out of all the models that meet the accuracy criteria set by the user.

Such a practice would allow users or autoML/NAS algorithms to choose accurate as well as robust DANNs keeping in view the application area of the machine learning model. In Supplementary Note 2.15, we have given an algorithm as a guideline for selecting the robust architecture of a DANN for a given design space.

We have focused our analyses in this work on the computer vision application. However, there are a plethora of other applications and tasks where robustness quantification of DANNs is important [805, 522, 246, 380]. A possible future direction is to extend the presented analysis to more complex applications (e.g., natural language processing, graph data, and Deep Chip industry [379, 821, 336, 679, 368]) and larger models (e.g., Transformers, and Vision Transformers [633, 813]). Given our current analysis, we anticipate that for the larger datasets, complex tasks, and huge models, the graph robustness measures will be even more relevant and will help users/autoML/NAS algorithms find robust DANN architectures.

## 2.5 Methods

We start by presenting the techniques we employed for generating random graphs in the graph theory domain. Next, we describe the graph-theoretical properties used in our experiments to study random graphs. These graph measures are needed to study the structural information of the random graphs. Next, we provide details on transformations for building DANN architectures from random graphs and training these DANNs for various computer vision classification tasks. Finally, we present the multiple conditions, including natural noise and adversarial attacks that we used to evaluate the trained DANNs and quantify their robustness.

### 2.5.1 Generating Random Graphs

Random graphs are extensively used in percolation studies, social sciences, brain studies, and deep learning to understand the behavior of natural systems and DANNs [208, 344, 71, 70, 757]. We used random graphs, called relational graphs, employed recently in deep learning [778].



A recent a study used relational graphs and showed that the performance of a DANN can be quantified using its graph properties such as clustering coefficient and path length [778]. The relational graphs are generated through the WS-flex graph generator. WS-flex is a generalized version of the WS model having same-degree constraint relaxed for all nodes. Parameterized by  $N$  nodes,  $K$  average degree, and  $P$  rewiring probability, we represent these graphs by WS-flex( $N, K, P$ ). For the graph generator, we use notation  $g(\theta, s)$ , where  $g$  is the generator (for example, WS-flex),  $\theta$  represents parameters ( $N, K, P$ ), and  $s$  is the random seed. It is important to note that WS-flex( $N, K, P$ ) graph generator encompasses the design space of all the graphs generated by the three classical families of random graph generators, including Watts-Strogatz (WS), Erdős Rényi (ER), and Barabási-Albert (BA) [742, 208, 23, 778].

## 2.5.2 Graph-Theoretic Measures

*Average Path Length ( $L$ )* is a global graph measure defined as the average shortest path distance between any pair of graph nodes. It depicts the efficiency of the graph with which information is transferred through the nodes [483]. Small values of  $L$  indicate that the graph is globally efficient, and the information is effectively exchanged across the whole network and vice versa. Let  $G$  be an unweighted directed graph having  $V$ , a set of  $n$  vertices  $\{v_1, v_2, \dots, v_n\} \in V$ . Let  $d(v_1, v_2)$  be the shortest distance between  $v_1, v_2$  and  $d(v_1, v_2) = 0$  if  $v_2$  is unreachable from  $v_1$ . Then, average path length  $L$  is defined as,

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} d(v_i, v_j). \quad (2.1)$$

*Clustering Coefficient ( $C$ )* is a measure of the local connectivity of a graph. For a given node  $i$  in a graph, the probability that all its neighbors are also neighbors to each other is called clustering coefficient. The more densely interconnected is the neighborhood of a node, the higher is its measure of  $C$ . Large value of  $C$  is linked with the resilience of the network against random network damage [655]. The small-worldness of networks is also assessed by  $C$  [473]. For a node  $i$  with

degree  $k_i$ , clustering coefficient  $C_i$  is defined as,

$$C_i = \frac{2d_i}{k_i(k_i - 1)}, \quad 0 \leq C_i \leq 1. \quad (2.2)$$

where  $d_i$  is the number of edges between the  $k_i$  neighbors of node  $i$ .

*Graph Spectral Measures* focus on eigenvalues and eigenvectors of the associated graph adjacency and Laplacian matrices. We will use topological entropy and Ollivier-Ricci curvature.

- *Topological Entropy* ( $H$ ) of graph  $G$  having adjacency matrix  $A_G$ , is the logarithm of the spectral radius of  $A_G$ , i.e., logarithm of the maximum of absolute values of the eigenvalues of  $A_G$  [139].

$$H = \log(\lambda_{A_G}). \quad (2.3)$$

- *Ollivier–Ricci Curvature* (ORC) is the discrete analog of the Ricci curvature [519, 520]. From the many alternatives of Ricci curvature [185], we use the definition presented by Farooq et al. [211] (see Fig. 6 of ref). Let  $(X, d)$  be a geodesic (a curve representing the shortest path between two points on a surface or in a Riemannian manifold) metric space having a family of probability measures  $\{p_x : x \in X\}$ . Then, ORC  $\kappa_{ORC}(x, y)$  along the geodesic connecting  $x$  and  $y$  is,

$$\kappa_{ORC}(x, y) = 1 - \frac{W_1(p_x, p_y)}{d(x, y)}, \quad (2.4)$$

where  $W_1$  is the earth mover's distance (Wasserstein-1 metric), and  $d$  is the geodesic distance on the space. Curvature is directly proportional to the robustness of the network. The larger the curvature, the faster will be the return to the original state after perturbation. Smaller curvature means slow return, which is also called fragility [211].

We now provide the notion of robustness and fragility used in this paper. Fluctuation theorem [171] gives the concept of measuring a network's potential of returning to its "relaxed" and unperturbed state when subjected to some random perturbation. Let  $p_{\gamma, \alpha}(t)$  denote the probability

that under some perturbation  $\gamma$  at time  $t$ , the observable mean deviation of the network from its relaxed state is greater than  $\alpha$ . The rate  $R$  at which a dynamic system returns to its original state after perturbation is given by the following function,

$$R := \lim_{t \rightarrow \infty} \left( -\frac{1}{t} \log p_{\gamma, \alpha}(t) \right). \quad (2.5)$$

Here, a large value of  $R$  denotes a prompt return to relaxed state after a small perturbation ( $\gamma$ ), called the network robustness, whereas, a small  $R$  means slow return from a large perturbation ( $\gamma$ ), called the network fragility. In the field of thermodynamics, entropy is closely related to the rate function  $R$  from large perturbations [171, 172]. Fluctuation theorem [171, 170] states that, given random perturbations to the network, change in system entropy  $\Delta H$  is positively correlated to change in robustness  $\Delta R$ , and negatively correlated to change in fragility  $\Delta F$ , (since  $\Delta R := -\Delta F$ ).

$$\Delta H \times \Delta R > 0, \quad (2.6)$$

$$\Delta H \times \Delta F \leq 0. \quad (2.7)$$

Entropy  $\Delta H$  and curvature  $\Delta \kappa_{ORC}$  are also positively correlated (see Equation (7) of Tannenbaum et al. [678]), that is,

$$\Delta H \times \Delta \kappa_{ORC} > 0. \quad (2.8)$$

From Equations (2.6) and (2.8), we see that graph curvature and robustness are also positively correlated,

$$\Delta \kappa_{ORC} \times \Delta R > 0. \quad (2.9)$$

Equations (2.6) and (2.9) are the primary motivation in this work to study the curvature and entropy of deep neural networks.

### 2.5.3 From Graphs to DANNs

Let  $G = (V, \varepsilon)$  be a graph having node-set  $V = \{v_1, v_2, \dots, v_n\}$ , where node  $v$  has feature vector  $\mathbf{x}_v$ , and edge set  $\varepsilon = \{(v_i, v_j) \mid v_i, v_j \in V\}$ . The neighborhood of node  $v$  is defined as  $N(v) = \{u \mid (u, v) \in \varepsilon\}$ . To transform the graphs into DANNs, we adopt the concept of neural networks as relational graphs [778]. In relational graph, a single node represents one input channel and one output channel. Edge in the relational graph represents a message exchange between the two nodes it connects. The message exchange is a message function having node feature  $\mathbf{x}_v$  as input and a message-aggregation function as output. The aggregation function takes a set of messages as input and gives an updated node feature as output. One iteration of this process is one round of message exchange. At each round, each node sends messages to its neighbors, receives messages from all the neighbors, and aggregates them. At each edge, message transformation occurs through a message function  $f(\cdot)$ , followed by summation at each node through an aggregation function  $F(\cdot)$ . The  $i$ -th message exchange round between nodes  $v$  and  $u$  can be expressed as,

$$\mathbf{x}_v^{(i+1)} = F^{(i)}(\{f_v^{(i)}(\mathbf{x}_u^{(i)}), \forall u \in N(v)\}). \quad (2.10)$$

You *et al.* have shown that Equation (2.10) is the general definition of message exchange that can be used to instantiate any neural architecture [778]. We generate MLP, CNN, ResNet-18, and ResNet-29 for each of the 54 random graphs generated from the WS-flex generator. We have illustrated the graph-to-DANN transformation for a 64-node complete graph, generated from the WS-flex generator, to a 5-layer MLP, 8-layer CNN, and ResNet-18 models in the Supplementary Note 2.7 and Supplementary Fig. 2.8.

The same 54 WS-flex random graphs were transformed into a total of 216 DANNs having 54 neural networks in each of the four categories (MLP, CNN, ResNet-18, and ResNet-29). MLPs were trained on CIFAR-10 dataset, whereas, the CNNs were used for training on CIFAR-10 and CIFAR-100 datasets. The same ResNets-18 were used for training on ImageNet and Tiny ImageNet datasets. The baseline architectures have a complete graph structure for each architecture category.

To ensure consistency of our results, we trained each MLP and CNN five times and ResNets one time on respective datasets. The results reported in this paper are average values calculated for thirty different inferences over random test inputs for each MLP and CNN, whereas, five random test inference runs for each ResNet. List of frameworks and hyperparameters used in our experiments are provided in Supplementary Note 2.12. The compute resources and wall clock times are given in Supplementary Note 2.13.

#### 2.5.4 Datasets

We used four different image classification datasets for our experiments that allowed us to train DANNs of different sizes on tasks that varied in their complexity. We used 10-class CIFAR-10 [384] dataset to train MLPs and CNNs. CIFAR-100 [384] dataset having 100 classes was used to train CNNs and ResNet-29. Both datasets have 50,000 training images and 10,000 validation images. To further scale our experiments, we trained ResNet-18 on the Tiny ImageNet [341] dataset having 200 classes. Each class in Tiny ImageNet has 500 training images and 50 validation images. We also trained ResNet-18 on the ImageNet [598] dataset having 1,000 classes, 1.2 M training images and 50,000 validation images.

#### 2.5.5 Robustness Analysis

We assessed the robustness of DANNs against natural additive noise and malicious noise (adversarial attacks). First, we evaluated the models using clean test images from respective datasets. Then we fed DANNs with different test images corrupted with additive noise and adversarial attacks. It is important to note that we chose the severity levels of adversarial attacks and additive noise so that the predictive performance of DANNs is at the minimum greater than 3%. We observed at higher levels of noise, the performance would naturally drop to 0%, which was not helpful in our analysis. Moreover, different severity levels work on different datasets owing to the inherent features and attributes of the data.

We evaluated DANNs using adversarial examples generated from three different types of attacks, (1) Fast Gradient Sign Method (FGSM) [254], (2) Projected Gradient Descent (PGD) [466], and (3) Carlini Wagner (CW) [116].

Consider a valid input  $x_0$  and a target class  $y_0$ . It is possible to find  $x$  through non-random perturbation to  $x_0$  that changes a DANN’s prediction to some other  $y$ ; such  $x$  is called an adversarial example. Given a loss function  $J(\mathbf{x}; \mathbf{w})$ ,  $\mathbf{x}_0$  be the input to the model having parameter  $\mathbf{w}$ , the adversarial example  $\mathbf{x}$  is created by the adversarial attack as,

$$\text{FGSM :} \quad x = x_0 + \epsilon \cdot \text{sign}(\nabla_x J(x_0; w)), \quad (2.11)$$

$$\text{PGD :} \quad x^{t+1} = \Pi_{x+B} \{x^t + \alpha \cdot \text{sign}(\nabla_x J(x^t; w))\}, \quad (2.12)$$

$$\text{CW :} \quad \min_x \|x - x_0\|^2 + c \cdot \max\left\{\left(\max_{i \neq j} \{g_j(x)\} - g_i(x)\right), 0\right\}. \quad (2.13)$$

In Equation (2.11),  $\epsilon$  is the severity level of the attack and should be small enough to make the perturbation undetectable. In Equation (2.12),  $x^t$  is an adversarial example after  $t$ -steps,  $\alpha$  is the step-size,  $\Pi_{x+B}$  refers to the projection operator for each input  $x$  having a set of allowed perturbations  $B$  chosen to capture the perceptual similarity between images. In Equation (2.13),  $c > 0$  is the attack magnitude,  $i$  is the input class, and  $j$  is the target class. FGSM and PGD have the  $l_\infty$ -distance metric, whereas CW, a regularization-based attack, has  $l_2$ -distance metric in our analysis.

For the FGSM attacks, we used eighteen severity levels,  $\epsilon=[0.0001, 0.0005, 0.001, 0.0015, 0.002, 0.0025, 0.003, 0.004, 0.005, 0.01, 0.015, 0.02, 0.025, 0.04, 0.045, 0.06, 0.08, 0.3]$ . For the PGD attacks on CIFAR datasets, we used  $\max(B) = 0.008$ ,  $\alpha = 2/255$ , and  $t = 7$ . For the Tiny ImageNet dataset, we used  $\max(B) = 0.002$ ,  $\alpha = 2/255$ , and  $t = 10$ , and for the ImageNet dataset, we used  $\max(B) = 0.001$ ,  $\alpha = 2/255$ , and  $t = 10$ . For the CW attacks on CIFAR datasets, we used  $c = 0.007$  and steps= 100. For the Tiny ImageNet dataset, we used  $c = 0.01$ , steps= 100, whereas for the ImageNet dataset, we used  $c = 5e - 7$  and steps= 100.

For testing under additive noise, we used three different types of noise to generate corrupt images for all the datasets, (1) Gaussian, (2) speckle, and (3) salt&pepper noise. For each noise type, we used different levels of corruption quantified by the variance and monitored the performance drop. The noise variance used in our experiments for the Gaussian and speckle noise types are  $\sigma^2 = [0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6]$ . For the salt&pepper noise type, we used the maximum ratio of *salt vs. pepper*=0.5, where salt changes a pixel value to 1 randomly and pepper changes a pixel value to 0 randomly, in the input image. Sample images for each dataset used in our experiments, with noise types and levels are shown in Supplementary Note 2.14, Supplementary Figs. 2.14 and 2.15.

### 2.5.6 Statistical Analysis

We conducted various statistical tests to ascertain the significance of our analysis. We computed the Pearson product-moment correlation coefficient to assess the relationship between adversarial accuracy and the graph robust structural properties. We also computed the Pearson product-moment correlation coefficient between different structural graph-theoretic measures as shown in Supplementary Fig. 2.13. For reference, Pearson product-moment correlation coefficient  $r$  ranges from  $-1$  to  $+1$ , where larger the absolute value of  $r$ , the higher is the degree of correlation and stronger is the relationship between variables, and vice versa. Specifically, the absolute values of  $r = 0$  indicates no relationship,  $0 < r \leq 0.3$  indicates weak relationship,  $0.3 < r \leq 0.4$  indicates moderate relationship,  $0.4 < r \leq 0.7$  indicates strong relationship,  $r > 0.7$  indicates very strong relationship, and  $r = 1$  indicates perfect relationship. We used the Student’s t-test to establish that average of the correlations between entropy and robustness for two types of datasets as well as two model types are statistically different. This analysis established how entropy is related to the increase in model size and task complexity. The significance level in all these analyses is set to 95%, i.e.,  $p < 0.05$  indicates statistically significant values.

## 2.6 Data and Code Availability

The datasets used in this study are publicly available on following links: CIFAR-10 and CIFAR-100<sup>1</sup>, Tiny ImageNet<sup>2</sup>, and ImageNet<sup>3</sup>. The source data of figures are given in Supplementary Data. Correspondence should be addressed to A.W. For the simulations in deep learning domain, we have used PyTorch machine learning library, primarily developed by Facebook’s AI Research lab. The base-code for relational graph experiments (<https://github.com/facebookresearch/graph2nn/>) is under the MIT License with copyright (c) Facebook, Inc. and its affiliates. The calculations in graph domain and graph theoretic measures have been coded in Matlab software. Both of these code-packages (in PyTorch and Matlab) are published in the GitHub repository associated with this paper (<https://github.com/Waasem/RobDanns>).

## 2.7 Graph to DANN Transformation

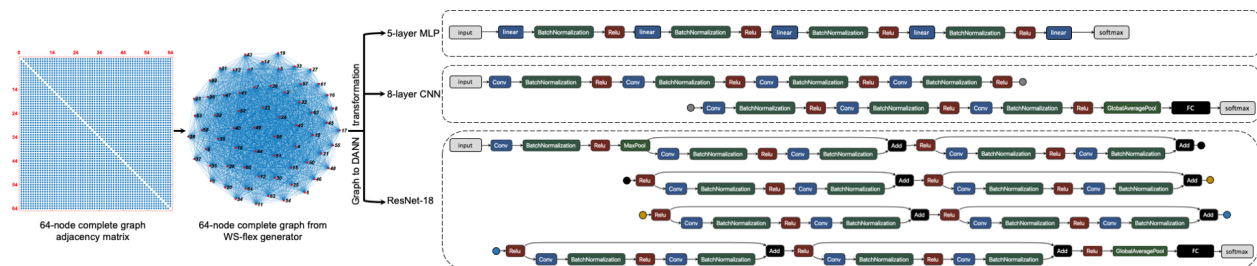


Figure 2.8 Schematic for graph to DANN transformation.

## 2.8 Further Results and Limitations of the Study

In our experiments with 5-layer MLPs trained and tested over CIFAR-10 dataset, we noticed that graph structural measures do not efficiently quantify the robustness of MLPs for low severity levels of adversarial attacks and additive noise. MLPs are very dense networks having no weight sharing. Each neuron in MLP has multiple edges across layers, making them fully connected (FC)

<sup>1</sup><https://www.cs.toronto.edu/~kriz/cifar.html>

<sup>2</sup><https://www.kaggle.com/c/tiny-imagenet/overview>

<sup>3</sup><https://www.image-net.org/>



networks. Under insults such as adversarial attack and natural noise, the MLPs are inherently robust because multiple neurons collectively contribute to the same task. MLPs depict superior accuracy for a given task than CNNs under a robust training regime [194]. We believe that because of the in-built robust nature of MLPs, the graph structural properties such as entropy and curvature do not significantly differentiate robust and fragile MLPs. Results of MLPs trained five times on the CIFAR-10 dataset and randomly evaluated thirty times are illustrated in Supplementary Fig. 2.9. The correlation between entropy and accuracy of MLPs is insignificant for most of the evaluation categories.

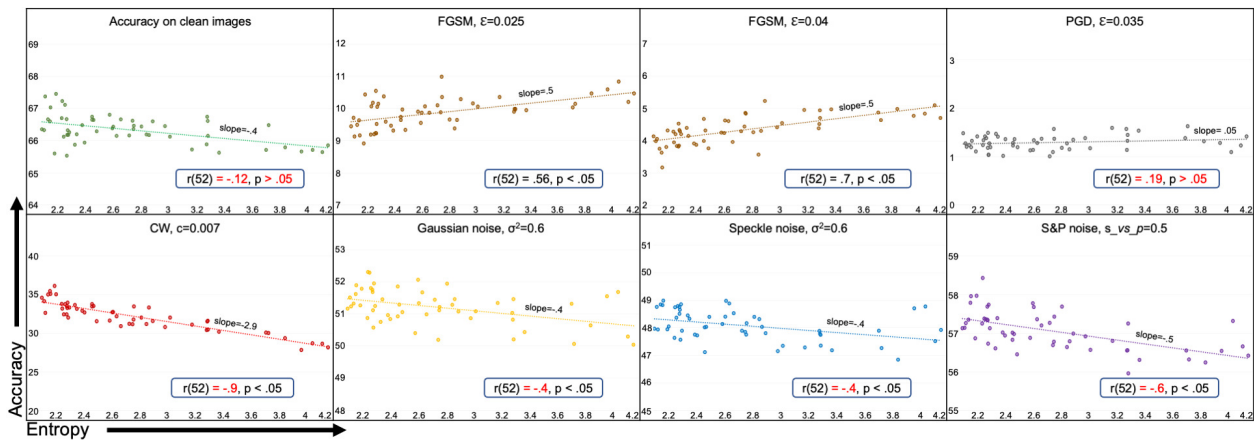


Figure 2.9 Test accuracy vs. entropy for MLPs trained on CIFAR-10 dataset.

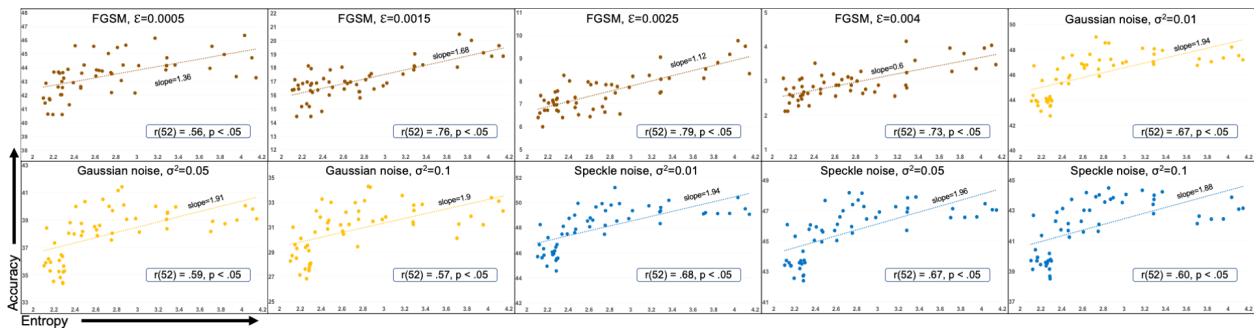


Figure 2.10 Additional results for ResNet-18 on ImageNet dataset.

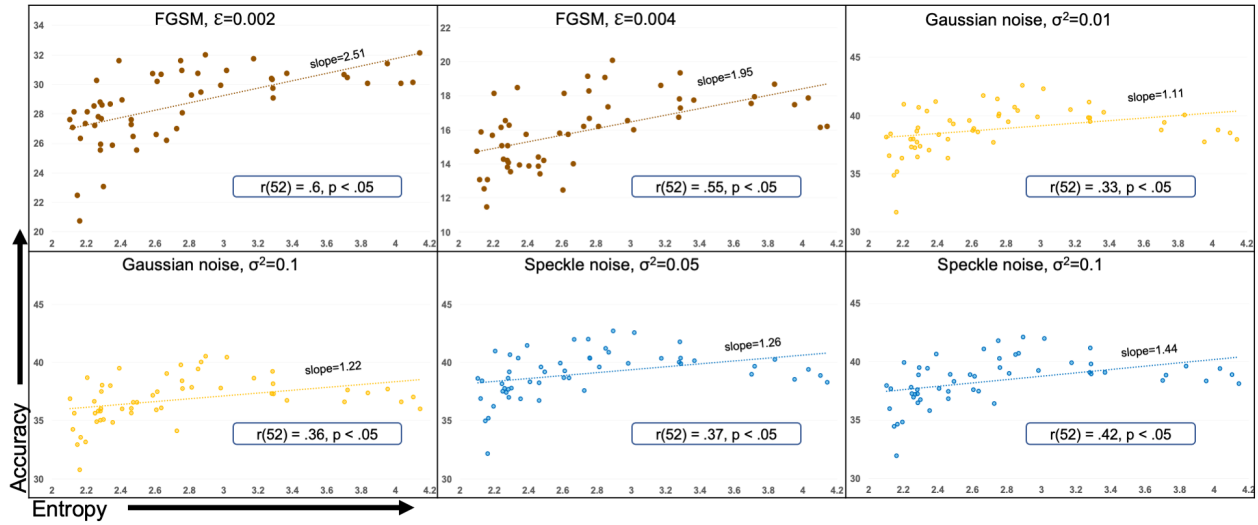


Figure 2.11 Results for ResNet-18 trained and evaluated on Tiny ImageNet dataset.

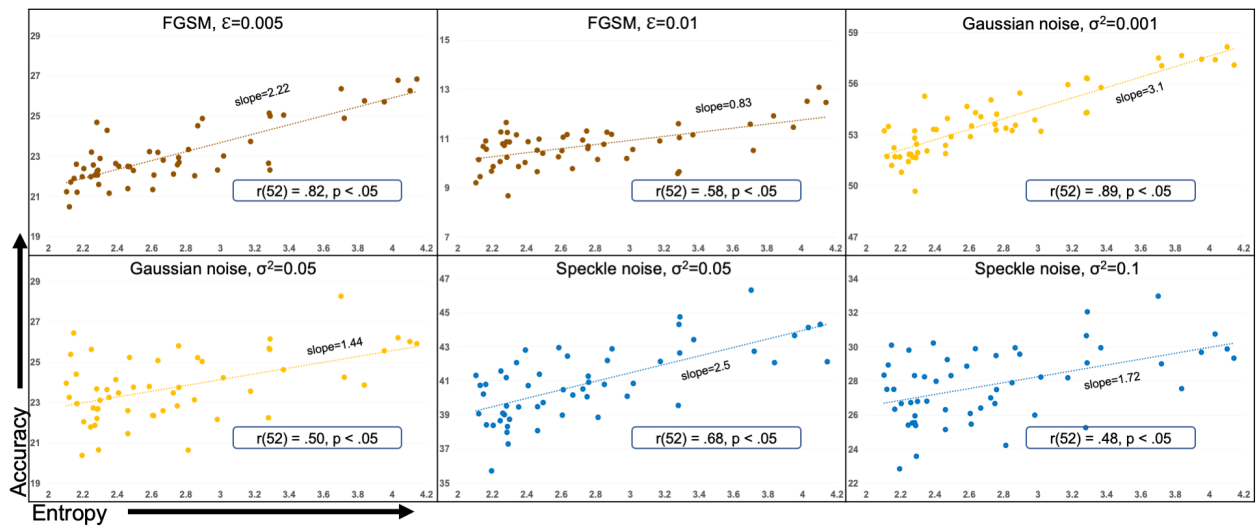


Figure 2.12 Results for robustness evaluation of CNN on CIFAR-100.

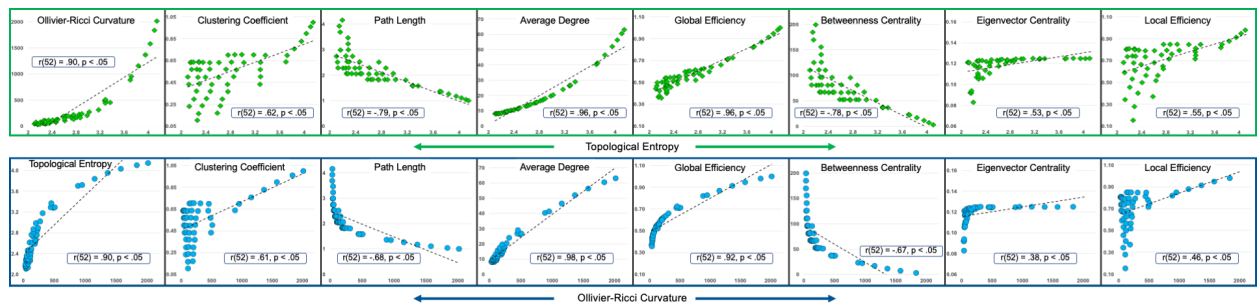


Figure 2.13 Correlation of entropy and curvature with other graph measures.

## 2.9 Curvature vs. Test Accuracy

Table 2.1 Curvature vs. test accuracy

	ResNet-18			CNN	
Metric	ImageNet	Tiny ImageNet	Metric	CIFAR-100	CIFAR-10
Clean Accuracy	0.50	<b>0.55</b>	Clean Accuracy	<b>0.83</b>	0.45
FGSM( $\epsilon=.001$ )	<b>0.56</b>	0.49	FGSM( $\epsilon=.0001$ )	<b>0.79</b>	0.42
FGSM( $\epsilon=.002$ )	<b>0.68</b>	0.45	FGSM( $\epsilon=.001$ )	<b>0.83</b>	0.46
FGSM( $\epsilon=.003$ )	<b>0.64</b>	0.35	FGSM( $\epsilon=.005$ )	<b>0.81</b>	0.43
FGSM( $\epsilon=.004$ )	<b>0.69</b>	0.35	FGSM( $\epsilon=.01$ )	<b>0.64</b>	0.44
PGD( $B=0.001$ )	<b>0.61</b>	-	FGSM( $\epsilon=.015$ )	0.45	<b>0.46</b>
PGD( $B=.002$ )	-	0.48	PGD( $B=.008$ )	0.52	0.07 <sup>†</sup>
CW( $c=5e-7$ )	<b>0.78</b>	-	CW( $c=.007$ )	<b>0.36</b>	-0.33
CW( $c=.1$ )	-	0.38			

	ResNet-18			CNN	
Metric	ImageNet	Tiny ImageNet	Metric	CIFAR-100	CIFAR-10
Gau( $\sigma^2=.001$ )	<b>0.49</b>	0.08 <sup>†</sup>	Gau( $\sigma^2=.001$ )	<b>0.80</b>	0.41
Gau( $\sigma^2=.01$ )	<b>0.45</b>	0.10 <sup>†</sup>	Gau( $\sigma^2=.01$ )	<b>0.66</b>	0.47
Gau( $\sigma^2=.05$ )	<b>0.38</b>	0.13 <sup>†</sup>	Gau( $\sigma^2=.05$ )	<b>0.50</b>	0.12 <sup>†</sup>
Gau( $\sigma^2=.1$ )	<b>0.37</b>	0.14 <sup>†</sup>	Spkl( $\sigma^2=.01$ )	<b>0.76</b>	0.44
Spkl( $\sigma^2=.001$ )	<b>0.49</b>	0.07 <sup>†</sup>	Spkl( $\sigma^2=.05$ )	<b>0.58</b>	0.47
Spkl( $\sigma^2=.01$ )	<b>0.46</b>	0.09 <sup>†</sup>	Spkl( $\sigma^2=.1$ )	<b>0.43</b>	0.23 <sup>†</sup>
Spkl( $\sigma^2=.05$ )	<b>0.45</b>	0.14 <sup>†</sup>	S&P( $ratio=.5$ )	<b>0.59</b>	0.15 <sup>†</sup>
Spkl( $\sigma^2=.1$ )	<b>0.36</b>	0.21 <sup>†</sup>			
S&P( $ratio=.5$ )	0.33	<b>0.34</b>			

Pearson correlation coefficient between graph curvature and test accuracy of DANNs. All values except (<sup>†</sup>) are significant,  $r(52)$ ,  $p < 0.05$ . The bold font indicates the better accuracy of the same DANN on one dataset compared to the other dataset. <sup>†</sup> denotes insignificant correlation values,  $r(52)$ ,  $p > 0.05$ . These results indicate that curvature can quantify the robustness of DANNs, especially in complex tasks and bigger models.

## 2.10 Average Degree vs. Test Accuracy

Table 2.2 Average degree vs. test accuracy

		ResNet-18		CNN	
Metric	ImageNet	Tiny ImageNet	Metric	CIFAR-100	CIFAR-10
Clean Accuracy	0.60	<b>0.63</b>	Clean Accuracy	<b>0.89</b>	0.47
FGSM( $\epsilon=.001$ )	<b>0.61</b>	0.54	FGSM( $\epsilon=.0001$ )	<b>0.85</b>	0.44
FGSM( $\epsilon=.002$ )	<b>0.71</b>	0.51	FGSM( $\epsilon=.001$ )	<b>0.88</b>	0.49
FGSM( $\epsilon=.003$ )	<b>0.67</b>	0.44	FGSM( $\epsilon=.005$ )	<b>0.83</b>	0.46
FGSM( $\epsilon=.004$ )	<b>0.72</b>	0.44	FGSM( $\epsilon=.01$ )	<b>0.63</b>	0.47
PGD( $B=0.001$ )	<b>0.65</b>	-	FGSM( $\epsilon=.015$ )	0.45	<b>0.49</b>
PGD( $B=.002$ )	-	0.55	PGD( $B=.008$ )	0.48	0.09 <sup>†</sup>
CW( $c=5e-7$ )	<b>0.82</b>	-	CW( $c=.007$ )	<b>0.37</b>	-0.33
CW( $c=.1$ )	-	0.45			

		ResNet-18		CNN	
Metric	ImageNet	Tiny ImageNet	Metric	CIFAR-100	CIFAR-10
Gau( $\sigma^2=.001$ )	<b>0.58</b>	0.15 <sup>†</sup>	Gau( $\sigma^2=.001$ )	<b>0.85</b>	0.44
Gau( $\sigma^2=.01$ )	<b>0.53</b>	0.17 <sup>†</sup>	Gau( $\sigma^2=.01$ )	<b>0.71</b>	0.51
Gau( $\sigma^2=.05$ )	<b>0.46</b>	0.20 <sup>†</sup>	Gau( $\sigma^2=.05$ )	<b>0.51</b>	0.14 <sup>†</sup>
Gau( $\sigma^2=.1$ )	<b>0.45</b>	0.22 <sup>†</sup>	Spkl( $\sigma^2=.01$ )	<b>0.82</b>	0.46
Spkl( $\sigma^2=.001$ )	<b>0.57</b>	0.15 <sup>†</sup>	Spkl( $\sigma^2=.05$ )	<b>0.63</b>	0.50
Spkl( $\sigma^2=.01$ )	<b>0.54</b>	0.16 <sup>†</sup>	Spkl( $\sigma^2=.1$ )	<b>0.46</b>	0.26 <sup>†</sup>
Spkl( $\sigma^2=.05$ )	<b>0.54</b>	0.22 <sup>†</sup>	S&P( $ratio=.5$ )	<b>0.56</b>	0.16 <sup>†</sup>
Spkl( $\sigma^2=.1$ )	<b>0.45</b>	0.28			
S&P( $ratio=.5$ )	<b>0.42</b>	<b>0.42</b>			

Pearson correlation coefficient between average degree of graphs and test accuracy of DANNs. All values except (<sup>†</sup>) are significant,  $r(52)$ ,  $p < 0.05$ . Bold text indicates better accuracy of the same DANN evaluated on one dataset compared to the other dataset. <sup>†</sup> mark the insignificant correlation values,  $r(52)$ ,  $p > 0.05$ . Average degree of graphs are also related to robustness of DANNs.

## 2.11 Global Efficiency vs. Test Accuracy

Table 2.3 Global efficiency vs. test accuracy

		ResNet-18		CNN	
Metric	ImageNet	Tiny ImageNet	Metric	CIFAR-100	CIFAR-10
Clean Accuracy	0.73	<b>0.78</b>	Clean Accuracy	<b>0.93</b>	0.45
FGSM( $\epsilon=.001$ )	<b>0.71</b>	0.68	FGSM( $\epsilon=.0001$ )	<b>0.90</b>	0.43
FGSM( $\epsilon=.002$ )	<b>0.80</b>	0.65	FGSM( $\epsilon=.001$ )	<b>0.90</b>	0.49
FGSM( $\epsilon=.003$ )	<b>0.67</b>	0.50	FGSM( $\epsilon=.005$ )	<b>0.83</b>	0.48
FGSM( $\epsilon=.004$ )	<b>0.75</b>	0.54	FGSM( $\epsilon=.01$ )	<b>0.58</b>	0.48
PGD( $B=0.001$ )	<b>0.71</b>	-	FGSM( $\epsilon=.015$ )	0.34	<b>0.50</b>
PGD( $B=.002$ )	-	0.66	PGD( $B=.008$ )	0.45	0.15 <sup>†</sup>
CW( $c=5e-7$ )	<b>0.85</b>	-	CW( $c=.007$ )	<b>0.43</b>	-0.27
CW( $c=.1$ )	-	0.52			

		ResNet-18		CNN	
Metric	ImageNet	Tiny ImageNet	Metric	CIFAR-100	CIFAR-10
Gau( $\sigma^2=.001$ )	<b>0.71</b>	0.36	Gau( $\sigma^2=.001$ )	<b>0.89</b>	0.43
Gau( $\sigma^2=.01$ )	<b>0.67</b>	0.38	Gau( $\sigma^2=.01$ )	<b>0.78</b>	0.52
Gau( $\sigma^2=.05$ )	<b>0.59</b>	0.39	Gau( $\sigma^2=.05$ )	<b>0.51</b>	0.17 <sup>†</sup>
Gau( $\sigma^2=.1$ )	<b>0.55</b>	0.39	Spkl( $\sigma^2=.01$ )	<b>0.88</b>	0.46
Spkl( $\sigma^2=.001$ )	<b>0.70</b>	0.35	Spkl( $\sigma^2=.05$ )	<b>0.70</b>	0.51
Spkl( $\sigma^2=.01$ )	<b>0.68</b>	0.37	Spkl( $\sigma^2=.1$ )	<b>0.51</b>	0.28
Spkl( $\sigma^2=.05$ )	<b>0.67</b>	0.42	S&P( $ratio=.5$ )	<b>0.48</b>	0.13 <sup>†</sup>
Spkl( $\sigma^2=.1$ )	<b>0.59</b>	0.46			
S&P( $ratio=.5$ )	<b>0.52</b>	<b>0.52</b>			

Pearson correlation coefficient between global efficiency measure of graphs and test accuracy of DANNs. All values except (<sup>†</sup>) are significant,  $r(52)$ ,  $p < 0.05$ . Bold text indicates better accuracy of the same DANN evaluated on one dataset compared to the other dataset. <sup>†</sup> mark the insignificant correlation values,  $r(52)$ ,  $p > 0.05$ . These results indicate that global efficiency of graphs is also related to robustness of DANNs.

## 2.12 Frameworks and Hyperparameters

Frameworks and corresponding packages used in our experiments are given in Supplementary Table 2.4. The hyperparameters used in the training and evaluation of DANNs are given in Supplementary Table 2.5. For the sake of procedural consistency and comparisons of results, the set of parameters other than the those mentioned in Supplementary Table 2.5 are kept the same as in original experiments for relational graphs by their respective authors [778].

Table 2.4 Frameworks and packages used in our codebase.

	<b>Package name</b>	<b>Version</b>
<b>Operating systems</b>	Ubuntu	20.04.3
	Windows	10
	macOS	11.6
<b>Programming languages</b>	Python	3.6.15
	Matlab	R2020a
<b>Deep learning framework</b>	Pytorch	1.4.0
	torchvision	0.5.0
<b>Adversarial library</b>	RobustBench	-
	torchattacks	3.2.1
	foolbox(optional)	3.3.1
	art(optional)	1.9.0
<b>Miscellaneous</b>	scikit-image	0.17.2
	scikit-learn	0.24.2
	scipy	1.4.1
	numpy	1.19.5
	networkx	2.3
	pyyaml	5.1.2
<b>Adversarial attacks</b> (torchattacks)	FGSM	-
	PGD	-
	CW	-
<b>Additive noise</b> (scikit-image)	Gaussian	-
	Speckle	-
	Salt & Pepper	-

Table 2.5 Training & eval hyperparameters for our DANN experiments.

	<b>CIFAR-10</b>		<b>CIFAR-100</b>		<b>TinyImageNet</b>	<b>ImageNet</b>
<b>Hyperparam</b>	<b>MLPs</b>	<b>CNNs</b>	<b>CNNs</b>	<b>ResNet29</b>	<b>ResNet18</b>	<b>ResNet18</b>
Epochs	200	100	350	150	75	75
Batch size	256	1024	32	512	256	450
Base lr	0.1	0.1	0.025	0.021	0.1	0.025
lr policy	Cosine					step=[0, 25, 50, 70]
Momentum	0.9					
Weight decay	0.0005	0.01	0.0005	0.01	0.006	0.0001
Drop out	-	-	FC: p=0.1	-	Conv:p=0.2, FC:p=0.5	-
Trg iterations	5	5	5	5	1	1
Eval iterations	30	30	30	30	5	5

### 2.13 Compute Resources and Wall-clock Times

Training time for a 5-layer MLP transformed from the WS-flex random graph on CIFAR-10 dataset is approximately 7 minutes on NVIDIA TITAN RTX GPU. Each MLP was trained five times with random seed, consuming approximately 40 minutes in training the model. On the NVIDIA TITAN RTX GPU, training of all 54 MLPs on CIFAR-10 dataset approximately took 3 days. For CIFAR-100 dataset, the 54 CNNs took approximately 5 days in training the DANNs, five times each. For Tiny ImageNet experiments on the 54 ResNet-18s, the baseline model took approximately 3 hours on TITAN RTX GPU, whereas, the longest training time for a ResNet-18 was approximately 18 hours. Total time for training 54 ResNet-18 models on Tiny ImageNet was approximately 22 days. Training the baseline model of ResNet-18 on ImageNet dataset took approximately 70 hours (3 days) on TITAN RTX GPU, the longest training time for a ResNet-18 model on ImageNet was approximately 123 hours (5 days). Total training time for 54 ResNet-18 models on ImageNet was approximately 3 months with parallel training on four GPUs. All the aforementioned training times include the inference times for FGSM, PGD, and CW adversarial attacks as well as Gaussian, Speckle, and Salt&Pepper additive noise. For tracking the experiments, visualization of results, and hyperparameter tuning, we used the Weights and Biases [81] which is a freely available performance visualization platform for machine learning tasks.



## 2.14 Sample Images

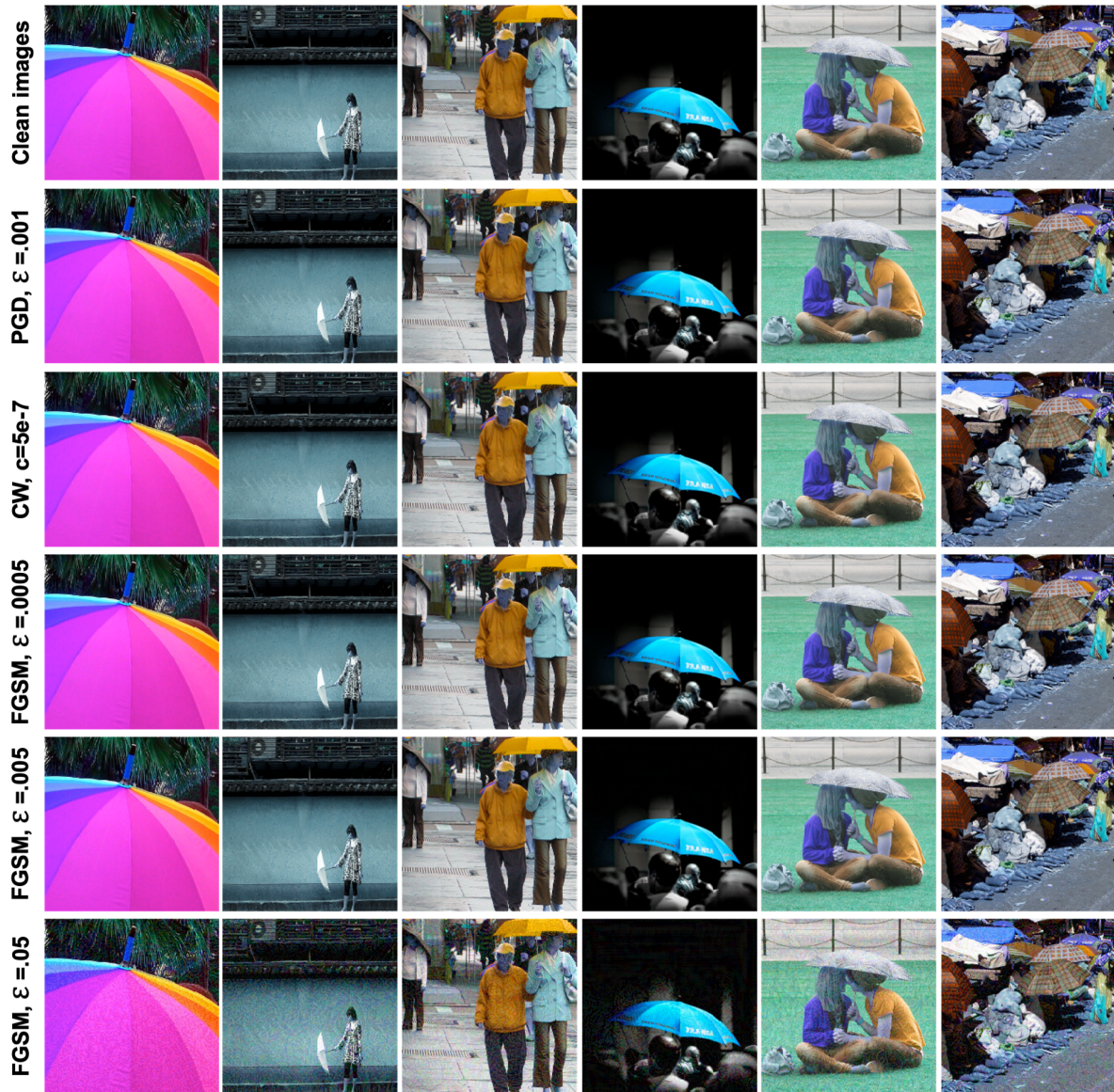


Figure 2.14 Comparison of clean images from ImageNet dataset with adversarial examples.

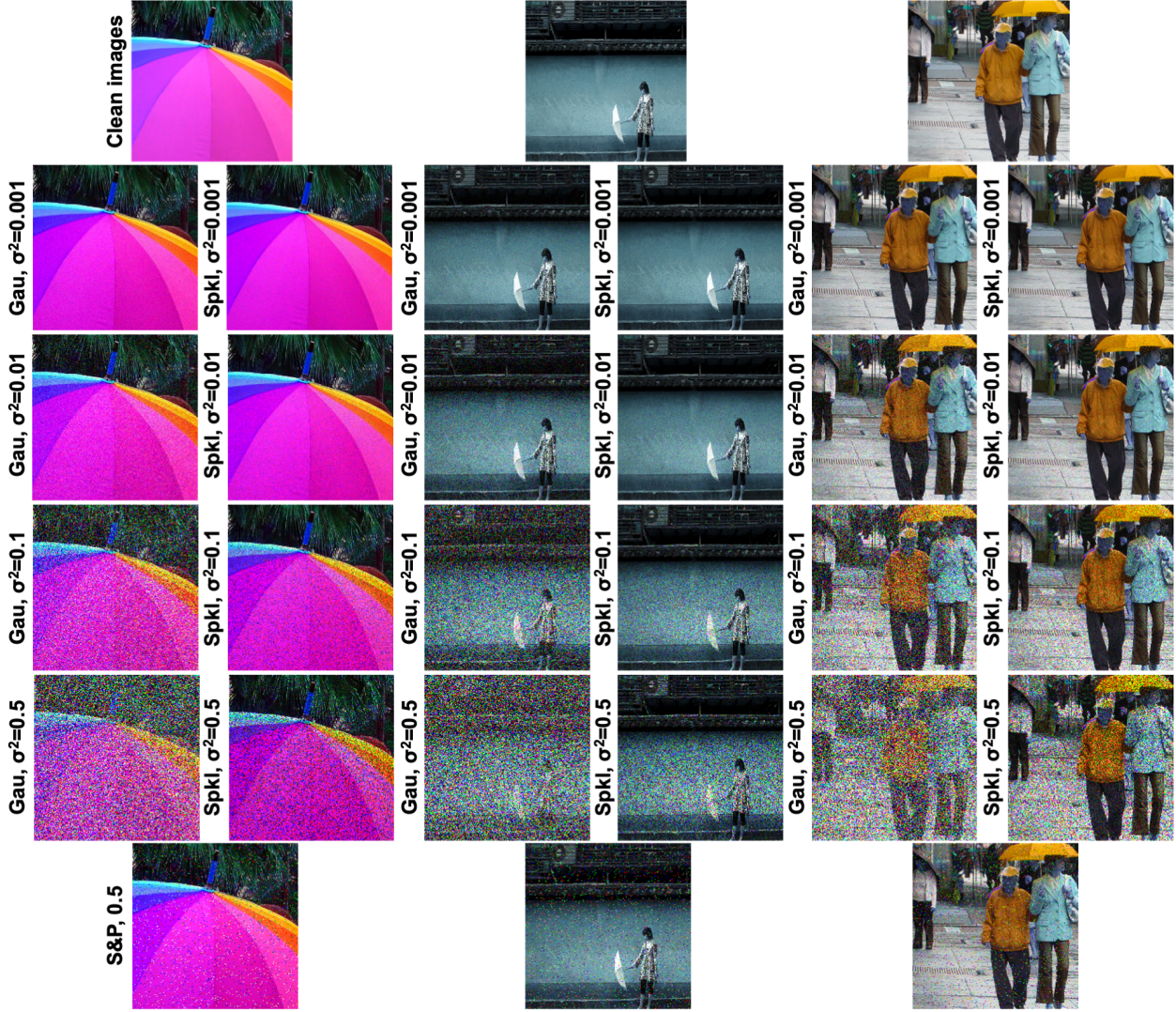


Figure 2.15 Comparison of clean images from ImageNet dataset with noisy images.

## 2.15 Algorithm for Robust Model Selection

We propose algorithm 2.1 as a guideline for selecting the robust architecture of DANN for a given task without the need to undergo the train-validate-test loop for all the choices of architectures. The input parameters of the algorithm include dataset  $x$ , task  $t$ , number of nodes  $n$ , the number of DANNs to be considered  $\alpha$ , and the binary flag about using ensemble of models. For all possible ( $\tau$ ) number of graphs in the given design space of  $n$ -node graphs, calculate curvature ( $\kappa_{ORC}$ ) and entropy ( $H$ ). Select  $\alpha$  number of graphs having highest  $\kappa_{ORC}$  and  $H$  values, generate DANNs from

these  $\alpha$  graphs, train DANNs for the downstream task  $t$ , and evaluate under natural and malicious noise inputs. All these  $\alpha$  DANNs have higher robustness than the rest of  $(\tau - \alpha)$  DANNs. Select the DANN with the highest test accuracy, or use ensemble of these  $\alpha$  number of DANNs. Selection of parameter  $\alpha$  is at the user's discretion as per availability of resources such as computational power and time, generally  $\alpha \leq 10$ .

---

Algorithm 2.1 Robust Model Selection

---

**Input:** data  $x$ , task  $t$ , nodes  $n$ , number of graphs to be selected  $\alpha$ , ensemble

**repeat**

**for** each  $\tau$  graph  $\in n$ -node design space **do**

find all graphs  $G_\tau(n)$

calculate  $H(G_\tau), \kappa_{OR}(G_\tau)$  (graph measures)

**end for**

**for**  $i \leftarrow 1$  **to**  $\alpha$  **do**

$G_i \leftarrow \max(G_\tau(H, \kappa_{OR}))$  ▷ graphs having highest entropy, curvature

$RG_\alpha \leftarrow G_i$  ▷ add to the list of robust graphs

**end for**

**until**  $\alpha$  robust graphs ( $RG_\alpha$ ) found  
{required number of graphs found}

**repeat**

convert robust graphs to neural networks,  $RG_\alpha \rightarrow NN_\alpha$

**for**  $j \leftarrow 1$  **to**  $\alpha$  **do**

train, validate  $NN_j(x, t)$

calculate test accuracy  $RobAcc(NN_j)$  with noisy, adversarial inputs

**end for**

**if** ensemble **then**

select Robust NN = avg( $RobAcc(NN_j)$ )

**else**

select Robust NN = max( $RobAcc(NN_j)$ )

**end if**

**until** Robust Neural Network found

---

## Chapter 3: Brain Tumor Segmentation and Surveillance with Deep Artificial Neural Networks

### 3.1 Note to Reader

This chapter has been previously published in Springer as: Waqas, A., Dera, D., Rasool, G., Bouaynaya, N.C., Fathallah-Shaykh, H.M. (2021). Brain Tumor Segmentation and Surveillance with Deep Artificial Neural Networks. In: Elloumi, M. (eds) Deep Learning for Biomedical Data Analysis. Springer, Cham., and has been reproduced with permission from Springer [734].

### 3.2 Introduction

The task of *brain tumor segmentation*, presented in this chapter, is the confluence of multiple techniques usually employed in diverse fields of science such as *Digital Image Processing* (DIP), *Computer Vision* (CV), and *Machine Learning* (ML). ML algorithms, specifically *Deep Artificial Neural Networks* (DNNs), have achieved state-of-the-art accuracy in CV related tasks, including image segmentation. DNNs are built using large stacks of individual artificial neurons, each of which performs mathematical operations of multiplication, summation, and non-linear operations. One of the key reasons for the success of DNNs is the ability to learn useful features automatically from the data as opposite to manual selection by expert humans [394, 363, 355]. Various architectures of DNNs employed for brain tumor segmentation have been discussed along with a case study of one of those architectures in detail. We discuss a new technique for quantifying uncertainty in the output decision of DNNs. We also present different techniques for tumor surveillance.

The rest of the chapter is organized as follows: In section 3.3, we touch upon the relevant theoretical background of the techniques involved in solving the tumor segmentation problem.

We discuss image segmentation in general and medical image segmentation, particularly, and the concept of surveillance in the medical sphere. Section 3.4 demonstrates brain tumor segmentation through DNNs. Section 3.5 presents the particularly suited *Inception* modules in *Deep Learning* (DL) for brain tumor segmentation. Section 3.6 explains the concept of uncertainty estimation in the decision made by DNNs. Finally in Section 3.7, we discuss tumor surveillance techniques supported by a case study followed by conclusion in Section 3.8.

### 3.3 Theoretical Background

The task of *brain tumor segmentation* using DNNs inherently involves various tasks. Therefore, it is imperative to imbibe some basic theoretical background about these concepts, which leads up to the task-at-hand.

#### 3.3.1 Image Segmentation

A picture is worth a thousand words. This is because a picture contains far more information in a few pixels that the human brain can process simultaneously as compared to the numerous words that can express the same amount of information sequentially. Thus, understanding the image and extracting useful information from it bears the central role in the fields of DIP and CV. Classification task, in particular, assigns a label or class to an input image. However, image classification does not provide pixel-level information, such as the location of objects in an image, objects' shapes and boundaries, information about which pixel belongs to which object, etc. For this purpose, images are segmented by assigning a specific label to pixels with similar characteristics in an image. *Segmentation* is a technique frequently used in DIP and CV fields for extracting useful information from images [29]. It is the process of partitioning an image into segments (having sets of pixels) representing various objects in the image. The purpose is to modify the representation of an image into a more elaborate format, which is easy to understand anatomically and helpful in extracting meaningful information for analysis. In usual practice, this process is used to locate objects of interest and draw boundaries/shapes conforming to these objects in an

image(s). Image segmentation has contributed to many spheres of human life, ranging from the film-making industry to the field of medicine [657]. For example, the green screens used in *Marvel* [472] movies employed segmentation to extract the foreground objects and place them on different backgrounds depicting dangerous real-life scenes, Fig. 3.1(a). An example of medical image segmentation includes the identification of multiple organs in the abdomen and thorax, as shown in Fig. 3.1(b,c). Various techniques are used for classical image segmentation, e.g., *region-based* [256](*threshold segmentation, regional growth segmentation*), *edge-detection* [786](*Sobel operator, Laplacian operator*), *clustering-based* [786](*K-means*), and *weak-supervised learning* [764] methods in CNN. Further details on classical image segmentation techniques are beyond the scope of this chapter, so we will confine ourselves to medical image segmentation, in general, with particular focus on brain tumor [474] segmentation.

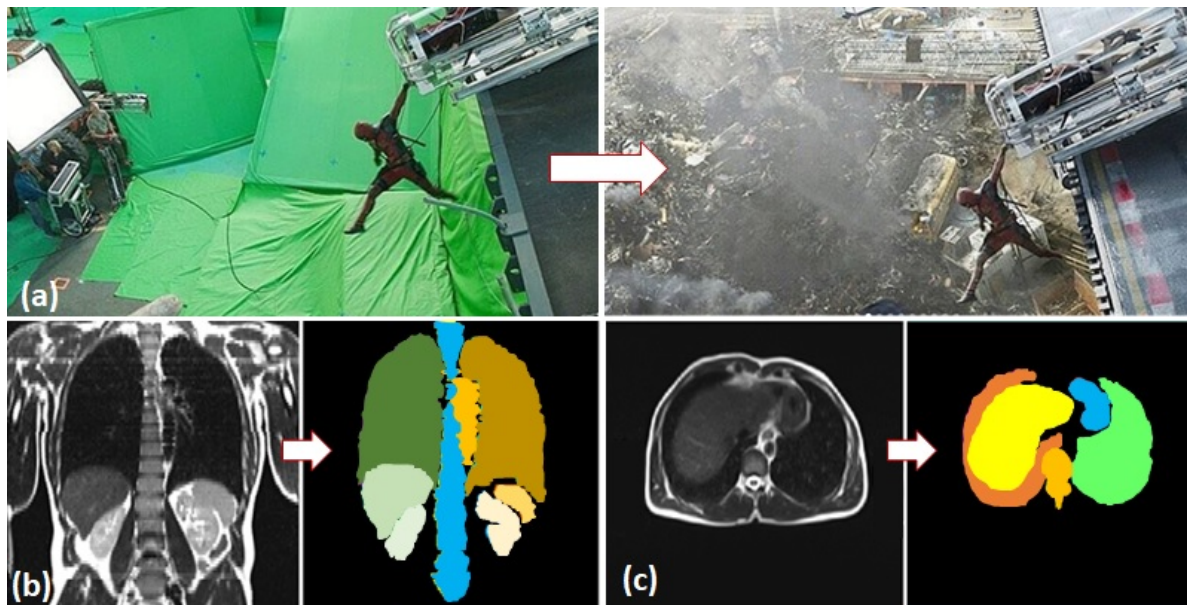


Figure 3.1 Use of image segmentation in Marvel movies and medical imaging.

### 3.3.2 Brain Tumor Segmentation

*Brain tumors* [474] are masses or growths of abnormal cells in the brain, categorized into *primary* and *secondary* or *metastatic* types. Primary brain tumors originate from brain cells, whereas secondary tumors metastasize into the brain from other organs [516]. The most common type of primary brain tumors are *gliomas* [553], which arise from brain glial cells, and can be of *Low-Grade* (LGG) or *High-Grade* (HGG) sub-types. HGGs are aggressively-growing and infiltrative malignant brain tumors, which usually require surgery or radiotherapy and have poor survival prognosis with the highest mortality rate and prevalence [73]. *Magnetic Resonance Imaging* (MRI) is a crucial diagnostic tool for brain tumor analysis, monitoring and surgery planning. Several complimentary 3D MRI modalities such as T1, *T1 with gadolinium-enhancing Contrast* (T1C), T2-weighted (T2), and *FLuid-Attenuated Inversion Recovery* (FLAIR) are acquired to emphasize different tissue properties and areas of tumor spread. For example, in T1C MRI modality, the contrast agent (e.g., gadolinium) emphasizes hyper-active tumor sub-regions.

*Brain tumor segmentation* [96, 363, 355, 624] is the technique of labeling tumor pixels in an MRI to distinguish them from normal brain tissues and artifacts. These MRI scans are the representation of the internal structure or function of the brain's anatomic region in the form of an array of picture elements called pixels or voxels. It is a discrete representation resulting from a sampling/reconstruction process that maps numerical values to positions of the space. The number of pixels used to describe the field-of-view of an acquisition modality is an expression of the detail with which the anatomy or function can be depicted. Depiction of the numerical value of pixel depends on the imaging modality, acquisition protocol, reconstruction, and post-processing. MRI scans come in various file formats and standards but the six commonly used are: *Analyze* [475], *Neuroimaging Informatics Technology Initiative* (NIfTI) [509], *Minc* [711], PAR/REC format used by Philips MRI scanners [616], *Nearly Raw Raster Data* (NRRD) [512], and *Digital imaging and communications in medicine* (Dicom) [182]. A comparison of characteristics of these formats is shown in Fig. 3.2. In clinical practice, the process of separating the tumor pixels from normal brain tissues provides useful information about existence, growth, diagnosis, surveillance and treatment

planning. The process of manual delineation requires anatomical knowledge by specially trained persons, whereas such manual practices are expensive, time-consuming, and are prone to errors due to human limitations. The process of automated segmentation of brain tumors from 3D images facilitates in overcoming these shortcomings [391].

	Dicom	Nifti	Minc	Analyze	NRRD	PAR/REC format
Stands for	Digital Imaging and Communications in Medicine	Neuroimaging Informatics Technology Initiative	Medical Imaging NetCDF Toolkit	Developed by Biomedical Imaging Resource (BIR) at Mayo Clinic	Nearly Raw Raster Data	Philips MRI scanner format
File extension	.dcm	.nii	.mnc	.img and .hdr	.nrrd	.PAR
Data type	Signed, unsigned integer (8, 16-bit, 32-bit only allowed for radiotherapy dose)	Signed, unsigned integer (8-64 bit),float (32-128 bit), complex(64-256 bit)	Signed, unsigned integer (8-32 bit),float (32, 64 bit), complex(32, 64 bit)	Unsigned integer (8-bit), signed integer(16, 32 bit), float(32-64 bit), complex(64 bit)	-	8 or 16 bit unsigned integers
Header Length	Variable length binary format	Fixed-length 352 byte binary format (348 byte in case of data stored as .img and .hdr)	Extensible binary format	Fixed-length: 348 byte binary format	Extensible with attached and detached	8 or 16 bit unsigned integers

Figure 3.2 Summary of medical imaging file formats

### 3.3.3 Tumor Surveillance

*National Cancer Institute* (NCI), part of the U.S. *National Institutes of Health* (NIH), defines tumor surveillance as *closely watching a patient's condition but not treating it unless there are changes in test results. Surveillance is also used to find early signs that a disease has come back. It may also be used for a person who has an increased risk of a disease, such as cancer* [311]. The process of surveillance regularly involves (scheduled) medical tests and examinations to track the growth of the tumor. The term has also been used in the realm of public health, wherein collective information of a disease, such as cancer, is recorded in a group of people belonging to a specific



category (ethnic, age, gender, regional, etc.) Active surveillance is extremely beneficial, especially for patients with low-risk cancer diagnoses. Apart from the routine biopsy, active surveillance is almost surgically noninvasive. It helps in the delay of more invasive treatments such as surgical removal of a tumor, sparing the patients from burdensome side effects and potential complications for as long as possible. Moreover, by deferring the invasive treatment to the point when the disease worsens, active surveillance enables cancer patients to maintain a quality of life. A case in point is the exceptionally beneficial surveillance of low-risk prostate cancer in men. The reason for this success is that almost 50% of prostate cancer diagnoses are categorized as low-risk with less possibility of spread, and few cases may never require advanced forms of treatment. Such cases do not immediately need to be aggressively treated in the absence of worsened disease while the specialists keep records of the tumor's growth over time. Surveillance allows the specialists to monitor the disease right from the onset, thus leveraging them the liberty to analyze the effects and progress of disease and determine the next course of action [585]. A study by Harvard researchers found that the aggressiveness of prostate cancer at diagnosis appears to remain stable over time for most men. If patients had chosen active surveillance, then this could make them feel more confident in their decision about treatment [552]. Early detection of the tumor through surveillance could assist both the patients and the specialists in taking more considered decisions about treatment.

#### 3.3.4 Deep Learning Segmentation Task

*Computer Vision (CV)* is the field of computer science that aims to replicate (to some extent) the complex nature of the human vision system into modern-day computers and machines. It endeavors to enable machines to visually gaining a high-level understanding of objects in the imagery in its quest to mimic the humans. A chronological insight in some of the most active topics of research in computer vision can be found in [673]. Most attractive topics in today's CV tasks include object classification (i.e., categorizing objects in an image), localization (i.e., spatially locating objects in an image), detection/ recognition and segmentation (i.e., identifying the category of each pixel in an image) as shown in Fig. 3.3.

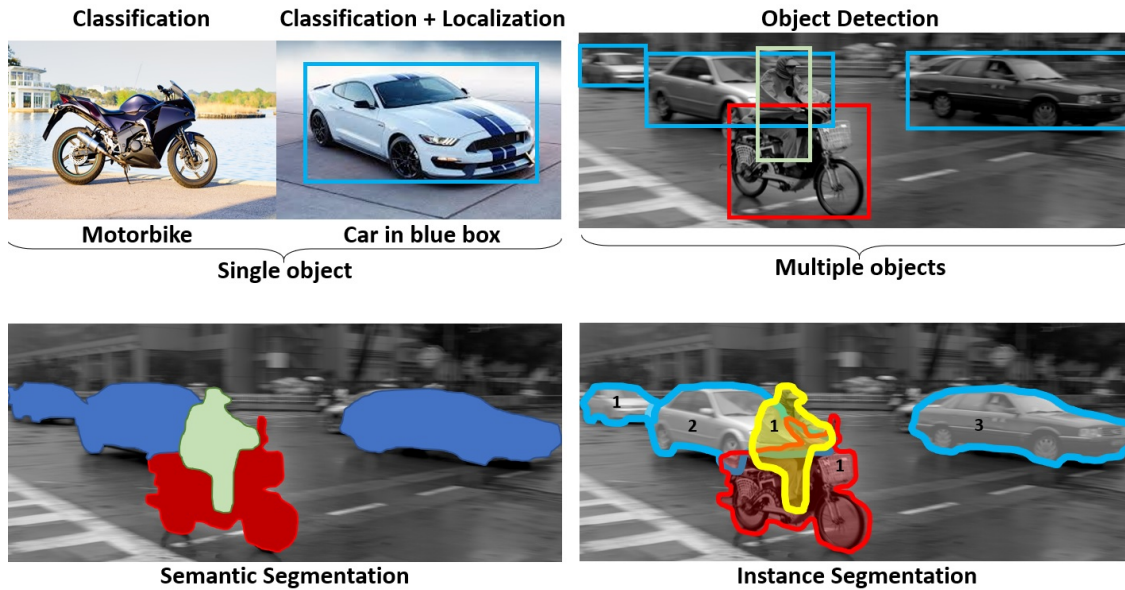


Figure 3.3 Important computer vision tasks.

The commonly used classical image segmentation techniques, described in section 3.2.1 above have been replaced by their more efficient counterparts in ML because of the former's inherently rigid algorithms and the need for human intervention. Image segmentation is the fundamental component of DL, which is part of a broader family of ML. Compared with other DL algorithms, CNNs have proven to be the more efficient selection for segmentation tasks from imagery. Image segmentation using CNN involves feeding the CNN with the desired image as an input and getting the labels of each pixel, i.e., labeled image as an output. Instead of processing the complete image at once, CNN deals with a fraction of image conforming to its *filter*, convolving and ultimately mapping over the entire image. To learn more on CNNs, a concise explanation supported by a visualization can be referred to at [199, 198].

### 3.3.5 Motivation

A large population suffers from fatalities caused by cancer, and brain tumors are one of the leading causes of death for cancer patients, especially children and adolescents. Brain tumors account for one in every 100 cancers diagnosed annually in the United States [640]. In 2019, the

*American Cancer Society* reported that 23,820 new brain cancer cases were discovered in the United States [637]. One of the most frequent primary brain tumors is glioma [292], which affects the glial cells of the brain as well as the surrounding tissues. The HGG or *GlioBlastoMa* (GBM) is the most common and aggressive type with a median survival rate of one to two years [103]. Although neurosurgery may be the only therapy for many brain tumors [400], other treatment methods such as radiotherapy and chemotherapy are also used to destroy the tumor cells that cannot be physically resected or to slow their growth. Before the treatment through chemotherapy, radiotherapy, or brain surgeries, there is a need for medical practitioners to confirm the boundaries and regions of the brain tumor and determine where exactly it is located and the exact affected area. Moreover, all of these invasive treatments face challenging practice conditions because of the structure and nature of the brain. These conditions make it very difficult to distinguish the tumor tissue from normal brain parenchyma for neurosurgeons based on visual inspection alone [488, 363, 355, 624].

Moreover, such manual-visual practices usually involve a group of clinical experts to define the location and the type of the tumor accurately. This lesion localization process is laborious, and its quality depends on the physicians' experience, skills, slice-by-slice decisions, and the results may still not be universally accepted among the clinicians. Treatment protocols for high-grade pediatric brain tumors and general low-grade tumors recommend regular follow-up imaging for up to 10 years. For these longitudinal studies, a comparison of the current MRI with all prior imaging takes a very long time, which is practically infeasible. Automated computer-based segmentation methods present an excellent solution to the challenges mentioned above by saving physician's time and providing reliable and accurate results while reducing the diagnosis efforts of surgeons on a single patient [347]. Brain tumor segmentation is motivated by assessing tumor growth, treatment responses, computer-based surgery, treatment of radiation therapy, and developing tumor growth models. Thus, the computer-assisted diagnostic system is meaningful in medical treatments to reduce the workload of doctors and to get accurate results.

### 3.3.6 Challenges

Segmentation of gliomas in pre-operative MRI scans — conventionally done by expert board-certified neuro-radiologists and other physicians — provides quantitative morphological characterization and measurement of glioma sub-regions. The quantitative analysis task is challenging due to the high variance in appearance and shape, ambiguous boundaries and imaging artifacts. Although computer-aided techniques have the advantage of fast speed, consistency in accuracy and immunity to fatigue [629], automatic segmentation of brain tumors in multi-modal MRI scans is still one of the most difficult tasks in medical image analysis and applications. Automatic segmentation involves dealing with a complicated and massive amount of data, artifacts due to patient’s motion, limited acquisition time, and soft tissue boundaries that are usually not well defined. Moreover, many classes of tumors have a variety of irregular shapes, sizes, and image intensities, especially the surrounding structures of tumors. Numerous attempts have been made in developing ML algorithms for segmenting normal and abnormal brain tissues using MRI images, which will be covered in detail in section 3.4. However, feature selection to enable automation is challenging and requires a combination of computer engineering and medical expertise. Thus, developing fully-automated brain tumor segmentation remains a challenging task, and a large part of the research community is currently involved in overcoming this challenge in bringing state-of-the-art ideas in this field into reality.

## 3.4 Brain Tumor Segmentation Using Deep Artificial Neural Networks

The task of segmenting brain tumor in MRI images has been adopted in DNNs from the image segmentation task in CV. This section focuses on these techniques imported in DL from CV and also gives an overview of the various DL architectures employed on brain MRI datasets.

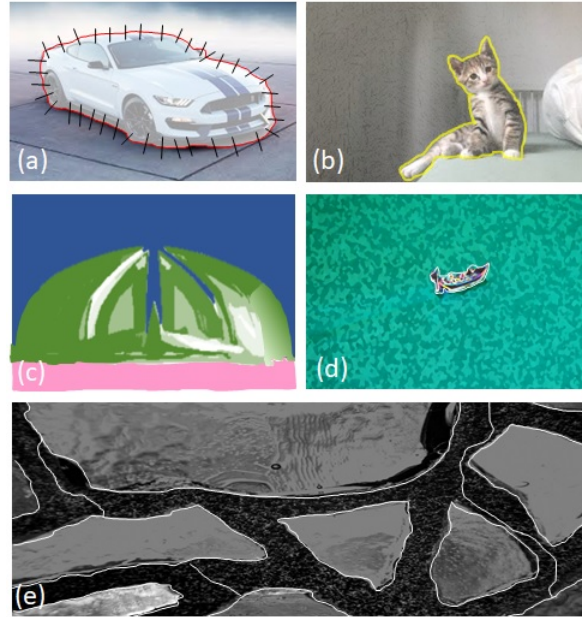


Figure 3.4 Few image segmentation techniques in computer vision.

### 3.4.1 Image Segmentation in Computer Vision Realm

Image segmentation is the task of finding groups/ clusters of pixels that belong to the same category. It divides an input image into *segments* to simplify image analysis. These segments represent objects or parts of objects and comprise sets of pixels belonging to each part. Practically, the segmentation sorts pixels into larger components, eliminating the need to consider individual pixels as units of observation. In statistics, this problem is known as *cluster analysis* and is a widely studied area with many different algorithms [349, 322, 321, 320]. In CV, image segmentation is one of the oldest and most extensively used problems dates back to the 1970s [99, 589, 547, 518, 594, 275]. Some of the most extensively known techniques developed for image segmentation are: (a) active contours [85]; (b) level sets [160]; (c) region splitting and graph-based merging [218]; (d) mean shift (mode finding) [152]; (e) normalized cuts (splitting based on pixel similarity metrics), as depicted in Fig. 3.4. The segmentation process itself has two forms, namely; *semantic*, and *instance* segmentation. The former classifies all the pixels of an image into meaningful or

*semantically interpretable* classes of objects and is usually referred to as *dense prediction*. The latter identifies each instance of each object in an image and differs from semantic segmentation in that it does not categorize every pixel. For example, in Fig. 3.3, semantic segmentation classified all cars, while instance segmentation identifies each one individually. Various metrics are used for performance evaluation of image segmentation including pixel accuracy  $P_{acc}$ , mean accuracy  $M_{acc}$ , *Intersection-over-Union* (IoU)  $M_{IU}$ , frequency weighted IoU  $F_{IU}$  and *Dice coefficient* [445]. Let  $n_{ij}$  indicate the number of pixels of class  $i$  predicted to belong to class  $j$ , where there are  $n_{cl}$  different classes, and let  $\mathbf{t}_i = \sum_j n_{ij}$  indicates the number of pixels of class  $i$ , then the performance evaluation terms mentioned above are defined by:

$$P_{acc} = \frac{\sum_i n_{ii}}{\sum_i \mathbf{t}_i}, \quad (3.1)$$

$$M_{acc} = \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{\mathbf{t}_i}, \quad (3.2)$$

$$M_{IU} = \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{\mathbf{t}_i + \sum_j n_{ji} - n_{ii}}, \quad (3.3)$$

$$F_{IU} = \frac{1}{\sum_k \mathbf{t}_k} \sum_i \frac{\mathbf{t}_i n_{ii}}{\mathbf{t}_i + \sum_j n_{ji} - n_{ii}}, \quad (3.4)$$

*Dice Similarity Coefficient* (DSC) has also been extensively used for evaluating segmentation algorithms in medical imaging applications [62]. DSC between a predicted binary image  $P$  and ground truth binary image  $G$ , both of size  $N \times M$  is given by:

$$DSC(P, G) = 2 \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P_{ij} G_{ij}}{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P_{ij} + \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} G_{ij}}, \quad (3.5)$$

where  $i$  and  $j$  represent pixel indices for the height  $N$  and width  $M$ . The range of DSC is  $[0, 1]$ , and a higher value of DSC corresponds to a better match between the predicted image  $P$  and the ground truth image  $G$ .

The application of image segmentation techniques in the medical imaging field opened a new frontier of knowledge with advances in the areas of diabetic retinopathy detection, skin cancer

classification, brain tumor segmentation and many more. In this chapter, we will restrict ourselves to brain tumor segmentation only and look at the various techniques employed in *Artificial Neural Networks* (ANNs) for brain tumor segmentation.

### 3.4.2 Deep Artificial Neural Networks and Image Segmentation

DNNs have achieved significant milestones in the CV field. DNNs have multiple layers between the input and output layers. The basic element of ANN, i.e., *artificial neuron*, has multiple inputs that are weighted and summed up, followed by a *transfer function* or *activation function*. Then the neuron outputs a scalar value. An example of ANN is illustrated in Fig. 3.5 [441]. Inspired by biological processes, ANNs use shared-weight architecture where the connectivity pattern between neurons mimics the organization of the brain visual cortex [231, 393]. ANNs imitate the concept of receptive fields where individual cortical neurons respond to stimuli only in a restricted field of view. Because of their shared-weight architecture and translation invariance characteristics, ANNs are shift or space-invariant. Due to the linear operations followed by the non-linear activations, ANNs are capable of extracting higher-level representative features [261] and can compute any function [507].

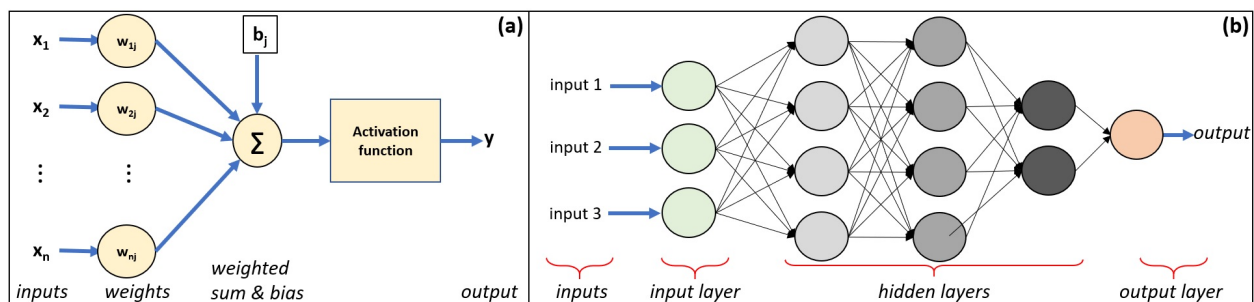


Figure 3.5 Artificial neuron model and ANN model.

### 3.4.3 DL-based Image Segmentation Architectures

Most prominent DL architectures used by the CV community include *Convolutional Neural Networks*(CNNs), *Recurrent Neural Networks* (RNNs) and *Long Short Term Memory* (LSTM), *encoder-decoders*, and *Generative Adversarial Networks* (GANs) [395, 291, 252, 253]. With our focus on CNNs, let us discuss the three important DL-based image segmentation architectures.

#### 3.4.3.1 Convolutional Neural Networks

Waibel *et al.* introduced *Convolutional Neural Networks* (CNN) that had weights shared among temporal receptive fields, and it had back-propagation training for phoneme recognition [716]. LeCun *et al.* developed a CNN architecture for document recognition as shown in Fig. 3.6a [395]. The three basic components/ layers of a CNN are: 1) *convolutional layer*, having a kernel (or filter) of weights convolved with the input image to extract features; 2) *nonlinear layer* having an element-wise activation function applied to feature maps; and 3) *pooling layer*, which reduces spatial resolution and replaces appropriate neighborhood of a feature map with some statistical information (mean, max, etc.) [486]. Deep CNNs have performed extremely well on a wide variety of medical imaging tasks, including diabetic retinopathy detection [263], skin cancer classification [209], and brain tumor segmentation [146, 316, 722, 664, 107]. Some of the most well-known CNN architectures include *AlexNet* [385], *VGGNet* [639], *ResNet* [284], *GoogLeNet* [670] which use *Inception* modules architecture (explained in detail in section 3), *MobileNet* [293], and *DenseNet* [298].

#### 3.4.3.2 Fully Convolutional Networks

*Fully Convolutional Networks* (FCNs), proposed by Long *et al.*, use convolutional layers to process varying input sizes [445]. It was one of the first DL models for semantic segmentation. As shown in Fig. 3.6b, the final output layer of FCN has a large receptive field and corresponds to the height and width of the image, while the number of channels corresponds to the number of classes. The convolutional layers classify every pixel to determine the context of the image, including the



location of objects. FCNs have been applied to a variety of segmentation problems, such as brain tumor segmentation [722], instance-aware semantic segmentation [420], skin lesion segmentation [784], and iris segmentation [436].

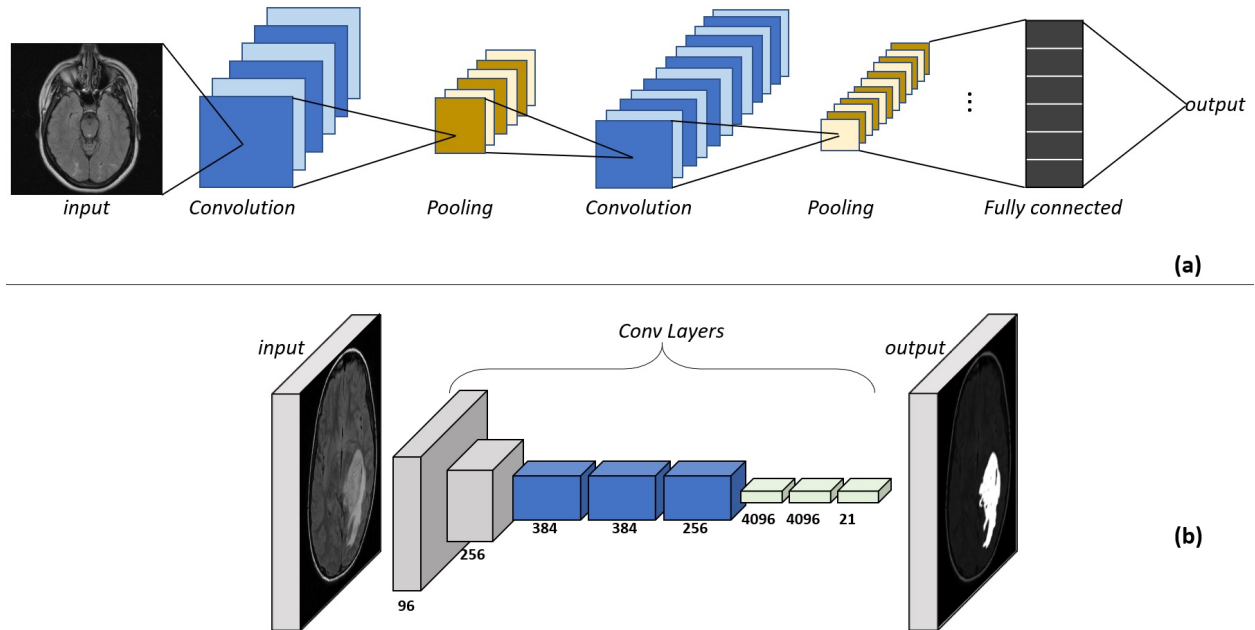


Figure 3.6 Basic architecture of CNN and FCN.

### 3.4.3.3 Encoder-Decoder Based Models

Encoder-decoder models are inspired by the FCNs, and the most well-known architecture of these models are *U-Net* and *V-Net* [593, 485]. *U-Net* was proposed for segmenting biological microscopy images, and it used the data augmentation technique to learn from the available annotated images more effectively. The *U-Net* architecture consists of two parts; a *contracting* or *down-sampling* path to capture the context, and a symmetric *expanding* or *up-sampling* path for localization of the captured context. The contracting path has FCN-like architecture that extracts features with  $3 \times 3$  convolutions while increasing the number of feature maps and reducing their dimensions. Contrarily, the expanding path carries out deconvolutions by reducing the number of feature maps while increasing their dimensions. The feature maps from the contracting path are concatenated to the expanding path to maintain the integrity of pattern information. Finally, a

segmentation map is generated from feature maps by  $1 \times 1$  convolution operation that categorizes each pixel of the input image. *U-Net* was trained on 30 transmitted light microscopy images, and it won the *International Symposium on Biomedical Imaging (ISBI)* cell tracking challenge in 2015 by a large margin. *V-Net* [485] is another well-known FCN-based model proposed for 3D medical image segmentation. It introduced a new objective function based on the Dice coefficient, which enabled the model to deal with strong class imbalance between the number of voxels in the foreground and the background. *V-Net* was trained end-to-end on MRI volumes depicting prostate, and it learned to predict segmentation for the whole volume at once.

#### 3.4.3.4 Other DL Models for Image Segmentation

In addition to the models described in previous sections, there are many families of DL architectures that are very popular for medical image segmentation. For example, convolutional graphical models (incorporating concepts of *Conditional Random Fields (CRFs)* and *Markov Random Field (MRFs)*), *Multi-scale pyramid network* models (*Feature Pyramid Network (FPN)*) [429], *Pyramid Scene Parsing Network (PSPN)* [802], *Regional CNN (R-CNN)* like *Fast R-CNN*, *Faster R-CNN*, and *Mask-RCNN*, dilated or atrous convolution (*DeepLab Family* [131]), RNN-based models (*ReNet* [713], *ReSeg* [712]), *Data-Associated RNNs (DA-RNNs)* [754]), and *attention-based* models (*OCNet* [785], *Expectation-Maximization Attention Network (EMANet)* [418], *Criss-Cross attention Network (CCNet)* [303]). Minaee *et al.* has presented an elaborate review reference of all of these models [486].

#### 3.4.4 Brain Tumor Segmentation Task Challenge

In this section, we discuss the brain tumor segmentation task in the realm of DL. Internationally held challenges on medical imaging analysis have become the standard for validation of the proposed methods. *Brain Tumor Segmentation (BraTS) Challenge* [515] is one such challenge that is held in conjunction with *Medical Imaging Computing and Computer-Assisted Intervention (MICCAI)* conference [260]. The first challenge workshop was held in 2012, followed by yearly

benchmarks held with MICCAI conferences. BraTS challenge evaluates state-of-the-art segmentation methods of brain tumors in MRI scans. It has a publicly available dataset (with accompanying expert delineations), which is used for benchmarking the submitted contenders for segmenting *multi-institutional pre-operative* MRI (mpMRI) scans having intrinsically heterogeneous (in appearance, shape, and histology) brain tumors, namely *gliomas*. In addition, this challenge also encompasses the survival prediction of the patient and evaluates the algorithmic uncertainty in tumor segmentation. The challenge evaluates segmentation of tumor sub-regions of *Enhancing Tumor* (ET), *Tumor Core* (TC), and *Whole Tumor* (WT) as shown in Fig. 3.7. ET are regions of hyper-intensity in T1C when compared to T1, and TC is the bulk of the tumor, that is typically resected. The TC involves ET and the necrotic (fluid-filled) and the non-enhancing (solid) parts of the tumor. WT is the complete extent of the disease, as it is comprised of the TC and the peritumoral *EDema* (ED), depicted by FLAIR. The dataset for the 2018 challenge consisted of a total of 542 patients, with 285 for training, 66 for validation, and 191 for testing scans having 210 *High-Grade Glioma* (HGG) and 75 *Low-Grade Glioma* (LGG) patients with annotations approved from experienced neuro-radiologists through a hierarchical majority vote. The data consists of clinically-acquired 3T multi-contrast MR scans from around 19 institutions, with ground truth labels by expert board-certified neuro-radiologists in NIfTI files (.nii.gz). *Dice coefficient* and *Hausdorff distance* (95%) have been used as evaluation schemes. Apart from these, *Sensitivity* and *Specificity* are also used as metrics. An assessment of state-of-the-art ML methods used for brain tumor segmentation under the BraTS challenge from the period 2012-2018 has been compiled by Bakas *et al.* [63].

### 3.5 Inception Modules in Brain Tumor Segmentation

After assimilating the brain tumor segmentation problem using DNNs, let us now look at one of the architectures that has appreciable accuracy in solving this problem. This architecture is based on *U-Net* with *Inception* modules.

### 3.5.1 Segmentation Using Inception/ Dilated Inception Modules

Cahall *et al.* proposed an image segmentation framework for tumor delineation that benefits from two state-of-the-art ML architectures in CV: *Inception* modules and *U-Net* [107, 670, 593]. This new framework includes two learning regimes, i.e., learning to segment intra-tumoral structures (necrotic and non-enhancing tumor core, peritumoral edema, and enhancing tumor) or learning to segment glioma sub-regions (WT, TC, ET). Both learning regimes are described in section 3.3.4 above. These learning regimes were incorporated into a modified loss function based on the DSC described in section 3.3.1 eq. 3.5 above.

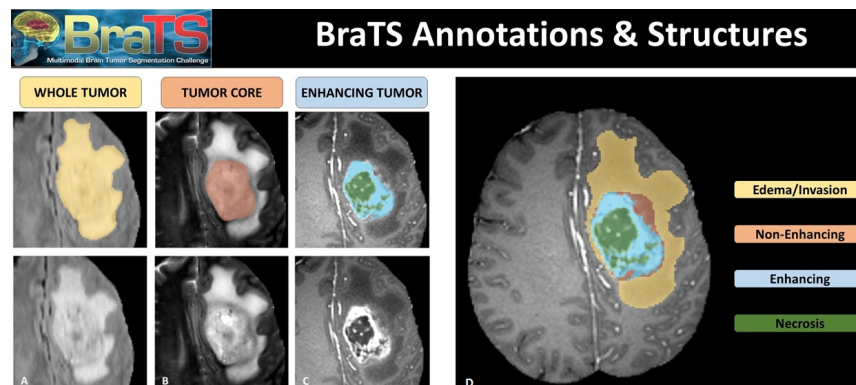


Figure 3.7 Image patches with annotated tumor (glioma) sub-regions.

*U-Net* was originally developed for cell tracking. However, it has been applied recently to other medical segmentation tasks, such as, brain vessel segmentation [444], brain tumor segmentation [188], and retinal segmentation [245]. To tackle different medical imaging segmentation problems, variations and extensions of *U-Net* have also been proposed, such as 3D *U-Net* [343, 608], H-DenseUNet [419], RIC-UNet [791], and Bayesian *U-Net* [526]. Cahall *et al.* used a cascade learning approach in which three different models were used first to learn the WT, then TC, and finally, ET resulting in a proposed end-to-end implementation for all tumor sub-types [107].

### 3.5.2 BraTS Dataset and Pre-processing

We used BraTS 2018 dataset, described in section 3.3.4, for experiments [480, 62, 61, 63]. The dataset contains four sequences for each patient’s MRI (T1, T1c, T2, and FLAIR) images. It also contains ground truth in the form of pixel-level manual segmentation markings for three intratumoral structures: necrotic and non-enhancing tumor core (labeled as 1), peritumoral edema (labeled as 2), and enhancing tumor (labeled as 4). The glioma sub-regions have been defined as WT having all three intratumoral structures (labeled as  $(1 \cup 2 \cup 4)$ ), TC containing all except peritumoral edema (labeled as  $(1 \cup 4)$ ), and ET (labeled as 4). Different sequences provide complementary information for identifying the intratumoral structures; FLAIR highlights the peritumoral edema, T1c distinguishes the ET, and T2 highlights the necrotic and non-enhancing tumor core. BraTS images have been pre-processed for skull-stripping, re-sampled to an isotropic  $1 \text{ mm}^3$  resolution, and co-registered all four modalities of each patient. Cahall *et al.* [107] performed additional pre-processing in the following order:

1. Discard excess background pixels from images by obtaining the bounding box of the brain and extracting the selected portion, effectively zooming on the brain.
2. Re-size the cropped image to  $128 \times 128$  pixels.
3. Drop the images having no tumor regions in the ground truth segmentation.
4. Apply intensity windowing function to each image such that the lowest 1% and the highest 99% pixels were mapped to 0 and 255, respectively.
5. Normalize images by subtracting the mean and dividing by the standard deviation of the dataset.

### 3.5.3 Deep Artificial Neural Network Architectures

In medical imaging, semantic segmentation’s accuracy depends on the ability to extract the local structural as well as global contextual information from MRI scans while training the model. For

this reason, many multi-path architectures in the context of medical imaging have been proposed, and all of them extract the structural and contextual information from input data at multiple scales [343, 280, 602]. This features extraction-aggregation concept at various scales was also done in *Inception* modules [670]. However, the feature extraction mechanism in the *Inception* module is different from the multi-path architectures. The *Inception* module applies filters of various sizes at each layer and concatenates resulting feature maps [670]. Cahall *et al.* [107] proposed a modified version of *Dilated Residual Inception* (DRI) [627] based on *U-Net* and factorized convolution *Inception* module [593, 670]. DRI's special blocks were inspired from *Inception* module [669] and dilated convolution [780]. DRI has fewer parameters than the original *Inception* module and employs residual connections to alleviate the vanishing gradients problem at a faster convergence rate [284]. *MultiResUNet* combined a *U-Net* with residual *Inception* modules for multi-scale feature extraction, applying the architecture to several multimodal medical imaging datasets [308]. Integration of *Inception* modules with *U-Net* has been evaluated for left atrial segmentation [719], liver and tumor segmentation [417], and brain tumor segmentation [407].

### 3.5.3.1 *Inception Module*

The convolutional layer in the proposed *Inception* module [107] in the original *U-Net* was replaced with an *Inception* module having multiple sets of  $3 \times 3$  convolutions,  $1 \times 1$  convolutions,  $3 \times 3$  max pooling, and cascaded  $3 \times 3$  convolutions as depicted in Fig. 3.8(B). At each layer on the contracting path, the height and width of the feature maps are halved, and the depth is doubled until reaching the bottleneck, i.e., the center of the *U*. On the corresponding expanding path at each layer, the height and width of feature maps are doubled, and the depth is halved until having the segmentation mask as the output. As with *U-Net*, feature maps generated on the contracting path are concatenated to the corresponding expanding path. The authors employed a *Rectified Linear Unit* (ReLU) as the activation function, with batch normalization [314] in each *Inception* module. The architecture setting receives an input image of size  $N \times M \times D$  and outputs an  $N \times M \times K$  tensor where  $N = M = 128$  pixels,  $D = 4$  represents the four MRI modalities (T1, T1c, T2, FLAIR),

and  $K = 3$  represents the segmentation classes (intra-tumoral structures or glioma sub-regions). The output image of  $K$  slices is a binary image representing the predicted segments for the  $i^{\text{th}}$  class ( $0 \leq i \leq K - 1$ ). Pixel-wise activation functions (sigmoid [513] for *glioma* and *softmax* [513] for intra-tumoral structures) are used to generate the output binary images.

### 3.5.3.2 Dilated Inception U-Net

Another useful architecture, called *Dilated Inception U-Net* (DIU-Net), integrates dilated or astrous convolutions [131] and *Inception* modules in the *U-Net* architecture [632] as shown in Fig. 3.8(A). Here, each *dilated Inception* module consists of three  $1 \times 1$  convolutional operations, followed by one  $l$ -dilated convolutional filter (with  $l = 1, 2, 3$ ), as illustrated in Fig. 3.8(C). The  $1 \times 1$  convolutional filters perform dimensionality reduction, while three  $l$ -dilated convolutional filters each of size  $3 \times 3$  implement atrous convolutions. In dilated convolutions, an image  $I$  of size  $m \times n$  and a discrete convolutional filter  $w$  of size  $k \times k$  are convolved by:

$$(I * w)(p) = \sum_s I[p + s] w[s]. \quad (3.6)$$

Simple convolution operation of eq. 3.6 can be generalized to  $l$ -dilated convolution ( $*_l$ ) as [780]:

$$(I *_l w)(p) = \sum_s I[p + l_s] w[s]. \quad (3.7)$$

For  $l = 1$ , we get the simple convolutional operation of eq. 3.6. For  $l > 1$ ,  $l-1$  zeroes are inserted between each filter element, creating a scaled and sparse filter of size  $k_s \times k_s$ , where  $k_s$  is defined by:

$$k_s = k + (k - 1)(l - 1), \quad (3.8)$$

$$= l(k - 1) + 1. \quad (3.9)$$

The scaling  $s$  increases the receptive field of the filter by a factor  $\frac{k_s}{k}$ .

$$\frac{k_s}{k} = \frac{k + (k - 1)(l - 1)}{k}, \quad (3.10)$$

$$= l + \left( \frac{-l + 1}{k} \right). \quad (3.11)$$

The receptive field of the filter increases linearly with  $l$ , while the number of elements ( $k \times k$ ) remains fixed.

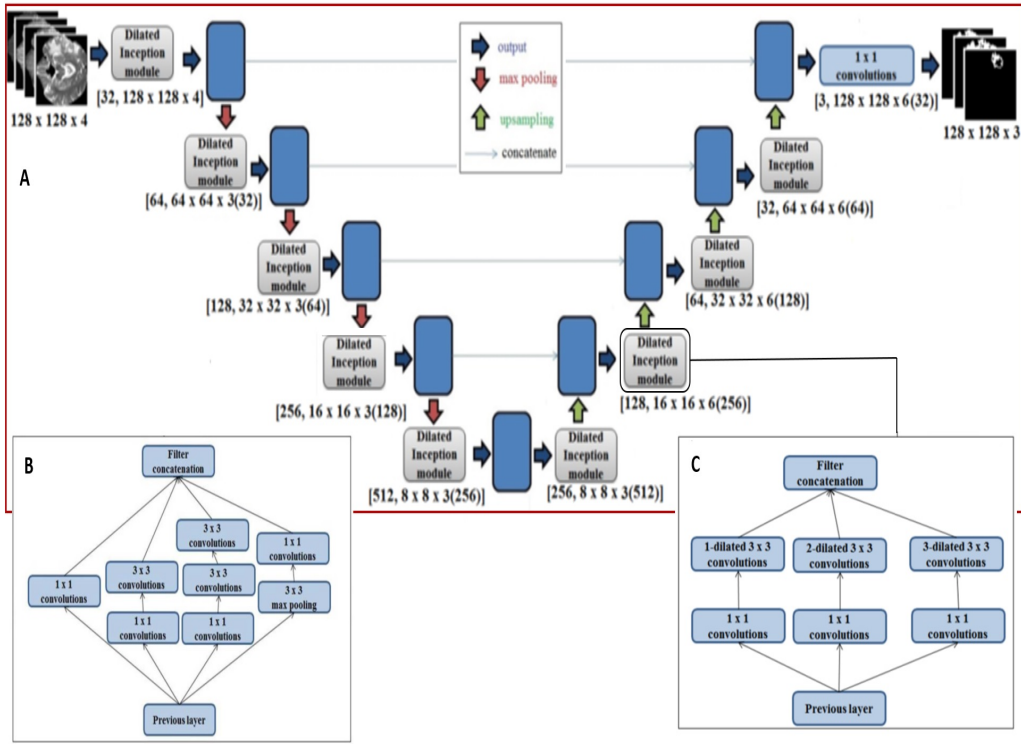


Figure 3.8 DIU-Net, Inception module, and Dilated Inception module architectures.

### 3.5.3.3 Modified DSC as Objective/Loss Function

Cahall *et al.* used a modified version of DSC (eq. 3.5) as an objective/ loss function, after incorporating three changes: (1) the sign of DSC was changed to convert it into a minimization



problem, (2) a log function was introduced, and (3) a new parameter  $\gamma$  was used to cater for extremely large values of the loss function [107]. From initial experiments, it was empirically observed that  $\gamma = 100$  provided the best segmentation performance. Modified DSC as a loss function for a binary class (tumor or not tumor) and multi-class (for  $K$  classes) are given in the following two equations:

$$\mathcal{L}_{DSC}(P, G) = -\log \left[ 2 \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P_{ij} G_{ij} + \gamma}{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P_{ij} + \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} G_{ij} + \gamma} \right], \quad (3.12)$$

$$\mathcal{L}_{DSC}(P, G) = -\log \left[ \frac{1}{K} \sum_{i=0}^{K-1} DSC(P_i, G_i) \right]. \quad (3.13)$$

### 3.5.4 Experimental Setup and Results

Four different models were trained by Cahall *et al.* [107], two for the *U-Net* architecture (intra-tumoral structures and glioma sub-regions), and two for the *U-Net* with *Inception* module (intra-tumoral structures and glioma sub-regions). All four models were trained using  $k$ -fold *cross-validation* on the dataset that was randomly split into  $k$  mutually exclusive subsets of equal or near-equal size. Each algorithm was run  $k$  times subsequently, and each time one of the  $k$  splits was taken as a validation subset and the rest as the training subset. *Stochastic gradient descent* [597] with an adaptive moment estimator (*Adam*[597]) was used for training all models and their variations [374]. With a batch size of 64 and 100 epochs, the learning rate was initially set to  $10^{-4}$ , which was exponentially decayed every 10 epochs. All learnable parameters (weights and biases) were initialized based on the He initialization method [282]. *Keras* [143] *Application Programming Interface* (API) with *TensorFlow* [1] backend was used for implementation, and all models were trained on a Google Cloud Compute [255] instance with 4 NVIDIA TESLA P100 *Graphical Processing Units* (GPUs).

#### 3.5.4.1 Results - Inception Modules

For intra-tumoral structures, the addition of *Inception* modules to *U-Net* resulted in statistically significant improvements in WT (DSC improved from 0.903 to 0.925,  $p < 0.001$ ), TC (0.938 to 0.952,  $p < 0.001$ ), and ET (0.937 to 0.948,  $p < 0.001$ ). For glioma sub-regions, significant improvements were also noticed in WT (0.898 to 0.918,  $p < 0.001$ ), TC (0.942 to 0.951,  $p = 0.001$ ), and ET (0.942 to 0.948,  $p = 0.002$ ). Changing the objective from intra-tumoral structures to glioma sub-regions learning in the *U-Net* resulted in no difference in performance for WT (0.903 to 0.898,  $p = 0.307$ ), TC (0.938 to 0.942,  $p = 0.284$ ), and ET (0.937 to 0.942,  $p = 0.098$ ). However, *U-Net* with *Inception* modules, which learned the intra-tumoral structures outperformed those which learned glioma sub-regions in WT (0.918 to 0.925,  $p = 0.007$ ), but there was no difference in the performance for TC (0.952 to 0.951,  $p = 0.597$ ) and ET (0.948 to 0.948,  $p = 0.402$ ). This implies that integrating *Inception* modules in the *U-Net* architecture resulted in statistically significant improvement in tumor segmentation performance that was quantified using  $k$ -fold cross-validation ( $p < 0.05$  for all three glioma sub-regions). The improvement in the validation accuracy can be attributed to the multiple convolutional filters of different sizes employed in each *Inception* module. These filters are able to capture and retain contextual information at multiple scales during the learning process, both in the contracting as well as expanding paths. We also consider that the improvement in the tumor segmentation accuracy is linked to the new loss function based on the modified DSC (eq. 3.13). DSC scores for *Inception* modules are comparable or exceed the results of No New-Net [317], which achieved second place in the BraTS 2018 competition, and the ensemble approach proposed in [317, 346, 793].

#### 3.5.4.2 Results - DIU-Net

DIU-Net showed significant improvement in the WT sub-region with an increase in the Dice score from 0.925 to 0.931 with  $p < 0.05$ . For the TC sub-region, the Dice score improved from 0.952 to 0.957 with  $p < 0.05$ . However, for the ET, the change was not statistically significant,  $p = 0.114$ . Interestingly, DIU-Net is computationally more efficient. DIU-Net has 2.5 million fewer

parameters than *U-Net* with *Inception* modules. In contrast, DIU-Net achieves significantly better results at a lesser computational cost (15% fewer parameters). The *Dice* scores for each glioma sub-region are comparable or exceed the results of other recently published architectures, including *No New-Net*, SDResU-Net and the ensemble approach proposed in [317, 346, 793].

### 3.6 Uncertainty Estimation in Brain Tumor Segmentation

As mentioned before, accurate segmentation of brain tumors is crucial for treatment planning and follow-up evaluations. Furthermore, the robustness and trustworthiness of the segmentation results are of particular interest in medical imaging and in the clinic for diagnosis and prognosis due to their link to human health. In this section, we propose a new DL framework, named *extended Variational Density Propagation* (exVDP), that can quantify uncertainty in the output decision [173]. In exVDP, we adopt the *Variational Inference* (VI) [86] framework and propagate the first two moments of the variational distribution through all ANN’s layers (convolution, max-pooling and fully-connected) and non-linearities. We use the first-order Taylor series linearization [539] to propagate the mean and covariance of the variational distribution through the non-linear activation functions in the DNNs.

We consider a CNN with a total of  $C$  convolutional layers and  $L$  fully-connected layers, where the convolutional kernels and the weights of the fully-connected layers are random tensors. A non-linear activation function follows every convolutional and fully-connected layer. Moreover, the ANN contains max-pooling layers. The ANN’s weights (and biases) are represented by  $\Omega = \{ \{ \{ \mathbf{W}^{(k_c)} \}_{k_c=1}^{K_c} \}_{c=1}^C, \{ \mathbf{W}^{(l)} \}_{l=1}^L \}$ , where  $\{ \{ \mathbf{W}^{(k_c)} \}_{k_c=1}^{K_c} \}_{c=1}^C$  is the set of  $K_c$  kernels in the  $c^{\text{th}}$  convolutional layer, and  $\{ \mathbf{W}^{(l)} \}_{l=1}^L$  is the set of weights in  $L$  fully-connected layers. We consider input tensor  $\mathcal{X} \in \mathbb{R}^{l_1 \times l_2 \times K}$ , where  $l_1$ ,  $l_2$ , and  $K$  represent image height, width, and number of channels, respectively.

### 3.6.1 Variational Learning

We introduce a prior distribution over ANN weights,  $\Omega \sim p(\Omega)$ . We assume that convolutional kernels are independent of each other within a layer as well as across different layers. This independence assumption is desirable as it promotes convolutional kernels to extract uncorrelated features within and across layers. Given the training data  $\mathcal{D} = \{\mathcal{X}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^N$  and the prior  $p(\Omega)$ , the posterior  $p(\Omega|\mathcal{D})$  is given through the Bayes' rule. However,  $p(\Omega|\mathcal{D})$  is typically intractable. VI methods approximate the true posterior  $p(\Omega|\mathcal{D})$  with a simpler parametrized variational distribution  $q_\phi(\Omega)$ . The optimal parameters of the variational posterior  $\phi^*$  are estimated by minimizing the *Kullback-Leibler* (KL) divergence between the approximate and the true posterior [84, 86].

$$\begin{aligned} \phi^* &= \operatorname{argmin} \operatorname{KL} [q_\phi(\Omega) \| p(\Omega|\mathcal{D})] \\ &= \operatorname{argmin} \int q_\phi(\Omega) \log \frac{q_\phi(\Omega)}{p(\Omega)p(\mathcal{D}|\otimes)} d\Omega \\ &= \operatorname{argmin} \operatorname{KL} [q_\phi(\Omega) \| p(\Omega)] - E_{q_\phi(\Omega)} \{\log p(\mathcal{D}|\otimes)\}. \end{aligned} \quad (3.14)$$

The optimization objective is given by the *Evidence Lower Bound* (ELBO)  $\mathfrak{L}(\phi; \mathbf{y}|\mathcal{X})$ :

$$\mathfrak{L}(\phi; \mathbf{y}|\mathcal{X}) = E_{q_\phi(\Omega)}(\log p(\mathbf{y}|\mathcal{X}, \Omega)) - \operatorname{KL}(q_\phi(\Omega) \| p(\Omega)). \quad (3.15)$$

ELBO consists of two parts, the expected log-likelihood of the training data given the weights and a regularization term, which can be re-written as:

$$\operatorname{KL}(q_\phi(\Omega) \| p(\Omega)) = \sum_{c=1}^C \sum_{k_c=1}^{K_c} \operatorname{KL}(q_\phi(\mathcal{W}^{(k_c)}) \| p(\mathcal{W}^{(k_c)})) - \sum_{l=1}^L \operatorname{KL}(q_\phi(\mathbf{W}^{(l)}) \| p(\mathbf{W}^{(l)})). \quad (3.16)$$

### 3.6.2 Variational Density Propagation

We propose to approximate the true unknown posterior  $p(\Omega|\mathcal{D})$  by a variational distribution  $q_\phi(\Omega)$ . We have defined *Gaussian* distribution as a prior over convolutional kernels and weights of the fully-connected layers [173]. The task is now to propagate the moments of the variational dis-

tribution  $q_\phi(\Omega)$  through various layers, i.e., convolution, activation, max-pooling, fully-connected, and *softmax*. It is important to note that in our settings, the convolutional kernels, resulting activations, extracted features, logits, and output of the *softmax* function are all random variables. Therefore, instead of performing algebraic operations on real numbers, we are confronted with operations on random variables, including (1) multiplication of a random variable with a constant, (2) multiplication of two random variables, and (3) non-linear transformations [173] operating over random variables. As a result of the multiplication of two Gaussian random variables [539] or non-linear transformation, the resulting random variables may not have *Gaussian distribution* [539]. Our goal is to propagate the mean and covariance of the variational distribution and later obtain the mean and covariance of the predictive distribution,  $p(\mathbf{y}|\mathcal{X}, \mathcal{D})$ . The mean of  $p(\mathbf{y}|\mathcal{X}, \mathcal{D})$  represents the ANN’s prediction, while the covariance matrix reflects the uncertainty associated with the output decision. An illustration of the proposed variational density propagation CNN with one convolutional layer, one max-pooling and one fully-connected layer is shown in Fig. 3.9.

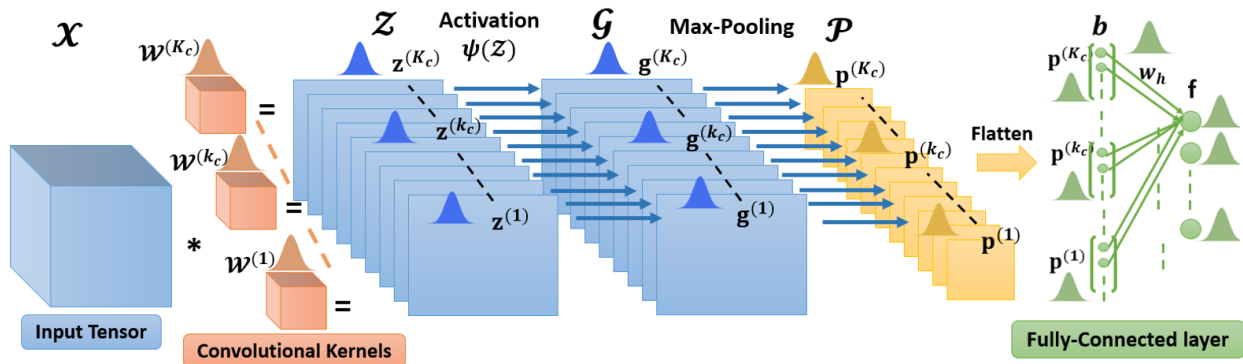


Figure 3.9 A schematic layout of the proposed variational density propagation CNN.

### 3.6.3 Extended Variational Density Propagation

We start with our mathematical results for the propagation of the mean and covariance of the variational distribution  $q_\phi(\Omega)$  through convolutional layers, activation functions, max-pooling, fully-connected layers, and the *softmax* function. We use first-order Taylor series for the approxi-

mation of the first two moments (mean and covariance) after a non-linear activation function and refer to this method as the exVDP [173].

### 3.6.3.1 First Convolutional Layer

The convolution operation between a set of kernels and the input tensor is formulated as a *matrix-vector multiplication*. We first form sub-tensors  $\mathcal{X}_{i:i+r_1-1,j:j+r_2-1}$  from the input tensor  $\mathcal{X}$ , having the same size as the kernels  $\mathcal{W}^{(k_c)} \in \mathbb{R}^{r_1 \times r_2 \times K}$ . These sub-tensors are subsequently vectorized and arranged as the rows of a matrix  $\tilde{\mathcal{X}}$ . Thus, convolving  $\mathcal{X}$  with the  $k_c^{\text{th}}$  kernel  $\mathcal{W}^{(k_c)}$  is equivalent to the multiplication of  $\tilde{\mathcal{X}}$  with  $\text{vec}(\mathcal{W}^{(k_c)})$ . Let

$$\mathbf{z}^{(k_c)} = \mathcal{X} * \mathcal{W}^{(k_c)} = \tilde{\mathcal{X}} \times \text{vec}(\mathcal{W}^{(k_c)}), \quad (3.17)$$

where  $*$  denotes the convolution operation and  $\times$  is a regular matrix-vector multiplication. We have defined a multivariate Gaussian distribution over the vectorized convolutional kernels, i.e.  $\text{vec}(\mathcal{W}^{(k_c)}) \sim \mathcal{N}(\mathbf{m}^{(k_c)}, \Sigma^{(k_c)})$ . It follows that:

$$\mathbf{z}^{(k_c)} \sim \mathcal{N}\left(\mu_{\mathbf{z}^{(k_c)}} = \tilde{\mathcal{X}}\mathbf{m}^{(k_c)}, \Sigma_{\mathbf{z}^{(k_c)}} = \tilde{\mathcal{X}}\Sigma^{(k_c)}\tilde{\mathcal{X}}^T\right). \quad (3.18)$$

### 3.6.3.2 Non-linear Activation Function

We approximate the mean and covariance after the non-linear activation function  $\psi$  using the first-order Taylor series approximation [539]). Let  $\mathbf{g}_i^{(k_c)} = \psi[\mathbf{z}_i^{(k_c)}]$  be the element-wise  $i^{\text{th}}$  output of  $\psi$ . We have  $\mu_{\mathbf{g}^{(k_c)}}$  and  $\Sigma_{\mathbf{g}^{(k_c)}}$ :

$$\begin{aligned} \mu_{\mathbf{g}_i^{(k_c)}} &\approx \psi(\mu_{\mathbf{z}_i^{(k_c)}}), \\ \Sigma_{\mathbf{g}^{(k_c)}} &\approx \begin{cases} \sigma_{\mathbf{z}_i^{(k_c)}}^2 \left(\frac{d\psi(\mu_{\mathbf{z}_i^{(k_c)}})}{d\mathbf{z}_i^{(k_c)}}\right)^2, & \text{if } i = j. \\ \sigma_{\mathbf{z}_i^{(k_c)}\mathbf{z}_j^{(k_c)}} \left(\frac{d\psi(\mu_{\mathbf{z}_i^{(k_c)}})}{d\mathbf{z}_i^{(k_c)}}\right) \left(\frac{d\psi(\mu_{\mathbf{z}_j^{(k_c)}})}{d\mathbf{z}_j^{(k_c)}}\right), & \text{if } i \neq j. \end{cases} \end{aligned} \quad (3.19)$$

### 3.6.3.3 Max-pooling Layer

For the max-pooling,  $\mu_{\mathbf{p}^{(k_c)}} = \text{pool}(\mu_{\mathbf{g}^{(k_c)}})$  and  $\Sigma_{\mathbf{p}^{(k_c)}} = \text{co-pool}(\Sigma_{\mathbf{g}^{(k_c)}})$ , where pool represents the max-pooling operation on the mean and co-pool represents down-sampling the covariance, i.e., we keep only the rows and columns of  $\Sigma_{\mathbf{g}^{(k_c)}}$  corresponding to the pooled mean elements.

### 3.6.3.4 Flattening Operation

The output tensor  $\mathcal{P}$  of the max-pooling layer is vectorized to form the input vector  $\mathbf{b}$  of the fully-connected layer such that,  $\mathbf{b} = \left[ \mathbf{p}^{(1)T}, \dots, \mathbf{p}^{(K_c)T} \right]^T$ . The mean and covariance matrix of  $\mathbf{b}$  are given by:

$$\mu_{\mathbf{b}} = \begin{bmatrix} \mu_{\mathbf{p}^{(1)}} \\ \vdots \\ \mu_{\mathbf{p}^{(K_c)}} \end{bmatrix}, \Sigma_{\mathbf{b}} = \begin{bmatrix} \Sigma_{\mathbf{p}^{(1)}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_{\mathbf{p}^{(K_c)}} \end{bmatrix}. \quad (3.20)$$

### 3.6.3.5 Fully-connected Layer

Let  $\mathbf{w}_h \sim \mathcal{N}(\mathbf{m}_h, \Sigma_h)$  be  $h^{\text{th}}$  weight vector of the fully-connected layer, where  $h = 1, \dots, H$ , and  $H$  is the number of output neurons. We note that  $f_h$  is the product of two independent random vectors  $\mathbf{b}$  and  $\mathbf{w}_h$ . Let  $\mathbf{f}$  be the output vector of the fully-connected layer, then the elements of  $\mu_{\mathbf{f}}$  and  $\Sigma_{\mathbf{f}}$  are derived by the following proposition:

**Proposition 1.**

$$\mu_{f_h} = \mathbf{m}_h^T \mu_{\mathbf{b}}, \quad (3.21)$$

$$\Sigma_{\mathbf{f}} = \begin{cases} \text{tr}(\Sigma_h \Sigma_{\mathbf{b}}) + \mathbf{m}_h^T \Sigma_{\mathbf{b}} \mathbf{m}_h + \mu_{\mathbf{b}}^T \Sigma_h \mu_{\mathbf{b}}, \\ \mathbf{m}_{h_1}^T \Sigma_{\mathbf{b}} \mathbf{m}_{h_2}, \end{cases} \quad h_1 \neq h_2, \quad (3.22)$$

where  $h_1, h_2 = 1, \dots, H$  represent any two weight vectors in the fully-connected layer.

### 3.6.3.6 Softmax Function

Let the output of the ANN be  $\mathbf{y} = \varphi(\mathbf{f})$ , where  $\varphi$  is the *softmax* function. Using the first-order Taylor series approximation, the mean and covariance of the output vector, i.e.,  $\mu_{\mathbf{y}}$  and  $\Sigma_{\mathbf{y}}$ , are derived as follows [638]:

$$\mu_{\mathbf{y}} \approx \varphi(\mu_{\mathbf{f}}); \Sigma_{\mathbf{y}} \approx \mathbf{J}_{\varphi} \Sigma_{\mathbf{f}} \mathbf{J}_{\varphi}^T, \quad (3.23)$$

where  $\mathbf{J}_{\varphi}$  is the *Jacobian* matrix of  $\varphi$  with respect to  $\mathbf{f}$  evaluated at  $\mu_{\mathbf{f}}$  [638].

### 3.6.3.7 Objective Function

Assuming a diagonal covariance matrix for the variational posterior distribution,  $N$  independently and identically distributed (iid) data points and using  $M$  Monte Carlo samples to approximate the expectation by a summation, the expected log-likelihood in the ELBO objective function is given as follows:

$$E_{q_{\phi}(\Omega)}(\log p(\mathbf{y}|\mathcal{X}, \Omega)) \approx -\frac{NH}{2} \log(2\pi) - \frac{1}{M} \sum_{m=1}^M \left[ \frac{N}{2} \log(|\Sigma_{\mathbf{y}}|) + \frac{1}{2} \sum_{i=1}^N (\mathbf{y}^{(i)} - \mu_{\mathbf{y}}^{(m)})^T (\Sigma_{\mathbf{y}}^{(m)})^{-1} (\mathbf{y}^{(i)} - \mu_{\mathbf{y}}^{(m)}) \right] \quad (3.24)$$

The regularization term in (3.16) is the KL-divergence between two multivariate Gaussian distributions [539]. If we have a CNN with one convolutional layer followed by the activation function, one max-pooling and one fully-connected layer, thus the regularization term in the ELBO objective function is derived as follows:



$$\begin{aligned}
\text{KL}(q_\phi(\Omega)\|p(\Omega)) = & \\
& \frac{1}{2} \sum_{k=1}^{K_1} \left( r_1 r_2 K \sigma_{r_1,k}^2 \sigma_{r_2,k}^2 \sigma_{K,k}^2 + \|\mathcal{M}^{(k)}\|_F^2 - r_1 r_2 K - r_1 r_2 K \left( \log(\sigma_{r_1,k}^2 \sigma_{r_2,k}^2 \sigma_{K,k}^2) \right) \right) \\
& + \frac{1}{2} \sum_{h=1}^H \left( n_f \sigma_h^2 + \|\mathbf{m}_h\|_F^2 - n_f - n_f \log \sigma_h^2 \right), \tag{3.25}
\end{aligned}$$

where  $(r_1 \times r_2 \times K)$  is the size of the kernels,  $K_1$  is the number of kernels in the convolutional layer,  $H$  is the number of output neurons and  $n_f$  is the length of the weight vector  $\mathbf{w}_h$  in the fully-connected layer.

### 3.6.3.8 Back-propagation

During back-propagation, we compute the gradient of the objective function  $\nabla_\phi \mathcal{L}(\phi; \mathcal{D})$  with respect to the variational parameters:

$$\phi = \left\{ \left\{ \left\{ \mathcal{M}^{(k_c)}, \sigma_{r_1,k_c}^2, \sigma_{r_2,k_c}^2, \sigma_{K_{c-1},k_c}^2 \right\}_{k_c=1}^{K_c} \right\}_{c=1}^C, \left\{ \mathbf{m}_h, \sigma_h^2 \right\}_{h=1}^H \right\}, \tag{3.26}$$

where  $(r_1 \times r_2 \times K_{c-1})$  is the size of the  $k_c^{\text{th}}$  kernel,  $K_c$  is the number of kernels in the  $c^{\text{th}}$  convolutional layer and  $H$  is the number of output neurons. We use  $\nabla_\phi \mathcal{L}(\phi; \mathcal{D})$  to update our parameters  $\phi$  using the gradient descent update rule.

## 3.6.4 Application to Brain Tumor Segmentation in MRI Images

The performance of the proposed exVDP model on the HGG brain tumor segmentation task using the BraTS 2015 dataset has been evaluated. The dataset consists of 5 classes, i.e. class 0 - normal tissue, class 1 - necrosis, class 2 - edema, class 3 - non-enhancing, and class 4 - enhancing tumor [480]. The evaluation of segmentation is based on three regions, (1) complete tumor (1, 2, 3 and 4), (2) tumor core (1, 3 and 4), and (3) enhancing tumor (class 4) [480].

Brain tumor segmentation has been formulated as a multi-class classification problem by randomly sampling patches from four MRI modalities, i.e., FLAIR, T1, T2 and T1c [554]. The label of each patch has been manually set to the label of the center pixel. The sampled patches are balanced over all classes, and a total of 100,000 patches of size  $33 \times 33$  are extracted from the BraTS data of 20 patients. These patches are divided into training and validation bins (95% for training and 5% for validation). The test set included randomly sampled 372 images, i.e., 43,264 patches, from each of the four modalities. The proposed exVDP model is compared with a deterministic CNN, presented in [554]. The following CNN architecture has been used: six convolution layers (all kernels were  $3 \times 3$ , and we had 32, 32, 64, 64, 128, 128 kernels in layers one to six, respectively, followed by ReLU activation), two max-pooling layers, and a fully-connected layer. The architecture is shown in Table (3.1).

DSC has been used to evaluate the segmentation results before and after adding Gaussian noise or targeted adversarial attack (targeted class is class 3, i.e., *non-enhancing tumor*). The evaluation of the proposed model on the BraTS dataset is done without doing any pre-processing or data augmentation techniques.

In Table (3.2), DSC values for three test cases have been presented, i.e., noise-free, Gaussian, and adversarial noise. We note that the DSC values of the proposed model are significantly higher than that of the deterministic CNN for all cases in general and adversarial noise in particular. Fig. 3.10 shows segmentation results for exVDP and a deterministic CNN for a representative HGG image (with and without adversarial noise). The uncertainty map associated with each segmentation is also presented for the exVDP model. The uncertainty map allows physicians to review the segmentation results quickly and, if needed, make corrections of tumor boundaries in the regions where the uncertainty is high.

### **3.7 Tumor Surveillance**

As defined in section 3.3.3, tumor surveillance is the process of monitoring patient's tumor in longitudinal studies to establish severity of the disease and planning treatment accordingly. It helps

Table 3.1 Architecture of the two models, i.e., exVDP and deterministic CNN

Layer	Type	Filter size	HGG stride	No. kernels	FC units	Input
1	Conv.	3×3	1×1	32	-	33×33×4
2	Conv.	3×3	1×1	32	-	33×33×32
3	Conv.	3×3	1×1	64	-	33×33×32
4	Max-pool.	3×3	2×2	-	-	33×33×64
5	Conv.	3×3	1×1	64	-	16×16×64
6	Conv.	3×3	1×1	128	-	16×16×64
7	Conv.	3×3	1×1	128	-	16×16×128
8	Max-pool.	3×3	2×2	-	-	16×16×128
9	FC	-	-	-	5	6272

identifying early signs of tumor occurrence which is critical especially in case of the cancerous tumors.

### 3.7.1 Rationale for Tumor Surveillance

Temporal medical imaging data is widely used in oncology as well as radiology for visual comparison of disease over an extended period of time. The 2D medical images (CT or MRI scans) are examined by the physicians to diagnose the disease in 4D (3D tumor volume over time) usually referred to as *change in volume over time*. A detection at an earlier stage of disease is more responsive to treatment, resulting in improved outcomes for the patient. Biological characteristics of various tumor types such as growth, location, and patterns of local as well as metastatic disease are the basis for surveillance scheduling, protocols, and selection of imaging techniques. Usually low-risk tumor is subjected to active surveillance as part of a patient’s treatment plan and an ideal candidate for active surveillance is the one which has one or more of these conditions: (1) disease has not spread, (2) tumor is small and growing slowly, or (3) patient exhibits no symptoms of specific cancer. The data from active surveillance is also used to look for trends and patterns over time in certain regions/ groups of people, and to see if preventive measures are making a

Table 3.2 Segmentation results using DSC for BraTS test dataset

Method	Tumor Regions	No noise	Adversarial noise	Gaussian noise
exVDP	Complete	<b>80.8%</b>	<b>77.4%</b>	<b>80.6%</b>
	Core	74.6%	72.6%	74.5%
	Enhancing	74.0%	69.8%	73.9%
Deterministic CNN	Complete	78.0%	43.4%	66.9%
	Core	65.0%	47.1%	51.9%
	Enhancing	75.0%	43.9%	55.7%

difference among the sample population. Apart from brain cancer, active surveillance has been widely employed in other forms of cancer diagnosis. It has been shown that use of *Prostate-Specific Antigen* (PSA) [310] testing as part of active surveillance of prostate cancer helps in understanding tumor progression and prognosis, enabling the patients diagnosed with lower grade disease feel more comfortable [552].

### 3.7.2 Surveillance Techniques

*Change-point detection* is the classical technique of detecting abrupt changes in sequential data, which focuses predominantly on datasets with a single observable. It has been a long-standing research area in statistics [72, 100], with applications in fields ranging from economics [371], bioinformatics [521, 233], and climatology [586], wherein it is dealt as the problem of detecting abrupt changes in temporal data. The objective is to determine if the observed time series is statistically homogeneous or otherwise to find the point in time when the change happens. There are two variations to the change-point detection technique: posterior and sequential. Posterior tests are done offline after entire data is collected, and a decision of homogeneity or change-point is made based on the analysis of all the collected data. Whereas, sequential tests are done on-the-fly as the data is presented sequentially, and the decisions are made online. *Gleason grade* is another technique used for pathological scoring of the differentiation of prostate cancer, and it has been

the most widely used grading system for prostate tumor differentiation and prognostic indicator for prostate cancer progression [44].

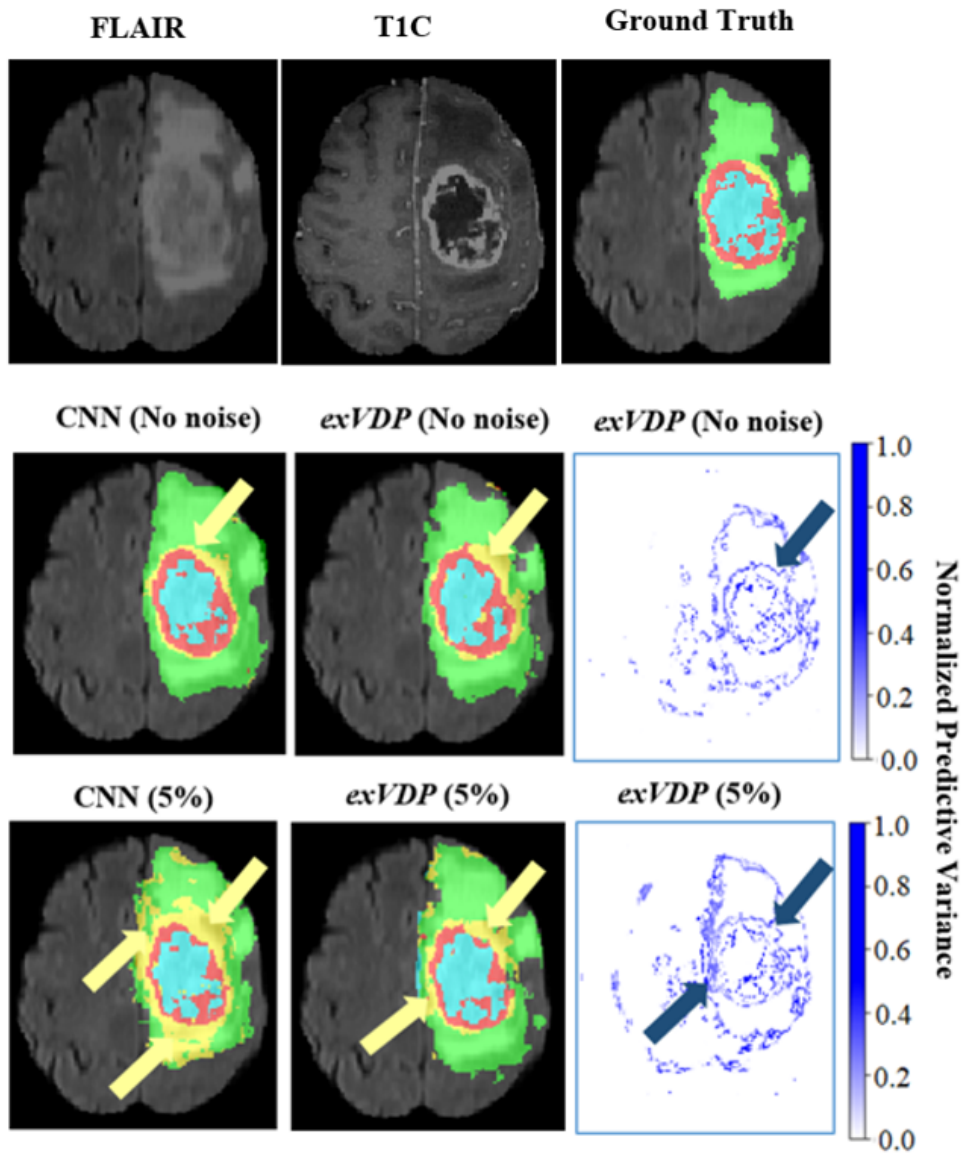


Figure 3.10 Segmentation results of exVDP and deterministic CNN.

### 3.7.3 Community-level Active Surveillance

Apart from the individual tumor surveillance, the term of active surveillance is also often used for collective cancer surveillance data and programs in the United States through *Cancer Registries*. A cancer registry is an information system designed for the collection, storage, and management of data on persons with cancer [312]. Data on cancer in the United States is collected through two types of registries: hospital registries, which are the part of a facility's cancer program, and population-based registries, usually tied to state health departments. Hospital registries provide patient's data on care within the hospital for evaluation. Population-based registries, under state health departments, collect information on all cases diagnosed within a certain geographic area from multiple reporting facilities, including hospitals, doctors' offices, nursing homes, pathology laboratories, radiation and chemotherapy treatment centers, etc. The collected data is used to build statistics like new cancer cases (incidence), death rates (mortality), cancer types related to types of jobs, cancer trends over time to keep an eye on age and racial groups that are most affected by different types of cancer. Registries are staffed under the *Certified Tumor Registrar* (CTR), having pre-defined standards of training, testing, and continuing education, and they compile timely, accurate, and complete cancer information to report to the registry. The major cancer surveillance programs in the United States are the *National Cancer Data Base* (NCDB) [83], *National Cancer Institute's* (NCI) *Surveillance, Epidemiology and End Results* (SEER) program [312], and *National Program of Cancer Registries* (NPCR) of the *Center for Disease Control and Prevention* (CDC) [118]. *Central Brain Tumor Registry of the United States* (CBTRUS) is a registry dedicated to collecting and disseminating statistical data on all primary benign and malignant brain tumors [117]. A recent study [389] tries to estimate excess mortality in people with cancer and multi-morbidity in the COVID-19 affected patients through analysis of surveillance data DATA-CAN, the UK National Health Data Research Hub for cancer emergency [281].

### 3.7.4 Surveillance of Brain Tumor

Tumors of the *Central Nervous System* (CNS) are the second most common tumors among children after leukemia. Treatment protocols for high-grade pediatric brain tumors recommend regular follow-up imaging for up to 10 years. Based on the surveillance data of high-grade childhood brain tumor patients, a review of maximal time to recurrence and minimal time to radiologically detectable long-term sequel such as secondary malignancies, vascular complications, and white matter disease found that there was no recurrence of the primary brain tumor, either local or distant, 10 years or more after the end of treatment in the reviewed literature and so the results do not justify routine screening to detect tumor recurrence more than 10 years after the end of treatment [530].

Tumor surveillance is being used for building statistical figures in a broad spectrum of ways, including adult glioma incidence and survival by race or ethnicity in the United States [527], county-level glioma incidence and survival variations [159], and accurate population-based statistics on the brain and other central nervous system tumors [387]. A CBTRUS statistical report on the primary brain and *Other Central Nervous System* (OCNS) tumors data diagnosed in the United States in the period 2011-2015 states that brain and OCNS tumors (both malignant and non-malignant) were the most common cancer types in persons age 0–14 years for both males and females. For age 15–39 years, these tumors were the second most common cancer in males and the third most common among females in this age group. For age 40+, these were the eighth most common cancer type, with males having eighth and females having the fifth most common brain cancer. These results were based on the NPCR data of 388,786 brain and OCNS tumors, and 16,633 tumor cases from SEER [528]. Tumor Surveillance among patients also enables the authorities to predict cancer cases and death in advance and respond in time to offset the predicted scores. A recent report by the *American Cancer Society* on cancer statistics in 2020 projected the number of new cancer cases and deaths that will occur in the United States. Incidence data from 2002 to 2017 were collected, and it was estimated that in 2020, 23,890 new cases and 18,020 deaths related to brain and ONS tumors were projected to occur in the United States [637]. Similarly, tumor surveillance is equally important to avoid the side effects of the aggressive forms of treatment. Patients treated for glioma,

meningioma, and brain metastases may develop side effects of treatment, including neuropathy (with visual loss), cataracts, hypopituitarism, cognitive decline, increased risk of stroke, and risk of secondary tumor occurring months or even years later. The surgical treatment causes immediate side effects, chemotherapy-caused side effects occur early after treatment (but infertility may not manifest itself until later), and radiotherapy's side effects occur months or even years after treatment. The risks vary depending on the technique used and the area of the brain treated. Surveillance enables the physicians to identify these potential late side-effects earlier which increases the length and quality of life for patients [224].

### 3.7.5 An Example of Surveillance Study

In this section, we will study an example of tumor surveillance, specifically of patients having low-grade gliomas [214]. Low-grade gliomas, constituting around 15% of all adult brain tumors, significantly affect neurological morbidity by brain invasion. Generally, there is no universally-accepted technique available for the detection of growth of low-grade gliomas in the clinical setting. Clinicians usually consider visual comparisons of two or more longitudinal radiological scans through subjective evaluation for detecting the growth of low-grade gliomas. The paper [214] suggests a *Computer-Assisted Diagnosis* (CAD) method to help physicians detect earlier growth of low-grade gliomas. This method consists of tumor segmentation, computing volumes, and pointing to growth by the online abrupt change-of-point method considering only past measurements. The study suggests that early growth detection of tumor sets the stage for future clinical studies to decide upon the type of treatment-path to be undertaken and whether early therapeutic interventions prolong survival and improve quality of life. Longitudinal (temporal) radiological studies of 63 patients were carried out with a median follow-up period of 33.6 months. These patients were diagnosed with grade 2 gliomas by expert physicians through manual (visual) procedures as well as detection of growth with that of the CAD method, and both detection methods were compared by 7 expert physicians [214]. Each patient had at least 4 MRI scans available for review either after the initial diagnosis or after the completion of chemotherapy with temozolomide (if applicable).



The researchers calculated the time to growth detection from the impressions of the radiological reports of these patients from 627 MRI scans. Unexpectedly, the study found large differences in growth detection by visual comparison and by physicians aided by the CAD method. The reasons for missing growth by the visual inspection can be attributed to one or more of these reasons: (1) interpreting a large number of prior studies by physicians takes a very long time, (2) the practice in vogue of comparing the current MRI to a couple of MRI scans immediately preceding it, (3) the lack of determination of baseline MRI, (4) small changes from one scan to the next, and (5) comparing single 2-D images overlooks the growth in the third dimension.

The study [214] showed that the CAD method helped physicians detect growth at earlier times and significantly smaller tumor volumes than the manual standard method. Moreover, physicians aided by the CAD method diagnosed tumor growth in 13 of 22 glioma patients labeled as clinically stable by the standard method. Fig. 3.11 shows the volume growth curves of grade 2 gliomas of two patients diagnosed with oligos, seen at the University of Alabama at Birmingham clinics between 1 July 2017 and 14 May 2018 [214]. The x-axis corresponds to the time interval from the baseline MRI, and the y-axis corresponds to the change in the volume of tumors from the baseline. The volume at each time step until the growth detected by CAD is colored in yellow, and the manual (visual) detection of change-point time is colored in red. The CAD for patient 1 detected a change-point in 20 months from the baseline, whereas visual detection by a physician was done in 80 months.

Similarly, CAD detection time for patient 2 was also around 20 months, where visual detection was in 150 months, primarily because this tumor did not grow at a faster pace. The detection of tumor volume growth in time enabled the researchers to identify tumors with nonlinear and non-homogeneous growth. Early growth detection holds the potential of lowering the morbidity, and perhaps mortality of patients with low-grade gliomas. The decision to treat a patient would be determined by the rate of growth and proximity to critical areas of the brain, once they have been measured. The study also suggested early interventions for cases where (1) the new growth is in the

proximity of key nonsurgical structures like the corpus callosum, (2) the rate of growth is elevated, or (3) the tumor is sensitive to chemotherapy.

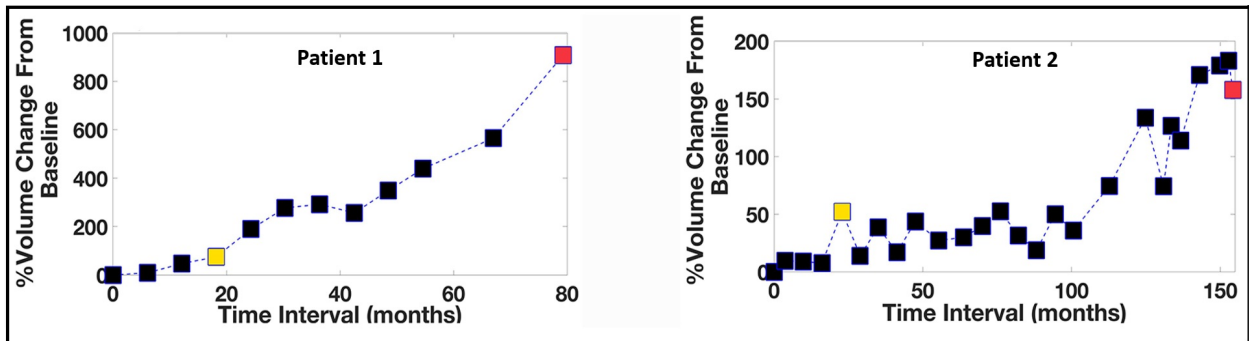


Figure 3.11 Volume growth curves of grade 2 gliomas of two patients diagnosed with oligos.

### 3.8 Conclusion

In this chapter, we have thoroughly reviewed the image segmentation task in the classical CV field and examined various techniques of CV employed in DL frameworks for brain tumor segmentation. We have also assessed multiple DL architectures having varying attributes that make them suit-to-task employment. We have looked into a case study for in-depth analysis of *U-Net* with *Inception* and *dilated Inception* modules in the context of brain tumor segmentation. A new DL framework, called exVDP, that can quantify uncertainty in the output decision of an ANN, has also been discussed. In the last section, we have discussed the concept of tumor surveillance, its rationale, techniques, and an example study on low-grade gliomas surveillance.

The brain tumor segmentation community has achieved substantial progress in the last decade because of the advances in DL. Although efforts have been made in commercializing the technology for clinicians, there is still a long way to make brain tumor segmentation a reliable and routine tool broadly applied to practical clinical decisions with minimal human interventions. This is due to the lack of existing methods in the face of adversarial examples and research-oriented frameworks that are not suited to production environments. Breakthrough is likely to come with the advent of

effective and scalable platforms by the ML community, and direction of research towards adversarial learning.

## **Chapter 4: Revolutionizing Digital Pathology with the Power of Generative Artificial Intelligence and Foundation Models**

### **4.1 Note to Reader**

This chapter has been previously published in Elsevier Laboratory Investigation as: Waqas, A., Bui, M. M., Glassy, E. F., El Naqa, I., Borkowski, P., Borkowski, A. A., & Rasool, G. (2023). Revolutionizing digital pathology with the power of generative artificial intelligence and foundation models. *Laboratory Investigation*, 100255, and has been reproduced with permission from Elsevier [733].

### **4.2 Introduction**

Conventional pathology methods have been crucial in diagnosing disease, heavily relying on examining tissue samples under a microscope. With technological advancements and a growing emphasis on precision medicine, digital pathology has emerged as a new approach for conducting precise quantitative assessments. Digital pathology involves utilizing whole slide imaging (WSI) to digitize and analyze tissue samples using a computer. Computational pathology further builds on it and incorporates artificial intelligence (AI) and machine learning to enable the extraction of information that goes beyond what the human eye can perceive. The clinical responsibilities of pathologists, such as providing precise diagnoses and quantifying biomarkers for diagnosis, prognosis, and predictions, may be strengthened in terms of precision, reproducibility, and scalability by using AI-driven analysis tools. AI can address the challenging problems in pathology workflow, including: (1) increasing workload and staff shortages leading to physician burnout, (2) growing diagnostic complexity, including ever-expanding cancer protocols and biomarkers, (3)

case variability, often involving rare diseases or overlapping morphological changes, (4) issues with the quality of slides due to artifacts introduced by tissue folding, staining inconsistencies, and compression artifacts, and (5) lack of standardization, which hinders interoperability between different laboratories, platforms, image formats, and analysis tools.

AI is a broad field focused on simulating human intelligence by creating models and algorithms to automate various tasks, such as recognizing objects in images, understanding and generating natural language text, or making predictions based on historical data [181]. Machine learning is a subset of AI that involves creating statistical and mathematical models and learning algorithms for recognizing patterns in the data [667]. Artificial neural networks that attempt to mimic the human brain's way of analyzing data have recently made significant progress [625]. The advancements made possible by artificial neural networks have revolutionized computer vision (a sub-field of AI that deals with image processing) and natural language processing (a sub-field of AI that deals with text and speech) [625]. Although the initial adoption of these technologies in medicine and healthcare was slow, recently, medical imaging has been transforming at an unprecedented rate. Digital and computational pathology are also rapidly evolving on the research front, with the industry offering new AI-enabled technologies [686, 210, 622, 533, 17].

Although task-specific traditional AI tools date back to the 1970s, the decade of 2010 saw a sharp rise in the research and development of narrow AI methods enabled by deep learning models, e.g., convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformers. These AI models eliminated the need for feature engineering using domain expertise, a defining characteristic of classical machine learning techniques widely known as pathomics in the pathology domain [265]. For a given task, the performance of these artificial neural network-based AI models surpassed previous AI techniques. Developing a task-specific AI starts with selecting a particular problem, e.g., counting mitosis in a histopathology image, then curating and annotating relevant historical data, and finally, training the model by learning optimal parameters (or weights). Annotating (or equivalently labeling) data require experts (pathologists) to carefully review each data sample and identify/define objects/patterns that help AI learn about the task during training.

The performance of task-specific AI with supervised learning techniques strongly depends on the availability of large, high-quality, annotated training datasets. Despite boasting above-human performance, task-specific AI suffers from significant limitations, including the requirement for a large amount of expert-annotated datasets, the lack of performance generalization (e.g., the AI may fail if used on images generated using a staining protocol different than the one used for generating training images), and the inability to use relevant data from other modalities, e.g., patient demographics, laboratory data, or their prior disease history cannot help the model improve its prediction accuracy [210, 686]. Analysis of task-specific AI in pathology through qualitative interviews of 24 professionals revealed such shortcomings in the existing tools, which hinder their broad integration in the decision-making processes of pathologists [195]. For further details, the reader is referred to the surveys reviewing the use of AI in pathology [373, 542].

The 2020s are witnessing the rise of foundation models and generative AI. Foundation models are very large task-agnostic AI models trained using unannotated (possibly multimodal) datasets and form the brain of generative AI [92, 493]. A trained foundation model can be adapted to perform many different tasks using a modest amount of task-specific annotated data [493]. Training a foundation model may not require manually annotating large amounts of data as these models use self-, semi-, or unsupervised learning techniques. Foundation models can consume data from various modalities, including images (e.g., WSIs), text (e.g., pathology reports), and tabular data (e.g., medical records). The well-known generative AI model, ChatGPT, is based on a foundation model called Generative Pre-training Transformer (GPT) [524, 523, 101, 568, 566, 531]. Foundation models hold much promise for quantitative image analysis, diagnosis and prognosis, pathology report generation, and question/answering with conversational use in pathology lab workflow [92, 493].

Section 4.3 provides a brief overview of AI and machine learning models and advances enabled by these task-specific AI models in computational pathology. We introduce various foundation models, their structure, characteristics, and limitation in Section 4.4. Section 4.5 outlines the transformative role that foundation models may play in the pathology laboratory workflow in the

near future. We provide use cases delineating how a pathology generative AI based on a foundation model could serve as an expert companion pathologist that assists in efficiently and objectively performing routine laboratory tasks, including image analysis, presenting and justifying findings, quantifying the analysis, generating reports, performing prognostics, and making predictions. We also outline the potential role that foundation models and generative AI can play in pathology education and training before concluding the review in Section 4.6.

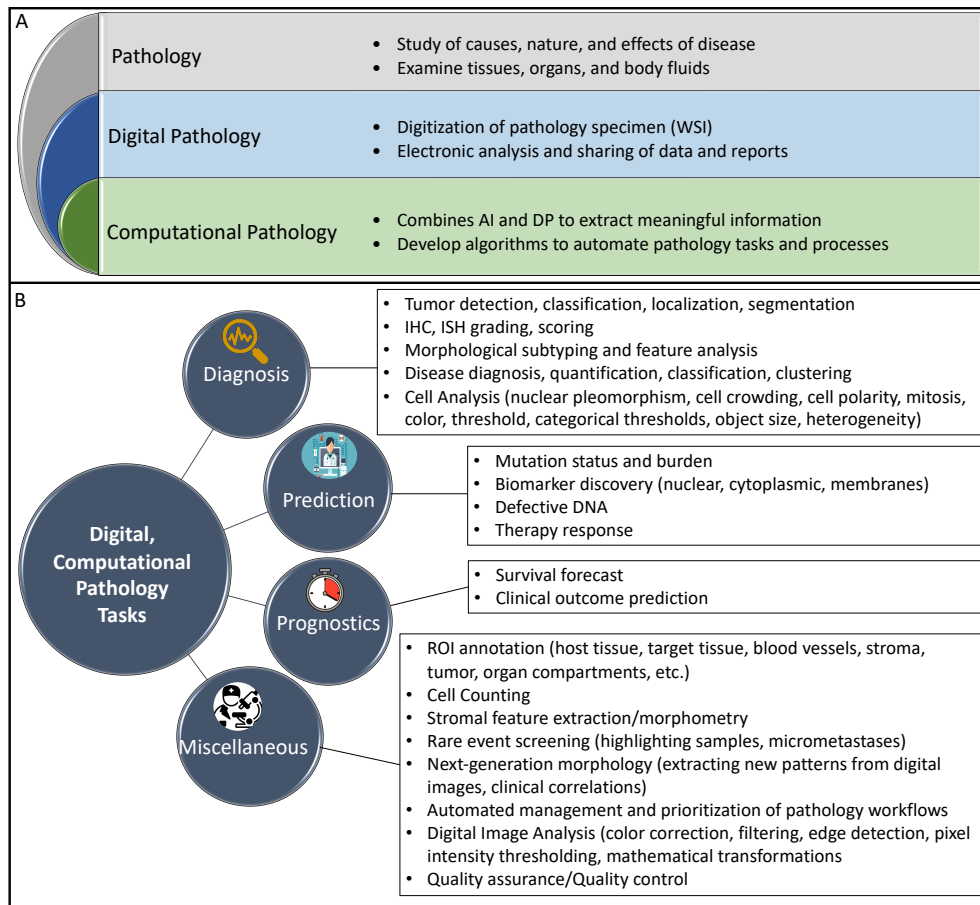


Figure 4.1 Pathology, digital pathology, and computational pathology - definitions and tasks.

### 4.3 AI in Digital Pathology

AI comprises computational methods, statistical and mathematical models, and the implementation of various algorithms to mimic human-style intelligence. AI-based technologies have enabled pathologists and researchers to analyze large amounts of data with greater accuracy and

Table 4.1 Definitions of key terminologies

Term	Definition
Digital pathology	A comprehensive term that includes various tools and systems to digitize pathology slides and associated meta-data, as well as their storage, review, analysis, and supporting infrastructure.
Computational pathology	A branch of pathology that utilizes computational techniques to analyze methods of studying disease through patient specimens. It may involve using AI methods to analyze data and extract meaningful information from digitized pathology images.
Artificial intelligence (AI)	The field of AI aims to simulate human intelligence in machines, allowing them to perform tasks such as learning, problem-solving, and decision-making.
Machine learning	It is a branch of AI that programs computers to optimize a performance criterion using sample data or past experience. It uses the theory of statistics to build learning models.
Artificial neurons	These are the fundamental building blocks of artificial neural networks. It is a mathematical function that receives one or more inputs, applies a weighted sum, adds a bias term, and applies a nonlinear activation function to the result. The output of the activation function is then passed on to the next layer of neurons.
Artificial neural network	A computational model inspired by the structure and function of biological neural networks in the brain. It is a network of interconnected artificial neurons that work together to process information and make predictions or decisions.
Neural network architecture	The architecture of a neural network refers to its structure, which is determined by the number and arrangement of its layers, the number of neurons in each layer, and the connections between the neurons.
AI training	The process of teaching an AI system to learn patterns from data and make accurate predictions or decisions. The training process involves feeding large amounts of data into the AI system and adjusting its internal parameters to optimize performance.
Supervised learning	AI training technique that uses annotated data, i.e., each data point is associated with a known target value. Goal is to learn a mapping between inputs and outputs such that trained AI can make accurate predictions on new, unlabeled data. If learning involves lesser labeled data compared to unlabeled samples, it is weakly-supervised learning.
Self-supervised learning	This technique of training AI does not require explicit data annotations. The AI learns to solve the given task using the inherent structure in the data as the supervisory signal.
Unsupervised learning	A learning technique for finding patterns, relationships, or structure in the data, such as clusters or groups of similar data points, without any knowledge of the ground truth. Unlike self-supervised learning, which uses a supervisory signal implicit in the data, unsupervised learning does not use any supervisory signal.
Computer vision	A field of AI that enables computers and systems to derive meaningful information from digital images, videos, and other visual data.
Natural language processing	An area of AI that deals with a wide range of computational methods and techniques for analyzing, understanding, and generating natural language text.
Multimodal AI	Multimodal AI refers to AI models that involve multiple data modalities, such as vision (images) and language (text), and require AI to integrate information across data modalities.
Convolutional neural networks (CNNs)	CNNs are types of artificial neural networks commonly used for image and video analysis. CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images by using multiple convolutional layers, followed by pooling layers and fully connected layers.
Recurrent neural networks (RNNs)	RNNs are specialized for processing sequential data, such as text, speech, or time series. RNNs are designed to capture context and dependencies between the elements of a sequence.
Graph neural networks (GNNs)	GNNs are neural networks that process data with a graph structure. GNNs analyze relationships between objects (nodes) and their mutual relationships (edges) by iteratively using message-passing algorithms to update the features, allowing the network to capture the relationships between nodes in the graph.
Transformers	Transformers are neural networks that use a self-attention mechanism (or equivalently scaled dot-product) to capture relationships between input elements, especially in long sequences. They can process and learn from all data types, including images, text, speech, etc.
Foundation models	Foundation models are an emerging class of AI trained on a vast quantity of unannotated data at scale resulting in a model that can be adapted to a wide range of downstream tasks with only a handful of annotated examples. They use Transformer architecture and are the workhorse of generative AI models.
Generative AI models	These are models specialized in generating new data similar to the training data, such as images or text. Examples include Bayesian networks, GANs, and foundation models such as ChatGPT, GPT-4, Stable Diffusion, and Dall-E 2.

speed, making the process of disease diagnosis faster and more precise [686, 78, 686, 26, 148]. AI has made it easier to identify patterns and biomarkers that were previously challenging to detect, leading to more personalized and targeted treatments [7].



#### 4.3.1 Digital Pathology

Digital pathology involves digitizing tissue specimens, allowing them to be analyzed and shared electronically. Digital pathology uses complex imaging systems to capture high-resolution images of tissue specimens, which can then be viewed and analyzed on a computer screen [169]. Digital pathology improves the accuracy and efficiency of pathology diagnoses by allowing pathologists to access and share images remotely, collaborate with other experts, and integrate computer-aided analysis tools. With the advent of digital pathology, the amount of data generated has increased exponentially, enabling the automation of time-consuming processes such as segmentation and mitotic counting [536]. Public data archives, such as The Cancer Genome Atlas (TCGA) [687], Clinical Proteomic Tumor Analysis Consortium (CPTAC) [206], and The Cancer Imaging Archive (TCIA) [149], host pathology image data for multiple cancer sites. This is possible only because of digital pathology and other advancements.

Whole Slide Imaging (WSI) is the technology that allows high-resolution digital images of entire microscope slides to be created and viewed on a computer screen. This process involves scanning glass slides containing tissue samples or other specimens using specialized digital scanners. WSIs can capture the entire slide at very high magnification, allowing users to zoom in and examine specific regions of interest in great detail [2]. WSIs are usually too large for contemporary computers to analyze directly, so they are tessellated into smaller tiles or patches, which serve as input for pathology AI workflows [2].

#### 4.3.2 Computational Pathology

Computational pathology combines digital pathology with AI, machine learning, and other computational techniques to extract meaningful information [2]. Often interchangeably called “histomics,” “pathomics,” or “tissue phenomics,” computational pathology aims to develop algorithms that can automatically detect and classify pathology images, predict disease outcomes, and identify new biomarkers for disease [370]. Computational pathology involves extracting many features from histopathology slides (called histomics) or pathology slides (called pathomics) and

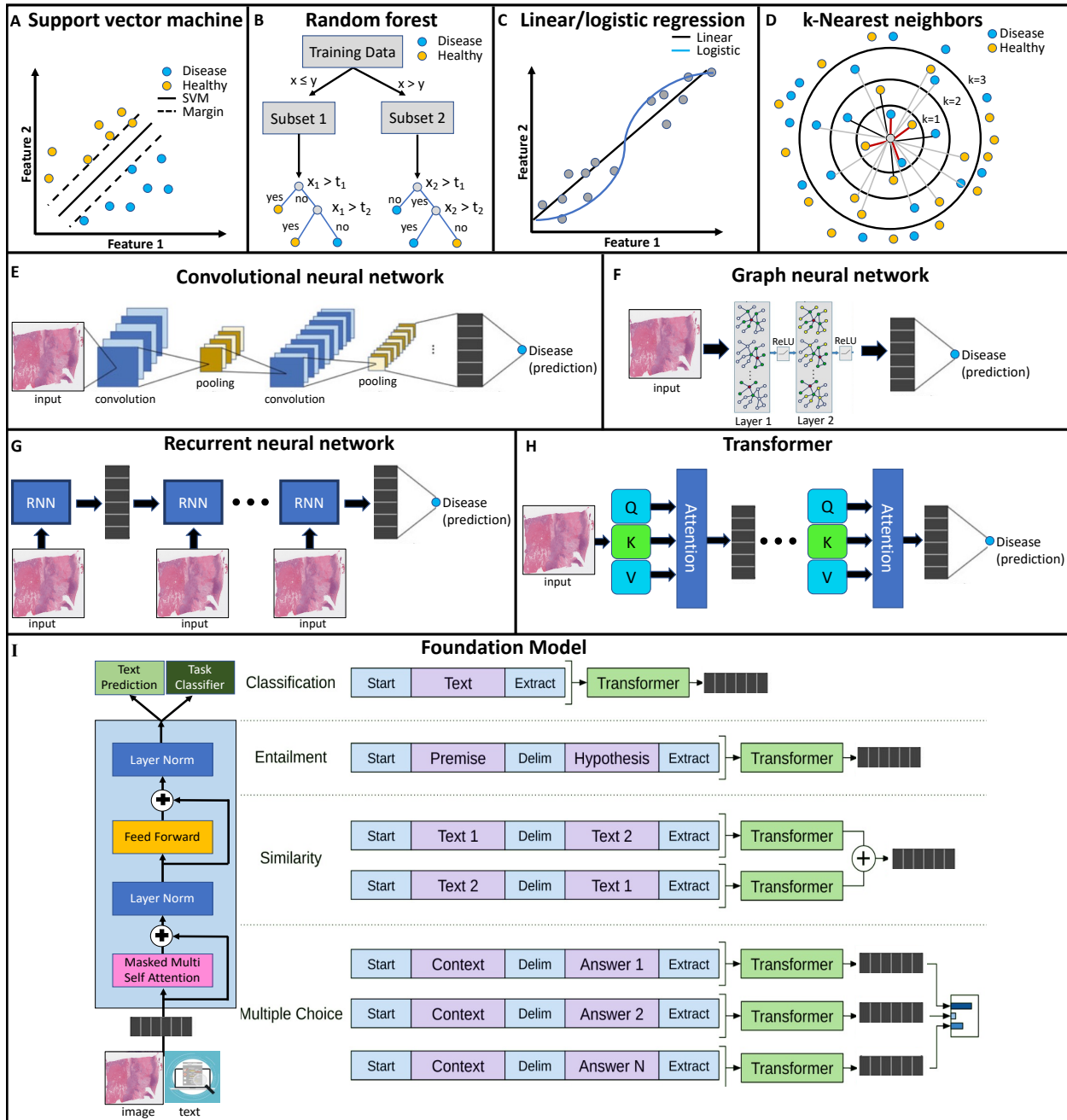


Figure 4.2 A schematic layout of various ML algorithms and AI models in digital pathology.

analyzing these features to relate to biological and clinical endpoints. Computational pathology also aims to standardize pathology diagnoses and reduce variability between pathologists [2]. Notwithstanding quality issues in digital pathology [59], computational pathology methods can perform well on tasks such as classification, segmentation, and analysis of digital pathology images, at times surpassing human-level performance [147, 200]. The definitions of pathology, digital pathology,

and computational pathology are illustrated in Fig. 4.1A, and their key tasks are illustrated in Fig. 4.1B.

#### 4.3.3 Classical Machine Learning in Digital Pathology

Classical machine learning consists of manually selecting informative features from the data by domain experts and then using these features for prediction, classification, or regression. The manual extraction and selection of features is also referred to as *feature engineering* using computer vision techniques based on morphology and texture, for instance. Classical machine learning has been extensively used in digital pathology for image segmentation and classification [667] using Support Vector Machines (SVMs), Random Forests, k-Nearest Neighbor (k-NN), Decision Trees, and others [667, 164]. A detailed review of the classical machine learning techniques in digital pathology is presented in [266] and [315]. Owing to the manual selection of usable and informative features, the applicability of classical machine learning methods is limited [266, 315].

#### 4.3.4 Task-specific AI in Digital Pathology

More recently, task-specific AI models based on artificial neural networks have been gaining popularity [734, 175, 735, 12]. Artificial neural networks use stacked layers of artificial neurons to process large amounts of data and identify underlying patterns. The model selects a set of useful and informative features based on the assigned task without any human intervention. These models include Convolutional Neural Networks (CNNs), variants of Recurrent Neural Networks (RNNs), Graph Neural Networks (GNNs), and Transformers, as illustrated in Fig. 4.2 [667]. We refer to these approaches as task-specific or narrow AI because of their limited scope. Also known as weak AI, they are incapable of general intelligence or human-like reasoning [22]. Developing a task-specific AI starts with selecting a particular task, followed by data collection and annotation. Finally, AI is supervised to learn patterns in the data by minimizing its prediction error. With the availability of digital slides and large computational power fueled by graphical processing units (GPUs; electronic circuits responsible for graphics manipulation and output) and tensor processing

units (TPUs; Google’s custom integrated circuits used to accelerate machine learning workloads), artificial neural networks-based task-specific AI models have found a strong foothold in digital pathology [686, 200, 630]. In the following discussion, we briefly introduce task-specific AI models and their essential components. In Table 4.2, we present a non-exhaustive list of various categories of task-specific AI models used in digital pathology. Interested readers are encouraged to explore the relevant works of interest.

#### 4.3.4.1 *Convolutional Neural Networks (CNNs)*

CNNs are specialized artificial neural networks for processing image data. CNNs are designed to automatically learn and extract features from images, such as lines, edges, corners, and textures, through the convolution operation. Convolution involves sliding a filter over an input image and computing dot products between the filter and the image pixels. The resulting features are used to classify or detect objects in the image. Based on the filter type, shape, size, and arrangement, various architectures of CNNs have been proposed.

#### 4.3.4.2 *Recurrent Neural Networks (RNNs)*

RNNs process sequential data like speech, text, or time series. RNNs are designed to capture temporal dependencies in the data by maintaining a hidden state that is updated at each time step. The hidden state encodes information from previous time steps and provides context for the current time step. Long Short-Term Memory networks (LSTMs) and Gated Recurrent Units (GRUs) are RNNs that help the model better capture long-term dependencies and avoid the vanishing gradient problem of RNNs using sequential data processing.

#### 4.3.4.3 *Graph Neural Networks (GNNs)*

GNNs process graph-structured data, such as social networks, molecular data, and knowledge graphs [739]. GNNs are designed to capture local and global graph structures by aggregating information from neighboring nodes and edges. GNNs typically operate on a fixed-size local

neighborhood around each node, allowing them to scale to large graphs. GNNs have shown promising results in various applications, including node classification, link prediction, and graph generation. GNNs have been used to analyze complex biological networks, drug discovery models, cell classification, tumor structures, and protein structures [8, 724, 135].

#### 4.3.4.4 *Transformers*

Transformers were initially introduced for language translation [705]. However, they have performed remarkably in various AI tasks, including computer vision and time series analysis [92, 14]. Unlike RNNs, Transformers do not require that the sequential data be processed in order. Instead, they are designed to process variable-length input sequences (such as words in a sentence) without recurrent connections. Transformers use a self-attention mechanism that allows each piece of input (or token) to attend to other tokens in the sequence, capturing long-range dependencies [705]. Transformers have achieved state-of-the-art results in various natural language processing, computer vision, and graph processing tasks [705, 14].

#### 4.3.5 AI-based Algorithms Used in Pathology

Interest in AI/ML-enabled medical devices has increased in recent years. The US Food and Drug Administration (FDA) has cleared more than 500 healthcare-related AI algorithms, four of which are for pathology [225]. Among them, two were introduced earlier, and the other two more recently. “PAPNET Testing System” was approved in 1995, and it was designed for rescreening negative Pap tests or as a primary screener [215]. “Pathwork Tissue of Origin Test” was approved in 2008, and it is a molecular diagnostic test developed to assist in diagnosing metastatic, poorly differentiated, and undifferentiated cancer [215]. “Tissue of Origin Test Kit FFPE” was approved in 2012, and it is an in vitro diagnostic to measure the degree of similarity between the RNA expression patterns in a patient’s formalin-fixed, paraffin-embedded (FFPE) tumor and the RNA expression patterns in a database of fifteen tumor types [215]. “Paige Prostate” was recently approved in 2021,

Table 4.2 Summary of AI models used in digital pathology.

DP Task	AI Model	Details	Ref
Diagnosis	CNN	MIDOG: Mitosis domain generalization challenge.	[54]
		Gleason grading and diagnosis of prostate cancer.	[104]
		Feature extraction to classify brain tumor grade.	[463]
		Mitosis detection in breast cancer.	[710]
		Segment nuclei in histology images using weakly-supervised training	[264]
	GAN	Nuclei segmentation on histopathology images.	[469]
		LSTM	4D medical image segmentation.
	GNN	Learn micro- and macro-structural features in H&E slides of breast cancer.	[42]
		Grading colorectal cancer in histology images.	[816]
		Classify healthy tissue from dysplastic gland areas in the colorectal cancer histology slides.	[661]
		Classify infiltrating ductal carcinoma (IDC) and ductal carcinoma in situ (DCIS) breast cancer and grade Gleason 3 and 4 prostate cancer.	[665]
Prediction	CNN	Disease outcome prediction in colorectal cancer.	[106]
	LSTM	Predicting sentiment, text categorization in records.	[201]
		Medical event prediction using a multi-channel fusion of EHR data.	[438]
	GNN	Stratify prostate cancer using tissue microarrays.	[724]
	Transformers	Predicting RNAseq expressions from kidney WSIs using multiple instance learning.	[26]
Predicting biomarkers from histopathology slides in colorectal cancer.		[504]	
Prognostics	CNN	Prediction of OS using Glioma multimodal data.	[97]
Misc	CNN	Correlations between true hypoxia fraction in histological and the approximated fraction in MRI scans.	[325]
	GAN	Similarity between virtually stained images (generated by AI model) & histochemically stained images.	[325]
	LSTM	Medical image denoising.	[573]
		De-identification of medical text.	[398]
	Transformers	WSI representations using unsupervised learning.	[715]
Review	CNN	Deep learning in digital pathology for breast cancer.	[307]
	GNN	GNN-based methods in cancer pathology.	[739, 15]
	Transformers	Transformers in Medical field.	[739, 299, 55, 753]

and it is a software device to assist pathologists in the detection of foci that are suspicious for cancer during the review of scanned WSI from prostate needle biopsies prepared from H&E stained FFPE tissue [215, 533]. The PaigeAI prostate algorithm and the Pathwork digital and AI platform are the pioneering algorithms that have significantly impacted pathology practices by aiding in diagnosing and characterizing various diseases.

Pathology has a branch of anatomic pathology (AP) and clinical pathology (CP). The four algorithms mentioned above exclusively pertain to AP, where the focus lies on examining tissue samples for diagnosing diseases such as cancer. It is worth mentioning that listing AI/ML-related algorithms in CP, which concentrates on the analysis of bodily fluids and laboratory tests, is beyond the purview of this specific review. Moreover, despite the noteworthy progress in pathology AI, to

the best of our knowledge, there has not yet been a generative AI algorithm developed for pathology, AP, or CP. Generative algorithms have the capability to create new data or images, potentially aiding in generating synthetic samples for training and research purposes. While such algorithms have seen success in other domains, their application in pathology, encompassing both anatomic and clinical aspects, has yet to be realized. The absence of generative AI in pathology presents a promising avenue for future research and exploration to unlock new possibilities and enhance the field's diagnostic and prognostic capabilities.

#### 4.3.6 Limitations of Task-Specific AI

Task-specific AI models have many limitations restricting their widespread use in digital pathology.

1. *Task specificity* refers to the fact that the trained AI performs well on a single task only, e.g., grading cancer sub-types in an organ using H&E slides. A change in the number of grades, organ type, or cancer type (same organ) will render the model useless (significantly reducing its accuracy with low reliability) and will require model retraining [175, 176].
2. *Distribution* of the input data should be the same. These AI models require the input data to have similar characteristics and follow the same probability distribution function of the input data (the mean and standard deviation and range of the pixel values of WSI pixels) [175, 113]. Adding natural or adversarial noise may significantly reduce AI's performance [241]. AI models are known to be fragile in the presence of noisy inputs, subtle changes in the data, or adversarial attacks [735, 175, 241]. These AI models cannot generalize to changes in data resulting from various common reasons, e.g., hardware, software, firmware upgrades in scanners, changes in the staining quality or the protocol, shifts in population demographics (e.g., a different geographical region), and changes in data patterns due to new diseases such as COVID [12, 176]. New representative data must be collected and annotated for each changing scenario to retrain (fine-tune) the AI models to be current and accurate.

3. *Requirement of large annotated task-specific datasets* for training. The success of these models depends mainly on the availability of large, task-specific annotated datasets. This requirement stems from the data-driven nature of these models, which learn to identify informative features from data without needing domain experts to engineer data features [28]. By leveraging vast amounts of annotated, independent, and identically distributed (i.i.d) data, models can uncover hidden patterns and sub-visual features that may be difficult for humans to detect. However, obtaining a large annotated dataset remains a critical challenge for AI models in digital pathology. These AI models cannot directly benefit from large amounts of unannotated datasets, e.g., WSIs, pathology reports, clinical notes, etc., and may require techniques such as weakly supervised learning, unsupervised learning, self-supervised learning, transfer learning, and continual learning [358, 16].
4. The task-specific AI models are generally restricted to processing one data modality only. Incorporating information from other modalities, e.g., the patient's medical data from medical records, omics data, or radiographs, into the AI decision-making is generally not straightforward [739]. Recently, some research efforts have focused on creating AI models that can process multimodal data to improve their predictive accuracy with moderate success [135, 90, 703].
5. *Knowledge accumulation* is important. The recent success of ChatGPT has shown that creating an internal general-purpose knowledge base is essential for successful and robust AI models [523, 101]. ChatGPT has a central repository of information created during model training using 570GB of data from books, web-based text, Wikipedia, articles, and other online writings [524, 523, 568]. There is no precedence for creating such models in digital pathology, medical imaging, or any area of medical data processing. Task-specific, narrow versions of AI models are built by individual academic labs or industries that do not contribute to reusable knowledge accumulation [739].



6. *Transparency and reproducibility* of AI models are a challenge that undermines the enormous potential of applying such methods to complex tasks. The lack of sufficient details regarding Methods and the unavailability of algorithm/code in a published work by the Google Health team on breast cancer screening [478] was recently raised [270, 477]. The research community is gradually transitioning to open-access, reproducible, and transparent methodologies.
7. *Explainability* of AI refers to the challenge of understanding how and why an AI makes a particular decision or prediction [505]. While AI can make accurate predictions or decisions, they often do so in ways that are opaque or difficult to understand for human beings. This lack of transparency can be problematic in scenarios where decisions made by AI have significant real-world consequences. Interpretable or explainable narrow AI models with attribution maps produce results humans can easily understand and interpret. However, these approaches come at the cost of reduced accuracy or increased model complexity [506].

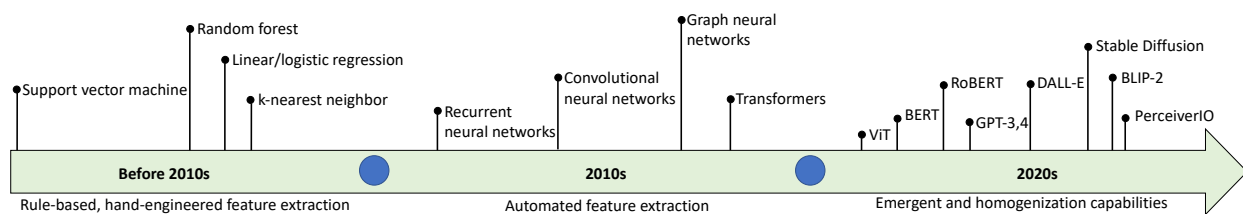


Figure 4.3 The evolution of machine learning models used in digital pathology.

#### 4.4 Foundation Models and Generative AI

The 2020s are witnessing the rise of foundation models - large AI models pre-trained using unannotated multimodal datasets (please refer to Figs. 4.2 and 4.3) [92]. A trained foundation model can be adapted (or fine-tuned) to perform different tasks using limited annotated examples, much less than required to train tasks-specific AI. In the following, we present our perspective on

how foundation models and generative AI that use these models can transform the digital pathology laboratory workflow. The pathology-specific foundation models can be created and fine-tuned to serve as a pathologist's expert assistant by performing quantitative image analysis for diagnosis, prognosis, disease grading, and prediction. It can then generate pathology reports based on the presented imaging data and converse with the pathologist to justify the findings presented in the generated reports.

#### 4.4.1 Foundation Models

The term "foundation models" was initially coined by Bommasani *et al.* to describe recently proposed models that have led to a paradigm shift in AI model design, development, and deployment processes [92]. Foundation models are huge models trained at scale using comprehensive unannotated data (possibly multimodal). Two relevant attributes of foundation models are their size, which refers to the number of learnable parameters or weights, and the number of compute operations quantified using Floating Point Operations (FLOPs) at the model testing or inference stage. Foundation models generally have billions or trillions of learnable parameters and billions of FLOPs [523, 101, 570, 617, 236, 179, 442]. The unannotated datasets may consist of billions of words (or tokens) and images from the internet without any labels assigned by human operators [236]. Foundation models leverage the existing concepts of pre-training, transfer learning, and unsupervised and self-supervised learning. However, their essence lies in *scaling* because of the following three factors: (i) the introduction of Transformer architecture [705] that supports training models with the number of learnable parameters in billions or trillions, (ii) the availability of thousands of GPUs, and (iii) availability of massive training datasets that can reach billions of tokens for natural language processing and hundreds of millions of images for computer vision tasks [92, 705].

Recently, a host of foundation models have been trained for language, vision, and joint language-vision (multimodal) tasks and shared via GitHub<sup>1</sup> and Hugging Face<sup>2</sup>. Some of the remarkable

---

<sup>1</sup><https://github.com/trending>

<sup>2</sup><https://huggingface.co>

works include BERT and RoBERTa in language processing [179, 442], Vision Transformers for image processing tasks [193], Mask2Former, OneFormer, and ClipSeg for image segmentation [140, 323, 457], Perceiver IO for multimodal (text, images, audio, and video) problems [319], ViperGPT for answering visual queries using code generation [666], LLaVA for visual instruction tuning [434], and BLIP-2 for image captioning, visual question-answering, and chat-based prompting [409].

#### 4.4.2 Characteristics of Foundation Models

Some distinguishing characteristics of foundation models are summarized as follows:

- *Expressivity* is the ability of foundation models to learn, capture, and represent the relevant information from data [92]. Foundation models are more expressive than their task-specific AI models as they exclusively use the Transformers architecture, which learns long-range relationships and higher-order interactions in the data using a self-attention mechanism [705]. There exists a trade-off between the model's expressivity and its efficiency. Increasing the model size may increase its expressivity at the cost of reduced efficiency [92]. Recently proposed foundation models such as Perceiver IO and GANformer attempt to offer a balance between efficiency and expressivity [319, 305].
- *Scalability* refers to the ability of a foundation model to efficiently consume large amounts of data [92]. With the ever-growing availability of data from diverse sources, the foundation model needs to be capable of further scaling while overcoming the challenges of failure and catastrophic forgetting [12, 358].
- *Multimodality* is the ability of the foundation model to learn relations among various modalities of the data [92]. Humans perceive knowledge through processing multimodal data. GPT-4 is a multimodal foundation model [524]. Other multimodal models include CLIP [565], ALIGN [328], SimVLM [732], Flamingo [21], CoCa [781], and CONCH [452].

- *Compositionality* is the ability of a foundation model to generalize to new tasks and contexts [92]. Compositionality helps foundation models achieve out-of-distribution generalization and perform in-context learning [92, 402].
- *Emergence* is the characteristic introduced by scaling the Transformer architecture with large datasets and computational resources [705, 92]. Emergence means that the behavior of the trained AI model is implicitly induced rather than explicitly constructed [92, 402]. In-context learning is an example of emergence in foundation models [402, 744].
- *Homogenization* is also introduced by scaling and refers to the consolidation of methodologies for building AI models across a wide range of applications [92]. For example, almost all language processing tasks can be performed by a single large language model, e.g., BERT [179], GPT [524, 101, 568, 566], T5 [570], or many others [811].
- *Transfer learning, adaptation, and fine-tuning* are the defining characteristics of foundation models [92, 749, 744]. These characteristics imply that the skills that AI may learn from one task will often transfer to new tasks. A foundation model may adapt to the new tasks without the need for any annotated examples, referred to as zero-shot learning. When a few examples are used to fine-tune the AI, we call this few-shot learning [92]. Generally, all foundational models are pre-trained using unannotated datasets and later adapted using small annotated datasets for specific downstream tasks. A recent survey reviews the various pre-training methods used in deep learning and foundation models on medical data [561].
- *In-context learning* is the ability of a trained foundation model to learn a new task or correct itself using demonstration and without updating the model's parameters which is usually done via gradient descent algorithm [101, 744, 437]. In-context learning is a scale-enabled emergent ability that allows foundation models to generalize to new tasks without having to re-train the AI model again. GPT-2, a relatively small model having 1.5 billion parameters, did not permit in-context learning [568]. It was GPT-3 with its 175 billion parameters that exhibited in-context learning [101]. However, in-context learning introduces the necessity

for *prompt engineering*, i.e., finding the most appropriate prompt to allow AI to solve the task at hand [437, 744]. A prompt is a piece of text, image, or symbols inserted in the input of AI so that the given task can be re-formulated as the original task for which the model was trained [101, 402, 744].

#### 4.4.3 Types of Foundation Models

Foundation models are an emerging area of AI that has shown great promise, e.g., ChatGPT, GPT-4, DALL-E 2, and Stable Diffusion are foundation models that can generate impressive text and images, provide concise summaries of large datasets, and help analyze unstructured data efficiently [523, 524, 590, 575]. These models can be further divided into large language models that tackle natural language processing tasks and vision-language models that handle multimodal learning jointly from images, text, and other data sources.

##### 4.4.3.1 Large Language AI Models

Large language models can handle various natural language processing tasks, including text generation, natural language understanding, sentiment analysis, question answering, information retrieval, reading comprehension, commonsense reasoning, natural language inferences, word sense disambiguation, and others [92]. With the introduction of *word embeddings*, where each word in a sentence was associated with a context-independent vector of real numbers [698], the natural language processing field has seen considerable progress [179]. Following the success of word embeddings, autoregressive language models were proposed to employ self-supervised or weakly-supervised learning to predict the next word in a sentence given the previous words [179]. Autoregressive models such as GPT, ELMo, and ULMFiT use the context of the words in representation embeddings [566, 555, 294]. The Transformer architecture enabled self-supervised learning at scale resulting in models like BERT, GPT, GPT-2, GPT-3, GPT-4, LLaMA, T5, and BART [179, 566, 568, 101, 524, 442, 570, 690, 405]. Most of these industry-sponsored models are not open-source for researchers [524, 523]. Recently, BLOOM, a 176B-parameter open-access

language model, was developed with the collaboration of hundreds of researchers [617]. BLOOM is a decoder-only Transformer language model trained on the ROOTS corpus, a dataset comprising hundreds of sources in 46 natural and 13 programming languages (59 in total) [617]. A comprehensive review of large language models is out of the scope of this article. For a comprehensive review of the large language models, please refer to review papers and blogs [811].

#### 4.4.3.2 *Vision-Language AI Models*

Vision-language AI can learn to perform various tasks involving images (or videos) and corresponding natural language text [236, 811]. The vision-language models are one step closer to how humans perceive the world, learn about it, and execute various tasks in it [236, 448]. In the following, we describe two types of imaging analysis tasks that vision-language models can perform:

- These tasks reside at the intersection of natural language processing and computer vision fields and consist of extracting information from images and natural language text and finding the relationships and patterns to link text and images [236, 184]. Image captioning, visual question answering, visual dialog, image or text retrieval given text or image, visual grounding, and image generation are a few image-text tasks undertaken by these foundation AI models [236, 304]. Visual question-answering tasks typically require a more detailed understanding of the image and complex reasoning than a system producing image captions [304]. The recent foundation models in image-text tasks include Contrastive Language-Image Pre-Training (CLIP), A Large-scale Image and Noisy-Text Embedding (ALIGN), SimVLM, Florence, Flamingo, CoCa, Clinical-BERT, and Contrastive learning from Captions for Histopathology (CONCH) [565, 328, 781, 732, 21, 766, 452].
- Image classification, object detection, and segmentation are the core visual recognition tasks in the field of computer vision. Traditionally, these tasks were considered pure vision problems without needing to include language information while learning these tasks. However,

CLIP and ALIGN models showed that language supervision could play an essential role in pre-training vision-language that can do various visual recognition tasks with zero-shot learning [565, 328]. CLIP and ALIGN use noisy image-text data from the internet to enable large-scale pre-training of vision encoders. The state-of-the-art foundation models include: (1) image classification - UniCL, CLIP, and ALIGN [767, 565, 328], (2) object detection in a given image - ViLD, RegionCLIP, GLIP, Detic, PromptDet, OWL-ViT, OV-DETR, and X-DERT [262, 809, 411, 815, 487, 787, 108], and (3) segmentation of different objects in a given image - LSeg, OpenSeg, CLIPSeg, MaskCLIP, DenseCLIP, and GroupViT [406, 242, 457, 812, 578, 760].

#### 4.4.4 Training Foundation Models and Generative AI

Foundation models employ two key techniques in training: self-supervised learning and generative training. The true potential of the enormous quantity of unannotated data is only possible with supervised learning, without the need to create annotations or labels using human effort. Examples of such data include (1) text, images, and videos available online or (2) medical records, diagnostic imaging, molecular data, and histopathology WSIs available in hospital databases. During the training of foundation models, the supervision signal is determined by the context of the input data, e.g., the BERT language model is trained to predict randomly removed words from sentences or fill in the blank [179]. Sometimes, the models are shown plausible and implausible pairs of images and corresponding texts. Thus, the model learns to associate image features with their correct text description [565]. This perspective generalizes the traditional close-set classification AI models to recognize unseen concepts in real-world applications, such as open-vocabulary object detection [236].

The generative training methods help foundation models learn the joint or conditional probability distributions over training input data [92]. That is, the trained foundation model will be able to accurately generate the input data pattern similar to the ones used for training it. Generative training is performed using one of two techniques, (1) de-noising or (2) auto-regressive. During the training

of the de-noising models, the input is corrupted with noise, and the model is expected to produce noise-free input patterns [570]. The auto-regressive models, after training, can generate the input data piece by piece, iteratively predicting the next element in a sequence given the previous elements [576].

#### 4.4.5 Challenges and Limitations of Foundation Models

Developing foundation models require massive datasets, computational resources, and technical expertise [92]. Owing to their massive size, it may not be possible to fit the parameters of a foundation model in the memory of the largest GPU or a single computer. For example, a recent large language model shared by Meta AI, LLaMA, has 65 Billion parameters and was trained using 1.4 trillion tokens [690]. The enormous computational operations inside foundation models can result in unrealistically long training and inference times. Foundation models require specialized software, hardware, and inference algorithms to train and use [646].

“Hallucination” is a known limitation of generative AI, which refers to mistakes in the generated text or images that are semantically, syntactically, or visually plausible but are, in fact, incorrect, nonsensical, and do not refer to any real-world concepts [397, 25, 524]. The accuracy and integrity of the generated text and images may be challenging to establish using factual data from verified sources [25]. One possible solution is to use an engineered system like *Bing Chat* that also generates links to the actual websites, articles, and reference material<sup>1</sup>. In some cases, the generative AI models can identify their own mistakes [397]. Furthermore, the generative models are sensitive to the form and choice of words, referred to as the “prompt.” A prompt may consist of text, image(s), or symbol(s) inserted in the input of generative AI so that the given task can be re-formulated as the original task for which the model was trained [437, 397]. The future generative AI models may be less sensitive to the precise prompt. However, the current models need “prompt engineering” to produce the best results [437, 397]. Therefore, effectively using a generative AI may require engineering an appropriate prompt by the human user. Foundation models and generative AI also

---

<sup>1</sup><https://www.bing.com/>



face other challenges similar to tasks-specific AI models, including explainability, robustness, and trustworthiness [505, 337, 92, 113, 114, 175, 241, 12, 25].

#### **4.5 Transformers, Foundation Models, and Digital Pathology**

This section presents recent work from the literature focused on using Transformers (the core component of foundation models) in digital pathology. We focus on the work where a single AI model based on Transformer architecture is trained using large, diverse datasets to perform multiple tasks. Later, we present our perspective on the potentially transformative role of foundation model-based AI in digital pathology. Because of foundation models' strong adaptation and scalability properties, they can be effectively trained once and modified infinite times to suit various digital pathology tasks. We also present our perspective on the trustworthiness and acceptability of generative AI and foundation models by pathologists. Figure 4.4 presents a prospective framework for utilizing foundation models and generative AI for various pathology tasks.

Transformers architecture has recently been modified to consume high-resolution gigapixel WSI data [133]. The authors used a self-supervised hierarchical learning mechanism on 33 cancer site data having approximately 105 million pathology images to predict nine slide-level tasks, including cancer subtyping, survival, and unique morphological phenotypes [133]. Although molecular procedures and analysis have led to remarkable discoveries, they are usually time-consuming, expensive, and require multiple tissue samples. Transformer-based foundation models can address these challenges by predicting the bulk RNA-seq directly from the whole slide images [26]. Transformers-based foundation model, CONCH, has shown state-of-the-art performance on multiple tasks including histology image classification, segmentation, captioning, text-to-image and image-to-text retrieval using task-agnostic pretraining on 1.17 million image-caption pairs [452]. Similarly, attention-based multiple-instance learning has accurately predicted biomarkers from cancer pathology slides in a self-supervised learning setting [504]. The authors showed the performance of an attention-based multiple-instance learning framework for predicting microsatellite instability and mutations in BRAF, KRAS, NRAS, and PIK3CA in colorectal cancer pathology

slides [504]. To address the interpretability challenge of the AI model's decisions, a probabilistic perspective on attention-based multiple-instance learning on WSI data has outperformed previous methods in matching the pathologists' annotations [165]. Such pre-foundation models can be scaled to predict biomarkers directly from the histopathology slides belonging to pan-cancer sites [634]. The Transformer model pre-trained on a large publicly available pathology dataset can be fine-tuned under a weakly-supervised contrastive learning scheme on smaller datasets. Wang *et al.* have shown that such a training framework can outperform the state-of-the-art WSI classification on three different tasks [729]. For the multimodal medical data analysis, modality co-attention Transformers have been shown to outperform other methods in survival predictions by fused learning on WSI data and genomic sequences [136]. Moreover, Transformers are far more robust to adversarial attacks and perturbations in digital pathology than CNNs because of the more robust latent representation of clinically relevant information [241]. The performance and robustness of Transformers-based models in various tasks and modality settings have shown the prospective utilization of a single foundation model for large-scale rollout involving multiple tasks.

Given the strong support for compositionality and multimodality and the modular nature of the foundation models, image and language models can be combined to share their learned representations as a larger foundation model. Thus, a Transformer trained to interpret WSIs can be combined with a trained language generation model (e.g., GPT) to create a vision-language model. Such a model will interpret and analyze WSIs and generate text reports based on the analysis. The same model can be augmented to annotate relevant areas on the input image to support its finding in the generated report. Finally, a conversational component can be added to allow the model to interact with the pathologist to answer their question about the model's output.

The authors believe that a multimodal pathology foundation model capable of processing WSIs and natural language can be created using data available in the public domain, such as the National Cancer Institute's The Cancer Genome Atlas (TCGA) for genetic data, Clinical Proteomic Tumor Analysis Consortium (CPTAC) for proteomics data, and The Cancer Imaging Archive (TCIA) for imaging data [687, 206, 149]. The base model can be trained with pan-cancer datasets and later

fine-tuned for various organs, cancer types, and use cases with only a few task-specific annotated examples. The base pathology foundation model can be shared with the community, eliminating the need to collect data, annotate, and train AI models from scratch for each use case. A recent synopsis explores AI techniques for multimodal data fusion and disease association discovery in oncology data [430]. Quantifying patterns across 17,355 H&E stained slides from 28 cancer types through deep learning accurately classified cancer types and correlated learned features with numerous recurrent genetic aberrations across considered cancer types [230]. In the following, we build on the idea of training and sharing a base pathology foundation model that can be adapted for research, clinical, laboratory, and educational use cases in digital pathology.

#### 4.5.1 Qualitative Image Analysis

A trained foundation AI model can be adapted for various pathology image analysis tasks. The adaptation may not require any annotated data (zero-shot learning) or may require only a handful of samples (few-shot learning). Examples include (1) separating the different types of cells in an image and identifying the regions of interest, (2) identifying and counting the number of cells in a given image, (3) categorizing cells into different types based on their appearance and features, (4) identifying the presence and extent of cancerous tissue in an image, (5) assessing the severity and extent of a disease by grading and staging tissue samples, (6) measuring the number of specific proteins or molecules in a tissue sample to determine their potential as biomarkers for disease, (7) predicting the likelihood of disease progression or patient outcome based on the analysis of tissue samples, or (8) immunohistochemistry scoring.

Apart from adapting the base pathology foundation model to various imaging tasks, we can use the same model for analyzing different types of stains, images from different scanners, and noisy slides containing different artifacts. Foundation models can leverage the multi-site cytology data (cervix, kidney, breast, lung, thyroid, bladder, bone marrow, skin, etc.) to perform various downstream tasks such as malignant cells classification, slide-level stratification, cells location in cytological smears, and cell components identification [331].

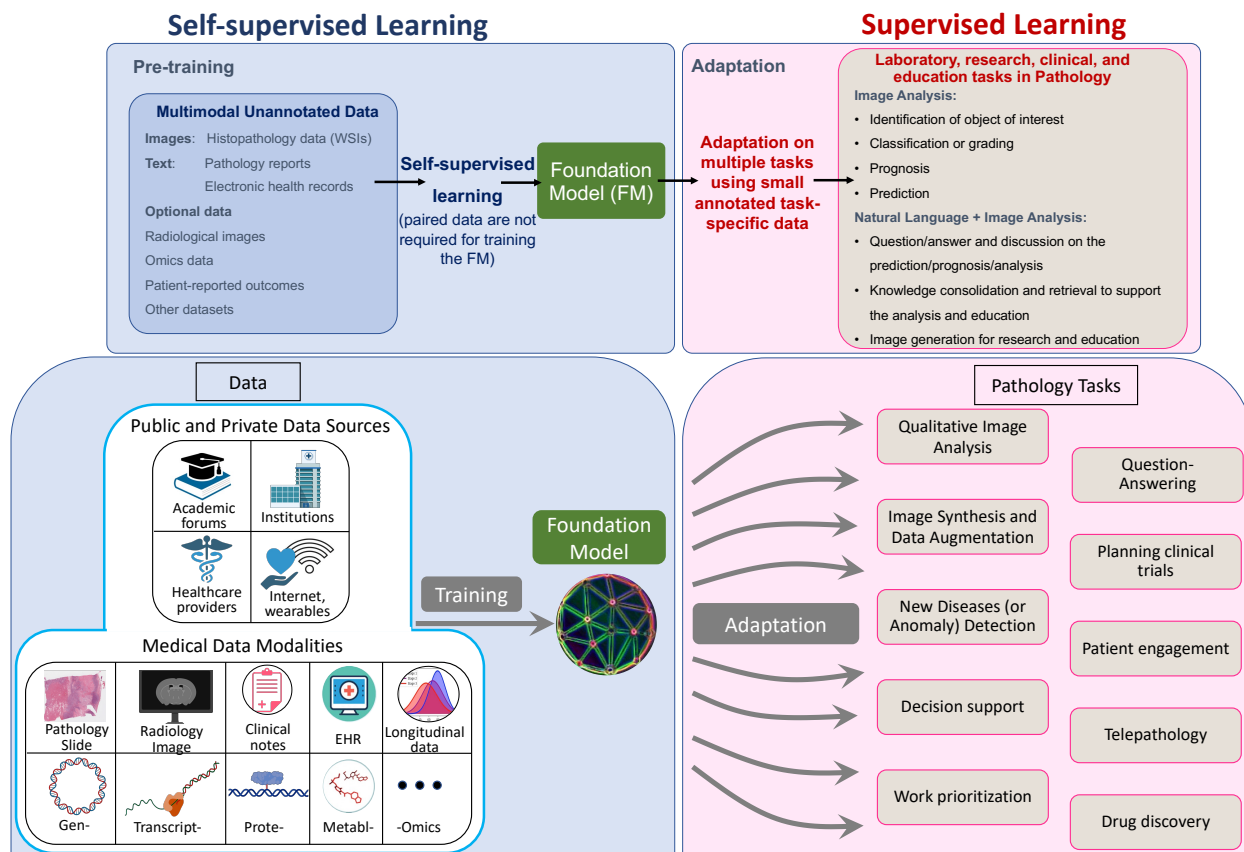


Figure 4.4 A prospective schematic layout of using foundation models and generative AI.

Going one step further, the image analysis performed by the AI can be internally fed to the generative AI, allowing pathology report generation directly from the image [461, 644]. Some sample pathology reports generated using ChatGPT (March 14 update) are presented in Fig. 4.5A, B, and C. These reports were generated by text prompt only without providing any image to ChatGPT as it cannot process image data.

AI supported by large models can reduce pathologists' workload and inter-rater variability while improving the quality and consistency of pathology reports [634]. The image analysis and report generation pipeline can serve as the "first pair of eyes" and potentially help pathology labs with workload and staffing shortages. AI model adaptation and fine-tuning allow it to learn from its mistakes and update itself under the guidance of the pathologist(s). Thus promoting the AI from

just a data processing pipeline to an assistant who will, over time, learn to help the users perform their tasks efficiently.

#### 4.5.2 Image Synthesis, De-noising, and Virtual Staining

Publicly available Generative AI has yet to show plausible pathology image generation capabilities. It has been recently shown that, despite being state-of-the-art at the time of assessment, the text-guided diffusion model (GUIDE) lacked a good depiction of the style and contents of medical images [348]. However, we argue that there are enough pathology image data in the public domain to train pathology image generation models using GUIDE, Stable Diffusion, or Dall-E 2, as the starting point. A well-trained pathology image generation AI can address various research and clinical challenges including (1) de-noising digitized slides to remove noise and artifacts and normalize the image to a standard color and tone, effectively making the task of image analysis pipeline easy and less prone to error (2) virtual staining - generating images with different staining techniques without requiring additional physical samples, helping pathologists compare and contrast the effects of various stains and facilitate more accurate diagnoses [59], (3) super-resolution imaging - using AI synthesis techniques to generate super-resolution images from low-quality and noisy digital slides, aiding pathologists in examining fine details and structures that may not be visible in the original images due to noisy or erroneous digitization process [210], (4) simulating disease progression - generating images simulating the progression or regression of pathological conditions, thus providing pathologists with a better understanding of disease evolution and enabling more informed treatment planning, (5) education and training - create diverse and realistic examples for educational purposes, thus help trainee pathologists gain experience in diagnosing a wide range of conditions and improve their diagnostic skills without relying on actual patient samples [279], and (6) synthesizing images to study the effects of various factors on disease presentation, such as genetic mutations, environmental factors, or treatment options, thus contribute to a better understanding of disease mechanisms and the development of more effective therapies.

<p>Patient Name: John Doe Date of Birth: 01/01/1960 Date of Procedure: 03/21/2023 Procedure: Craniotomy for Resection of Brain Tumor</p> <p><b>Brain Surgery Report A</b></p> <p><b>Clinical History:</b> The patient presented with a several-month history of progressive headaches, nausea, and dizziness. Imaging studies revealed a 4 cm mass in the left frontal lobe of the brain, which was suspicious for a malignant tumor. The patient underwent a craniotomy for resection of the mass.</p> <p><b>Gross Description:</b> Received in formalin is a 4 x 3.5 x 3.0 cm piece of tan-pink soft tissue. The specimen is labeled with a suture indicating the superior aspect of the tumor. The tumor is well-circumscribed and has a soft, friable consistency. Multiple sections are submitted for histologic evaluation.</p> <p><b>Microscopic Description:</b> The histologic sections show a high-grade astrocytoma, consistent with a grade IV glioblastoma. The tumor is characterized by pleomorphic, hyperchromatic nuclei with frequent mitotic figures and areas of necrosis. The tumor cells infiltrate the adjacent brain parenchyma and extend into the subarachnoid space. There is no evidence of vascular invasion.</p> <p><b>Immunohistochemical Stains:</b> Immunohistochemical staining shows positive expression of glial fibrillary acidic protein (GFAP) and negative expression of IDH1 R132H mutation.</p> <p><b>Final Diagnosis:</b></p> <ul style="list-style-type: none"> <li>• Left frontal lobe brain tumor: Grade IV glioblastoma (high-grade astrocytoma).</li> </ul> <p><b>Comments:</b> The findings are consistent with a diagnosis of grade IV glioblastoma, which is a high-grade astrocytoma. The tumor is well-circumscribed and has a soft, friable consistency. Multiple sections were submitted for histologic evaluation, and there is no evidence of vascular invasion. Further treatment planning, including radiation and chemotherapy, is recommended.</p> <p><b>Pathologist:</b> Dr. Sarah Lee MD, Board Certified in Anatomic and Clinical Pathology</p>	<p>Patient Name: Jane Smith Date of Birth: 05/10/1975 Date of Procedure: 03/15/2023 Procedure: Excision of Left Breast Mass</p> <p><b>Breast Biopsy Report B</b></p> <p><b>Clinical History:</b> The patient presented with a palpable mass in her left breast. Imaging studies revealed a suspicious mass, and a biopsy was performed, which showed invasive ductal carcinoma. The patient underwent a left breast lumpectomy for definitive treatment.</p> <p><b>Gross Description:</b> Received in formalin is a left breast lumpectomy specimen measuring 3.5 x 2.5 x 2.0 cm. The specimen is oriented with sutures indicating the superior, inferior, medial, and lateral margins. The inked margins are negative for tumor. The mass is well-circumscribed and measures 1.5 cm in greatest dimension. The tumor is firm, white, and tan.</p> <p><b>Microscopic Description:</b> The lumpectomy specimen shows a well-circumscribed, invasive ductal carcinoma measuring 1.5 cm in greatest dimension. The tumor is predominantly composed of small, uniform cells arranged in glandular and solid patterns. There is no evidence of lymphovascular invasion. The surgical margins are free of tumor.</p> <p><b>Immunohistochemical Stains:</b> Immunohistochemical staining is positive for estrogen receptor (ER), progesterone receptor (PR), and HER2/neu.</p> <p><b>Final Diagnosis:</b></p> <ul style="list-style-type: none"> <li>• Left breast lumpectomy: Invasive ductal carcinoma, well-circumscribed, measuring 1.5 cm in greatest dimension.</li> <li>• Tumor markers: Positive for ER, PR, and HER2/neu.</li> </ul> <p><b>Comments:</b> The findings are consistent with a diagnosis of invasive ductal carcinoma of the left breast. The tumor is well-circumscribed and measures 1.5 cm in greatest dimension. The surgical margins are free of tumor. Further staging workup and treatment planning are recommended.</p> <p><b>Pathologist:</b> Dr. John Smith MD, Board Certified in Anatomic and Clinical Pathology</p>
<p>Patient Name: John Doe Date of Birth: 01/01/1960 Date of Procedure: 03/01/2023 Procedure: Pancreatic Biopsy</p> <p><b>Pancreatic Biopsy Report C</b></p> <p><b>Clinical History:</b> The patient presented with persistent abdominal pain, loss of appetite, and weight loss. Imaging studies revealed a pancreatic mass suspicious for malignancy. A biopsy was performed for further evaluation.</p> <p><b>Gross Description:</b> Received in formalin is a core biopsy measuring 1.2 cm in length and 0.3 cm in diameter. The specimen is tan-white and firm.</p> <p><b>Microscopic Description:</b> The biopsy shows infiltration of pancreatic tissue by poorly differentiated adenocarcinoma. The tumor cells are arranged in glandular and solid patterns and have large, irregular nuclei with prominent nucleoli. Mitotic activity is high. No normal pancreatic tissue is identified in the biopsy specimen.</p> <p><b>Immunohistochemical Stains:</b> Immunohistochemical staining is positive for CK7 and CA19-9 and negative for CK20, CDX2, and TTF-1.</p> <p><b>Final Diagnosis:</b></p> <ul style="list-style-type: none"> <li>• Pancreatic biopsy: Poorly differentiated adenocarcinoma</li> <li>• Tumor markers: Positive for CK7 and CA19-9, negative for CK20, CDX2, and TTF-1.</li> </ul> <p><b>Comments:</b> The findings are consistent with a diagnosis of poorly differentiated adenocarcinoma of the pancreas. Further staging workup and treatment planning are recommended.</p>	<p>Biopsy Site: Pancreas <b>Pancreatic Biopsy Description D</b></p> <p>The biopsy was taken from the pancreas, which is an organ in the abdomen involved in digestion and producing hormones like insulin.</p> <p><b>Clinical Information:</b> The patient came to the doctor with belly pain and yellowing of the skin and eyes (jaundice). Imaging tests found a mass (abnormal growth) in the pancreas. A biopsy (taking a small tissue sample) was performed for further examination.</p> <p><b>Macroscopic Description:</b> The biopsy sample was preserved in a chemical called formalin. It's a small, firm, tan-white piece of tissue about 1.5 cm long and 0.5 cm wide.</p> <p><b>Microscopic Description:</b> Looking at the tissue under a microscope, it contains normal pancreas components, including acinar cells (which produce digestive enzymes), ductal cells (forming ducts that transport the enzymes), and fibrous stroma (connective tissue). However, there's also a well-developed cancer called adenocarcinoma invading the nearby pancreatic tissue. The cancer cells form gland-like structures and can be seen moving into the neighboring ducts. The cancer cells are also found invading nerves (perineural invasion). The tumor is moderately differentiated, meaning it somewhat resembles normal tissue. There are 10 cell divisions (mitoses) seen in each square millimeter of the tumor, which helps determine how fast the tumor is growing.</p> <p><b>Immunohistochemical Analysis:</b> Special tests called immunohistochemistry were performed, showing the cancer cells produce specific proteins (cytokeratin 19 and carcinoembryonic antigen), confirming the diagnosis of pancreatic adenocarcinoma.</p> <p><b>Diagnosis:</b></p> <ol style="list-style-type: none"> <li>1. Pancreatic adenocarcinoma, moderately differentiated: This is a type of pancreatic cancer that is moderately similar to normal tissue.</li> <li>2. Mitotic count of 10 per square millimeter: This is a measure of how fast the tumor cells are dividing and growing.</li> <li>3. Perineural invasion present: Cancer cells are invading the nerves, which is an important finding that may affect treatment and prognosis.</li> </ol>

Figure 4.5 Surgery and biopsy reports generated by ChatGPT.

### 4.5.3 Detecting Zebras (New Disease Identification)

Foundation models can be adapted to identify deviations from the norm, which may indicate potential anomalies, such as abnormal cell structures, lesions, or other abnormalities. Anomaly detection of finding zebras goes beyond regular tasks of identifying disease sub-type or grading. This use case aims to identify and report patterns never seen in the training data to improve the accuracy and efficiency of identifying unusual or unexpected events. Transformer-based models can learn to directly predict the bulk RNA-seq from WSI and simultaneously output the WSI representation [26]. Such models can augment pathologists' expertise and provide more accurate and timely diagnoses.

### 4.5.4 Patient Engagement

Generative AI can help pathologists, who are the “doctor’s doctor,” engage directly with the patients by bringing them to the front line without additional time or resource commitment. Language models can generate more approachable and accurate descriptions and explanations of the pathologist’s findings for the patients. Image generation models can create annotated images to depict the disease visually. In Fig. 4.5D, we present the description of a biopsy report generated by GPT-4. The text is aimed to explain the pathology biopsy report (presented in Fig. 4.5C) to a non-medical person. In addition, they can educate the patient about the disease entity just diagnosed by the pathologist and possible treatment options.

### 4.5.5 Education and Training

Pathology education is currently powered and driven by virtual and digital transformations and is swiftly adapting to the advancements offered by AI [279]. Generative AI can retrieve and integrate knowledge from various sources, such as textbooks, and scientific articles, providing a comprehensive view of the state of knowledge. Pathology-focused ChatGPT-like models can answer pedagogical questions quickly, such as the definition of terms or recent advancements reported in the literature.

Conversational AI, such as ChatGPT and its variants, can solve higher-order reasoning questions. ChatGPT has the comparative relational level of accuracy in pathology, as noted by the responses shown in Figure 4.5. Hence, students and academicians have the opportunity to adapt to this emerging technology and use it for solving reasoning-type questions. Further evolution of such conversational tools needs to be critically analyzed by the specialists, such as pathologists, for their efficacy and acceptability.

#### 4.5.6 AI-Driven Standardization in Digital Pathology Workflow

Foundation models and generative AI can help standardize digital pathology by addressing various aspects of the diagnostic process, such as image acquisition, analysis, interpretation, and reporting.

1. AI models can correct for inconsistencies in image acquisition, such as variations in lighting, staining, and scanning parameters. By automatically adjusting for these factors, AI can ensure that images are more consistent and comparable across laboratories and scanners.
2. AI-based tools can extract and quantify relevant features in images in a standardized and reproducible manner. This can include cell counting, morphological measurements, and biomarker quantification, reducing the variability that may arise from manual or semi-automated methods. A human operator will need to approve AI-generated features.
3. AI-driven algorithms can provide a second reader opinion or decision support for pathologists, reducing diagnostic variability and errors. By learning from large datasets and incorporating best practices, AI can help standardize the diagnostic process and improve the overall quality of diagnoses.
4. AI can help identify inconsistencies in staining techniques, equipment, and reporting protocols, enabling better standardization and quality assurance across laboratories. By monitoring and benchmarking these factors, AI can improve the overall quality of digital pathology services.



5. AI models can process and summarize the patient visits and interventions spread over multiple time points in the form of patient timeline and EMR summary of care. These models may also generate synoptic reports culled from the non-structured data in the pathology reports.
6. AI-driven language models can assist in the standardized extraction of information from pathology reports and facilitate the integration of this information with other clinical and research data. The AI model can also provide a degree of certainty to the diagnostic information extracted from pathology reports [243]. This can help improve the consistency, certainty, and comprehensiveness of data available for decision-making and research purposes.
7. AI can create standardized training materials and assessment tools for pathologists, ensuring that they are educated and evaluated based on best practices and the latest advancements in the field.
8. AI can facilitate better communication and collaboration among laboratories and healthcare providers by providing a common platform for data analysis, visualization, and decision support. The AI language models can provide the translation of a pathology report between English and other languages for communication and collaboration among pathologists in different regions of the world. This can contribute to standardizing workflows and practices across the digital pathology ecosystem.

#### 4.5.7 Trustworthiness and Acceptability of AI by Pathologists

AI will not replace Pathologists but will help them in augmenting their tasks. Pathologists' trust in AI-based technologies is pivotal for successfully incorporating these tools into laboratory practice. The main ingredients for developing such trust are the validation of AI models before deployment in pathology laboratory workflow, performance monitoring after deployment, and continuous interaction between pathologists and the developers of the AI systems. The question of whether pathologists should trust AI is complex. We have demonstrated AI's potential to revolutionize pathology in the near future through innovative tools for helping in diagnosis, prognosis,

and making accurate predictions swiftly. Nevertheless, concerns persist about the reliability and trustworthiness of AI, in part owing to the biases and missingness in training datasets and the inability of algorithms to tackle such limitations.

Additionally, AI models are vulnerable to adversarial examples that can deceive the AI model to drastically change its output with only a subtle change in the input [241, 222]. Groups such as Trustworthy Software Foundation (TSFdn) [313] in the UK and Coalition for Health AI (CHAI) [79] in the US are addressing trust issues by advocating for credible, fair, and transparent AI systems in healthcare. TSFdn defined five facets of software trustworthiness: safety, reliability, availability, resilience, and security [313]. CHAI has recently published a blueprint for trustworthy AI implementation in healthcare and defined seven key elements of trustworthy AI in healthcare: useful, safe, accountable and transparent, explainable and interpretable, fair, secure and resilient, and privacy-enhanced [151]. Furthermore, Dorr *et al.* proposed to create a “Code of Conduct for AI in Health Care” that aims to harmonize standards and ensure responsible AI usage [190]. The potential benefits, limitations, and risks of generative AI systems and foundation models, such as GPT-4, have recently been highlighted, emphasizing their cautious use in clinical settings [397]. Rajpurkar *et al.* presented the generalization checks for AI systems as transparency, clinician-AI collaboration, and post-deployment monitoring [574]. Nakagawa *et al.* introduced various challenges in digitizing medical workflows, including data biases, privacy concerns, and algorithm fragility, while emphasizing the need to carefully consider AI’s impact on pathologists [499]. AI model cards are an important documentation framework for understanding, sharing, and improving machine learning models [491, 532]. We summarize the strategies that may contribute to building the trust of pathologists in AI systems as follows:

1. Users, such as practicing pathologists, should understand the basic functionality of AI models and their reasoning or logic for a certain decision, just like radiologists understand the core principles that are used by CT scanners, MRI machines or X-rays to generate a certain type of images. Therefore, AI developers need to prioritize the use of transparent and explainable models that provide insights into their functionality and decision-making process.

2. Equally important is the engagement of pathologists in AI model development and validation processes, encouraging valuable feedback from both pathology and AI domain experts in AI development and deployment. Collaborative efforts and continuous feedback allow for performance improvement, ensuring the long-term utility of the models.
3. There is value in educating and training medical professionals (practicing pathologists in this case) on the topics related to AI model development and the potential use of these technologies for solving various problems in healthcare. We have developed a hands-on machine learning course for medical professionals that aims to provide the specialists such as pathologists with a practical acquaintance of AI concepts [580].
4. Before deploying AI models in a clinical workflow, it is essential to perform rigorous evaluation and validation in the clinical settings where the model is intended to be used in a sandbox environment. This rigorous assessment will help identify and fix modeling issues, software bugs, and data and algorithm interoperability challenges. Such validation and evaluation processes will instill users' trust in AI models, assuring their suitability for real-world implementation and sustained use.
5. During AI model development, validation, and deployment, it is crucial to address ethical considerations to ensure fairness and reduce bias in the models' decisions. Validating and reporting the model's output for various groups based on age, race, and sex should be considered an essential part of model development and deployment [491].

## **4.6 Conclusion**

Foundation models and generative AI have the potential to transform digital pathology, leading to faster and more accurate diagnoses, improved patient outcomes, and a better understanding of disease mechanisms, along with reducing workload for pathologists, helping standardize lab workflow, and contributing to the education and training. These powerful technologies will not replace pathologists but augment and streamline their skill sets. This review presented an overview of

generative AI and foundation models and their potential role in digital pathology. We demonstrated how AI as a field has grown from a narrow problem-solving technique to a comprehensive tool for language understanding, image analysis, data generation, question-answering, and conversation. Finally, we present our perspective on the future role of generative AI and foundation models in digital pathology and future use cases where generative and conversational AI and foundation models can have a transformative impact in digital pathology. Adapting and integrating generative foundation models in traditional diagnostic methods can provide a more comprehensive and accurate assessment of pathology specimens while enabling the development of personalized treatments for patients. However, generative AI and foundation models have associated challenges and limitations. Further research and development efforts are needed to fully realize the current AI wave's potential to ensure their safe and effective implementation in clinical practice.

## **Chapter 5: Digital Pathology and Multimodal Learning on Oncology Data**

### **5.1 Note to Reader**

This chapter is under review at BJR AI, and is being reproduced with permission from BJR AI.

### **5.2 Introduction**

The examination of tissue samples under a microscope by pathologists is the standard procedure for disease diagnosis in clinical settings. However, technological advancements in the last couple of decades have ushered in a new era in the field of pathology called digital pathology (DP). Some of the key factors that have enabled the transformation from traditional pathology to DP include: (1) Whole Slide Imaging (WSI), which scans entire tissue sections on glass slides at high resolution and stores them as digital slides that can be analyzed computationally [788], (2) Image analysis algorithms from the fields of computer vision and machine learning that can detect, segment, classify, and quantify morphological features, cells, tissues, biomarkers from digitized images [9], (3) Advanced imaging techniques such as computational staining and multiplexed imaging [59], (4) Molecular assays that generate quantitative molecular data from tissues using techniques such as polymerase chain reaction, microarrays, and next-generation sequencing [105], and (5) Enterprise imaging that includes pathology informatics [150].

As the field of pathology evolves, so too does our understanding of cancer. Cancer is a complex disease that can manifest in different forms, each with its distinct characteristics and implications for screening, diagnosis, prognosis, and treatment. Over time, our understanding of cancer has evolved. Initially thought to be primarily a result of genetic mutations, it is now recognized as a complex interplay of genetic, environmental, and lifestyle factors. Advances in

genomics, proteomics, bioinformatics, and imaging have led to a deeper understanding of cancer, revealing that even cancers originating in the same organ can have vastly different characteristics and responses to treatment. This complexity underscores the necessity for personalized medicine approaches tailored to the unique genetic makeup of each cancer and each patient.

Transitioning from the complexity of cancer biology to the use of advanced technologies, the success of artificial intelligence (AI) and machine learning (ML) in different domains of science, especially those focused on quantitative imaging analysis, has attracted anatomical pathology researchers and practitioners. As a result, computational pathology (CPATH) emerged when AI and ML models were introduced for various types of quantitative analysis in DP [2]. CPATH uses AI and ML models to extract information from digital images that may extend beyond immediate human capabilities. By offering AI/ML-powered analytical tools, CPATH has the potential to enhance the roles of pathologists in clinical settings by making their work more efficient, accurate, and reproducible. This enhancement spans various aspects, including improving the precision, reproducibility, and scalability of tasks related to disease diagnosis and the quantification of biomarkers for prognosis and prediction.

However, while the integration of AI and ML into DP clinical workflow promises significant advances, it also presents several challenges. These challenges include a shortage of staff trained in both AI and DP that may temporarily lead to increased workloads. Nevertheless, over time, the increased use of AI and its incorporation into the training curriculum will gradually improve the number of AI-qualified individuals in the DP/CPATH field. The growing complexity of diagnostics due to evolving protocols and new biomarkers, as well as variations in rare diseases and morphological similarities, also pose challenges. Additionally, quality issues arising from tissue artifacts and staining/imaging inconsistencies, along with a lack of standardized practices and protocols, may impede interoperability and pose reproducibility challenges.

Table 5.1 Key pathology terminologies

Term	Definition
Pathology	Stems from the Greek word <i>pathología</i> , meaning the study of suffering, it is a medical specialty investigating the origin, progression, and alterations in the structure, function, and natural course of diseases [232]. There are two main branches of pathology: anatomical pathology (AP) focuses on studying tissue samples to diagnose diseases such as cancer, and clinical pathology (CP) analyzes bodily fluids and conducts laboratory tests [733]. Molecular pathology is an emerging branch that studies diseases at the molecular level by examining tissue or body fluids.
Histology	Microscopic examination of tissues and organs by sectioning, applying stains, and observing the prepared sections under a microscope. This process enables the visualization of tissue structures and any distinctive alterations the tissue might have undergone. Hematoxylin (a basic dye that stains cell nuclei in a purplish blue color) and Eosin (an acidic dye that stains cytoplasm in a pinkish red color), referred to as H&E, are commonly used together to stain structures of the cells to define intracellular organelles and proteins [267].
Histopathology	It extends the utility of histological analysis by focusing on the study of disease indicators through the microscopic examination of processed and fixed glass slides containing sections from biopsies or surgical specimens. To visualize various tissue components under a microscope, these specimens are dyed with one or more stains depending on the investigational question being asked [267, 266].
Immunohistochemistry (IHC)	A valuable tool extensively employed in pathology for cell classification and diagnosis. Commonly performed on formalin-fixed paraffin-embedded (FFPE) tissue, IHC targets specific antigens in tissues and cells using antibodies to determine cell type and organ of origin. IHC usage has lately expanded to study predictive and prognostic biomarkers in oncological settings, making it a useful technique in modern pathology practice [467].
Whole slide imaging (WSI)	The process of digitizing glass slides at multiple magnifications and focal levels to create digital images for observation and image analysis. WSI aims to replicate the experience of traditional light microscopy through digital means and is often known as virtual microscopy [538].
Cytology slides	These refer to glass slides that contain specimens of individual cells or cell clusters obtained from various bodily fluids, tissues, or fine needle aspiration (FNA) biopsy. Unlike histology, which focuses on the function of tissues, cytology slides are prepared for microscopic examination to analyze cellular morphology, structure, and characteristics [372].
Digital pathology (DP)	DP includes converting histopathology, IHC, or cytology slides into digital formats through whole-slide scanners, followed by applying computer-aided analytical tools for objective analysis. DP has improved pathology diagnoses through multiple techniques as discussed in this review.
Computational pathology (CPATH)	CPATH is the sub-branch of DP that involves computational analysis of digital slides. CPATH aims to analyze patient specimens by extracting information from digitized pathology images in combination with their associated meta-data, typically using AI methods to gain valuable insights into disease processes [733, 2, 229].
Radiomics	Radiomics is a sub-field of radiology, where radiological images are converted into high-dimensional mine-able data for AI-based analysis to uncover biomarkers for diagnosis, prognosis, and treatment response prediction [495].
Pathomics	Pathomics is a sub-field of pathology, that covers a wide variety of data captured from image analyses in the form of features that describe multiple phenotypic features of tissues in WSIs. Using AI, pathomics provides a quantitative assessment of identified structures and features to complement traditional histopathologic evaluation by pathologists [265].

Amid these challenges, multimodal learning in oncology represents a cutting-edge approach to cancer research and treatment, leveraging the integration of various data types to gain a comprehensive understanding of cancer’s complexities [87, 136, 740]. This approach combines diverse datasets, including genomic, proteomic, imaging, clinical, and demographic data, to develop a holistic view of cancer biology. By harnessing advanced computational techniques offered by AI/ML, multimodal learning enables analyzing these disparate data types jointly in a unified framework [740, 733]. This integration is paramount for identifying novel biomarkers, elucidating tumor heterogeneity, predicting treatment responses, and ultimately facilitating the development of more effective, personalized therapies [204].

Building upon these developments, DP and CPATH have paved the way for exploring novel methods for cancer diagnosis, prognosis, treatment planning, and also for research into a deeper understanding of cancer biology [74, 748, 634, 82]. The innovative algorithms within CPATH, especially those employing differential privacy and federated learning, enable the development of robust analytical models. These models are crucial for safeguarding patient privacy, fostering collaboration among institutions, and paving the way for the adoption of clinical-grade systems [454]. Such advancements underscore the critical role of AI and ML in transforming cancer care, making it more precise, personalized, and effective.

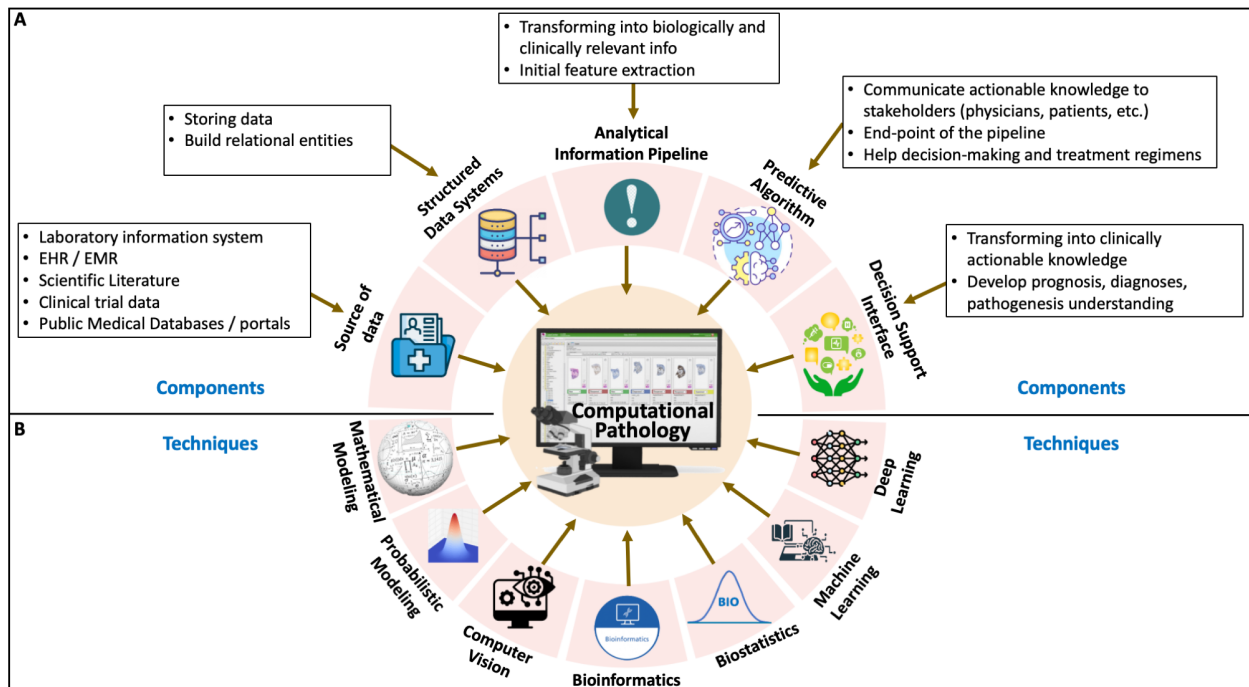


Figure 5.1 A template of the future road map for DP and CPATH.

### 5.3 From Conventional to Digital Pathology (DP)

Conventionally, pathologists observe tissue samples on the glass slides under a microscope [733]. Such a traditional approach involves physical preparation and storage of glass slides and local consultations only, limiting the efficiency, reproducibility, scalability, and accessibility of the



technique. These limitations, along with the advent of modern computers, imaging systems, and quantitative image processing techniques, have prompted a shift towards DP [376]. The journey of digitization in pathology began with the development of the first virtual microscope during 1996-1998, where high-resolution images replicating the glass slides were produced using imaging scanners [537]. The system was later advanced to support data caching and pre-computed image pyramids [537]. In the last two and half decades, the emergence and commercialization of WSI and glass slide scanners have revolutionized clinical pathology practices [537, 615]. Presently, many commercial WSI scanners are available (Philips, Leica, and Hamamatsu, etc.), and many open-source WSI systems (caMicroscope, Digital Slide Archive, Sedeen viewer, QuPath) are supported actively [268, 471, 603, 65]. The advent of DP, powered by WSI, has revolutionized medical diagnostics [376, 746]. Improvements in technology by turning glass slides into digital formats and decreasing storage costs have increased the use of WSIs [601]. The adoption of DP by pathologists has seen significant growth, increasing from 30% in 2013 to 64% in 2021, according to a report by the College of American Pathologists [376]. In Fig. 5.1, we have presented a template for the future roadmap for DP and CPATH that may have the necessary components and methods in a unified framework. With a multitude of advantages, DP has brought a revolution in high-resolution magnification analysis, electronic storage, automation of processes, quality assurance and control, telepathology and enhanced diagnostic efficiency, streamlined diagnostic workflow, accessible measurements and quantification, cancer pre-screening, intra-operative consultation, histopathological second opinion, and education and training.

#### **5.4 Computational Pathology (CPATH) - AI in DP**

More recently, the advancements in AI and ML have started revolutionizing DP, which has been a helpful tool in histopathology image analysis in the last decade [376]. Digital Pathology Association (DPA) defines CPATH as, “the big-data approach to pathology,” where patient’s pathology images and associated meta-data are combined to extract patterns and analyze image features through AI techniques [2]. Based on the recent advances in CPATH, the College of American Pathologists

Table 5.2 Key AI terminologies

Term	Definition
Artificial Intelligence (AI)	Initially started as a simple theory of human intelligence being exhibited by machines, AI aims to simulate human intelligence in machines by developing computers that can engage in learning, reasoning, self-correction, problem-solving, and decision making [381].
Machine Learning (ML)	ML is a subset of AI that programs computers to process natural data and solve an optimization problem using data samples or past experience. It is an engineering discipline that is closely related to mathematical statistics to build learning models [144].
Deep Learning (DL)	DL is a subfield of ML and refers to models developed using multiple layers of artificial neural networks. The deep neural networks focus on compositional learning using multiple (hence the word “deep”) layers of representations, where each layer is built using multiple artificial neurons [144, 734, 735, 12, 357]. Various types of DL models have been developed, including Convolutional Neural networks (CNNs), Recurrent Neural networks (RNNs), Graph Neural Networks (GNNs), Transformers, and many others.
Convolutional Neural Networks (CNNs)	CNNs are specialized DL models for analyzing image or grid data [144, 252, 653]. CNNs use convolution operations and can learn spatial correlations, features, and relationships from the input data [252, 177].
Recurrent Neural Networks (RNNs)	RNNs are a family of DL models used for processing sequential data like speech, text, or time-series data (e.g., patient vital signs recorded over time) [252, 174, 13]. RNNs are used to model time/space-based relationships between sequences of input data and to learn patterns based on the order (in time or space) [252]. Sub-categories of RNNs called Long short-term memory network (LSTM) and gated recurrent unit (GRU) models are better at learning long-term dependencies, making them great for real-world use cases, e.g., predicting medical events and detecting anomalies in time series data [174].
Graph Neural Networks (GNNs)	GNNs are DL models designed for analyzing graph data, like social relationships or biological pathways [735]. GNNs pass messages between neighboring nodes to aggregate local and global information. By modeling these relationships, GNNs can make predictions about nodes, edges, and sub-graphs in the network, where each entity may represent patients, drugs, proteins, etc. GNNs enable new insights in fields like precision medicine and drug discovery [8, 665, 724].
Transformers	Transformers are DL models that have shown promising results on language translation and image recognition tasks [706, 191]. Their key innovation is a technique called self-attention. Unlike RNNs, which process data sequentially, self-attention allows Transformers to learn relationships between any parts of the input, regardless of their position in the sequence. For example, a Transformer can learn that the symptoms “cough” and “fever” are strongly related even if they are distant in the patient’s medical history. The model learns which connections are important by assigning an “attention” score between input sequences (also referred to as tokens) [13, 706, 191].
Foundation Models (FMs)	FMs are a new class of DL models characterized by their large scale and ability to adapt to new tasks [92]. FMs are trained on extensive datasets, allowing them to develop a broad understanding of various topics [733, 450]. The key characteristics of FMs include scalability in learning from diverse data sources, multimodal learning, compositionally in terms of generalization, and the emergence of implicit learning. Transfer learning, adaptation, in-context learning, and fine-tuning make FMs exceptionally useful across numerous fields, including the medical domain [453]. A recent review paper covers existing FMs and their prospective use in DP [733].
Multimodal Learning	Multimodal Learning involves integrating data from multiple modalities for an ML task, e.g., prediction, classification, or regression. A modality refers to specific data input or source types, such as text, images, audio, video, or other data. The core idea is to leverage the complementary information from different modalities to improve the performance and accuracy of learning algorithms. Multimodal learning aims to achieve a more comprehensive understanding of the problem, leading to better decision-making than using a single modality [740, 481].

announced the inclusion of 30 new current procedural terminologies (CPTs) related to DP for use by pathologists, effective from January 1, 2024 [514]. The integration of AI in DP has facilitated precise diagnosis by processing information that may not be perceived by the human eye at times [733, 537]. The cardinal tasks of AI in DP include the development and validation of models and algorithms that can detect different patterns and features from data presented in the form of WSIs. The AI models can identify regions of interest (ROIs) from WSIs, extract and classify patches, classify entire WSIs, generate labels, and identify various objects of interest [537]. Examples of how various AI and ML models have been proposed to advance various DP tasks include,

diagnosis support, enhanced image analysis, improving prognostics and biomarkers predictions, post-treatment analysis and disease surveillance, and cross-modality learning [2, 376, 733, 558, 740].

## **5.5 DP, CPATH, and Multimodal AI**

Advancements such as the digitization of patient health records, the development of new diagnostic technologies, innovative laboratory tests, and the establishment of robust information technology infrastructure in hospitals have led to an exponential increase in the volume of medical data collected for each patient [740]. Recent progress in multimodal learning, driven by deep neural networks, has demonstrated a remarkable ability to learn from varied data modalities, such as computer vision and natural language processing [740, 92]. Multimodal foundation models such as OpenAI's Contrastive Language-Image Pretraining (CLIP) and Generative Pretraining Transformer (GPT-4) have established new standards in the field, while the Foundational Language And Vision Alignment Model (FLAVA) marks a notable advancement by integrating vision and language representation learning to enable multimodal reasoning [5, 563, 641]. The success of multimodal AI models in other domains has inspired cancer researchers to explore the potential of multimodal AI in healthcare [599]. The multimodal AI models being developed in the medical domain can integrate diverse modalities, including radiological images, genetic profiles, and clinical information, within the DP workflow, enhancing diagnostic precision and enabling personalized treatments [276, 736]. Multimodal AI models are poised to empower healthcare teams by offering them a holistic view of the patient's disease state and paving the way for truly personalized cancer care [740, 481, 276].

In our context, "modality" refers to a specific data source or type. For example, EHR text descriptions of surgical procedures or post-visit reports represent one modality. Lab test values, vital measurements, and demographic information form another. Histopathology slides (stored as WSI), radiological scans (X-ray, MRI, PET, CT), and molecular data (genomics, proteomics) are

separate modalities [315, 734, 736]. Some modalities have sub-categories, such as MRI scans with FLAIR, T1, T2, and CT scans with various views and contrast protocols.

In multimodal learning, researchers use various modality-specific data processing methods and data fusion mechanisms. The generally defined five main stages are: (1) preprocessing, (2) feature extraction, (3) data fusion, (4) primary learner, and (5) final classifier [740]. Data fusion from different modalities can occur at various levels. Early fusion combines features before feeding data into ML models [736, 693]. Intermediate fusion occurs at the model level, while late fusion merges processed data at the decision level [740]. Explainability in deep learning involves techniques to make models' decisions, behaviors, and outputs transparent and understandable [505, 680]. This includes building interpretable models or generating post hoc explanations [733, 505]. In DP, which uses high-resolution WSIs, explainability is crucial. Techniques include heatmaps showing areas of significance, case-based explanations comparing with established predictions, and layer-wise relevance propagation to trace outputs back to inputs [269].

In the following, we provide a chronological overview of various applications of multimodal AI/ML techniques with a focus on histopathology data as the essential ingredient integrated with one or more other modalities (e.g., radiology, genomics, or clinical data). A schematic layout of the multimodal AI in DP and CPATH is provided in Fig. 5.2. We also provide a summary of references for each category of multimodal learning in Table 5.3, along with the explainability method of these works where applicable.

### 5.5.1 2019

Cheerla *et al.* developed a multimodal model to predict the survival of patients for 20 different cancer types by fusing three different data modalities that included clinical data, mRNA and microRNA expressions, and WSIs [126]. The proposed multimodal AI model was able to predict the overall survival (OS) with the concordance index (C-index) of 0.78 [126].

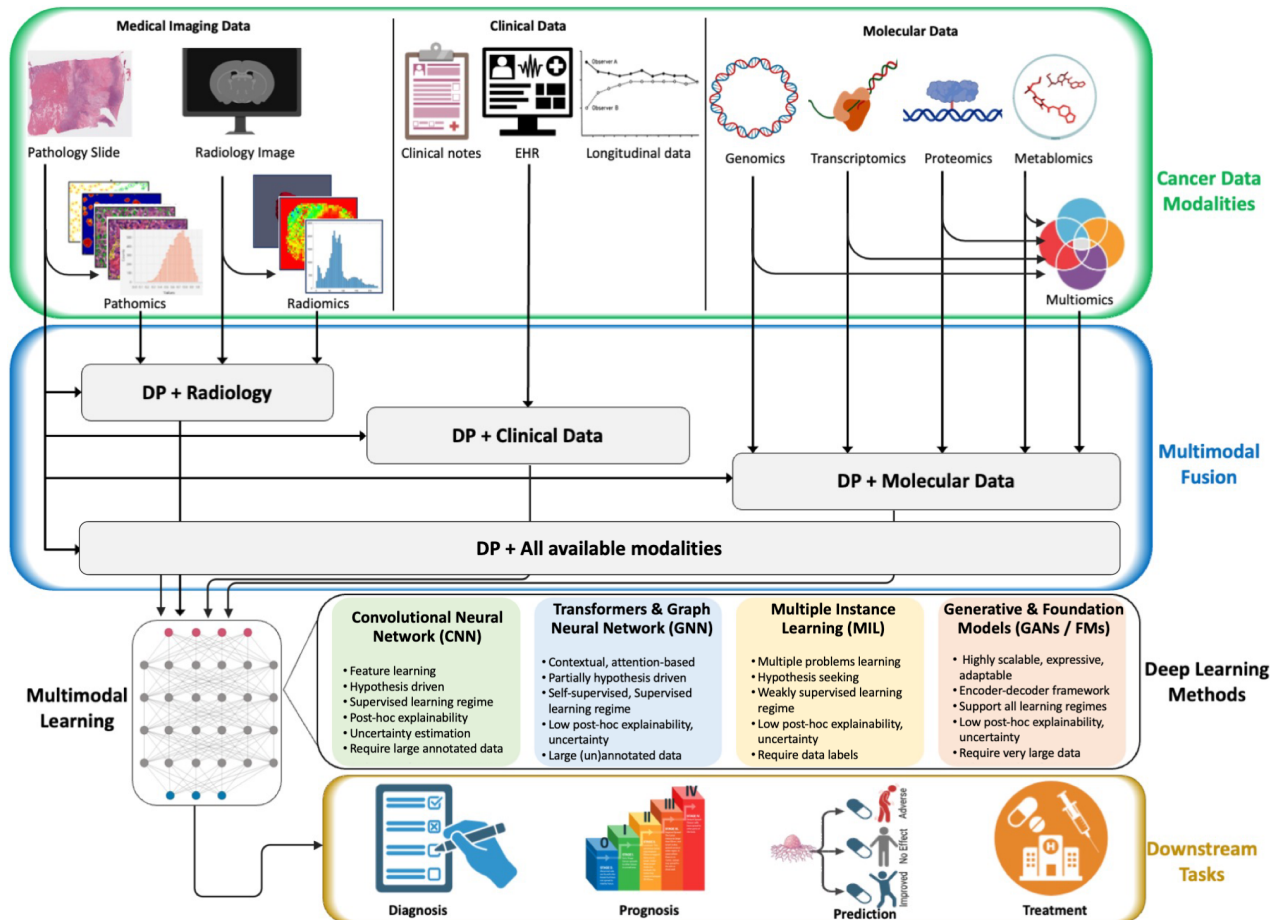


Figure 5.2 A schematic layout of multimodal learning in cancer care.

### 5.5.2 2020

Chen *et al.* proposed an end-to-end multimodal framework called “Pathomic Fusion” that predicted survival outcomes from WSIs and genomic (mutations, CNV, RNA-Seq) features of glioma and clear cell renal cell carcinoma (CCRCC) [135]. For the glioma patients, Pathomic Fusion outperformed the WHO paradigm and previous state-of-the-art with 6.31% and 5.76% improvements, respectively, and reaching a C-index of 0.826 [135]. For the CCRCC, similar results were observed, with tri-modal Pathomic Fusion achieving a C-index of 0.720 [135].

Microsatellite Instability (MSI) is an approved pan-cancer biomarker for immune checkpoint blockade (ICB) therapy. A multiple-instance AI model, called Ensembled Patch Likelihood Aggre-

gation (EPLA), was proposed to integrate pathological, genomic, and transcriptomic phenotypes and predict MSI [111]. The model also provided an interpretation of its results for effective use in ICB therapy. The proposed EPLA model got an AUC of 0.8848 and 0.8504 in the TCGA-COAD and Asian-CRC external validation, respectively [111]. EPLA was also able to capture correlation between poor differentiation and MSI [111]. The model was also able to identify imaging signatures associated with mutations burden, DNA damage, and antitumor immunity [111].

Rathore *et al.* used support vector machine (SVM) model to distinguish high-grade gliomas (HGG) from low-grade gliomas (LGG) by combining textual features with conventional imaging and clinical markers [582]. Their model identified the microvascular proliferation level, mitotic activity, presence of necrosis, and nuclear atypia in WSIs [582]. The texture features were successfully validated on the glioma patients in 10-fold cross-validation (accuracy = 75.12%, AUC = 0.652). Adding texture features to clinical and conventional imaging features significantly improved grade prediction compared to the models trained on clinical and conventional imaging features alone [582].

Classical ML models have been shown to capture the multi-scale association between H&E and CT images, which was used to classify lung adenocarcinoma from squamous cell carcinoma in NSCLC [27]. Alvarez *et al.* identified discriminant pathomic patterns for cancer classification and established correlations between multimodal features of 171 patients, depicting significant cross-scale associations between cell density statistics and CT intensity [27].

### 5.5.3 2021

The Multimodal Co-Attention Transformer (MCAT), an AI model leveraging the Transformer architecture and cross-modality attention, has outperformed all unimodal models and the more commonly used late fusion methods in survival prediction for five distinct types of cancers, including Bladder Urothelial Carcinoma (BLCA), Breast Invasive Carcinoma (BRCA), Glioblastoma & Lower Grade Glioma (GBMLGG), Lung Adenocarcinoma (LUAD), and Uterine Corpus Endometrial Carcinoma (UCEC) [136]. MCAT outperformed state-of-the-art methods on all benchmarks, with

a C-index increase of 6.35% and 12.4% compared to Attention Multi-Instance Learning (MIL) and Deep Attention Multiple Instance Survival Learning (DeepAttn-MISL) methods, respectively. The success of the MCAT model indicated that multimodal learning could be used in general for any survival outcome prediction task [136].

Khosravi *et al.* proposed AI-biopsy, an AI model based on the Inception-V1 architecture, which was pre-trained on a non-medical ImageNet dataset and transfer learned on MRI scans labeled with pathology assessments like Gleason scores and grade groups [369]. The proposed early data-fusion technique exhibited a higher agreement with biopsy results when compared to the Prostate Imaging Reporting and Data System (PI-RADS) score, indicating promising advancements in diagnostic precision [369]. AI-biopsy achieved AUCs of 0.78 (95% CI: [0.74–0.82]) and 0.89 (95% CI:[0.86–0.92]) in 400 patients on the tasks of distinguishing high-risk from low-risk cancers and cancerous vs. benign prostate disease, respectively.

WSI, demographic, genomic, and clinical data collected from three data sources, The Cancer Genome Atlas (TCGA), Clinical Proteomic Tumor Analysis Consortium (CPTAC), and Brigham and Women's Hospital, were used to develop an AI model for various tasks related to renal cell carcinoma [470]. The trained models were able to detect malignancy (AUC = 0.964–0.985), diagnose renal cancer subtypes (AUCs = 0.953–0.993), and predict survival outcomes [470]. The model also identified histopathology image features that were indicative of biomarkers such as copy-number alterations (CNAs) and tumor mutation burden (TMB) [470].

In general, single-omics data modality may not be able to provide a comprehensive view of tumor-related molecular alterations. In such cases, integrating multiple data modalities even within the -omics regime can help improve diagnosis, personalized treatments, and predict outcomes more accurately [688]. Tong *et al.* developed an encoder-decoder AI model for cancer classification among four cancer types (lung adenocarcinoma (LUAD), Kidney Renal Clear Cell Carcinoma (KIRC), Lung Squamous Cell Carcinoma (LUSC), and pancreatic adenocarcinoma (PAAD)) and patient survival prediction in two cancer types (Breast Invasive Carcinoma (BRCA), Ovarian Carcinoma (OV)) [688]. Two data integration methods were evaluated for four types of -omics data

that include gene expression, DNA methylation, miRNA expression, and copy number variations [688]. The multimodal AI models outperformed single-omics models for both tasks, cancer classification and survival prediction. Within the two data fusion methods, divergence-based consensus networks performed better than concatenation-based integration networks [688].

Multimodal learning using clinical data, H&E images, genomics, and transcriptomic features were shown to predict pathological complete response to treatment and improve outcomes in 168 breast cancer patients treated with chemotherapy with or without HER2-targeted therapy before surgery [604]. The multi-omic AI model predicted pathological complete response in an external validation cohort of 75 patients with an AUC of 0.87. The final AI model consisted of an ensemble of traditional ML models, including logistic regression, support vector machine, and random forest [604].

Chelebian *et al.* developed an ensemble CNN to provide interpretations that matched pathologists' manual assessments [127]. The resulting interpretations were more informative when the underlying genes were considered [127]. The framework paired H&E slides with spatially resolved gene expressions from the spatial transcriptomics technique, predicting the spatial variation of individual genes [127].

Braman *et al.* [97] proposed a Deep Orthogonal Fusion (DOF) model to learn from multi-parametric MRI scans, H&E slides, DNA sequencing, and clinical features into a multimodal risk score [97]. The embeddings (or representations) from individual data modalities were learned and fused through attention-gated tensor fusion. The authors introduced multimodal orthogonalization (MMO) learning loss that incentivized each embedding to be more complementary [97]. DOF predicted OS in 176 glioma patients with a median C-index of  $0.788 \pm 0.067$ , compared to the best performing unimodal model (C-index of  $0.718 \pm 0.064$ ) [97].

#### 5.5.4 2022

Compared with unimodal data, combining H&E WSIs with the molecular assay improved survival prediction reliability in multiple cancers, including BRCA, COADREAD, LUAD, PAAD,



and others [138]. Chen *et al.* proposed a deep learning framework, multimodal fusion (MMF), which achieved an overall C-index of 0.644 across the 14 cancer types, whereas the MIL model using WSI and SNN on molecular data had overall C-Indices of 0.578 and 0.606, respectively. On survival AUC, a similar improvement for the multimodal model was reported with an overall performance of 0.662 compared with unimodal models [138].

Often referred to as a “virtual biopsy,” radiomics features enhance conventional diagnostic radiologic workflows when combined with the pathomic features from H&E slides [98]. Brancato *et al.* employed different feature generation techniques to extract 91 radiomics features from MRI scans and 156 pathomic features from H&E images for 48 Glioblastoma multiforme (GBM) patients from the CPTAC data portal. The proposed radiomics-pathomics fusion technique showed significant cross-scale associations of features among the GBM patients [98].

Multimodal integration of radiology, pathology, and genomics data modalities improved insights into response to immunotherapy PD-L1 blockade in patients with NSCLC [703]. Vanguri *et al.* explored the multimodal integration of CT images, digitized programmed death ligand-1 IHC slides, genomics data including somatic mutations, copy number alterations and gene fusions, and known outcomes to immunotherapy to predict immunotherapy response among 247 patients with advanced NSCLC. The multimodal model called DyAM showed an AUC of 0.80, 95% CI [0.74–0.86] on immunotherapy response prediction task, that outperformed unimodal methods, including TMB (AUC = 0.61, 95% CI [0.52–0.70]) and programmed death ligand-1 IHC score (AUC = 0.73, 95% CI [0.65–0.81]) [703].

Ensemble of a range of ML regression models, including support vector regression, AdaBoost, gradient boost, and random forest, were shown to provide better survival outcomes prediction for 171 Glioma patients using integrated features from MRI (T1-GD, T1, T2, and T2-FLAIR) and H&E images [581]. The proposed ensemble fusion approach outperformed several other techniques based solely on radiology features, gene expression data, or pathology images [581].

Combining pretreatment MRI (T2, T1-CE, DWI) and H&E-stained biopsy slides using AI models can predict pathological complete response in patients with locally advanced rectal cancer [219].

The proposed framework, RAdioPathomics Integrated preDiction System (RAPIDS), showed a significant improvement in the prediction of pathological complete response with a combined AUC of 0.812 compared to the unimodal models for prospective validation study [219]. RAPIDS also performed better in the retrospective study training cohort (AUC 0.868 [95% CI 0.825–0.912]), validation cohort 1 (0.860 CI [0.828–0.892]) and validation cohort 2 (0.872 CI [0.810–0.934]) [219].

Harnessing AI-driven insights with radiomics-pathomics has led to personalized treatment plans, optimizing therapies while minimizing adverse effects [726]. An innovative fusion of pathomics, radiomics, and immunoscore data identified high-risk patients with colorectal cancer (CRC) lung metastasis. The combined nomogram model predicted the OS with an AUC of 0.860 and disease-free survival (DFS) with an AUC of 0.875. This integration offers insights about aggressive treatment strategies and tailored follow-up plans for high-risk patients [726].

Wan *et al.* developed a multiscale framework that harnessed the power of radiomics features extracted from MRI (T2WI, DWI) data, pathomics features derived from H&E images, and clinical variables to predict pathological good responses (pGR) in 153 patients with locally advanced rectal cancer who underwent neoadjuvant chemoradiotherapy (nCRT) [718]. This macroscopic-microscopic scale fusion identified the patients with down-staging to stage ypT0-1N0 after nCRT, a critical factor in guiding organ-preserving strategies to minimize invasion and uphold the quality of life. By integrating these diverse modalities, the multiscale AI model outperformed the traditional clinical-radiological model with an AUC of 0.90 (95% CI: [0.78–1.00]) compared to 0.68 (95% CI: [0.46–0.91]) in the validation group. These results highlight the complementary nature of information present in radiomics and pathomics features for assessing treatment response [718].

Boehm *et al.* used a multimodal dataset of 444 patients with ovarian cancer and integrated radiological data (CT scans), histopathological images (H&E), clinical, and genomic modalities to stratify patients based on risk [88]. The multimodal AI model stratified patients in the training and the test cohorts with the C-indices of 0.55 (95% CI: [0.549–0.554]) and 0.53 (95% CI:

[0.517–0.547]), respectively. The authors also discovered quantitative features, such as tumor nuclear size on H&E and omental texture on CT, associated with prognosis [88].

#### 5.5.5 2023

Palmal *et al.* employed a graph convolutional network (GCN) and Choquet fuzzy ensemble consisting of classical ML models (Logistic Regression, Random Forest, and SVM) to fuse the multi-omics and clinical data [534]. 1980 samples of breast cancer patients were classified as short-term or long-term survivors using different combinations of gene expression (24,000 features), copy number alteration (26,000 features), and clinical data (27 features) [534]. The performance metrics of accuracy, Matthews correlation coefficient, precision, sensitivity, specificity, balanced accuracy, and F1-measure showed significant result scores as 0.820, 0.528, 0.630, 0.666, 0.871, 0.769, and 0.647, respectively [534].

Tsai *et al.* developed a Multi-omics Multi-cohort Assessment (MOMA) framework to fuse WSIs with genomics, proteomics, and molecular data to predict OS, progression-free survival, and other variables for colorectal cancer patients [697]. MOMA was able to identify cross-scale relationships among histologic patterns, multi-omics, and clinical profiles in 1888 patients. MOMA predicted OS, progression-free survival, and MSI in held-out test datasets of stage I, II, and III colorectal cancer patients better than other state-of-the-art methods [697].

An AI model that fuses CT images with WSIs was shown to accurately predict lymph node metastasis stage (LNMs) in patients with gastric cancer [807]. Features extracted from CT images using a CNN (ResNet-50) and from WSI using the Vision Transformer model (ViT) were fused to learn the prediction task of five LNM stages [807]. The proposed radio-pathologic integrated method achieved 84.0% accuracy and significantly improved AUCs in predicting five LNM stages compared to single-modality models [807]. The integrated model also achieved an AUC of 0.978 (95% CI: [0.912–1.0]) in metastasis prediction task [807].

An AI model called Deep Learning Radio-Pathomics Model (DLRPM) seamlessly integrated radiomics, pathomics, clinical, and pathological features using ResNet-50 and SVM on 211 breast

cancer patients who underwent neoadjuvant chemotherapy [794]. DLRPM demonstrated remarkable capabilities for predicting pathological complete response in the validation dataset with an AUC of 0.927 (95% CI: [0.858–0.996]), significantly outperforming models that used individual data modalities and features, such as radiomics and pathomics [794].

The fusion of radiomics features with pathomics features and molecular data may provide a more comprehensive view of the intricate tumor microenvironment (TME), enabling a deeper understanding of its composition and tumor heterogeneity [345]. Fusing radiological and pathomics data may offer precise spatial mapping of histological features, facilitating detailed TME examination. This holistic approach may unravel the complex interactions between tumor cells and the surrounding microenvironment, advancing our understanding of cancer biology [345].

Recently, Tang et al. designed a new fusion method to combine dermatological images and patient clinical metadata for skin cancer classification using three public skin lesion datasets with more than 3,300 patients [677]. The proposed AI model, called joint-individual fusion with multimodal mutual fusion attention (JIF-MMFA), learns the shared features of multimodal data by employing a fusion attention module to enhance the relevant features from both data modalities [677]. JIF-MMFA showed the effectiveness of using patients' clinical data in improving the classification of skin cancer [677].

## **5.6 Discussion**

The global market for AI-based DP solutions is projected to grow at a compound annual growth rate of 8.59% from 2024 to 2031, with North America leading due to advanced healthcare infrastructure, higher income levels, and research facilities [3]. AI patent filings in DP have sharply increased in major patent offices, indicating growing innovation [18]. There is a significant push towards open-source and reproducible AI, with initiatives like Allen AI contributing through platforms like Hugging Face [620]. AI is increasingly used to predict patient responses to IO by analyzing histopathological and molecular data [550]. The integration of AI in DP promotes

Table 5.3 Summary of multimodal AI models for DP in this review

Modalities	Ref	Data type	Metrics	# patients	Task
Pathology, Molecular	[138]	H&E, Multiomics (mutation status, CNV, RNAseq expr)	5-fold cv, C-index, AUC	5,720	Survival outcome prediction and prognostic features correlations.
	[135]	H&E, Multiomics (mutation status, CNV, RNAseq expr)	15-fold cv, C-index	1186	Survival outcome prediction.
	[136]	WSIs, genomic features	5-fold cv, C-index	3,523	Overall Survival
	[111]	WSIs, Multiomics (DNA mutation, mRNA expr)	AUC	1,214	Microsatellite status prediction
	[127]	H&E slides, spatially resolved gene expression	Accuracy	7	Cancer detection, predict spatial variation of individual genes
	[697]	WSI, Multiomics (genomics, proteomics, and molecular subtyping)	C-index	1,888	Survival prediction, MSI prediction
Pathology, Clinical	[582]	WSIs, Pathomic features, Clinical	10-fold cv, Acc, AUC	735	Tumor grade prediction
	[677]	Dermatology images, Clinical	Balanced accuracy, AUC	>3,300	Cancer classification
Pathology, Radiology	[369]	WSI, MRI	NPV, PPV, accuracy, AUC, Cohen's kappa	400	Tumor classification 1) benign/cancerous, 2) high-risk/ low-risk tumor
	[98]	Radiomics (from MRI), Pathomic features (from H&E)	Spearman's correlation	48	Association among radiomics and pathomics features
	[581]	MRI (T1-GD, T1, T2, and T2-FLAIR), H&E images	5-fold cv, Regression	171	Survival prediction
	[219]	MRI (T2, T1-CE, DWI), H&E images	AUC, PPV, NPV	1033	Predict pathological complete response
	[718]	MRI (T2WI, DWI), H&E images	AUC	153	Predict pathological good responses
	[27]	CT, H&E images	AUC, accuracy	171	Cancer classification
	[807]	CT, WSIs	AUC, accuracy	252	Prediction of LNM stage
	Pathology, Radiology, Molecular	[703]	Radiology (CT), pathology (IHC), Multiomics (somatic mutations, CNAs, gene fusions)	10-fold cv, AUC	247
Pathology, Radiology, Clinical	[726]	Pathomics, Radiomics features, Immunoscore, Clinical factors	AUC	103	Survival prediction
	[794]	Pathomics, Radiomics, Clinico-pathological features	AUC	211	Predicting pathological complete response
Pathology, Molecular, Clinical	[126]	WSIs, Multiomics (mRNA, miRNA expr), Clinical	C-index	11,160	Overall Survival
	[470]	WSIs, Multiomics (CNA, somatic genetic mutations), Clinical	AUC	1,281	Detect malignancy, Diagnose cancer subtypes, Predict survival outcomes
	[604]	H&E, Multiomics, Clinical	AUC	168	Response to treatment
Path, Rad, Molec, Clin	[88]	H&E, CT, Multiomics, Clinical	C-index	444	Overall survival
	[97]	H&E, MRI, Multiomics (DNA seq), Clinical	15-fold Monte Carlo, C-index	176	Overall survival

collaboration among pathologists, oncologists, radiologists, and bioinformaticians, leading to a holistic approach to patient care [247].

Large language models (LLMs) like ChatGPT enhance pathology reports by ensuring consistency and translating complex medical data for clinicians and patients. LLMs assist pathologists by providing quick access to relevant literature and summarizing research findings [258, 733]. While most multimodal works focus on 2D WSIs, incorporating 3D data from techniques like microscopy or medical imaging could provide additional spatial context, improving model performance [373]. Combining WSIs with text reports, genomic data, or medical videos in multimodal learning is emerging as a way to enhance diagnostic capabilities [476]. Models like CLIP, which integrate text and image data, can be adapted for DP tasks [451]. Integrating histopathological, radiomic, genomic, and clinical data using a late-fusion approach has been shown to improve cancer patient risk stratification [88]. Multimodal AI is used across various domains, including biomedicine for disease diagnosis and treatment planning, autonomous vehicles for real-time decision-making, and augmented generative AI for content creation and multimodal translation [648, 161, 525, 257, 295]. Below we discuss some of the opportunities and challenges of multimodal learning in DP.

## **5.7 Opportunities**

### **5.7.1 Multimodal Patient Stratification**

Integrating various data types, including genomics, transcriptomics, epigenomics, proteomics, histopathology, and radiology, allows more precise patient stratification [87]. AI models can identify distinct patient subgroups with different phenotypes and treatment responses. This stratification can lead to more personalized and effective treatment strategies, enhancing patient care.

### **5.7.2 Non-Invasive Alternatives, Personalized Medicine**

AI models enable the prediction of histology subtypes or grades from radiomics features [430]. This capability will open many opportunities for developing non-invasive surrogates for existing

biomarkers, reducing the need for invasive procedures in the future. Additionally, AI models demonstrate promise in predicting clinical outcomes, including survival, treatment response, and recurrence, enabling personalized medicine approaches and better patient management.

### 5.7.3 Patient Empowerment

Integrating different modalities can create user-friendly digital solutions that empower patients to become active partners in managing their health [481]. Patients can access their own health records, diagnostic reports, and integrated biomarker information, allowing them to make informed decisions about their care. This level of patient engagement and access to integrated health information can enhance the patient experience, improve communication with healthcare providers, and improve health outcomes.

### 5.7.4 Improving Outcomes

Multimodal data integration, especially fusing histopathology and genomics data, has shown promise in improving the accuracy of outcome predictions for cancer patients [87]. Researchers and clinicians can improve overall/progression-free survival analysis by leveraging multimodal AI techniques, enabling more precise prognostication beyond traditional clinical factors. This can facilitate the development of more effective treatment regimens for high-risk patients and improve treatment outcomes.

### 5.7.5 Morphological Associations, Biomarker Discovery

Multimodal AI models can potentially discover morphological associations spread across data modalities, including the relationship between genetic mutations and tissue morphology [430]. Such findings can help exploratory studies and reduce the search space for potential biomarker candidates. By identifying morphological associates, AI can support cost-efficient biomarker discovery and aid in cancer screening and therapeutic target discovery.

### 5.7.6 Integrative Diagnostics

ID is defined to be the convergence of radiology and pathology imaging, clinical laboratory medicine, and advanced computational methods [75]. ID, especially with a focus on radiology and pathology data fusion, offers many benefits and untapped potential in oncology. Yet, despite the anticipation of numerous benefits, a noticeable scarcity of comprehensive literature surrounds this emerging field. ID can potentially identify appropriate therapies and suggest treatment changes when needed, decreasing morbidity and costs and improving patient outcomes [75].

### 5.7.7 Correlational Studies

Multimodal learning may hold enormous potential for conducting correlational studies and improving oncology patient diagnoses. Using histopathological and molecular data, AI models have already been shown to identify novel biomarkers and underlying tuberculosis mechanisms in susceptible individuals [49].

### 5.7.8 Medical Report Integration

Despite recent advancements related to digitizing medical data in EHR and radiological and laboratory imaging systems, there are apparently no efforts focused on correlating the imaging data (histopathology and radiology) with the textual reports (e.g., radiology reports, pathology reports, and other clinical notes) at the data storage level. One obvious reason is related to the siloed storage and data management systems. Modern AI systems based on vector databases can provide the necessary building blocks for data integration at the storage level [80].

## 5.8 Challenges

### 5.8.1 Data Availability, Integration, Missingness

Obtaining and integrating diverse medical data types, such as images and clinical records, is an extremely challenging task [691]. Data security and governance issues, data scarcity, unstructured



formats, and varying data identifiers pose significant hurdles in this process [345, 87]. Moreover, the issue of missing or incomplete data within each data modality further complicates the training of AI models [691, 430]. Strategies like synthetic data generation, dropout-based methods, and embedding models are being explored to address this challenge, but concerns about the accuracy and representativeness of synthetic data still remain [736, 693]

### 5.8.2 Interoperability

Ensuring that different medical devices, systems, and data formats can seamlessly communicate and exchange information is a considerable challenge [481, 345]. Data standards, protocols, and technologies differ across vendors, providers, and sometimes healthcare institutions and systems. Efforts such as Fast Healthcare Interoperability Resources (FHIR) for the EHR to overcome this challenge are essential for integrating diverse modalities and facilitating comprehensive patient care.

### 5.8.3 Data Modeling and Analysis

This challenge arises when integrating data from different sources or institutions because of institution-specific biases stemming from variations in data acquisition methods, clinical data standards, and the catchment area demographics [87]. Ensuring model generalizability in the presence of heterogeneous data modalities is another aspect of this challenge [87].

### 5.8.4 Transparency and Clinical Adoption

While multimodal AI models hold great promise in oncology, their adoption in clinical practice faces several hurdles. Ensuring transparency and interpretability of these models is challenging, as they often use abstract feature representations [430]. Prospective clinical trials and rigorous validation studies will be needed to demonstrate the added value of AI models in real-world clinical settings. Regulatory guidelines and addressing issues like fairness and dataset shifts are equally important for clinical adoption [430].

### 5.8.5 Personnel Training

The primary users of AI systems in DP are clinicians and pathologists. Training the administrative, technical, and professional personnel on AI systems is a secondary but essential challenge that needs to be addressed for the success of AI systems [274].

### 5.8.6 Explainability, Trust, and Reproducibility

AI models are known to be complex; this impacts the ability of pathologists and clinicians to understand the internal workings of these models [505]. Similarly, the reproducibility of model output and performance is another challenge that makes their adoption difficult. Having an interactive graphical user interface is essential in AI systems, but such interfaces are either not available entirely or do not provide output visualizations for the intermediate layers of the model. All these challenges reduce users' trust in AI systems, especially in the case of DP, where decisions directly impact diagnosis and clinical outcomes [274].

## 5.9 Conclusion

The integration of multimodal oncology data using advanced computational methods represents a pivotal advancement in the realm of DP and CPATH. The advancements in AI and ML models have enabled a more nuanced understanding of cancer's multifaceted nature. By synergistically combining clinical, radiological, pathology, and molecular data, researchers and clinicians can gain unprecedented insights into the complex dynamics of cancer, leading to more effective and personalized therapeutic strategies. However, there are multiple associated challenges to multimodal learning. Looking forward, DP stands on the brink of a new era where continuous innovation and interdisciplinary collaboration are key to harnessing the full potential of multimodal data integration. This will undoubtedly be crucial in advancing our understanding of cancer and opening new avenues for patient care and research.

## **Chapter 6: Multimodal Data Integration for Oncology in the Era of Deep Neural Networks: A Review**

### **6.1 Note to Reader**

This chapter is under review at *Frontiers in AI*, and is being reproduced with permission from *Frontiers*.

### **6.2 Introduction**

Cancer represents a significant global health challenge, characterized by the uncontrolled growth of abnormal cells, leading to millions of deaths annually. In 2023, the United States had around 1.9 million new cancer diagnoses, with cancer being the second leading cause of death and anticipated to result in approximately 1670 deaths daily [637]. However, advancements in oncology research hold the promise of preventing nearly 42% of these cases through early detection and lifestyle modifications. The complexity of cancer, involving intricate changes at both the microscopic and macroscopic levels, requires innovative approaches to its understanding and management. In recent years, the application of machine learning (ML) techniques, especially deep learning (DL), has emerged as a transformative force in oncology. DL employs deep neural networks to analyze vast datasets, offering unprecedented insights into cancer's development and progression [109, 128, 636, 497, 674]. This approach has led to the development of computer-aided diagnostic systems capable of detecting and classifying cancerous tissues in medical images, such as mammograms and MRI scans, with increasing accuracy. Beyond imaging, DL also plays a crucial role in analyzing molecular data, aiding in the prediction of treatment responses, and the identification of new biomarkers [704, 367, 496, 175, 176, 734, 68]. As the volume of oncology data continues to

grow, DL stands at the forefront of this field, enhancing our understanding of cancer, improving diagnostic precision, predicting clinical outcomes, and paving the way for innovative treatments. This review explores the latest advancements in DL applications within oncology, highlighting its potential to revolutionize cancer care [240, 122, 692, 306].

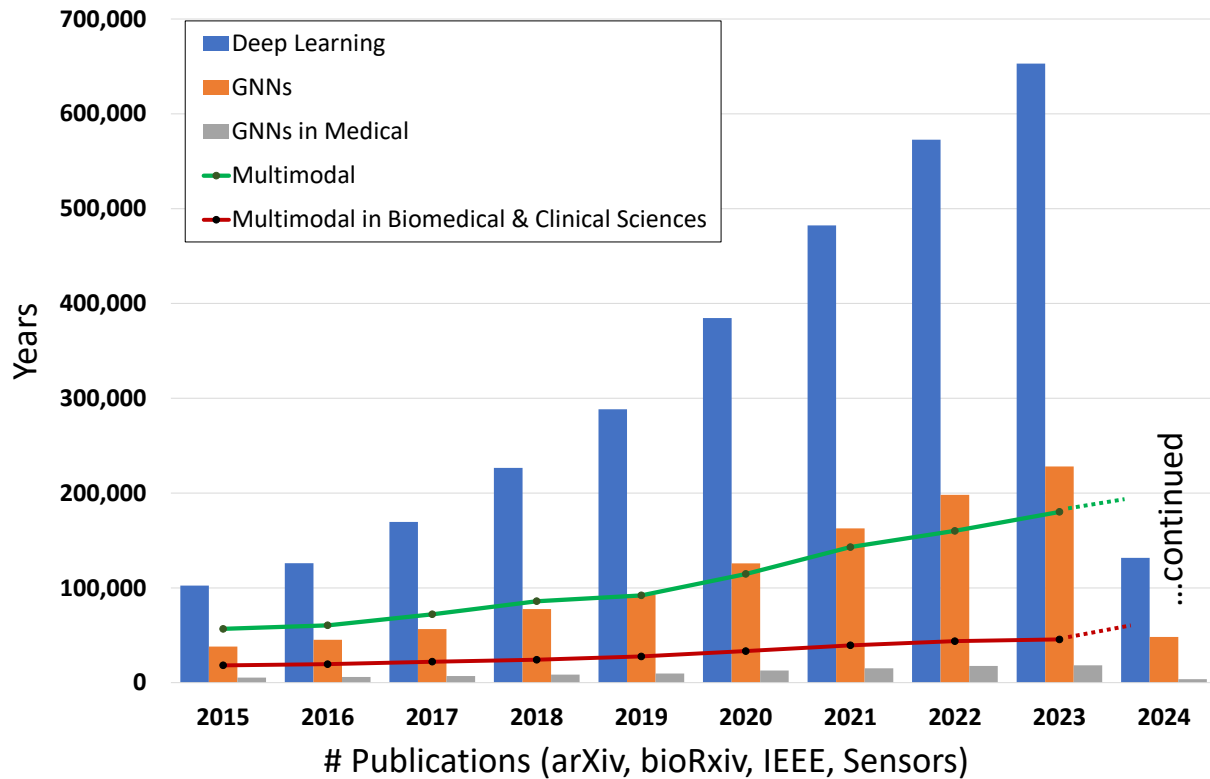


Figure 6.1 Publications involving DL, GNNs, GNNs in medical domain, and MML.

Multimodal Learning (MML) enhances task accuracy and reliability by leveraging information from various data sources or modalities [301]. This approach has witnessed a surge in popularity, as indicated by the growing body of MML-related publications (see Figure 6.1). By facilitating the fusion of multimodal data, such as radiological images, digitized pathology slides, molecular data, and electronic health records (EHR), MML offers a richer understanding of complex problems [694]. It enables the extraction and integration of relevant features that might be overlooked when analyzing data modalities separately. Recent advancements in MML, powered by Deep Neural Networks (DNNs), have shown remarkable capability in learning from diverse data sources,

including computer vision (CV) and natural language processing (NLP) [524, 91]. Prominent multimodal foundation models such as Contrastive Language-Image Pretraining (CLIP) and Generative Pretraining Transformer (GPT-4) by OpenAI have set new benchmarks in the field [564, 524]. Additionally, the Foundational Language And Vision Alignment Model (FLAVA) represents another significant stride, merging vision and language representation learning to facilitate multimodal reasoning [642]. Within the realm of oncology, innovative applications of MML are emerging. The RadGenNets model, for instance, integrates clinical and genomics data with Positron Emission Tomography (PET) scans and gene mutation data, employing a combination of Convolutional Neural Networks (CNNs) and Dense Neural Networks to predict gene mutations in Non-small cell lung cancer (NSCLC) patients [695]. Moreover, GNNs and Transformers are being explored for a variety of oncology-related tasks, such as tumor classification [365], prognosis prediction [621], and assessing treatment response [338].

Recent literature has seen an influx of survey and review articles exploring MML [90, 762, 64, 203, 276]. These works have provided valuable insights into the evolving landscape of MML, charting key trends and challenges within the field. Despite this growing body of knowledge, there remains a notable gap in the literature regarding the application of advanced multimodal DL models, such as Graph Neural Networks (GNNs) and Transformers, in the domain of oncology. Our article aims to fill this gap by offering the following contributions:

1. *Identifying large-scale MML approaches in oncology.* We provide an overview of the state-of-the-art MML with a special focus on GNNs and Transformers for multimodal data fusion in oncology.
2. *Highlighting the challenges and limitations of MML in oncology data fusion.* We discuss the challenges and limitations of implementing multimodal data-fusion models in oncology, including the need for large datasets, the complexity of integrating diverse data types, data alignment, and missing data modalities and samples.

3. *Providing a taxonomy for describing multimodal architectures.* We present a comprehensive taxonomy for describing MML architectures, including both traditional ML and DL, to facilitate future research in this area.
4. *Identifying future directions for multimodal data fusion in oncology.* We identify GNNs and Transformers as potential solutions for comprehensive multimodal integration and present the associated challenges.

By addressing these aspects, our article seeks to advance the understanding of MML’s potential in oncology, paving the way for innovative solutions that could revolutionize cancer diagnosis and treatment through comprehensive data integration. Section 6.3 covers the fundamentals of MML, including data modalities, taxonomy, data fusion stages, and neural network architectures. Section 6.4 focuses on GNNs in MML, explaining graph data, learning on graphs, architectures, and applications to unimodal and multimodal oncology datasets. Section 6.5 discusses Transformers in MML, including architecture, multimodal Transformers, applications to oncology datasets, and methods of fusing data modalities. Section 6.6 highlights challenges in MML, such as data availability, alignment, generalization, missing data, explainability, and others, followed by the multimodal data sources.

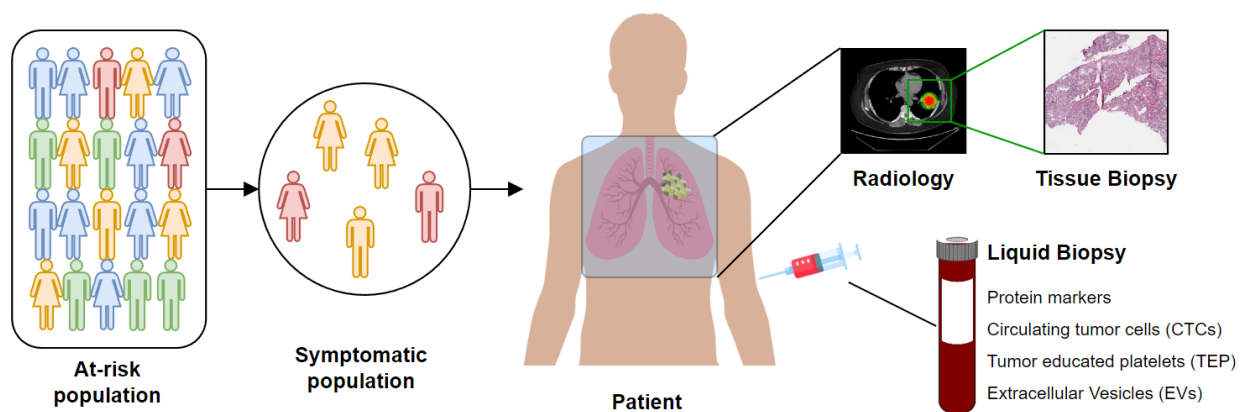


Figure 6.2 An overview of data collected from population to a tissue.

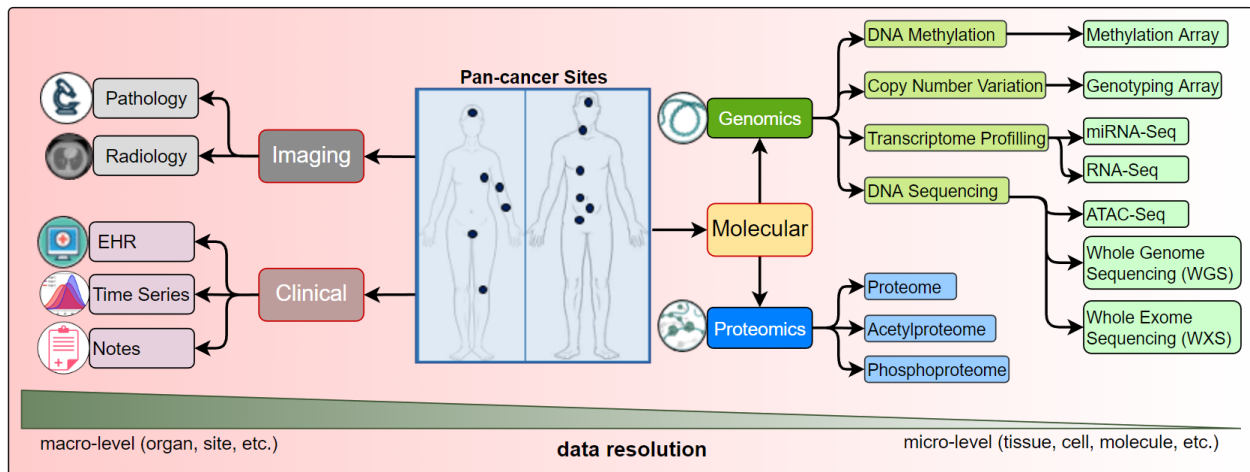


Figure 6.3 Detailed look into data modalities acquired for cancer care.

## 6.3 Fundamentals of Multimodal Learning (MML)

### 6.3.1 Data Modalities in Oncology

A data *modality* represents the expression of an entity or a particular form of sensory perception, such as the characters' visual actions, sounds of spoken dialogues, or the background music [645]. A collective notion of these modalities is called *multi-modality* [64]. Traditional data analysis and ML methods to study cancer data use single data modalities (e.g., EHR [489], radiology [734], pathology [432], or molecular, including genomics [46], transcriptomics [779], proteomics [727], etc.). However, the data is inherently multimodal, as it includes information from multiple sources or modalities that are related in many ways. Figure 6.2 provides a view of multiple modalities of cancer at various scales, from the population level to single-cell analysis. Oncology data can be broadly classified into 3 categories: clinical, molecular, and imaging, where each category provides complementary information about the patient's disease. Figure 6.3 highlights different clinical, molecular, and imaging modalities. Multimodal analysis endeavors to gain holistic insights into the disease process using multimodal data.

### 6.3.1.1 Molecular Data

Molecular data modalities provide information about the underlying genetic changes and alterations in the cancer cells [435]. Efforts toward integrating molecular data resulted in the *multi-omics* research field [737]. Two principal areas of molecular analysis in oncology are proteomics and genomics. *Proteomics* is the study of proteins and their changes in response to cancer, and it provides information about the biological processes taking place in cancer cells. *Genomics* is the study of the entire genome of cancer cells, including changes in DNA sequence, gene expression, and structural variations [90]. Other molecular modalities include transcriptomics, pathomics, radiomics and their combinations, radiogenomics, and proteogenomics. Many publicly available datasets provide access to molecular data, including the Proteomics Data Commons for proteomics data and the Genome Data Commons for genetic data [682, 259].

### 6.3.1.2 Imaging Data

Imaging modalities play a crucial role in diagnosing and monitoring cancer. The imaging category can be divided into 2 main categories: (1) radiological imaging and (2) digitized histopathology slides, referred to as Whole Slide Imaging (WSI). *Radiological* imaging encompasses various techniques such as X-rays, CT scans, MRI, PET, and others, which provide information about the location and extent of cancer within the body. These images can be used to determine the size and shape of a tumor, monitor its growth, and assess the effectiveness of treatments. *Histopathological* imaging is the examination of tissue samples obtained through biopsy or surgery [595, 733]. Digitized slides, saved as WSIs, provide detailed information about the micro-structural changes in cancer cells and can be used to diagnose cancer and determine its subtype.

### 6.3.1.3 Clinical Data

Clinical data provides information about the patient's medical history, physical examination, and laboratory results, saved in the patient's electronic health records (EHR) at the clinic. EHR consists of digital records of a patient's health information stored in a centralized database. These



records provide a comprehensive view of a patient’s medical history, past diagnoses, treatments, laboratory test results, and other information, which helps clinicians understand the disease [51]. Within EHR, time-series data may refer to the clinical data recorded over time, such as repeated blood tests, lab values, or physical attributes. Such data informs the changes in the patient’s condition and monitors the disease progression [562].

### 6.3.2 Taxonomy of MML

We follow the taxonomy proposed by William *et al.* [645] (see Figure 6.4), which defines 5 main stages of multimodal classification: preprocessing, feature extraction, data fusion, primary learner, and final classifier, as given below:

#### 6.3.2.1 *Pre-processing*

Pre-processing involves modifying the input data to a suitable format before feeding it into the model for training. It includes data cleaning, normalization, class balancing, and augmentation. Data cleaning removes unwanted noise or bias, errors, and missing data points [19]. Normalization scales the input data within a specific range to ensure that each modality contributes equally to the training [251]. Class balancing is done in cases where one class may have a significantly larger number of samples than another, resulting in a model bias toward the dominant class. Data augmentation artificially increases the size of the dataset by generating new samples based on the existing data to improve the model’s robustness and generalizability [19].

#### 6.3.2.2 *Feature Extraction*

Different data modalities may have different features, and extracting relevant features may improve model learning. Several manual and automated feature engineering techniques generate representations (or *embeddings*) for each data modality. Feature engineering involves designing features relevant to the task and extracting them from the input data. This can be time-consuming but may allow the model to incorporate prior knowledge about the problem. Text encoding techniques,

such as bag-of-words, word embeddings, and topic models [180, 819], transform textual data into a numerical representation, which can be used as input to an ML model [720]. In DL, feature extraction is learned automatically during model training[166].

#### 6.3.2.3 *Data Fusion*

Data fusion combines raw features, extracted features, or class prediction vectors from multiple modalities to create a single data representation. Fusion enables the model to use the complementary information provided by each modality and improve its learning. Data fusion can be done using early, late, or intermediate fusion. Section 6.3.3 discusses these fusion stages. The choice of fusion technique depends on the characteristics of the data and the specific problem being addressed [332].

#### 6.3.2.4 *Primary Learner*

The primary learner stage is training the model on the pre-processed data or extracted features. Depending on the problem and data, the primary learner can be implemented using various ML techniques. DNNs are a popular choice for primary learners in MML because they can automatically learn high-level representations from the input data and have demonstrated state-of-the-art performance in many applications. CNNs are often used for image and video data, while recurrent neural networks (RNNs) and Transformers are commonly used for text and sequential data. The primary learner can be implemented independently for each modality or shared between modalities, depending on the problem and data.

#### 6.3.2.5 *Final Classifier*

The final stage of MML is the classifier, which produces category labels or class scores and can be trained on the output of the primary learner or the fused data. The final classifier can be implemented using a shallow neural network, a decision tree, or an ensemble model [645]. Ensemble methods, such as stacking or boosting, are often used to improve and robustify the performance of the final classifier. Stacking involves training multiple models and then combining

their predictions at the output stage, while boosting involves repeatedly training weak learners and adjusting their weights based on the errors made by previous learners [94].

### 6.3.3 Data Fusion Strategies

Fusion in MML can be performed at different levels, including early (feature level), intermediate (model level), or late (decision level) stages, as illustrated in figure 6.4. Each fusion stage has its advantages and challenges, and the choice of fusion stage depends on the characteristics of the data and the task.

#### 6.3.3.1 *Early Fusion*

The early fusion involves merging features extracted from different data modalities into a single feature vector before model training. The feature vectors of the different modalities are combined into a single vector, which is used as the input to the ML model [645]. This approach can be used when the modalities have complementary information and can be easily aligned, such as combining visual and audio features in a video analysis application. The main challenge with early fusion is ensuring that the features extracted from different modalities are compatible and provide complementary information.

#### 6.3.3.2 *Intermediate Fusion*

Intermediate fusion involves training separate models for each data modality and then combining the outputs of these models for inference/prediction [645]. This approach is suitable when the data modalities are independent of each other and cannot be easily combined at the feature level using average, weighted average, or other methods. The main challenge with intermediate fusion is selecting an appropriate method for combining the output of different models.

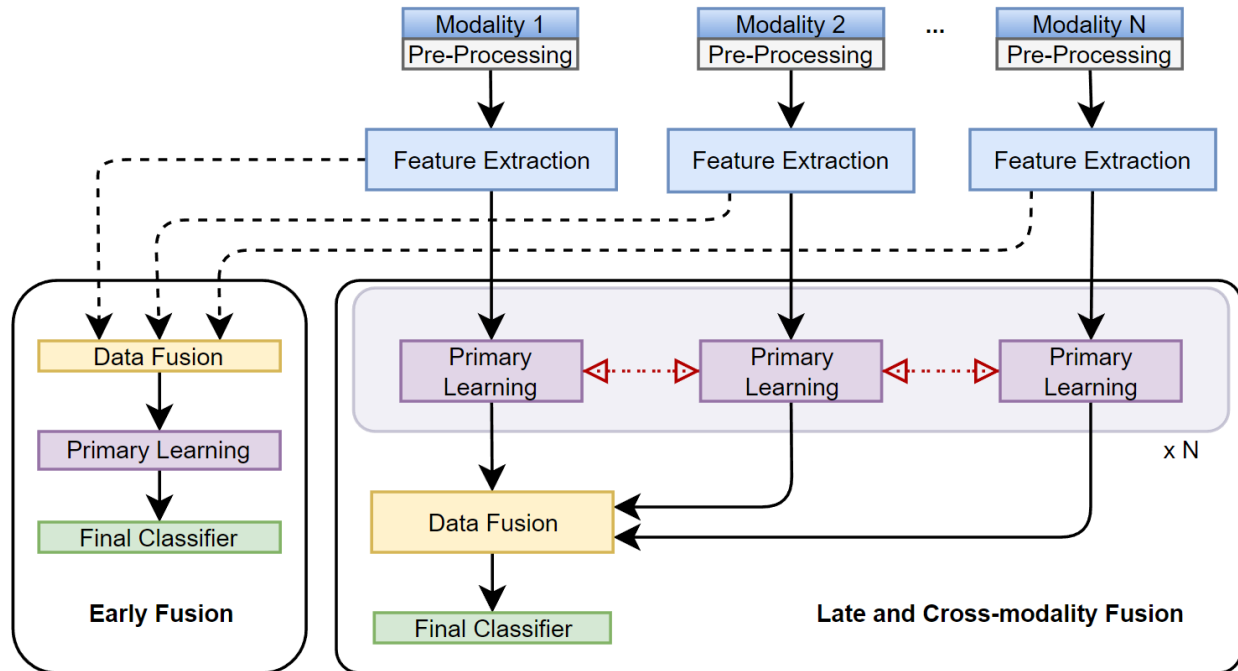


Figure 6.4 Taxonomy, stages, and techniques of multimodal data fusion.

### 6.3.3.3 Late Fusion

In late fusion, the output of each modality-specific model is used to make a decision independently. All decisions are later combined to make a final decision. This approach is suitable when the modalities provide complementary information but are not necessarily independent of each other. The main challenge with late fusion is selecting an appropriate method for combining individual predictions. This can be done using majority voting, weighted voting, or employing other ML models.

### 6.3.4 MML for Oncology Datasets

Syed *et al.* [668] used a Random Forest classifier to fuse radiology image representations learned from the singular value decomposition method with the textual annotation representation learned from the fastText algorithm for prostate and lung cancer patients. Liu *et al.* [439]

proposed a hybrid DL framework for combining breast cancer patients' genomic and pathology data using fully-connected (FC) network for genomic data, CNN for radiology data and a Simulated Annealing algorithm for late fusion. Multiview multimodal network (MVMM-Net) [649] combined 2 different modalities (low-energy and dual-energy subtracted) from contrast-enhanced spectral mammography images, each learned through CNN and late-fusion through FC network in breast cancer detection task. Yap *et al.* [774] used a late-fusion method to fuse image representations from ResNet50 and clinical representations from a random forest model for a multimodal skin lesion classification task. An award-winning work [463] on brain tumor grade classification adopted the late-fusion method (concatenation) for fusing outputs from two CNNs (radiology and pathology images).

The single-cell unimodal data alignment is one technique in MML. Jansen *et al.* devised an approach (SOMatic) to combine ATAC-seq regions with RNA-seq genes using self-organizing maps [324]. Single-Cell data Integration via Matching (SCIM) [658] matched cells in multiple datasets in low-dimensional latent space using autoencoder (AEs). Graph-linked unified embedding (GLUE) [112] model learned regulatory interactions across omics layers and aligned the cells using variational AEs. These aforementioned methods cannot incorporate high-order interactions among cells or different modalities. Single-cell data integration using multiple modalities is mostly based on AEs (scDART [800], Cross-modal Autoencoders [768], Mutual Information Learning for Integration of Single Cell Omics Data (SMILE) [765]).

## **6.4 Graph Neural Networks (GNNs) in Multimodal Learning**

Graphs are commonly used to represent the relational connectivity of any system that has interacting entities [412]. Graphs have been used in various fields, such as to study brain networks [211], analyze driving maps [178], and explore the structure of DNNs themselves [735]. GNNs are specifically designed to process data represented as a graph [717], which makes them well-suited for analyzing multimodal oncology data as each data modality (or sub-modality) can

be considered as a single node and the structures/patterns that exist between data modalities can be modeled as edges [203].

#### 6.4.1 The Graph Data

A graph is represented as  $G=(V, E)$  having node-set  $V=\{v_1, v_2, \dots, v_n\}$ , where node  $v$  has feature vector  $\mathbf{x}_v$ , and edge set  $E=\{(v_i, v_j) \mid v_i, v_j \in V\}$ . The neighborhood of node  $v$  is defined as  $N(v)=\{u \mid (u, v) \in E\}$ .

##### 6.4.1.1 Graph Types

As illustrated in figure 6.5(a), the common types of graphs include undirected, directed, homogeneous, heterogeneous, static, dynamic, unattributed, and attributed. *Undirected graphs* comprise undirected edges, i.e., the direction of relation is not important between any ordered pair of nodes. In the *directed graphs*, the nodes have a directional relationship(s). Homogeneous graphs have the same type of nodes, whereas heterogeneous graphs have different types of nodes within a single graph [771]. Static graphs do not change over time with respect to the existence of edges and nodes. In contrast, dynamic graphs change over time, resulting in changes in structure, attributes, and node relationships. *Unattributed graphs* have unweighted edges, indicating that the weighted value for all edges in a graph is the same, i.e., 1 if present, 0 if absent. *Attributed graphs* have different edge weights that capture the strength of relational importance [717].

##### 6.4.1.2 Tasks for Graph Data

In figure 6.5(b), we present 3 major types of tasks defined on graphs, including (1) *node-level tasks* - these may include node classification, regression, clustering, attributions, and generation, (2) *edge-level task* - edge classification and prediction (presence or absence) are 2 common edge-level tasks, (3) *graph-level tasks* - these tasks involve predictions on the graph level, such as graph classification and generation.

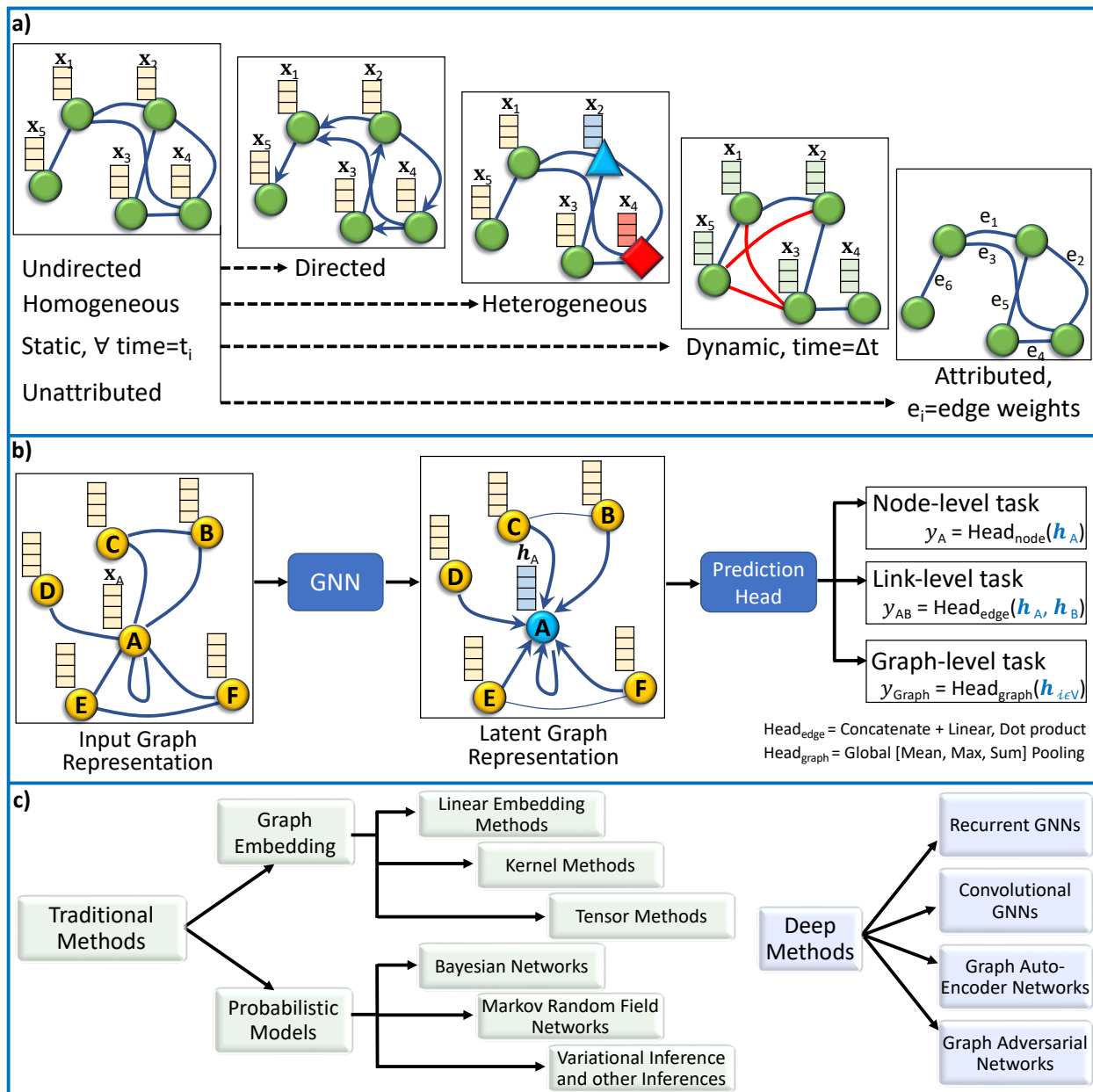


Figure 6.5 Common graphs, graph learning tasks, and types of representation learning.

### 6.4.2 ML for Graph Data

Representing data as graphs can enable capturing and encoding the relationships among entities of the samples [752]. Based on the way the nodes are encoded, representation learning on graphs

can be categorized into the traditional (or shallow) and DNN-based methods, as illustrated in Figure 6.5(c) [334, 752].

#### 6.4.2.1 *Traditional (Shallow) Methods*

These methods usually employ classical ML methods, and their two categories commonly found in the literature are *graph embedding* and *probabilistic methods*. Graph embedding methods represent a graph with low-dimensional vectors (graph embedding and node embedding), preserving the structural properties of the graph. The learning tasks in graph embedding usually involve dimensionality reduction through linear (principal component or discriminant analysis), kernel (nonlinear mapping), or tensor (higher-order structures) methods [334]. Probabilistic graphical methods use graph data to represent probability distribution, where nodes are considered random variables, and edges depict the probability relations among nodes [334]. Bayesian networks, Markov's networks, variational inference, variable elimination, and others are used in probabilistic methods [334].

#### 6.4.2.2 *DNN-based Methods - GNNs*

GNNs are gaining popularity in the ML community, as evident from figure 6.1. In GNNs, the information aggregation from the neighborhood is fused into a node's representation. Traditional DL methods such as CNNs and their variants have shown remarkable success in processing the data in Euclidean space; however, they fail to perform well when faced with non-Euclidean or relational datasets. Compared to CNNs, where the locality of the nodes in the input is fixed, GNNs have no canonical ordering of the neighborhood of a node. They can learn the given task for any permutation of the input data, as depicted in figure 6.6. GNNs often employ a message-passing mechanism in which a node's representation is derived from its neighbors' representations via a



recursive computation. The message passing for a GNN is given as follows:

$$\mathbf{h}_v^{(l+1)} = \sigma \left( W_l \sum_{u \in N(v)} \frac{\mathbf{h}_u^{(l)}}{|N(v)|} + B_l \mathbf{h}_v^{(l)} \right) \quad (6.1)$$

where  $h_v^{(l+1)}$  is the updated embedding of node  $v$  after  $l+1$  layer,  $\sigma$  is the non-linear function (e.g., rectified linear unit or ReLU),  $h_u^{(l)}$  and  $h_v^{(l)}$  represent the embeddings of nodes  $u$  and  $v$  at layer  $l$ .  $W_l$  and  $B_l$  are the trainable weight matrices for neighborhood aggregation and (self)hidden vector transformation, respectively. The message passing can encode high-order structural information in node embedding through multiple aggregation layers. GNNs smooth the features by aggregating neighbors' embedding and filter eigenvalues of graph Laplacian, which provides an extra denoising mechanism [464]. GNNs comprise multiple permutation equivariant and invariant functions, and they can handle heterogeneous data [335]. As described earlier, traditional ML models deal with Euclidean data. In oncology data, the correlations may not exist in Euclidean space; instead, its features may be highly correlated in the non-Euclidean space [775]. Based on the information fusion and aggregation methodology, GNNs-based deep methods are classified into the following:

*Recurrent GNNs* (RecGNNs) are built on top of the standard Recurrent Neural Network (RNN) by combining with GNN. RecGNNs can operate on graphs with variable sizes and topologies. The recurrent component of the RecGNN captures temporal dependencies and learns latent states over time, whereas the GNN component captures the local structure. The information fusion process is repeated a fixed number of times until an equilibrium or the desired state is achieved [271]. RecGNNs employ the model given by:

$$\mathbf{h}_v^{(l+1)} = \text{RecNN} \left( \mathbf{h}_v^{(l)}, \mathbf{Msg}_{N(v)}^{(l)} \right), \quad (6.2)$$

where, RecNN is any RNN, and  $Msg_{N(v)}^{(l)}$  is the neighborhood message-passing at layer  $l$ .

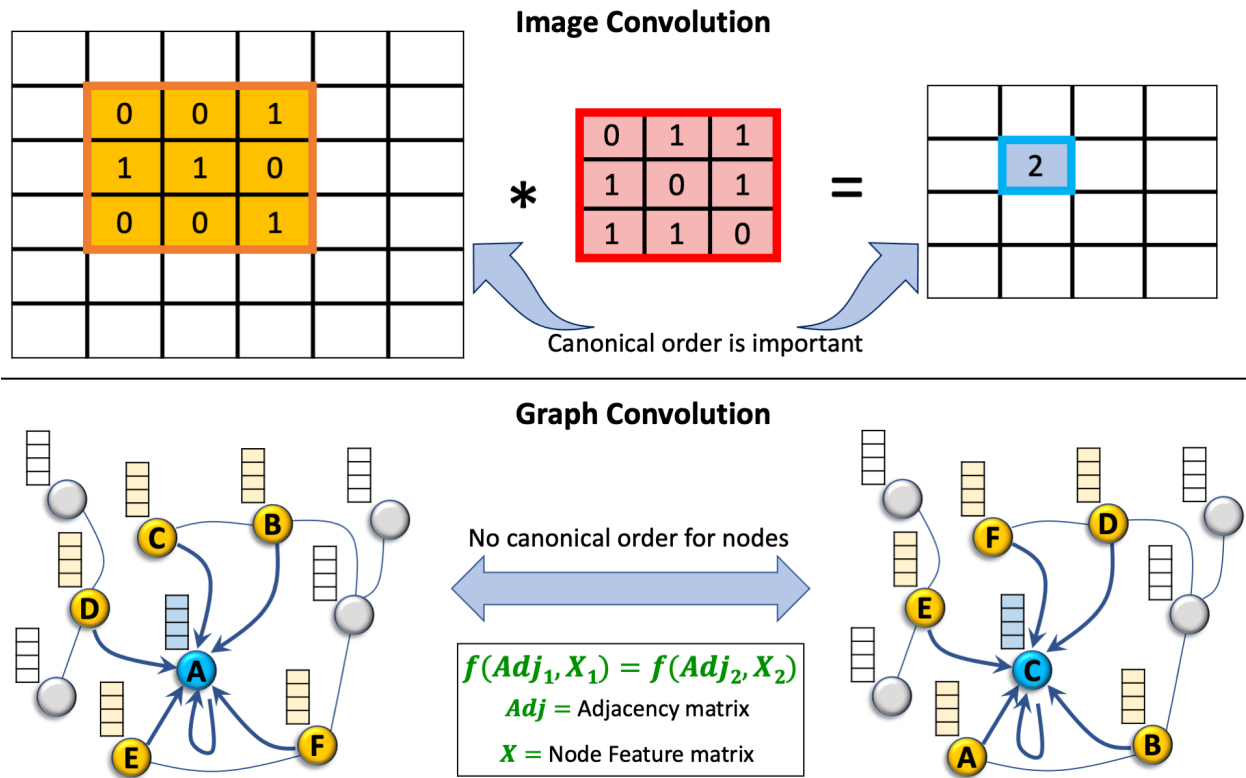


Figure 6.6 Convolution operation for graphs vs. image data.

*Convolutional GNNs* (ConvGNNs) undertake the convolution operation on graphs by aggregating neighboring nodes' embeddings through a stack of multiple layers. ConvGNNs use the symmetric and normalized summation of the neighborhood and self-loops for updating the node embeddings given by:

$$\mathbf{h}_v^{(l+1)} = \sigma \left( W_l \sum_{u \in N(v) \cup v} \frac{\mathbf{h}_u}{\sqrt{|N(v)||N(u)|}} \right). \quad (6.3)$$

The ConvGNN can be spatial or spectral, depending on the type of convolution they implement. Convolution in spatial ConvGNNs involves taking a weighted average of the neighboring vertices. Examples of spatial ConvGNNs include GraphSAGE[271], Message Passing Neural Network (MPNN)[244], and Graph Attention Network (GAT) [708]. The spectral ConvGNNs operate in the spectral domain by using the eigendecomposition of the graph Laplacian matrix. The convolution

operation is performed on the eigenvalues, which can be high-dimensional. Popular spectral ConvGNNs are ChebNet [168] and Graph Convolutional Network (GCN)[375]. An interesting aspect of these approaches is representational containment, which is defined as: convolution  $\subseteq$  attention  $\subseteq$  message passing.

*Graph Auto-Encoder Networks* (GAEs) are unsupervised graph learning networks for dimensionality reduction, anomaly detection, and graph generation. They are built on top of the standard AEs to work with graph data. The encoder component of the GAE maps the input graph to a low-dimensional latent space, while the decoder component maps the latent space back to the original graph [540].

*Graph Adversarial Networks* (GraphANs) are based on Generative Adversarial Networks and designed to work with graph-structured data and can learn to generate new graphs with similar properties to the input data. The generator component of the GraphAN maps a random noise vector to a new graph, while the discriminator component tries to distinguish between the generated vs. the actual input. The generator generates graphs to fool the discriminator, while the discriminator tries to classify the given graph as real or generated.

*Other GNNs* may include scalable GNNs [460], dynamic GNNs [609], hypergraph GNNs [60], heterogeneous GNNs [745], and many others [465].

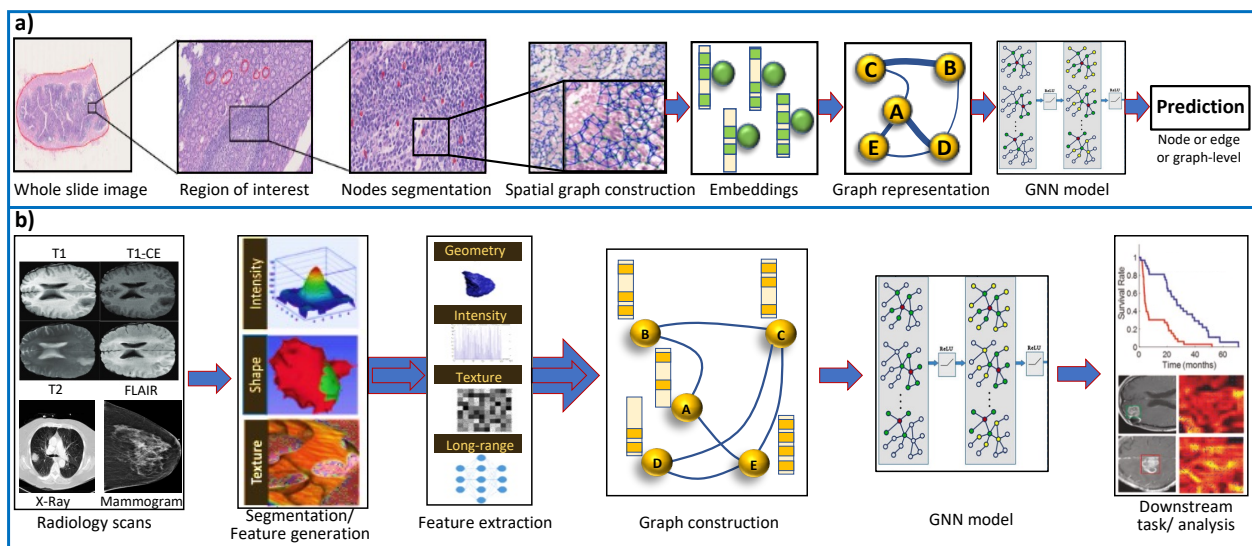


Figure 6.7 Data processing pipeline for medical images using GNNs.

### 6.4.3 GNNs and ML Using Unimodal Oncology Datasets

#### 6.4.3.1 Pathology Datasets

Traditionally, CNN-based models are used to learn features from digital pathology data. However, unlike GNNs, CNNs fail to capture the global contextual information important in the tissue phenotypical and structural micro and macro environment [15]. For using histology images in GNNs, the cells, tissue regions, or image patches are depicted as nodes. The relations and interactions among these nodes are represented as (un)weighted edges. Usually, a graph of the patient histology slide is used along with a patient-level label for training a GNN, as illustrated in Figure 6.7(a). Here, we review a few GNN-based pathology publications representative of a trove of works in this field. Histograms [42] used breast cancer histology data to distinguish cancerous and non-cancerous images. Pre-trained VGG-UNet was used for nuclei detection, micro-features of the nuclei were used as node features, and Euclidean distance among nuclei was incorporated as edge features. The resulting cell graphs were used to train the GCN-based robust spatial filtering (RSF) model, which performed superior to the CNN-based classification. Wang *et al.* [724] analyzed grade classification in tissue micro-arrays of prostate cancer using the weakly-supervised technique on a variant of GraphSAGE with self-attention pooling (SAGPool). Cell-Graph Signature ( $CG_{signature}$ ) [730] predicted patient survival in gastric cancer using cell-graphs of multiplexed immunohistochemistry images processed through two types of GNNs (GCNs and GINs) with two types of pooling (SAGPool, TopKPool). Besides the above-mentioned cell graphs, there is an elaborate review of GNN-based tissue graphs or patch-graphs methods implemented on unimodal pathology cancer data given in [15]. Instead of individual cell- and tissue-graphs, a combination of the multilevel information in histology slides can help understand the intrinsic features of the disease.

### 6.4.3.2 Radiology Datasets

GNNs have been used in radiology-based cancer data for segmentation, classification, and prediction tasks, especially on X-rays, mammograms, MRI, PET, and CT scans. Figure 6.7(b) illustrates a general pipeline of using radiology-based data to train GNNs. Here we give a non-exhaustive review of GNNs-based works on radiological oncology data as a single modality input. Mo *et al.* [492] proposed a framework that improved the liver cancer lesion segmentation in the MRI-T1WI scans through guided learning of MRI-T2WI modality priors. Learned embeddings from fully convolutional networks on separate MRI modalities are projected into the graph domain for learning by GCNs through the co-attention mechanism and finally to get the refined segmentation by re-projection. Radiologists usually review radiology images by zooming into the region of interest (ROIs) on high-resolution monitors. Du *et al.* [196] used a hierarchical GNN framework to automatically zoom into the abnormal lesion region of the mammograms and classify breast cancer. The pre-trained CNN model extracts image features, whereas a GAT model is used to classify the nodes for deciding whether to zoom in or not based on whether it is benign or malignant. Based on the established knowledge that lymph nodes (LNs) have connected lymphatic system and LNs cancer cells spread on certain pathways, Chao *et al.* [123] proposed a lymph node gross tumor volume learning framework. The framework was able to delineate the LN appearance as well as the inter-LN relationship. The end-to-end learning framework was superior to the state-of-the-art on esophageal cancer radiotherapy dataset. Tian *et al.* [685] suggested interactive segmentation of MRI scans of prostate cancer patients through a combination of CNN and two GCNs. CNN model outputs a segmentation feature map of MRI, and the GCNs predict the prostate contour from this feature map. Saueressig *et al.* [613] used GNNs to segment brain tumors in 3D MRI images, formed by stacking different modalities of MRI (T1, T2, T1-CE, FLAIR) and representing them as supervoxel graph. The authors reported that GraphSAGE-pool was best for segmenting brain tumors. Besides radiology, a parallel field of radiomics has recently gained attraction. Radiomics is the automated extraction of quantitative features from radiology scans. A survey of radiomics and radiogenomic analysis on brain tumors is presented in [643].

### 6.4.3.3 Molecular Datasets

Graphs are a natural choice for representing molecular data such as omic-centric (DNA, RNA, or proteins) or single-cell centric. Individual modalities are processed separately to generate graph representations that are then processed through GNNs followed by the classifier to predict the downstream task, as illustrated in Figure 6.8. One method of representing proteins as graphs is to depict the amino acid residue in the protein as the node and the relationship between residues denoted by edge [226]. The residue information is depicted as node embedding, whereas the relational information between two residues is represented as the edge feature vector. Fout *et al.* [226] used spatial ConvGNNs to predict interfaces between proteins which is important in drug discovery problems. Deep predictor of drug-drug interactions (DPDDI) predicted the drug-drug interactions using GCN followed by a 5-layer classical neural network [220]. Molecular pre-training graph net (MPG) [415] is a powerful framework based on GNN and Bidirectional Encoder Representations from Transformers (BERT) to learn drug-drug and drug-target interactions. Graph-based Attention Model (GRAM) [142] handled the data inefficiency by supplementing EHRs with hierarchical knowledge in the medical ontology. A few recent works have applied GNNs to single-cell data. scGCN [652] is a knowledge transfer framework in single-cell omics data such as mRNA or DNA. scGNN [725] processed cell-cell relations through GNNs for the task of missing-data imputation and cell clustering on single-cell RNA sequencing (scRNA-seq) data.

### 6.4.4 MML - Data Fusion at Pre-Learning Stage

The first and most primitive form of MML is the pre-learning fusion (see Figure 6.4), where features extracted from individual modalities of data are merged, and the fused representations are then used for training the multimodal primary learner model. In the context of GNNs being the primary learning model, the extraction step of individual modality representations can be hand-engineered (e.g., dimensionality reduction) or learned by DL models (e.g., CNNs, Transformers). Cui *et al.* [163] proposed a GNN-based early fusion framework to learn latent representations from radiological and clinical modalities for Lymph node metastasis (LNM) prediction in esophageal

squamous cell carcinoma (ESCC). The extracted features from the two modalities using UNet and CNN-based encoders were fused together with category-wise attention as node representation. The message passing from conventional GAT and correlation-based GAT learned the neighborhood weights. The attention attributes were used to update the final node features before classification by a 3-layer fully connected network. For Autism spectrum disorder, Alzheimer’s disease, and ocular diseases, a multimodal learning framework called Edge-Variational GCN (EV-GCN) [300] fuses the radiology features extracted from fMRI images with clinical feature vectors for each patient. An MLP-based pairwise association encoder is used to fuse the input feature vectors and to generate the edge weights of the population graph. The partially labeled population graph is then processed through GCN layers to generate the diagnostic graph of patients.

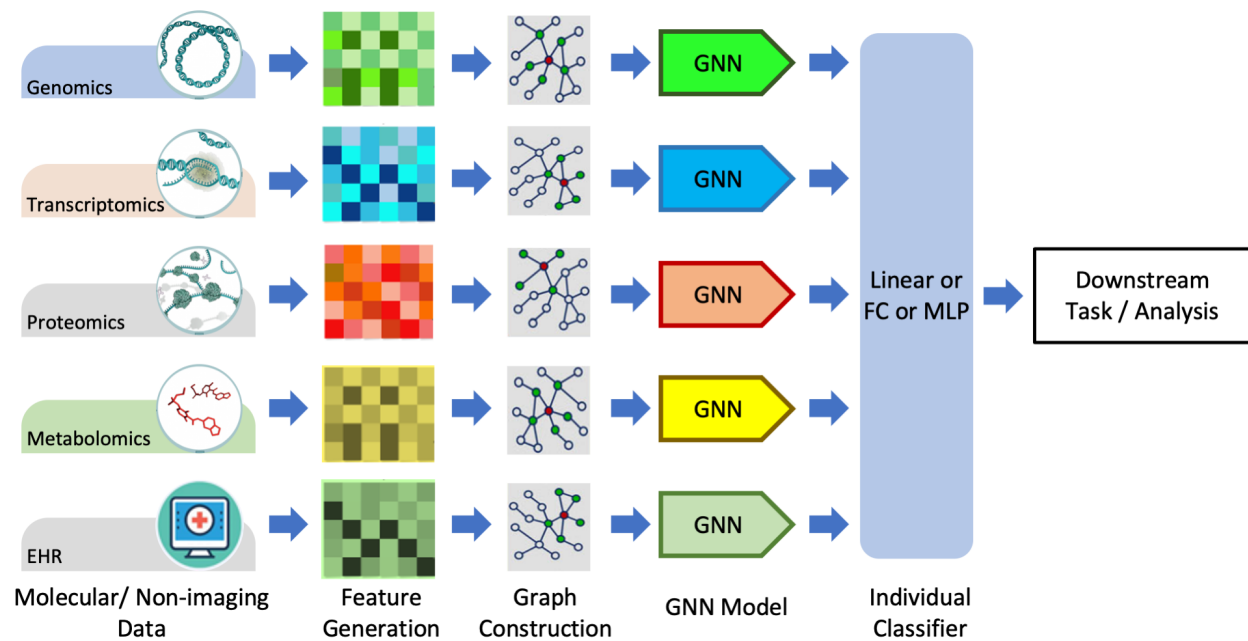


Figure 6.8 Graph data processing pipeline for molecular and textual data.

#### 6.4.5 MML - Data Fusion Using Cross-Modality Learning

Cross-MML involves intermediate fusion and/or cross-links among the models being trained on individual modalities (see Figure 6.4). For this survey, we consider the GNN-based hierarchical

learning mechanisms as the cross-MML methods. Hierarchical frameworks involve learning for one modality and using the learned latent embeddings in tandem with other data modalities sequentially to get the final desired low-dimensional representations. Lian *et al.* [423] used a sequential learning framework where tumor features learned from CT images using the ViT model were used as node features of the patient population graph for subsequent processing by the GraphSAGE model. The hierarchical learning from radiological and clinical data using Transformer-GNN outperformed the ResNet-Graph framework in survival prediction of early-stage NSCLC. scMoGNN [747] is the first method to apply GNNs in multimodal single-cell data integration using a cross-learning fusion-based GNN framework. Officially winning first place in modality prediction task at the NeurIPS 2021 competition, scMoGNN showed superior performance on various tasks by using paired data to generate cell-feature graphs. Hierarchical cell-to-tissue-graph network (HACT-Net) combined the low-level cell-graph features with the high-level tissue-graph features through two hierarchical GINs on breast cancer multi-class prediction [543]. Data imputation, a method of populating the missing values or false zero counts in single-cell data mostly done using DL autoencoders (AE) architecture, has recently been accomplished using GNNs. scGNN [725] used imputation AE and graph AE in an iterative manner for imputation, and GraphSCI [577] used GCN with AE to impute the single-cell RNA-seq data using the cross-learning fusion between the GCN and the AE networks. Clustering is a method of characterizing cell types within a tissue sample. Graph-SCC [790] clustered cells based on scRNA-seq data through self-supervised cross-learning between GCN and a denoising AE network. Recently, a multilayer GNN framework, Explainable Multilayer GNN (EMGNN), has been proposed for cancer gene prediction tasks using multi-omics data from 16 different cancer types [125].

#### 6.4.6 MML - Data Fusion in Post-Learning Regime

Post-learning fusion methods include processing individual data modalities and later fusing them for the downstream predictive task [689]. In the post-learning fusion paradigm, the hand-crafted features perform better than the deep features when the dimensionality of input data is



low, and vice versa [689]. Many interesting GNN-based works involving the post-learning fusion mechanism have recently been published. Decagon [820] used a multimodal approach on GCNs using proteins and drug interactions to predict exact side effects as a multi-relational link prediction task. Drug–target affinity (DTA) [503] experimented with four different flavors of GNNs (GCN, GAT, GIN, GAT-GCN) along with a CNN to fuse together molecular embeddings and protein sequences for predicting drug-target affinity. PathomicFusion [135] combined the morphological features extracted from image patches (using CNNs), cell-graph features from cell-graphs of histology images (GraphSAGE-based GCNs), and genomic features (using a feed-forward network) for survival prediction on glioma and clear cell renal cell carcinoma. Shi *et al.* [631] proposed a late-fusion technique to study screening of cervical cancer at early stages by using CNNs to extract features from histology images, followed by K-means clustering to generate graphs which are processed through two-layer GCN. BDR-CNN-GCN (batch normalized, dropout, rank-based pooling) [798] used the same mammographic images to extract image-level features using CNN and relation-aware features using GCN. The two feature sets are fused using a dot product followed by a trainable linear projection for breast cancer classification. Under the umbrella of multi-omics data, many GNN-based frameworks have been proposed recently. Molecular omics network(MOOMIN)[596], a multi-modal heterogeneous GNN to predict oncology drug combinations, processed molecular structure, protein features, and cell lines through GCN-based encoders, followed by late-fusion using a bipartite drug-protein interaction graph. Multi-omics graph convolutional networks (MOGONET) [728] used a GCN-GAN late fusion technique for the classification of four different diseases, including three cancer types: breast, kidney, and glioma. Leng *et al.* [401] extended MOGONET to benchmark three multi-omics datasets on two different tasks using sixteen DL networks and concluded that GAT-based GNN had the best classification performance. Multi-Omics Graph Contrastive Learner(MOGCL) [572] used graph structure and contrastive learning information to generate representations for improved downstream classification tasks on the breast cancer multi-omics dataset using late-fusion. Similar to MOGCL, Park *et al.* [541] developed a

GNN-based multi-omics model that integrated mRNA expression, DNA methylation, and DNA sequencing data for NSCLC diagnosis.

## 6.5 Transformers in MML

Transformers are attention-based DNN models originally proposed for NLP [706]. Transformers implement scaled dot-product of the input with itself and can process various types of data in parallel [706]. Transformers can handle sequential data and learn long-range dependencies, making them well-suited for tasks such as language translation, language modeling, question answering, and many more [529]. Unlike Recurrent Neural Networks (RNNs) and CNNs, Transformers use self-attention operations to weigh the importance of different input tokens (or embeddings) at each time step. This allows them to handle sequences of arbitrary length and to capture dependencies between input tokens that are far apart in the sequence [706]. Transformers can be viewed as a type of GNN [762]. Transformers are used to process other data types, such as images [192], audio [795], and time-series analysis [14], resulting in a new wave of multi-modal applications. Transformers can handle input sequences of different modalities in a unified way, using the same self-attention mechanism, which processes the inputs as a fully connected graph [762]. This allows Transformers to capture complex dependencies between different modalities, such as visual and textual information in visual question-answering (VQA) tasks [459].

Pre-training Transformers on large amounts of data, using unsupervised or self-supervised learning, and then fine-tuning for specific downstream tasks, has led to the development of foundation models [90], such as BERT [180], GPT [567], RoBERTa [819], CLIP [564], T5 [571], BART [404], BLOOM [618], ALIGN [329], CoCa [782] and more. Multimodal Transformers are a recent development in the field of MML, which extends the capabilities of traditional Transformers to handle multiple data modalities. The inter-modality dependencies are captured by the cross-attention mechanism in multimodal Transformers, allowing the model to jointly reason and extract rich data representations. There are various types of multimodal Transformers, such as Unified Transformer

(UniT) [296], Multi-way Multimodal Transformer (MMT) [676], CLIP [564], Flamingo [20], CoCa [782], Perceiver IO [318], and GPT-4[524].

### 6.5.1 Model Architecture

The original Transformer (Figure 6.9) was composed of multiple encoder and decoder blocks, each made up of several layers of self-attention and feed-forward neural networks. The encoder takes the input sequence and generates hidden representations, which are then fed to the decoder. The decoder generates the output sequence by attending to the encoder’s hidden representations and the previous tokens (i.e., auto-regressive). The self-attention operation (or scaled dot-product) is a crucial component of the Transformer. It determines the significance of each element in the input sequence with respect to the whole input. Self-attention operates by computing a weighted sum of the input sequence’s hidden representations, where the weights are determined by the dot product between the *query* vector and the *key* vector, followed by a scaling operation to stabilize the gradients. The resulting weighted sum is multiplied by a *value* vector to obtain the output of the self-attention operation. There has been a tremendous amount of work on various facets of Transformer architecture. The readers are referred to relevant review papers [529, 762, 272, 235].

### 6.5.2 Multimodal Transformers

Self-attention allows a Transformer model to process each input as a fully connected graph and attend to (or equivalently learn from) the global patterns present in the input. This makes Transformers compatible with various data modalities by treating each token (or its embedding) as a node in the graph. To use Transformers for a data modality, we need to tokenize the input and select an embedding space for the tokens. Tokenization and embedding selections are flexible and can be done at multiple granularity levels, such as using raw features, ML-extracted features, patches from the input image, or graph nodes. Table 6.1 summarizes some common practices used for various data types in oncology datasets. Handling inter-modality interactions is the main challenge in developing multimodal Transformer models. Usually, it is done through one of these

Table 6.1 Oncology data tokenization and embeddings selection techniques

Data Modalities	Tokenization Level	Token Embeddings Model
Pathology images	Patch	CNNs [137]
Radiology images	Patch	CNNs [759]
EHR data	ICD code	GNNs [626], ML models [579]
-Omics	Graphs K-mers	GNNs [340] ML model [327]
Clinical notes	Word	BERT [180], RoBERTa [819], BioBERT [396]

fusion methods: *early fusion* of data modalities, *cross-attention*, *hierarchical attention*, and *late fusion*, as illustrated in Figure 6.10. In the following, we present and compare data processing steps for these four methods using two data modalities as an example. The same analysis can be extended to multiple modalities.

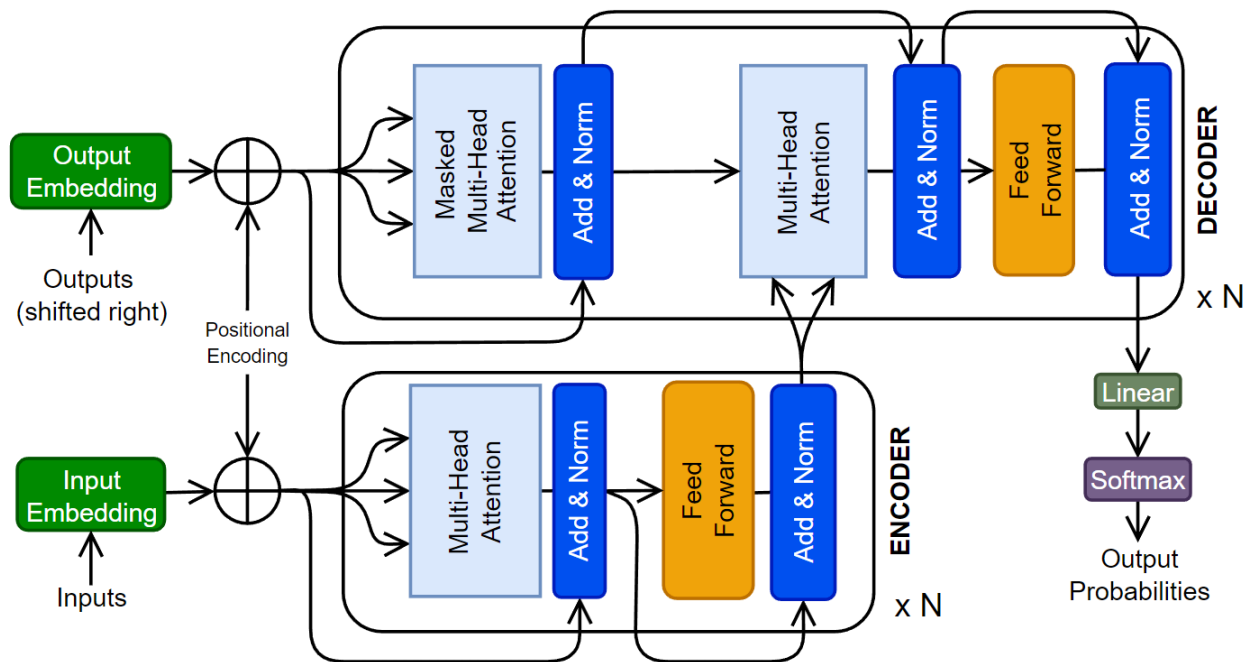


Figure 6.9 The original Transformer architecture is presented.

### 6.5.2.1 Early Fusion

Early fusion is the simplest way to combine data from multiple modalities. The data from different modalities are concatenated to a single input before being fed to the Transformer model, which processes the input as a single entity. Mathematically, the concatenation operation is represented as  $x_{cat}=[x_1, x_2]$ , where  $x_1$  and  $x_2$  are the inputs from two data modalities, and  $x_{cat}$  is the concatenated input to the model. Early fusion is simple and efficient. However, it assumes that all modalities are equally important and relevant for the task at hand [342], which may not always be practically true [810].

### 6.5.2.2 Cross-Attention Fusion

Cross-attention is a relatively more flexible approach to modeling the interactions between data modalities and learning their joint representations. The self-attention layers attend to different modalities at different stages of data processing. Cross-attention allows the model to selectively attend to different modalities based on their relevance to the task [414] and capture complex interactions between the modalities [591].

### 6.5.2.3 Hierarchical Fusion

Hierarchical fusion is a complex approach to combining multiple modalities. For instance, the Depth-supervised Fusion Transformer for Salient Object Detection (DFTR) employs hierarchical feature extraction to improve salient object detection performance by fusing low-level spatial features and high-level semantic features from different scales [817]. Yang *et al.*[769] introduced a hierarchical approach to fine-grained classification using a fusion Transformer. Furthermore, the Hierarchical Multimodal Transformer (HMT) for video summarization can capture global dependencies and multi-hop relationships among video frames [801].

### 6.5.2.4 Late Fusion

In late fusion, each data modality is processed independently by its own Transformer model, the branch outputs are concatenated and passed through the final classifier. Late fusion allows the model to capture the unique features of each modality while still learning their joint representation. Sun *et al.* proposed a multi-modal adaptive late fusion Transformer network for estimating the levels of depression [663]. Their model extracts long-term temporal information from audio and visual data independently and then fuses weights at the end to learn a joint representation of data.

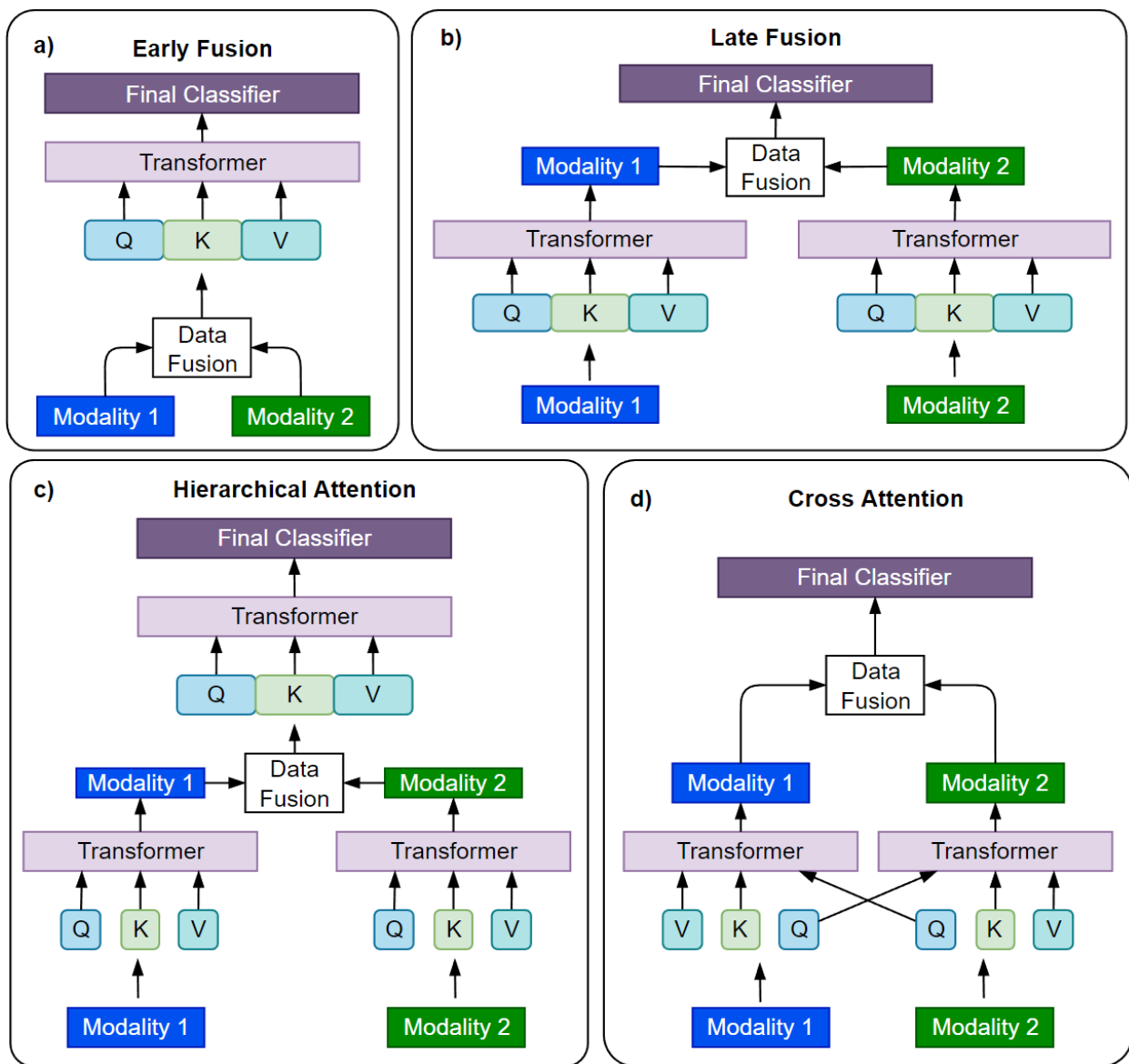


Figure 6.10 Four different fusion strategies using multimodal Transformers.

### 6.5.3 Transformers for Processing Oncology Datasets

Transformers have been successfully applied to various tasks in oncology, including cancer screening, diagnosis, prognosis, treatment selection, and prediction of clinical variables [90, 628, 424, 423, 137]. For instance, a Transformer-based model was used to predict the presence and grade of breast cancer using a combination of imaging and genomics data [90]. TransMIL [628], a Transformer model, was proposed to process histopathology images using self-attention to learn and classify histopathology slides by overcoming the challenges faced by multi-instance learning (MIL). Recently, a Transformer and convolution parallel network, TransConv [424], was proposed for automatic brain tumor segmentation using MRI data. Transformers and GNNs have also been combined in MML for early-stage NSCLC prognostic prediction using the patient’s clinical and pathological features and by modeling the patient’s physiological network [423]. Similarly, a multimodal co-attention Transformer was proposed for survival prediction using WSIs and genomic sequences [137]. The authors used a co-attention mechanism to learn the interactions between the two data modalities.

## 6.6 MML - Challenges and Opportunities

Learning from multimodal oncology data is a complex and rapidly growing field that presents both challenges and opportunities. While MML has shown significant promise, there are many challenges owing to the inductive biases of the ML models [202]. In this context, we present major challenges of MML in oncology settings that, if addressed, could unlock the full potential of this emerging field.

### 6.6.1 Large Amounts of High-quality Data

DL models are traditionally trained on large datasets with enough samples for training, validation, and testing, such as JFT-300M [662] and YFCC100M [683], which are not available in the cancer domain. For example, the largest genomics data repository, the Gene Expression Omnibus

(GEO) database, has approximately 1.1 million samples with the keyword ‘cancer’ compared to 3 billion images in JFT-300M [333]. Annotating medical and oncology data is a time-consuming and manual process that requires significant expertise in many different areas of medical sciences. Factors like heterogeneity of the disease, noise in data recording, background, and training of medical professionals leading to inter- and intra-operator variability cause lack of reproducibility and inconsistent clinical outcomes [431].

### 6.6.2 Data Registration and Alignment

Data alignment and registration refer to the process of combining and aligning data from different modalities in a useful manner [803]. In multimodal oncology data, this process involves aligning data from multiple modalities such as CT, MRI, PET, and WSIs, as well as genomics, transcriptomics, and clinical records. Data registration involves aligning the data modalities to a common reference frame and may involve identifying common landmarks or fiducial markers. If the data is not registered or aligned correctly, it may be difficult to fuse the information from different modalities accurately [425].

### 6.6.3 Pan-Cancer Generalization and Transference

Transference in MML aims to transfer knowledge between modalities and their representations to improve the performance of a model trained on a primary modality [425]. Because of the unique characteristics of each cancer type and site, it is challenging to develop models that can generalize across different cancer sites. Furthermore, the models trained on a specific modality, such as radiology images, will not perform well with other imaging modalities, such as histopathology slides. Fine-tuning the model on a secondary modality, multimodal co-learning, and model induction are techniques to achieve transference and generalization [743]. To overcome this challenge, mechanisms for improved universality of ML models need to be devised.



#### 6.6.4 Missing Data Samples and Modalities

The unavailability of one or more modalities or the absence of samples in a modality affects the model learning, as most of the existing DL models cannot process the “missing information”. This requirement, in turn, constrains the already insufficient size of datasets in oncology. Almost all publicly available oncology datasets have missing data for a large number of samples [333]. Various approaches for handling missing data samples and modalities are gradually gaining traction. However, this is still an open challenge [490].

#### 6.6.5 Imbalanced Data

Class imbalance refers to the phenomenon when one class (e.g., cancer negative/positive) is represented significantly more in the data than another class. Class imbalance is common in oncology data [490]. DL models struggle to classify underrepresented classes accurately. Techniques such as data augmentation, ensemble, continual learning, and transfer learning are used to counter the class imbalance challenge [490].

#### 6.6.6 Explainability and Trustworthiness

The explainability in DL, e.g., how GNNs and Transformers make a specific decision, is still an area of active research [413, 505]. GNNExplainer [777], PGM-Explainer [714], and SubgraphX [783] are some attempts to explain the decision-making process of GNNs. The explainability methods for Transformers have been analyzed in [587]. Existing efforts and a roadmap to improve the trustworthiness of GNNs have been presented in the latest survey [792]. However, the explainability and trustworthiness of multimodal GNNs and Transformers is an open challenge.

#### 6.6.7 Over-smoothing in GNNs

One particular challenge in using GNNs is over-smoothing, which occurs when the GNN is trained for too long, causing the node representations to become almost similar [752]. This leads to a loss of information, a decrease in the model’s performance, and a lack of generalization

[702]. Regularization techniques such as dropout, weight decay, skip-connection, and incorporating higher-order structures, such as motifs and graphlets, have been proposed. However, building deep architectures that can scale and adapt to varying structural patterns of graphs is still an open challenge.

#### 6.6.8 Modality Collapse

Modality collapse is a phenomenon that occurs in MML, where a model trained on multiple modalities may become over-reliant on a single modality, to the point where it ignores or neglects the other modalities [326]. Recent work explored the reasons and theoretical understanding of modality collapse [302]. However, the counter-actions needed to balance model dependence on data modalities require active investigation by the ML community.

#### 6.6.9 Dynamic and Temporal Data

Dynamic and temporal data refers to the data that changes over time [752]. Tumor surveillance is a well-known technique to study longitudinal cancer growth over multiple data modalities [734]. Spatio-temporal methods such as multiple instance learning, GNNs, and hybrid of multiple models can capture complex change in the data relationships over time; however, learning from multimodal dynamic data is very challenging and an active area of research [228].

#### 6.6.10 Data Privacy and Federated Learning

With the increased concern for the privacy of data, especially in medical settings, MML techniques need to adapt to local data processing and remote federation. Federated learning can help train large multimodal models across various sites without sharing data [544].

#### 6.6.11 Other Challenges

MML requires extensive computational resources to train models on a variety of datasets and tasks. Robustness and failure detection [12] are critical aspects of MML, particularly in applications

such as oncology. Uncertainty quantification techniques, such as Bayesian neural networks [175], are still under-explored avenues in the MML. By addressing these challenges, it is possible to develop MML models that are able to surpass the performance offered by single-modality models.

## **6.7 Conclusion**

Existing research into the integration of data across various modalities has already yielded promising outcomes, highlighting the potential for significant advancements in cancer research. However, the lack of a comprehensive framework capable of encompassing the full spectrum of cancer dataset modalities presents a notable challenge. The synergy between diverse methodologies and data across different scales could unlock deeper insights into cancer, potentially leading to more accurate prognostic and predictive models than what is possible through single data modalities alone. In our survey, we have explored the landscape of multimodal learning applied to oncology datasets and the specific tasks they can address. Looking ahead, the key to advancing this field lies in the development of robust, deployment-ready MML frameworks. These frameworks must not only scale efficiently across all modalities of cancer data but also incorporate capabilities for uncertainty quantification, interpretability, and generalizability. Such advancements will be critical for effectively integrating oncology data across multiple scales, modalities, and resolutions. The journey towards achieving these goals is complex, yet essential for the next leaps in cancer research. By focusing on these areas, future research has the potential to significantly enhance our understanding of cancer, leading to improved outcomes for patients through more informed and personalized treatment strategies.

## **Chapter 7: Building Flexible, Scalable, and Machine Learning-Ready Multimodal Oncology Datasets**

### **7.1 Note to Reader**

This chapter has been previously published in MDPI Sensors as: Tripathi A, Waqas A, Venkatesan K, Yilmaz Y, Rasool G. Building Flexible, Scalable, and Machine Learning-Ready Multimodal Oncology Datasets. *Sensors*. 2024; 24(5):1634, and has been reproduced with permission from MDPI [691].

### **7.2 Introduction**

Clinicians routinely gather data from multiple sources to gain a deeper insight into patients' health and provide tailored medical care. The reliance on multiple data sources for clinical decision-making makes medicine inherently multimodal, where the data modality refers to the form of data [87, 734, 203]. Each modality in such multimodal data may have a different resolution and scale due to its own data collection, recording, or generation process [430, 739]. The data modalities may include (i) -omics information from genome, proteome, transcriptome, epigenome, and microbiome, (ii) radiological images from computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), ultrasound scanners or X-ray machines, (iii) digitized histopathology slides created using tissue samples and stored as gigapixel whole slide images (WSI), and (iv) electronic health record (EHR) that houses structured information consisting of demographic data, age, ethnicity, sex, race, smoking history, etc. and unstructured data such as discharge notes or medical reports [87, 430, 203].

Integrating data from heterogeneous modalities can create a unified, richer view of cancer, potentially more informational and complete than the individual modalities [204, 87]. The multimodal medical data holds great potential to advance our understanding of complex diseases and help develop effective and tailored treatments [45, 739]. The advent of high-throughput multi-omics technologies like next-generation sequencing (NGS), high-resolution radiological and histopathology imaging, and the rapid digitization of medical records has led to an explosion of diverse, multimodal data [221]. This data deluge has been a boon for machine learning, where abundant training data has directly enabled significant breakthroughs [763, 733]. For example, the rise of large general-purpose datasets like Common Crawl for natural language processing (NLP) has fueled advances in language models and Artificial Intelligence (AI) assistants [153]. One may hope that extensive, standardized, and representative multimodal datasets in the medical domain would provide a fertile ground for developing advanced translational machine learning models. Machine learning thrives on massive, high-quality datasets; however, assembling such resources in healthcare poses unique challenges [95, 366]. First, multimodal medical data is inherently heterogeneous and noisy, spanning structured (demographics, medications, billing codes), semi-structured (physician notes), and unstructured data (medical images). Aggregating such heterogeneous data requires extensive harmonization and manual processing. Second, reliability, robustness, and accuracy are critical for all medical applications [12, 174, 735]. However, real-world clinical data is often incomplete, sparse, and contains errors, which makes building robust and reliable models more challenging. Meticulous quality control and manual curation of these datasets are essential to train machine learning models [76, 653]. Finally, strict data privacy and security considerations arise in healthcare, where data contain protected health information (PHI) that must be redacted, de-identified, and access controlled per the Health Insurance Portability and Accountability Act (HIPAA) [157, 517].

Traditionally, vast amounts of multimodal data are generated during clinical trials and research studies where raw data undergoes initial processing and quality control by researchers. The data is then transmitted to standardization pipelines such as the National Cancer Institute's (NCI) Center for

Cancer Genomics (CCG) Genome Characterization pipeline [501], where the data is systematically annotated, formatted, and quality-controlled before being deposited into centralized biobanks. For example, NGS data from cancer genomic studies is standardized by CCG and deposited into the NCI's Genomic Data Commons (GDC) [259]. However, imaging data from the same studies, consisting of CT, MRI, and PET scans, follow a different path and may end up in imaging archive like The Cancer Imaging Archive (TCIA) [149]. This leads to fragmentation of data across multiple disconnected databases. To address this, integrated data commons like the NCI Cancer Research Data Commons (CRDC) have been proposed [288]. The CRDC aims to link datasets from diverse sources using Findable, Accessible, Interoperable, and Reusable (FAIR) principles to enhance interoperability [709].

However, significant challenges remain in unifying multimodal data dispersed across repositories with heterogeneous interfaces, formats, and query systems. For example, a researcher studying lung cancer requires integrating clinical, imaging, and genomic data for their cohort across the GDC, TCIA, and other databases. But each has different application programming interfaces (APIs), schemas, and querying methods. Piecing together data manually across these silos is painstakingly difficult. There is a lack of unified interfaces and analytical tools that can work seamlessly across multiple cancer data repositories. This leads to isolated data silos and hampers easy access for multimodal data analysis. To address the limitations and fragmentation of current oncology data systems, we propose a novel solution called the “Multimodal Integration of Oncology Data System”, abbreviated as MINDS. MINDS is a scalable, cost-effective data lakehouse architecture that can consolidate dispersed multimodal datasets into a unified platform for streamlined analysis. To illustrate this, let's consider the process of developing a machine-learning model using a limited dataset with and without the use of MINDS in Figure 7.1.

### 7.2.1 Contributions of MINDS

MINDS makes several key objective contributions toward effectively managing and analyzing multimodal oncology data:

1. Integrating siloed multimodal data into a unified access point. By consolidating dispersed datasets across repositories and modalities, MINDS delivers a single unified interface for accessing integrated data. This overcomes fragmentation across disconnected silos.
2. Implementing robust data security and access control while supporting reproducibility. Strict access policies and controls safeguard sensitive data while enabling reproducibility via dataset versioning tied to cohort definitions.
3. An automated system to accommodate new data continually. Automated pipelines ingest updates and additions, ensuring analysts always have access to the latest data.
4. Enabling efficient, scalable multimodal machine learning. Cloud-based storage and compute scale elastically to handle growing data volumes while optimized warehousing delivers high-performance model training.

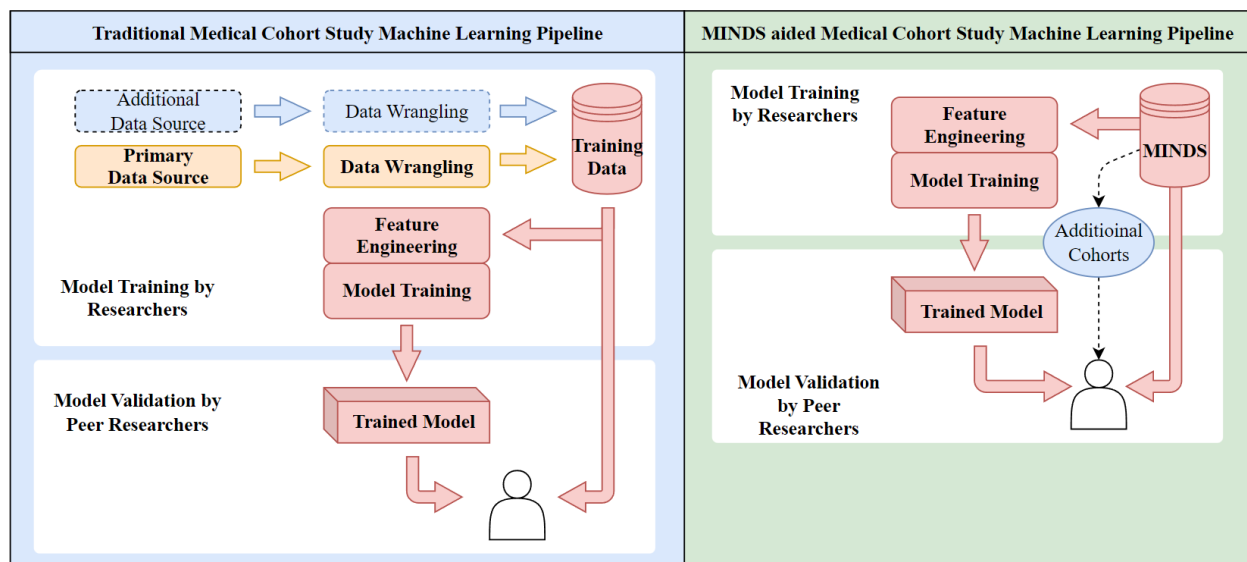


Figure 7.1 Comparison of developing ML model using dataset with/ without using MINDS.

Apart from the above-mentioned achievements, MINDS has several novel aspects. The unprecedented scale of heterogeneous data consolidation enables new analysis paradigms. The cohort

diversity in MINDS also surpasses existing systems. Tight integration between cohort definition and on-demand multimodal data assembly, not offered in current platforms. An industrial-strength cloud-native architecture delivers advanced translational informatics over a browser. Support for reproducibility via dataset versioning based on user cohort queries. This allows regenerating the same data even with newer updates. Option to build vector databases capturing data embeddings instead of actual data. This eliminates storage needs while ensuring patient privacy.

In this paper, Section 7.3 provides necessary background of the existing landscape of the multimodal heterogeneous datasets in oncology, from collection and processing to distribution. Section 7.4 delves into the methodology used to build the proposed data lakehouse architecture and discusses the project's technical aspects in detail. In Section 7.5, we discuss the implementation results and the study's potential implications on cancer research and clinics. Finally, Section 7.6 concludes with recommendations for future research.

### **7.3 Background and Literature Review**

The rapid growth of biomedical data has created immense opportunities for translational research and significant data management challenges. Pioneering efforts have paved the way within this crucial domain by establishing needed infrastructure and principles over the past decades. These include caBIG [388] in 2004, interconnecting cancer researchers via an ambitious grid architecture, tranSMART [619] enabling customized cohort investigation, and i2b2 [498] spearheading flexible clinical data warehousing with temporal abstractions. However, as data scales intensify, core capabilities around scalability, provenance tracking, standardized metadata assimilation, and customizable cohort building, have created substantive yet addressable headroom for enhancements.

Emerging techniques like high-dimensional multimodal assay fusion [481, 353, 363, 430] and multimodal data warehouses [610] have created new demands for consolidation platforms. By striving to synthesize the strengths of the seminal prior work while enhancing key dimensions like flexibility, replicability, and scalability, MINDS aims to stand on the shoulders of giants in pushing meaningful progress in addressing the constraints hampering reliable integrative modeling.



Such demands motivate the development of new solutions to effectively consolidate, integrate, and analyze exponentially growing heterogeneous data types while accounting for the crucial lineage of achievements that collectively established the foundation. Below we discuss the existing methods of oncology data integration.

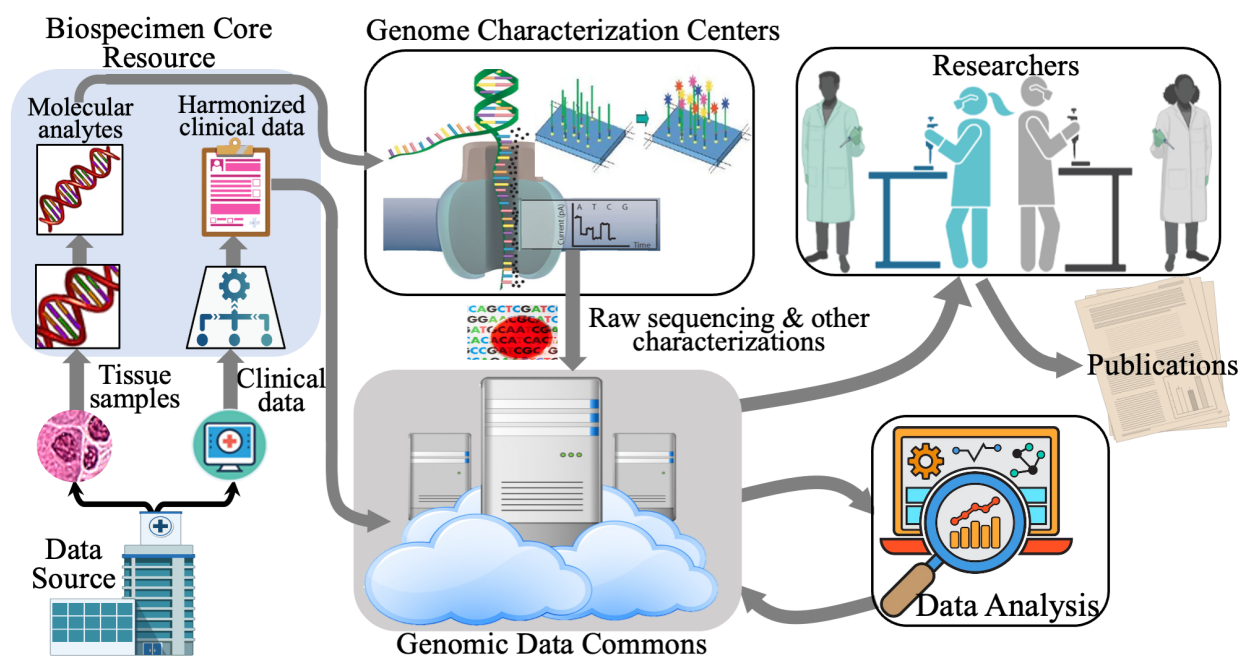


Figure 7.2 Genome Characterization Pipeline as an example of data characterization.

### 7.3.1 Data Characterization Pipeline

Standardized data characterization pipelines are vital in transforming raw biological samples into usable multimodal datasets. A sample data pipeline for gathering genomic modality from CCG for the GDC [259] is illustrated in Figure 7.2. The presented pipeline involves several stages, including tissue collection and processing, genome characterization, genomic data analysis, and data sharing and discovery. The NCI has adopted similar pipelines for medical images, referred to as the Imaging Data Commons [217] or IDC and Proteomics Data Commons or PDC [682].

- *Tissue collection and processing* include clinical trials and community oncology groups, collect tumor tissue samples and normal tissue from participating patients. These samples are either formalin-fixed paraffin-embedded (FFPE) tissues or frozen tissue. In CCG, Biospecimen Core Resource (BCR) is responsible for collecting and processing these samples and collecting, harmonizing, and curating clinical data [501].
- *Genome characterization* stage involves generating data from the collected samples. At CCG, the Genome Characterization Centers (GCCs) generate data from the samples received from the BCR. Each GCC supports distinct genomic or epigenomic pipelines, including whole genome sequencing, total RNA and microRNA sequencing, methylation arrays, and single-cell sequencing [501].
- *Genomic data analysis* is the stage where raw data from the previous stage is transformed into meaningful biological information at this stage. In CCG, the Genomic Data Analysis Network (GDAN) transforms the raw data output from the GCCs into biological insights. The GDAN has a wide range of expertise, from identifying genomic abnormalities to integrating and visualizing multi-omics data [501].
- *Data sharing and discovery* stage involves insightful genomic data processing, sharing, and unification at a central location. The NCI's Genomic Data Commons (GDC) harmonizes genomic data by applying a standardized set of data processing protocols and bioinformatic pipelines. The data generated by the Genome Characterization Pipeline are made available to the public via the GDC [501, 259].

### 7.3.2 Traditional Data Management - BioBanks

Traditionally, medical data modalities are stored and managed separately in biobanks. These biobanks are the repositories that store biological samples for use in research and by clinicians for reference. Today, such biobanks have become an essential resource in medical and oncological facilities [53]. They provide researchers access to various medical samples and associated clinical

and demographic data, which is used to study disease progression, identify biomarkers, and develop personalized and new treatments. However, traditional data management using biobanks has several limitations, enumerated below:

- One of the main issues is that data from different sources are often stored in separate biobanks, leading to fragmentation of information [10]. This makes integrating and analyzing data across different modalities difficult, limiting the potential for comprehensive, multi-dimensional analysis of patient data [53].
- How data is stored, formatted, and organized often varies significantly across biobanks, even for the same patient. For example, clinical data may be encoded differently, imaging data may use proprietary formats, and terminology can differ across systems. This heterogeneity and lack of unified standards make aggregating and analyzing data across multiple biobanks challenging [53].
- Over time, patient data stored in separate biobanks tends to go out of sync as patients undergo new tests and treatments, adding new data to different silos uncoordinatedly [53]. Piecing together a patient's history timeline requires extensive manual effort to sync disparate records across systems [53].
- The increasing prevalence of bio-banking has sparked an extensive discussion regarding the ethical, legal, and social implications (ELSI) of utilizing vast quantities of human biological samples and their associated personal data [392]. Ensuring and safeguarding the fundamental ethical and legal principles concerning research involving human data in Biobanks becomes significantly more intricate and challenging than conducting ethical reviews for specific research projects [392].

### 7.3.3 Data Commons

The concept of data commons has emerged to address the challenges faced by biobanks. A data commons is a shared virtual space where researchers can work with and use data from multiple

sources. The NCI has developed the CRDC, which integrates different data types, including genomic, proteomic, imaging, and clinical data, into a unified, accessible platform [288]. The CRDC provides researchers access to various data repositories, including the GDC, PDC, and IDC. Each of these repositories hosts a specific data type, and together, they form a comprehensive platform for multimodal data analysis. While the CRDC has made significant strides in integrating diverse data types, it still faces challenges. One of the main issues is the difficulty in harmonizing data from different sources. Due to the differences in data formats, standards, and quality control measures across data sites and modalities, it takes significant effort by the researchers to conform the data to uniform quality standards. The Cancer Data Aggregator (CDA) was developed to address this issue and facilitate data integration across different data commons. CDA provides an aggregated search interface across major NCI repositories, including the PDC, GDC, and IDC. It allows unified querying of core entities like subjects, research participants, specimens, files, mutations, diagnoses, and treatments, facilitating access across different data types [508]. CDA has limitations, like static outdated mapping and the inability to incorporate external repositories. This motivates the need for more robust integrative platforms. The proposed MINDS system aims to overcome these challenges in several key ways:

- CDA's mapping of the CRDC data is not real-time. For example, as of September of 2023, when querying patients with the primary diagnosis site being lung, only 4870 cases are present, despite there being 12,267 cases present in the GDC data portal. MINDS pulls source data directly from repositories like GDC to ensure real-time, up-to-date mapping of all cases.
- MINDS is designed as an end-to-end platform for users to build integrated multimodal datasets themselves rather than a fixed service. The open methodology enables full replication of huge multi-source datasets. To this end, anyone can replicate our method to generate the exact copy of over 40,000 public case data on their infrastructure.

- MINDS is flexible and incorporates diverse repositories and data sources, not just CRDC resources. Our proposed architecture can integrate new repositories as needed, unlike CDA, which is constrained to CRDC-managed data. For example, the cBioPortal for Cancer Genomics, a widely used platform for exploring, visualizing, and analyzing cancer genomics data, has its own data management and storage system separate from the CDA [121, 237]. The data stored in cBioPortal cannot be directly queried or accessed through CDA, limiting the potential for integrated data analysis across platforms.

#### 7.3.4 The Big Data Approach

We have used the Big Data approach in our work [95, 366]. Among the recent advancements in healthcare data management, the big-data approach is the most prominent and feasible solution [45, 221, 750]. The rapid technological progress has led to an unparalleled utilization of computer networks, multimedia, internet of things, social media, and cloud computing, resulting in an overwhelming generation of “big data” [500]. Effectively collecting, managing, and analyzing vast amount of healthcare data through big data processing has become crucial. The big data processing involves various techniques, such as data mining, leveraging data management, machine learning, high-performance computing, statistics, and pattern recognition to extract knowledge from extensive datasets. These datasets possess distinctive characteristics, often called the seven Vs of big data, as shown in Figure 7.3. The Big Data approach guides data handling strategies. By considering each of these aspects, we can effectively manage oncology data and, in turn, build better, effective models. We use two primary data management systems to facilitate our big data approach: Data Warehouses and Data Lakes.

Data warehouses represent a foundational pillar of the big data paradigm. A data warehouse integrates heterogeneous data from diverse sources into a centralized, well-organized repository to enable analysis. This repository provides a highly structured environment explicitly optimized for analytics, reporting, and deriving insights across vast information [500]. By fulfilling this role, data warehouses deliver immense value in informed decision-making. The process of as-

sembling data into warehouses is called data warehousing. “Schema-on-write” is the core concept employed, where the warehouse schema is predefined to meet specific analytical needs before data is loaded. This upfront structural optimization makes warehouses ideal for handling structured data. Supervised machine learning thrives in warehouses, as structured, consistent data facilitates training algorithmic models. Moreover, the innate high degree of organization enables fast, efficient querying to uncover trends and patterns through predictive analytics [500]. Overall, by structuring varied data sources into a unified environment purpose-built for analytics, data warehouses provide the backbone for deriving value from big data across many domains.

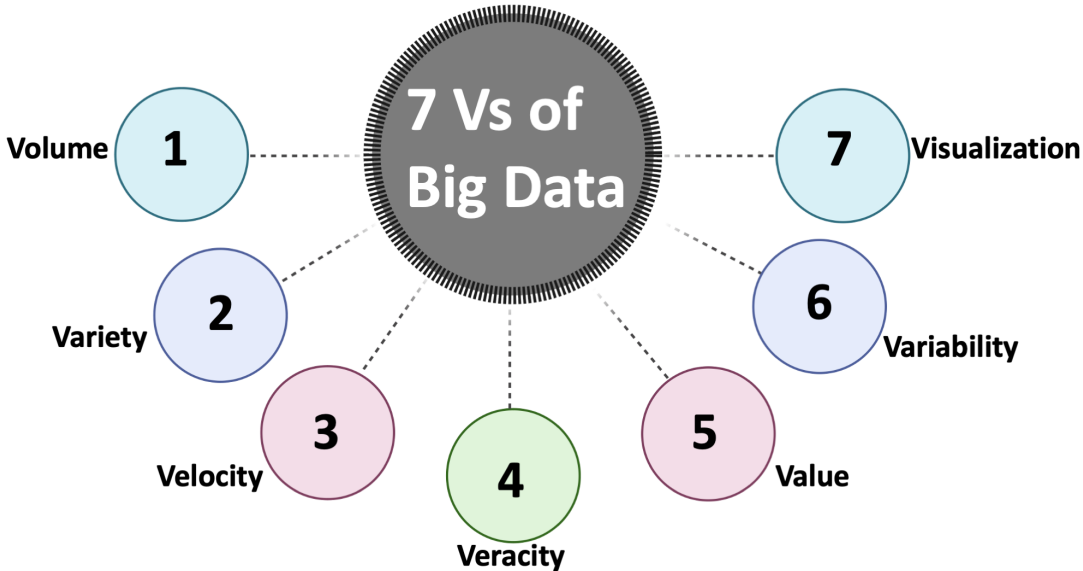


Figure 7.3 The 7 Vs of Big Data.

Data lakes complement the warehouses by providing centralized but low-structure storage to accumulate expansive, heterogeneous data in raw form. In contrast to “schema-on-write”, data lakes employ “schema-on-read”, which only defines structure when data is queried. This provides flexibility to modify analytics on-demand [500]. With their innate tolerance for storing original, unprocessed data, lakes accommodate structured, semi-structured, and unstructured data types. The lack of enforced structure enables rapid scaling to meet growing analytics demands. The dual architectures of data warehouses and data lakes provide structured refinement and raw

accommodating capabilities to put big data into action. Lakes aggregate heterogeneous datasets, while warehouses prepare refined data for analysis. This symbiotic combination ultimately enables MINDS to derive maximal value from oncology’s multidimensional data landscape.

### 7.3.5 Summary of Gaps in Existing Methods

While prior work has laid crucial foundations, several persistent constraints around consolidation, interoperability, scalability, provenance, and security have encumbered reliable integrative modeling on multimodal data. Biobanks carry siloed modalities with heterogeneous formats, creating barriers to unification and requiring extensive manual synchronization effort. Data commons combined various data types into unified platforms but lack harmonization of diverse data sources. Static mappings fail to reflect repositories’ real-time state, while disjoint querying systems limit holistic analysis across databases. Fundamentally, past efforts centered on aggregating principally structured sources, lacking the breadth to effectively harness the heterogeneity spanning images, assays, text, and sensors. With data volumes intensifying across these manifold streams, inflexible on-premises systems strain to provide needed scalability. Reproducibility suffers from dynamic dataset derivation as model provenance linkages fade. Finally, while ethical rigor grows in importance with scale, most architectures offer worryingly coarse-grained control over access policies. By tackling this multiplicity of the persistent challenges through enhancements leveraging the prior seminal achievements, MINDS aims to advance reliable, responsible multimodal modeling on big oncology data. The key limitations that constrain multimodal integrative modeling through existing approaches are summarized below:

- Prior consolidation is limited to structured data. Most prior efforts, like CDA, focused on consolidating structured clinical records. Support for aggregating unstructured imaging, -omics, and pathology data is lacking.
- Query interfaces have limited standardization. Different repositories have proprietary APIs and schemas. Unified interfaces for federated querying are needed.

- Scalability is constrained for large data. On-premises systems restrict scaling storage and compute for exponentially growing heterogeneous data.
- Minimal reproducibility without versioning. Dynamic dataset extracts make precise tracking of model data versions difficult, hampering reproducibility.
- Coarse-grained access controls. Most systems have limited options for fine-grained data access policies tailored to users.

Addressing these gaps is pivotal to unlocking translational applications of multimodal oncology data through enhanced consolidation, standardization, scalability, provenance, and security. By tackling each limitation, MINDS aims to overcome persistent bottlenecks that have hitherto encumbered reliable integrative modeling on heterogeneous big data.

## **7.4 Methodology**

This section details the technical implementation of the proposed MINDS architecture. We begin by presenting the high-level requirements that informed key architectural decisions. We then dive into the three-stage architecture of MINDS, describing each component and its role in enabling scalable and secure management of multimodal oncology data. Next, we provide deployment options for MINDS, including details on implementing the system in the cloud across diverse platforms and on-premise infrastructure. Finally, we outline key use cases and user interactions with MINDS.

### **7.4.1 Requirements of Data Management System**

To handle the complexities, scales, and heterogeneity in the structure and function of oncology data, the data management system design has to be comprehensive, scalable, and interoperable. The primary goal of this system is to cater to the needs of machine learning engineering, which requires a robust and efficient data management infrastructure to build accurate and reliable models. We set



off with the aim to design and build a data management system with the following requirements in mind:

- Requirement 1 is to minimize large-scale unstructured data storage whenever possible. This requirement ensures the efficient use of storage resources and allows the user to access the data directly from the data provider.
- Requirement 2 is that system should be horizontally and vertically scalable. Satisfying this requirement is crucial to handle the increasing volume of oncology data and ensure the system can accommodate data size and complexity growth.
- Requirement 3 is that system should be interoperable, allowing for the easy integration of new data sources. This is important in oncology, where data is often distributed across various databases and systems.
- Requirement 4 is that the system should track data from the point of ingestion to the point of training. This ensures reproducibility, a key requirement in scientific research and machine learning.
- Requirement 5 is to incorporate audit checkpoints in the data collection, pre-processing, storage, processing, and analysis stages of the data pipeline. This ensures data integrity, the prime consideration in delivering reliable machine learning outcomes.

#### 7.4.2 MINDS Architecture

Considering the above-mentioned requirements, we have built a Multimodal Integration of the Oncology Data System (MINDS). The MINDS system design adopts a common two-tier data architecture, a data lake, and a data warehouse [500] to process data and derive meaningful insights efficiently. Figure 7.4 illustrates the architecture of MINDS, which is divided into three primary stages: (1) Data Acquisition, (2) Data Processing, and (3) Data Serving. Key goals include scalability of individual components, interoperability via standardized APIs and schemas, security

leveraging authentication and encryption, and usability across interactive and programmatic access patterns. To meet this requirement MINDS is built using the cloud-based technology of Amazon Web Services (AWS), the cloud-based architecture allows us to scale up or down easily based on the data volume requirements and the required computational resources. It also provides a wide range of tools and services that can be leveraged to build, deploy, and manage a data management system. While the current MINDS implementation leverages AWS, the architecture is designed to enable deployment across different cloud platforms, not just AWS. The core methodology centers on interfacing with managed cloud services, abstracting the underlying infrastructure through common programmatic interfaces. This service-oriented approach enhances portability and avoids extensive customization tied to a single provider. For example, the S3 storage layer could be replaced with Google Cloud Storage buckets, AWS Glue with Azure Data Factory, RDS and Redshift with Snowflake’s data platform, and Lambda with Cloud Functions. The overall system architecture would remain consistent while swapping the provider services. When migrating platforms, trade-offs exist around performance, access controls, and other factors. But by using managed services with standard APIs, MINDS aims for platform-independent portability. Figure 7.5 provides a detailed layout of technical components at each stage using AWS cloud infrastructure and the tools utilized to actualize the system. Definitions of these technical components are summarized in Box 7.1.

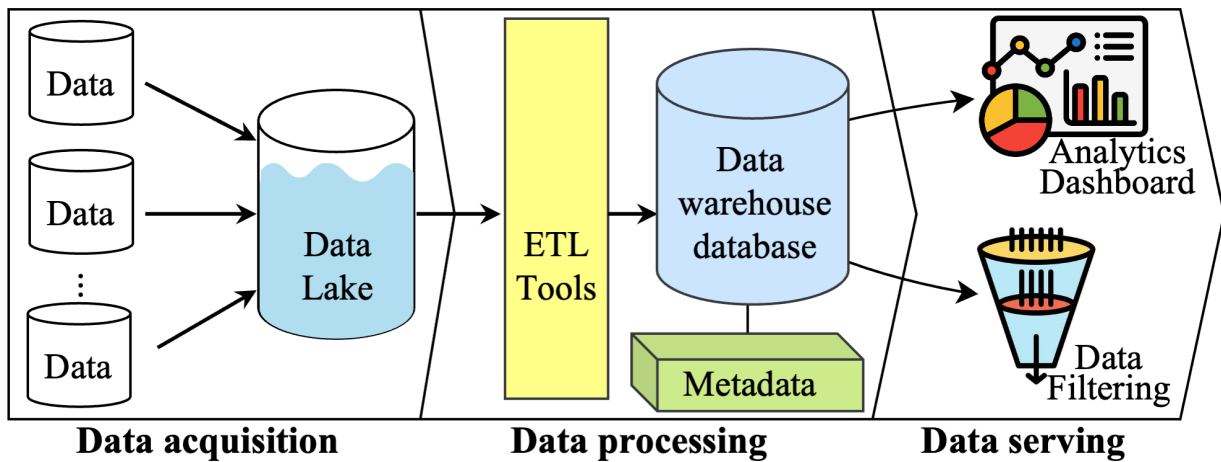


Figure 7.4 MINDS architecture implements a 3-stage pipeline.

Table 7.1 Definitions of key cloud components

Component	Definition
Amazon S3 Ingest Bucket	Object storage bucket for staging raw data before loading into a data lake.
Amazon Web Services (AWS)	A cloud platform that provides scalable computing, storage, analytics, and machine learning services.
AWS Athena	Serverless interactive query service to analyze data in Amazon S3 using standard SQL.
AWS Big Data Analytics	Suite of services for processing and analyzing big data across storage, compute, and databases.
AWS Data Lake Formation	Service to set up and manage data lakes with indexing, security, and data governance.
AWS Data Warehouse	Fully-managed data warehousing service for analytics using standard SQL.
AWS Glue Crawler	Discovers data via classifiers and populates the AWS Glue Data Catalog.
AWS Glue Data Catalog	Central metadata store on AWS for datasets, schemas, and mappings.
AWS Lambda	Serverless compute to run code without managing infrastructure.
AWS QuickSight	Business intelligence service for easy visualizations and dashboards.
AWS RDS	Amazon Relational Database Service is a managed relational database service that handles database administration tasks like backup, patching, failure detection, and recovery. Including RDS MySQL, a managed relational database optimized for online transaction processing.
AWS Redshift	Petabyte-scale data warehouse for analytics and business intelligence.
JDBC	JDBC (Java Database Connectivity) is a standard API for connecting to traditional relational databases from Java. The JDBC was released as part of the Java Development Kit (JDK) in version 1.1 in 1997 and has since been part of every Java edition.

#### 7.4.2.1 Stage-1: Data Acquisition

Data acquisition is the first and crucial step in building the MINDS platform. This process involves gathering all publicly available structured and semi-structured data from the data sources. As mentioned earlier, the CRDC and other oncology data management initiatives host vast amounts of patient information, and we use them as the primary data sources for our system. These sources primarily include the three data commons portals, GDC, IDC, and PDC. Additionally, we use the CRDC’s Cancer Data Aggregator (CDA) tool to map all the patient information across the commons into one cohesive database. This database then expands to accommodate the patient data stored across other portals, such as the cBioPortal, Xena, and other relevant data sources [121, 237, 249]. It is pertinent to mention that we do not store any unstructured data in MINDS, such as whole

slide images or radiology scans. MINDS instead pulls the unstructured data from their respective data commons based on the cohort the users want to build and the modalities they require for processing through the portal APIs. Hence, we are not required to store large unstructured data such as gigabyte pathology images in our database.

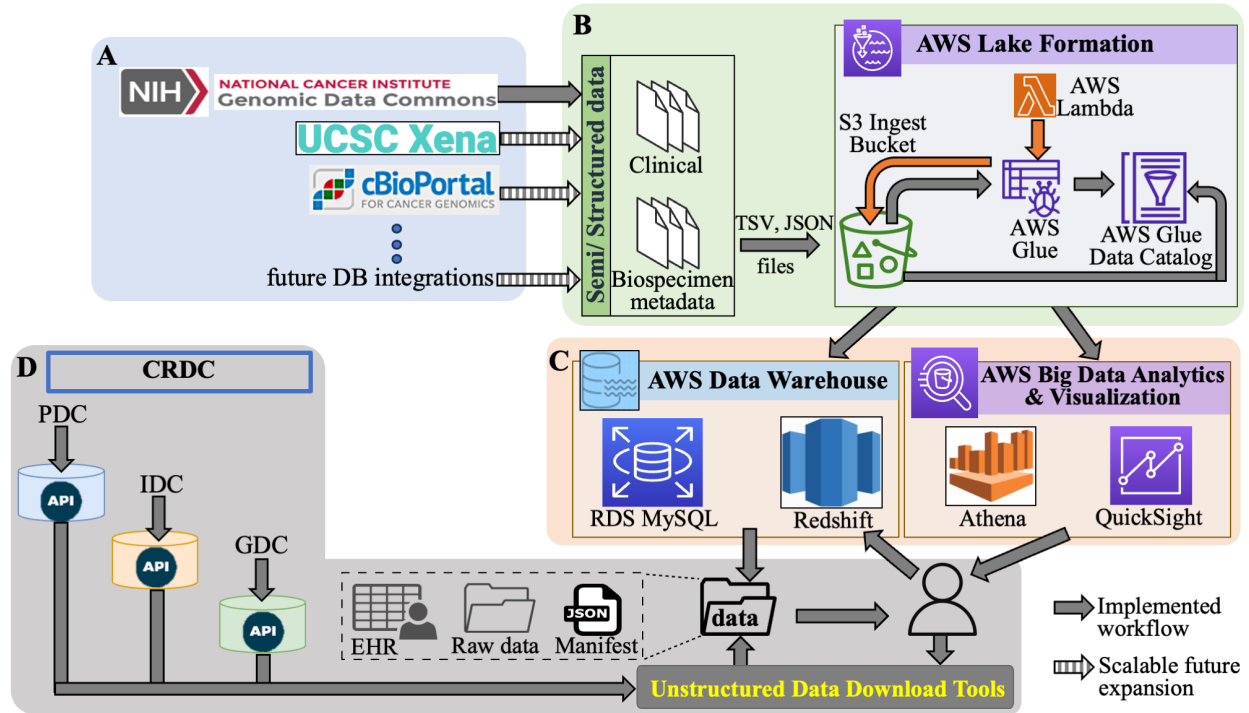


Figure 7.5 Overview of the MINDS architecture implemented on AWS.

For the initial version of MINDS, we leverage the GDC as the primary data source due to its comprehensive collection of up-to-date, publicly available oncology data. The GDC portal contains clinical, biospecimen, and molecular data across diverse cancer studies, representing over 86,000 cases spanning 78 projects. The GDC has the most extensive public data holdings out of the three NCI data commons. As of 2023, it hosts over 3 petabytes of genomic and clinical data from the NCI programs like The Cancer Genome Atlas (TCGA) and Therapeutically Applicable Research to Generate Effective Treatments (TARGET). The GDC also has a well-designed and detailed data model that structures and connects the clinical, biospecimen, and molecular data domains. The availability of this robust data dictionary and schema metadata makes the ingestion and integration

of new GDC datasets simpler and more consistent. Leveraging thousands of richly annotated multi-omic cancer profiles, we can develop integrative and predictive models by utilizing all the public cases in the GDC for MINDS initial deployment. The breadth of tumor types enables the building of generalized models applicable across different cancers. As the MINDS data repository expands to incorporate more primary sources beyond GDC, the experience of integrating the GDC data provides a solid foundation to build upon. The tooling ETL workflows developed to ingest and harmonize GDC data can be extended to transform and connect new oncology datasets into the MINDS knowledge system.

#### *7.4.2.2 Stage 2: Data Processing*

A foundation of the MINDS architecture is ingesting petabytes of structured clinical, biospecimen, and molecular data from cancer genomic repositories like the GDC. This raw metadata arrives in heterogeneous formats including JSON documents, CSV exports, and XML messages conveying case details, lab assays, pathology reports, and tissue sample attributes. While information-rich, effectively using this disjointed data to drive integrative insights requires significant wrangling. We leverage the GDC common data model as an integration schema to streamline aggregation and analysis. This model structures entities like cases, files, and read groups into a normalized graph representation, with nodes denoting key objects and edges linking related records. For example, a case entity may reference constituent pathology reports, sequencing files, or tissue aliquots to provide a unified view spanning this network of connected data. The GDC data dictionary rigorously defines properties and relationships to provide semantic consistency. Aligning raw data to this canonical representation enables unified storage, queries, and computational pipelines. However, the raw downloads natively arrive in varied shapes. JSON clinical documents describe patient characteristics differently than TSV biospecimen exports. Our first challenge is flexible parsing and mapping.

The need to integrate data from multiple sources is further pronounced in complex diseases such as cancer when considering efforts such as precision medicine and personalized treatments.

However, interoperability remains a major challenge in practice despite extensive standards development. Many clinical, genomic, imaging, and literature databases use disjoint interfaces, formats, and terminologies, thus hampering unified analytics. Several domain-agnostic standards have emerged to promote harmonization:

- Health Level 7 (HL7) defines structures and semantics for messaging healthcare data between computer systems, including Clinical Document Architecture and Fast Healthcare Interoperability Resources (FHIR) specifications [187, 289].
- Fast Healthcare Interoperability Resources (FHIR) specifies RESTful APIs, schemas, profiles, and formats for exchanging clinical, genomic, imaging, and other healthcare data. Offers platform-agnostic interconnection [289].
- Clinical Data Interchange Standards Consortium (CDISC) develops data models, terminologies, and protocols focused specifically on clinical research and FDA submissions, including the Study Data Tabulation Model (SDTM) and the Clinical Data Acquisition Standards Harmonization (CDASH) [119].

However, adopting these standards remains inconsistent, and significant translator development is required to bridge entities [57]. The tight coupling of databases to proprietary representations threatens interoperability. Furthermore, medical ontologies and terminologies such as those given below play a crucial role in promoting both human and machine-readable shared understanding:

- Systematized Nomenclature of MEDicine Clinical Terms (SNOMED CT) provides consistent clinical terminology and codes for electronic health records. Enables semantic interoperability [510].
- National Cancer Institute (NCI) Thesaurus models cancer research domain semantics with 33 distinct hierarchies and 54,000 classes/properties. Binds related concepts for knowledge discovery [502].

The GDC data model and dictionary enhance interoperability by structuring and defining entities, properties, and relationships standardized. When ingesting data, the AWS Glue crawler leverages these common semantics to map input elements into the unified representation. This semantic alignment enables integrated analysis despite originating heterogeneity.

GDC structures clinical, biospecimen, and molecular data using a consolidated data model that interconnects related entities into a directed acyclic graph (DAG) representation. This data model underpins the organization and semantics of the petabyte-scale GDC dataset. The model comprises a network of nodes representing key data objects (cases, samples, reads, etc.) linked through edges denoting relationships. Nodes have properties like type, age, and tumor\_stage, while edges characterize affiliations like a sample derived\_from a case. Robust semantic definitions specify permitted nodes, their properties, associated data types, and linkage rules. This ontology ensures consistency critical for downstream interoperability. Technically, the data model utilizes a mix of JSON and YAML schemas coupled with Python 3 and SQL codebases to architect domain representations. Schemas define valid elements and constraints serialized into JSON documents. Codebases ingest and query documents while preserving compliance. The GDC dictionary elaborates on metadata driving consistency. For example, the sample entity has documented required fields like sample\_type and permissible values like Solid Tissue. The authentication service verifies submitted entities to satisfy specs. At the infrastructure layer, dictionaries transform into SQLite representations. Indexed tables track datasets while optimized queries fetch connections. Although decentralized, federated services coalesce distributed systems into an integrated data collaboration. The presentation tier visualizes linkages via an entity-relationship diagram highlighting cardinality rules (one-to-many mappings, etc.). Users traverse graphs accessing constituent records through REST APIs. By providing rigorous blueprints governing content packaging, exchange, and interpretation, GDC data models power external consistency unlocking unified workflows spanning partners—enabling interconnected explorations.

### 7.4.2.3 Stage 3: Data Serving

We provide two core methods for researchers to consume processed oncology data based on workflows. We built interactive dashboards for interactive visualization and cohort analysis that use the data stored in the warehouse. Additionally, for developers and computational researchers needing to ingest data into pipelines, we provide an open-source Python toolkit as part of MINDS that programmatically downloads the unstructured multimodal data from disparate databases.

MINDS provides a dashboard. At data consumption stage, structured data in the data warehouse is utilized for various purposes. The data consumption process is designed to provide users with an interactive and intuitive interface for exploring, visualizing, and analyzing the data. This is achieved through a dashboard built on Amazon QuickSight [32], a fully managed business intelligence service that enables data visualization and interactive analysis. Users can interact with the dashboard to explore various aspects of the data and identify trends, patterns, or correlations using QuickSight's machine learning-powered insights.

Figure 7.6 presents sample visualizations enabled by the MINDS analytics dashboard, allowing researchers to explore different data attributes like the cause of death and tumor subtype distributions. For example, the death date graph reveals a peculiar underreporting anomaly between 2014–2017 that may warrant investigation into potential data quality issues. Meanwhile, tumor classification breakdown identifies pancreatic cancer as the most represented diagnosis, informing potential studies targeting prevalent categories.

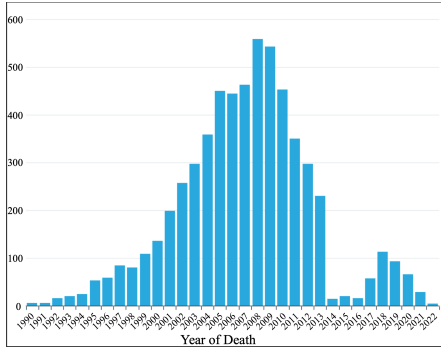
Beyond distributions, the interactive dashboards may also catalyze discoveries by empowering explorations into relationships between clinical factors, assays, and outcomes. As illustrated, researchers could assess survival trends across cancer subtypes to uncover prognostic biases. Recurrence patterns may be analyzed with modalities like genetic mutations and treatment regimens to reveal predictive biomarkers or personalized medicine insights. Apart from the analytical categories depicted in Figure 7.6, the MINDS analytics dashboard allows the researchers to filter data based on any clinical or biological fields such as age, gender, ethnicity, tumor grade, treatment type, year of diagnosis, survival, etc.



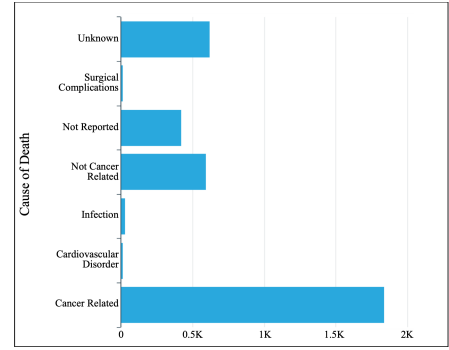
MINDS enables users to build focused, multimodal datasets for targeted analysis by combining warehouse-driven cohort queries with automated unstructured data collection. Patient cohorts are defined by querying the database directly through SQL. The case IDs can be extracted from the cohort, and the resulting list of case IDs is used to retrieve all related unstructured data from the GDC, IDC, and PDC portals using their respective API interfaces. As part of the MINDS toolkit, we provide a Python utility that accepts the case ID list as input and programmatically calls the APIs to bulk download images, pathology, -omics, and other files for those specific cases. The downloaded data is organized into a folder structure with a top-level “/raw” folder containing subfolders for each case ID. Each case folder contains the unstructured data objects from GDC, IDC, and PDC for that case. JSON manifest files are also generated to capture metadata like file IDs, types, and sources. This enables easy indexing and querying of the unstructured data extracts.

#### 7.4.3 Cloud Deployment

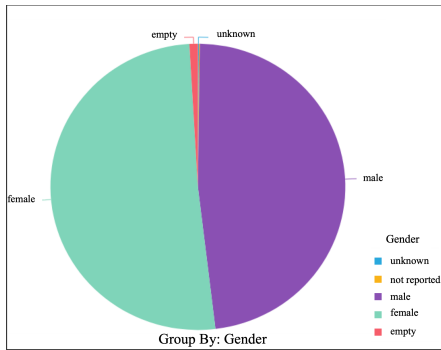
This section outlines the AWS cloud implementation of MINDS, leveraging core infrastructure services to enable scalable data aggregation, processing, and unified access. Our approach incorporated several key big data techniques essential to the MINDS architecture. We utilized Amazon S3 for distributed storage, creating a data lake environment capable of handling petabytes of heterogeneous data. AWS Redshift and EMR were employed for large-scale data warehousing and distributed SQL and Spark processing, respectively. These services enabled the building of high-performance SQL query engines and the efficient processing of large data volumes. AWS Glue played a critical role in machine learning-powered ETL, allowing for the transformation and structuring of data for analysis. Serverless computing using AWS Lambda was instrumental in managing scalable workloads, preventing server overflow, and ensuring system responsiveness. Together, these components formed a robust foundation essential for addressing the challenges of volume, variety, and velocity inherent in the multimodal oncology data within MINDS. While the current system deployment utilizes Amazon Web Services, the underlying architecture is designed for portability across cloud platforms. By interfacing with common storage, database, analytics, and



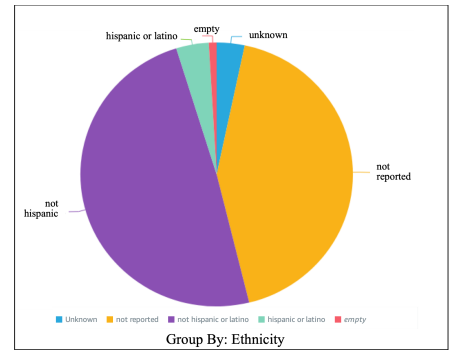
(a)



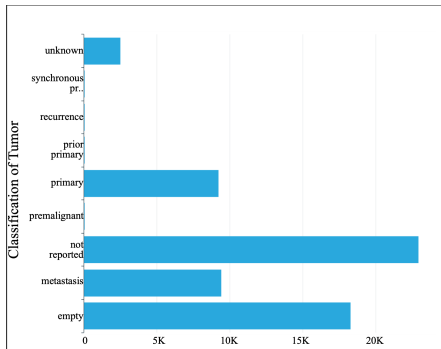
(b)



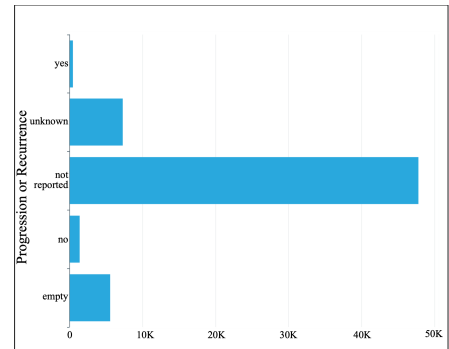
(c)



(d)



(e)



(f)

Figure 7.6 Quicksight analytics and visualization generated using clinical data from MINDS.

machine learning modules rather than low-level servers or virtual machines, much of the MINDS technology stack can be replicated on alternate providers.

#### 7.4.3.1 *MINDS Infrastructure on AWS*

We pull all semi-structured and structured data from the GDC data portal for all public cases, including TSV and JSON files containing various clinical (clinical, exposure, family history, follow-up, and pathology detail) and metadata of biospecimen (aliquot, analyte, portion, sample, and slide) information. This data is then uploaded into an Amazon S3 Ingest Bucket [35]. This bucket acts as the staging storage for the data before it is uploaded to the data lake. To orchestrate the full data lake setup, we utilize the AWS Data Lake Formation tool [37], which automates the transformation of the semi-structured data stored in the S3 bucket into a queryable data lake using AWS Glue crawlers to catalog the data and store it in data tables [38].

The data acquisition is not a one-time event but a continuous process that must be updated regularly to ensure the data lake is always up-to-date with the latest data. The new data is not uploaded arbitrarily but rather arrives through scheduled ETL routines that run every 12 h to poll source repositories like GDC using their APIs. For example, scripts leverage the GDC REST API to query for newly added cases, files or metadata since the last update based on a timestamp. The incremental changes are downloaded via the API and uploaded to the S3 bucket on a Linux-based cron schedule, such as daily at 9 AM UTC. This polling pattern is tailored for each integrated data source and its API capabilities. Explicitly tracking data provenance through structured ingestion and ETL ensures the S3 bucket receives only authorized data uploads, avoiding random additions. We use AWS Lambda serverless compute [41] to trigger Glue crawlers automatically whenever new data lands in the S3 bucket. This ensures our data lake is always up-to-date with the latest data without explicit manual synchronization. This also helps reduce the data transfer rates because the system updates the data lake only with the delta between the bucket and the data lake. The data acquisition process is designed to be robust and scalable, capable of handling the increasing

volume of oncology data. It also ensures the safety and integrity of the data by establishing secure connections to the databases from which data needs to be extracted.

Once the data is acquired, the next step is to clean, process, and aggregate this data. At this stage, the data is extracted from the data lake, transformed into a more structured format, and loaded into the data warehouse. This is done using Amazon AWS Glue 4.0 [36], which ensures consistency and compatibility across data types and sources. AWS Glue performs the ETL actions using the AWS Glue crawler [38]. The crawler works in a series of steps to ensure the data is appropriately cataloged and ready for analysis. Figure 7.7 shows the internal workings of the AWS crawler that ensure the data is properly processed and ready for analysis, making it easier for users to extract valuable insights from the data. The steps involved in the AWS crawler workflow are as follows:

1. Establish access-controlled database connections. The crawler first establishes secure connections to the databases from which data needs to be extracted. This ensures the safety and integrity of the data in transit.
2. Use custom classifiers. If any custom classifiers are defined, they catalog the data lake and generate the necessary metadata. These classifiers help in identifying the type and structure of the data.
3. Use built-in classifiers for ETL. AWS's built-in classifiers perform ETL tasks for the rest of the data. This process involves extracting data from the source, transforming it into a more suitable format, and loading it into the data warehouse.
4. Merge catalog tables into a database. The catalog tables created from the previous steps are merged into a single database. During this process, any conflicts in the data are resolved to ensure consistency and deduplication.
5. Upload catalog to a data store. Finally, the catalog is uploaded to a data store to be accessed and utilized for analytics. This data store is a central repository where all the processed and cataloged data is stored.

When ingesting data, the AWS Glue crawler parses source elements into this consolidated model by mapping input fields into the GDC dictionary. For instance, a Read Group JSON would have its metadata properties (like ID, library name, etc.) inserted as columns into the standardized Read Group table definition used across MINDS while retaining references to the parent Case/File IDs to recreate linkages. The unified representation enables joining and analysis across interconnected data domains related to biospecimen, sequencing, diagnoses, etc., even if originating formats vary. This ensures interoperability among diverse data sources through a common but fast health interoperable resource. To incorporate emerging repositories into this existing data model, we extract salient clinical and experimental metadata based on publication schemas and use the flexible AWS Glue schema evolution tools to extend existing definitions or spawn new tables aligned with import sources if needed. Templated mapping configurations adjust for input heterogeneity while producing consistent MINDS representations to power integrated SQL queries across past and future data partners - avoiding isolated silos or reengineering efforts when onboarding additional cohorts. Hence, MINDS has built-in scalability supported by interoperable functions. The crawler uses the GDC node schema definitions in YAML files to parse the JSON documents and infer the schema. The GDC case entity is defined with properties like `case_id`, `disease_type`, `demographic`, `diagnoses`, etc. When the crawler processes a case JSON document from the GDC portal, it maps the JSON properties to columns in a Glue table definition based on the GDC data model. This way, the GDC model's underlying graph structure transforms relationships into a relational view. The Glue crawler output is a table definition in the AWS Glue Data Catalog. Users can directly query and join with other clinical, biospecimen, and genomic tables ingested from GDC. The dictionaries also provide metadata like each property's data types and code lists. When creating data definition language (DDL) for the tables, the crawler leverages this to assign appropriate column types, formats, and validations. This helps maintain data integrity and consistency during the transformation process.

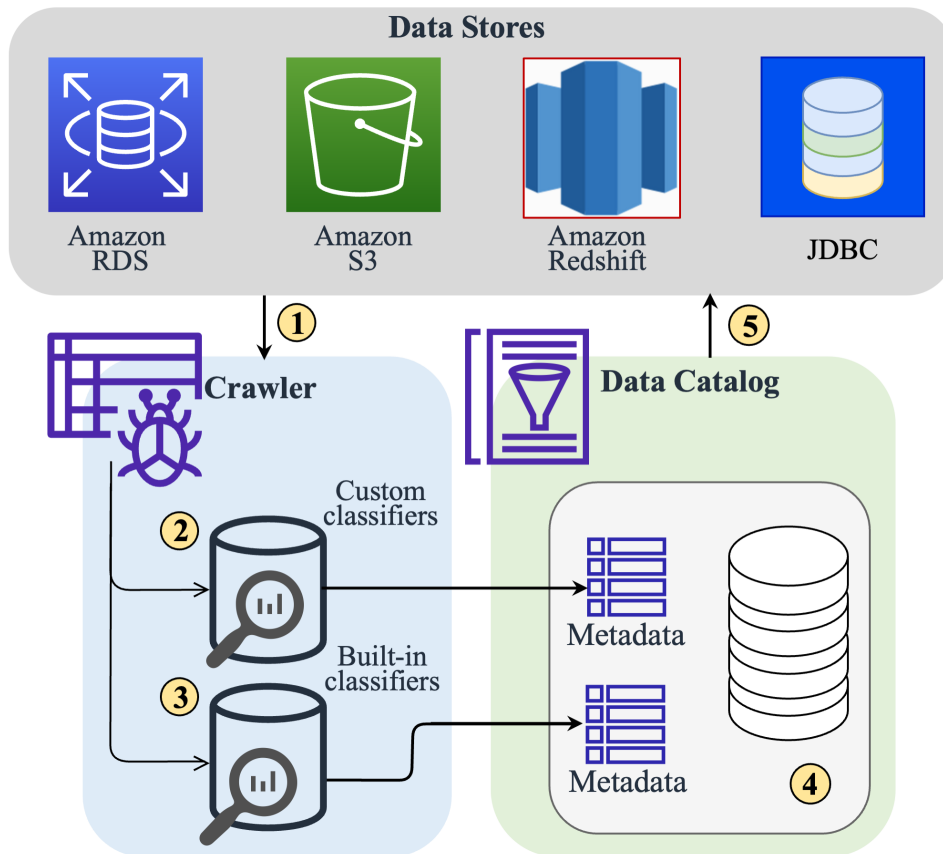


Figure 7.7 The AWS Glue crawler automates ETL in MINDS through a 5-step workflow.

The normalized clinical, biospecimen, and molecular data cataloged by the AWS Glue crawler undergoes loading into Amazon Redshift, which serves as the primary data warehouse for enabling high-performance analytics. With this structured data, we also populate an Amazon RDS MySQL cluster to support efficient inserts and updates as new data arrives from source systems. However, given its optimization for such read-heavy workloads, analytical queries are routed directly to Redshift [34]. As a petabyte-scale massively parallel processing (MPP) data warehouse service, Amazon Redshift employs advanced query processing, adaptive machine learning optimizers, and columnar storage layouts purpose-built for complex aggregations, filters, and joins across huge datasets. By leveraging separate data warehouse and transactional database environments, MINDS supports fluid exploration without impacting critical path operations that rely on consistent low-latency database performance unaffected by ad hoc analysis. We incrementally load the Glue-

cataloged data into Redshift using high-throughput COPY commands to enable fast bulk data movement from S3 object storage. Redshift Spectrum interfaces create external tables pointing directly at structured datasets in S3 buckets, providing direct access without loading the data into local warehouse storage. This allows interactive SQL analytics directly on raw JSON, CSV, and TSV objects with automatic inferencing of schema and transformations to perform as data is read at query runtime. The centralized AWS Glue Data Catalog manages table definitions, schemas, partitions, and mappings across these disparate storage and processing environments - serving as the primary metadata store and enabling unified access to explore and visualize data across tools like Amazon QuickSight, Amazon Athena [30], and Amazon SageMaker. We leverage Athena's serverless SQL query engine to enable users to analyze the consolidated data using standard ANSI SQL without needing to connect to the underlying data stores, enhancing accessibility directly.

#### *7.4.3.2 Benefits of Cloud as a PaaS Platform*

In Platform as a Service (PaaS), the cloud's intrinsic security features are not just an add-on; they form the bedrock of a comprehensive data protection strategy. MINDS utilizes the built-in security of cloud platforms to protect data. We implement several security services from AWS to ensure our data storage and processing are safe and private. This includes security, management, and backup mechanisms.

Security and management are critical aspects of any data management system, especially when dealing with sensitive medical data. In MINDS, we employ several AWS security services and best practices to ensure the highest data security and privacy level. Amazon S3, where our data lake resides, provides robust security capabilities. We have enhanced these with network traffic encryption using TLS 1.2 and enforcing data integrity with HTTPS. All data in S3 is encrypted at rest using 256-bit Advanced Encryption Standard (AES) keys managed through AWS Key Management Service (KMS). Additionally, we use Identity and Access Management (IAM) policies to precisely manage access at both the resource and action levels, utilizing temporary credential chains to avoid

exposure to raw secrets. Our setup includes Virtual Private Cloud (VPC) endpoints to prevent public exposure of the data.

Our data warehouse, Amazon Redshift [34], incorporates multiple layers of security to protect sensitive oncology data. It integrates with AWS IAM, allowing fine-grained access control to resources. Data in transit to and from Redshift is protected using SSL connections. For data at rest, Redshift employs encryption using Key Management Service (KMS) and Hardware Security Module (HSM) encryption for large volumes exceeding terabytes. Redshift also enforces strict SQL-based authorization to ensure secure data access [39]. Furthermore, we utilize features like Virtual Private Cloud (VPC) for network isolation and comprehensive audit logging and compliance certifications for enhanced security and accountability.

We adhere to stringent security practices in the context of data processing and ETL with AWS Glue. AWS Glue is integrated with AWS Lake Formation, which allows for fine-grained, column-level access control, ensuring that only authorized personnel can access sensitive data. AWS Glue ETL jobs run in a secure and isolated environment, with all necessary resources provided by AWS Glue [40]. This is complemented by regular updates to server security groups, operating system patches, and adherence to the Center for Internet Security (CIS) hardening guidelines. For data consumption, Amazon QuickSight employs AWS IAM and AWS Lake Formation for robust access control, supporting both encryption at rest via AWS KMS and encryption in transit using SSL. Additionally, AWS CloudTrail provides detailed audit logs, enabling effective incident investigation and response.

In addition to the above-mentioned security measures, we also employ monitoring and logging using AWS CloudTrail and Amazon CloudWatch [31]. These services provide visibility into user activity and API usage, allowing us to detect unusual or unauthorized activities. This helps build audit trails and trigger security events in case of an undesired action. We also use Amazon RDS Multi-AZ deployments for redundancy, high availability, and failover support for database instances. Multi-AZ creates a primary RDS instance with a synchronous secondary standby instance in another Availability Zone (AZ) for enhanced redundancy and faster failover.



MINDS leverages AWS services' robust backup, redundancy, and disaster recovery capabilities to maximize system availability and protect against data loss. Amazon S3 buckets are versioned, with all object modifications saved as new versions. This allows restoring to any previous version. Cross-region replication sends object replicas to geographically distant regions to mitigate region-level failures. S3 object lock prevents accidental deletions during a specified retention period. RDS clusters run as Multi-AZ deployments with a standby replica in a secondary AZ for high availability, automatic failover, and fast recovery. Point-in-time restore rolls back to previous database states using retained backups. Database snapshots are stored in S3 for long-term durability. Redshift distributes replicas across nodes for local redundancy. It replicates snapshots and transaction logs to S3 to protect against node failures. Snapshots can restore clusters to any point in time. Combining versioning, redundancy, failover capabilities, and recovery automation, MINDS provides resilience against failures and minimizes disruption. Robust security protects against data loss from malicious events.

#### *7.4.3.3 Scalability Across Different Platforms*

While the current MINDS implementation leverages AWS, the architecture is designed to enable deployment across different cloud platforms, not just AWS. The core methodology centers on interfacing with managed cloud services, abstracting the underlying infrastructure through common programmatic interfaces. This service-oriented approach enhances portability and avoids extensive customization tied to a single provider. For example, as shown in Figure 7.8, the S3 storage layer could be replaced with Google Cloud Storage buckets, AWS Glue with Azure Data Factory, RDS and Redshift with Snowflake's data platform, and Lambda with Cloud Functions. The overall system architecture would remain consistent while swapping the provider services. When migrating platforms, trade-offs exist around performance, access controls, and other factors. But by using managed services with standard APIs, MINDS aims for platform-independent portability. The MINDS architecture can be replicated to the Google Cloud Platform to demonstrate feasibility through the following replacement and compatibilities.

- Employing Cloud Data Fusion for data integration in place of AWS Glue
- Leveraging BigQuery for data warehousing rather than Redshift
- Using Cloud SQL over RDS for relational data
- Adopting Cloud Functions and Cloud Run for serverless compute instead of Lambda.
- Visualizing with Looker as an alternative to QuickSight
- Applying Cloud Data Loss Prevention for security rather than AWS options

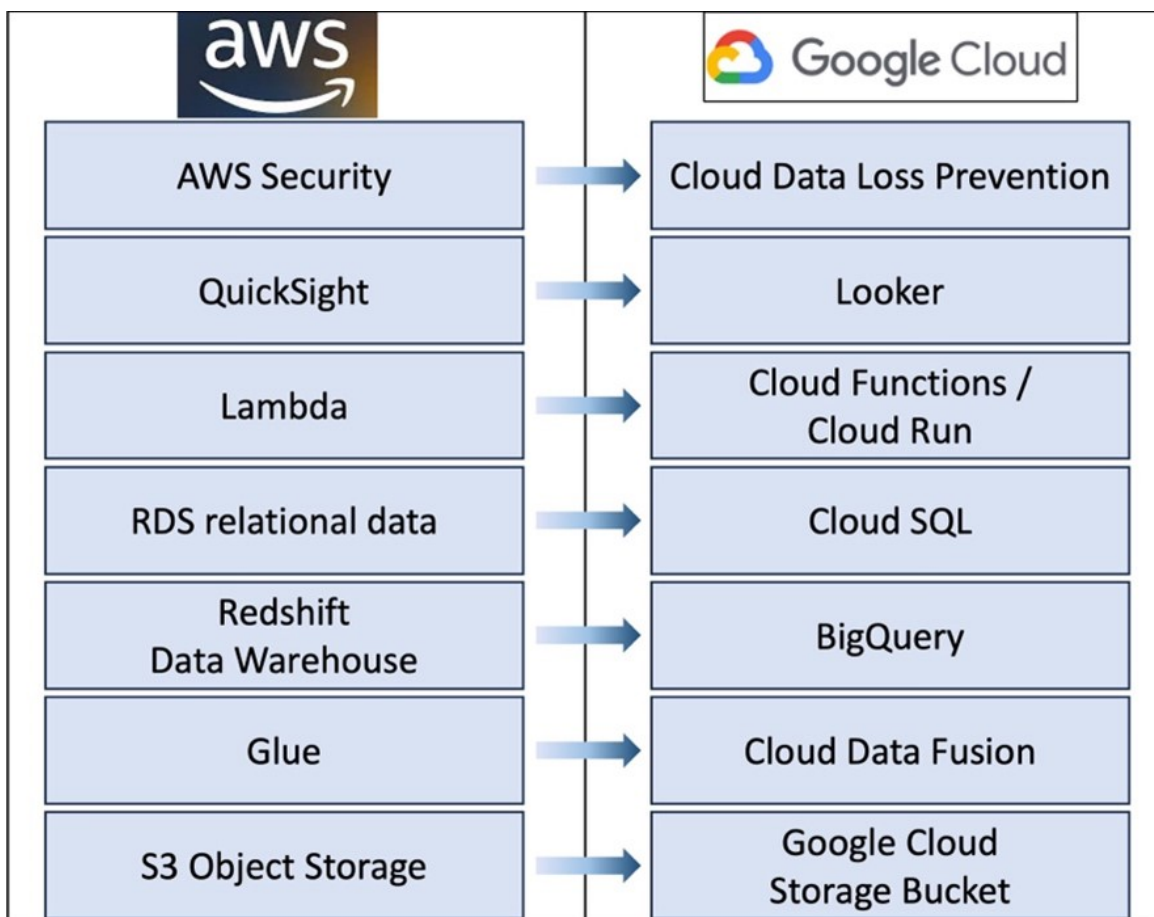


Figure 7.8 Demonstrating the feasibility of deploying MINDS across cloud platforms.

#### 7.4.4 On-Premise Deployment

While the cloud delivery model provides advantages like elastic scalability, hands-off management, and usage-based costing, some organizations may prefer on-premise deployment of MINDS due to data sovereignty, customization, or latency constraints. Despite extensive security protections, regulated data may mandate localized processing. Custom modules like augmented analytics dashboards may also require internal hosting. We provide an open-source Python toolkit for configurable local installations to address these needs while retaining MINDS' consistent methodology.

The MINDS library abstracts the orchestration of storage, databases, and web services into simple commands. A Docker container runs the setup scripts to bootstrap a production-ready environment. This generates a MySQL database pre-populated with the consolidated clinical data schema. The library emulates S3's file layout to organize unstructured dataset downloads. A lightweight Flask web application replaces interactive dashboards for cohort queries and drilling into associated multimodal records. Python notebooks connect natively to the local database for flexible ad-hoc analysis.

While foregoing autoscaling capabilities, on-premise deployment grants organizations direct control to modify pipelines, incorporate sensitive data, and reduce external network dependencies. The toolkit ensures feature parity while unlocking customizations. The same MySQL structure retains compatibility with predictive models trained in the cloud. Consistent metadata schemas, entity definitions, and configurability guard against lock-in across deployments. By supporting flexible topologies, MINDS balances sovereign data management with scalable cloud analytics.

#### 7.4.5 User Application

The MINDS platform aims to support users across academia, industry, and clinical settings by enabling scalable and secure access to integrated multimodal oncology data. Researchers can leverage MINDS to store, organize, search, and analyze large volumes of heterogeneous data spanning modalities like imaging, sequencing, pathology, and EHRs. For example, a lung cancer researcher may want to analyze treatment response biomarkers across a substantial cohort of lung

adenocarcinoma patients. However, gathering sufficient cases poses barriers, as relevant data resides in siloed repositories and trial databases. Each data source may only have a few hundred labeled lung adenocarcinoma cases that meet the desired criteria. Using MINDS, the researcher can easily construct an expansive harmonized analysis cohort. They can perform an SQL query against the aggregated clinical data warehouse to select all lung adenocarcinoma cases across the 95,000+ aggregated case database. This unified view allows for the efficient building of a cohort of over 7000 consolidated lung adenocarcinoma cases—a scale far beyond what any individual source provides. MINDS data processing pipelines will have mapped the clinical data from diverse sources like TCGA, TARGET, GENIE, and clinical trials into a standardized representation aligned with the GDC data model. This harmonizes heterogeneity and structures cohorts for analysis. The researcher can feed this harmonized case ID list into the unstructured data download clients. The tool will automatically retrieve all raw sequencing, imaging, and pathology data objects associated with each case from connected GDC, IDC, and PDC repositories. The researcher now has a turnkey dataset with thousands of consistently structured lung cancer cases annotated with multimodal data. This fuels large-scale integrative experiments to uncover treatment response biomarkers that drive outcomes. By aggregating and standardizing dispersed data into a centralized warehouse, MINDS created an augmented lung cancer cohort at a far larger scale and faster pace than otherwise feasible. This accelerates the discovery process through transformational access to interconnected big data.

## **7.5 Results and Discussion**

This section presents the results of implementing the proposed MINDS architecture for integrated multimodal oncology data management. We demonstrate MINDS' cohort building and data tracking capabilities and present its advantages over current solutions.

### **7.5.1 Multimodal Data Consolidation**

A fundamental challenge in developing integrated multimodal learning models is assembling the highly heterogeneous and fragmented data from myriad sources into unified datasets at sufficient

Table 7.2 Comparison of the storage size

<b>Data Source</b>	<b>Storage Size</b>	<b># of Cases</b>
MINDS	25.85 MB	41,499
PDC	36 TB	3,081
GDC	3.78 PB (17.68 TB public)	86,962
IDC	40.96 TB	63,788

scale. As shown in Table 7.2, MINDS directly addresses this by consolidating over 41,000 open-access cancer case profiles spanning diverse research programs into a structured 25.85 MB extract. This aggregated dataset encompasses clinical, molecular, and pathological data elements, providing a multifaceted view of each patient. Compared to petabyte-scale source systems, the extreme compression enables single-node processing and complex SQL analytics that are infeasible on individual repositories. The storage sizes reported for the GDC, PDC, and IDC refer to the total data contained in each repository. However, only a subset of cases in these repositories are open-access and available for research without access restrictions. For example, the GDC contains over 3 petabytes of genomic, imaging, and clinical data overall, but only 17.68 terabytes are associated with open-access cases that can be freely downloaded and analyzed. The 41,499 cases consolidated in MINDS are derived from these open repositories for unencumbered research use.

As shown in Table 7.3, the consolidated cases represent a comprehensive amalgamation of historical and contemporary research initiatives, vital for maintaining the relevance and accuracy of downstream analytical models in the face of evolving technologies. For example, the 11,315 cases from The Cancer Genome Atlas (TCGA) provide invaluable high-throughput molecular profiling using earlier genomic microarray platforms. In contrast, the 18,004 cases from Foundation Medicine incorporate the latest in contemporary genomic assays, such as next-generation sequencing (NGS) techniques. This strategic blending of data spanning different technological eras—from classic projects like TARGET to modern Foundation Medicine NGS panels—is critical for mitigating chronological biases and batch effects. By integrating this temporally diverse data through MINDS’ heterogeneous integration framework, we proactively inoculate our models against chronological distortions. This approach ensures that the algorithms focus on learning

durable, biological patterns that are generalizable across technological shifts rather than transient, platform-specific technical artifacts. Consequently, this temporal synthesis strategy enhances the generalizability of the machine learning models and future-proofs them against inevitable progress in profiling techniques. Access to such a rich and varied dataset is indispensable for training machine learning models, as it provides the large sample sizes necessary for deep learning and helps avoid statistical biases and spurious correlations that often arise from analyzing isolated datasets.

Table 7.3 Number of cases by programs from GDC open cases present in MINDS.

<b>Program</b>	<b># of Cases</b>
Foundation Medicine (FM)	18,004
The Cancer Genome Atlas (TCGA)	11,315
Therapeutically Applicable Research to Generate Effective Treatments (TARGET)	6542
Clinical Proteomic Tumor Analysis Consortium (CPTAC)	1526
Multiple Myeloma Research Foundation (MMRF)	995
BEATAML1.0	756
NCI Center for Cancer Research (NCICCR)	481
REBC	440
Cancer Genome Characterization Initiatives (CGCI)	371
Count Me In (CMI)	296
Human Cancer Model Initiative (HCMI)	228
West Coast Prostrate Cancer Dream Team (WCDT)	99
Applied Proteogenomics Organizational Learning and Outcomes (APOLLO)	87
EXCEPTIONAL RESPONDERS	84
Oregon Health and Science University (OHSU)	80
The Molecular Profiling to Predict Response to Treatment (MP2PRT)	52
Environment And Genetics in Lung Cancer Etiology (EAGLE)	50
ORGANOID	49
Clinical Trials Sequencing Project (CTSP)	44

### 7.5.2 Storage Optimization

By selectively assimilating solely essential clinical, biospecimen, and assay metadata instead of complete image pixel repositories, the MINDS structured ingestion approach reduces storage footprints from original petabyte scales down to a consolidated 25.85 MB extract. This approximate

1000× storage optimization maintains versatile multivariate cohort filtering capabilities across the 41,499 case corpus. Concretely, archiving TCGA, TARGET, and Foundation Medicine oncology profiles requires only MBs—facilitating responsive analytics from single commodity hardware, otherwise impossible at native TB+ scales. MINDS shifts the storage complexity curve through this strategic assimilation to unlock unified exploration. Deferring transfers of raw pixels and nucleic acid sequences until specifically requested for focused analysis prevents excessive upstream overheads. By directly handling initial cohort filtering on structured metadata upstream, MINDS right-sizes infrastructure economics to enable cloud-scale interactivity. Only specifically tailored subsets subsequently retrieve associated imagery. This optimized 2-stage architecture minimizes waste for targeted investigation. We further dissected storage contributions across warehoused clinical, biospecimen, and molecular categories, with cases consuming 10.24MB, followed by genomic variant calls and read groups. This proportional breakdown spotlights metadata categories benefiting from the greatest compression—guiding potential raw assimilation. Measurable storage optimization unlocks interactive analysis otherwise hampered by extreme technical costs, demonstrating quantifiable efficiencies.

### 7.5.3 Horizontal Scalability

Given that cohort queries constitute read-only analytical workloads, MINDS can scale underlying AWS Redshift compute capacity horizontally by adding managed nodes to meet surging analysis demands transparently. We empirically demonstrate corresponding latency reductions by doubling the cluster nodes, which directly halved runtimes for intensive 8-table cohort investigative queries, proving straightforward scaling. As increasingly complex algorithmic analysis workloads like multimodal federated learning and neural network training expand against MINDS unified corpus, decoupled storage from flexible computing facilitates economic growth, avoiding over-provisioning. This configurable capacity directly fulfills emerging surge requirements without architectural redesign. By empirically plotting reductions in query latencies resulting from MINDS-scaled infrastructure, we substantiate real-world horizontal scalability vital for cloud

viability. Similar economical scaling approaches apply when running MINDS datasets through downstream machine learning toolchains. Distributed training frameworks like XGBoost or PyTorch natively support decentralized parallel and vectorized execution pipelines across GPU grids. Maintaining unified data formats ensures interoperability with leading computational platforms.

By harmonizing dispersed data silos into a unified resource, MINDS effectively addresses the primary bottleneck in large-scale multimodal healthcare machine learning model development—a sufficiently large, heterogeneous, and representative dataset for training and validation of models.

#### 7.5.4 Cohort Building

Once aggregated data has been consolidated, tailored cohort extraction is needed to develop optimal machine learning training and test sets. Simple random sampling often fails to provide adequate cohort stratification along key variables. MINDS enables researchers to construct customized cohorts flexibly by querying the unified clinical data using performant SQL.

MINDS implements a flexible end-to-end workflow that allows users to submit analytical cohort queries and receive customized structured or unstructured data extracts. Figure 7.9 provides an overview of the MINDS system and all the data and query interactions with the user. The process begins with users formulating SQL-based queries that specify criteria to define a cohort of interest. These parameterized queries filter over patient attributes and allow the inclusion of any desired clinical, molecular, or demographic factors. For structured data, the submitted SQL query executes against MINDS' consolidated EHR database containing harmonized patient profiles. This filtered extraction returns a Pandas data frame containing detailed clinical records for all patients matching the cohort criteria. Alternatively, users can request unstructured data for their defined cohort. In this case, MINDS first extracts a list of unique patient case IDs for those meeting the criteria based on the SQL query parameters. These case IDs are then used to retrieve all associated unstructured medical objects related to those patients from connected repositories. This includes digital pathology slides, medical images like CT/MRI scans, -omics assay files, and other multimodal data assets. This flexible yet automated workflow allows researchers to obtain



structured medical records from the EHR or full multimodal datasets matching customized cohorts simply by submitting analytical SQL queries. The tight integration between cohort definition and data extraction enables the on-demand assembly of tailored data corpora for various biomedical applications.

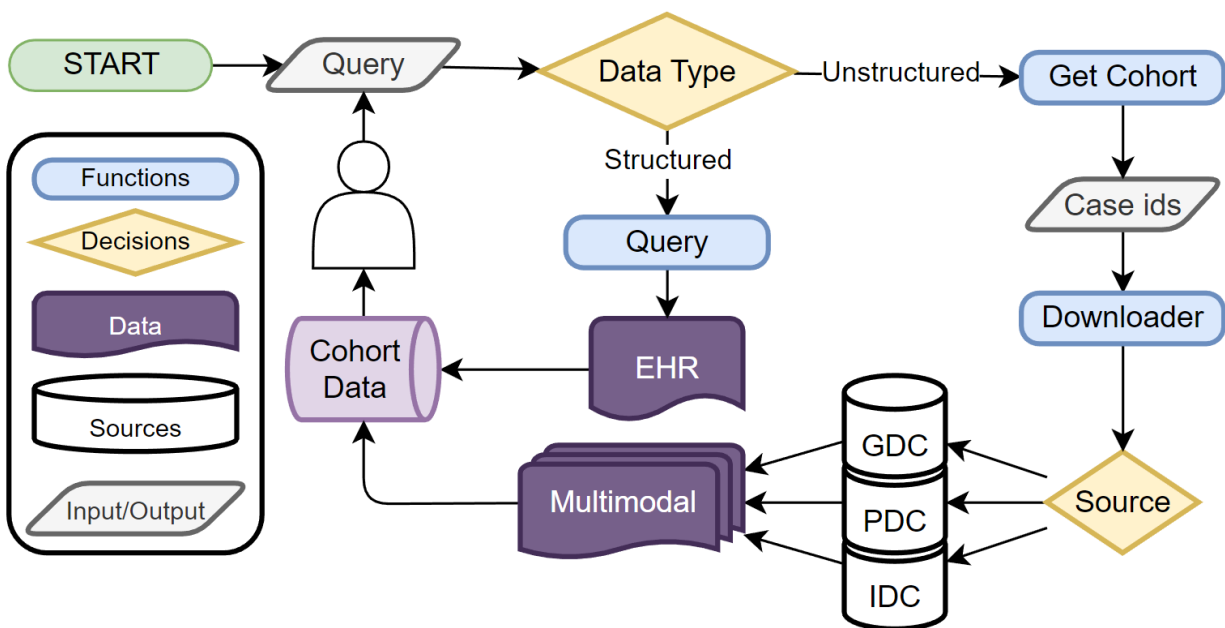


Figure 7.9 Overview of the workflow in MINDS.

Preliminary experiments demonstrate interactive cohort construction, with simple queries on a single clinical factor completed on average in 3–5 s. Even multidimensional queries joining clinical, molecular, and outcome data across tables are completed within 15 s. This enables rapid, iterative refinement of cohort criteria during model development.

### 7.5.5 Query Responsiveness

To quantify system performance, we extensively measured SQL query latencies over diverse criteria ranging from simple filters to multidimensional predicates across interconnected data domains. These complex joins emulate realistic exploratory analysis patterns that investigators conduct to uncover relationships within and across data types. Rigorously quantified wall-clock timings over

24,000+ SQL invocations reveal consistent sub-5 s average response times for typical single table queries. More complex multidimensional queries encompassing 8+ tables are completed within 15 s despite traversing metadata for thousands of cases. Minor fluctuations arise based on query types, but critically, sub-15-s overheads enable practically interactive cohort investigation workflows, allowing analysts to rapidly iterate without experiencing disruptions commonplace in legacy repositories. By maximizing fluidity, MINDS facilitates discovering underlying correlations that otherwise remain obscured in fragmented systems.

Researchers have full flexibility to extract customized sets for training algorithms by simply adjusting Boolean logic combining clinical, molecular, or biospecimen factors in the SQL queries. No system constraints are imposed. The ability to interactively construct bespoke cohorts by piping SQL queries directly on consolidated records has several key advantages for multimodal machine learning:

- MINDS allows researchers to build cohorts tailored to the problem. This prevents sampling biases linked to the availability of pre-defined cohorts.
- SQL combines and consolidates disparate clinical, molecular, and outcomes data from the entire period of medical treatment. This provides a complete view of each patient.
- Version IDs uniquely label dataset variants to enable precise tracking of changes during iterative model development. Researchers can pinpoint the exact dataset used to generate each model version.
- JSON manifests comprehensively log the dataset composition, including the originating queries, data sources, and extraction workflows. This provides full documentation of the data provenance.

Aligning emerging systems like the Multimodal Integration of Oncology Data System (MINDS) with such technologies is vital to avoid isolated silos and enable integrated analytics over clinical and research data. This demands extensive use of their common formats, unique identifiers, controlled vocabularies plus considerable translator development.

### 7.5.6 Data Tracking and Reproducibility

MINDS further simplifies multimodal analysis by automating the rebuild of full datasets tailored to each cohort. APIs and utilities extract images, -omics, and other unstructured data linked to cohort cases from connected repositories like GDC. Consistent organization and JSON manifest document datasets ready for consumption by machine learning models. To ensure reproducibility, MINDS assigns unique version IDs to cohort datasets. Any changes trigger new versions, enabling precise data tracking to develop different model variants. Comprehensive data provenance from EHR queries to unstructured set regeneration enhances reproducibility in machine learning training pipelines.

### 7.5.7 Integrated Analytics

Once unified datasets have been constructed, interactive analytics and visualizations are needed to explore cohort characteristics, correlations, and model outputs. MINDS delivers rapid analysis over aggregated multimodal data through integrated dashboards powered by Amazon QuickSight. Optimized cloud data warehousing components like Amazon Redshift enable ad-hoc exploration across thousands of variables without performance lags. QuickSight's advanced machine learning-driven insights uncover subtle trends and patterns. User-defined charts visualize model performance metrics across various cohorts. Key advantages of integrated analytics include:

- Rapid hypothesis testing during exploratory analysis to refine cohorts and features.
- Understanding model performance across cohorts reveals generalization capabilities.
- Uncovering correlations between clinical factors, assays, and predictions guides feature engineering.
- Visualizations build trust by providing direct views into model behaviors.

### 7.5.8 Limitations and Future Improvements

While MINDS has demonstrated significant benefits, there are several areas where the system could be improved. Including controlled data, a local deployment option, and enhanced analytics and visualization capabilities represent exciting directions for future work on MINDS. These improvements would increase the amount of data available in MINDS and enhance its utility for oncology research. Another future extension to this work could be to replicate MINDS on the Google Cloud Platform or Microsoft Azure platform. While there would be specific technical differences across providers, the high-level design focused on abstracted services ensures the seamless prevention of vendor lock-in. Multi-cloud deployments ensure MINDS provides flexible, portable data management capabilities spanning diverse infrastructures. To track the addition, deletions, and modifications to data, webhooks and event notifications can be implemented to achieve more real-time incremental updates. For example, an event trigger could invoke our ingest handler when new data is added to the remote platform. This event-driven approach avoids excessive API polling. Webhooks allow registering listeners to be notified immediately of data changes.

Additionally, though initially focused on cancer data, MINDS's flexible and modular design makes it well-suited for application across medical specialties. For example, the infrastructure could readily incorporate COVID-19 data types such as clinical outcomes, chest CT scans, and immunological biomarkers from initiatives like the Medical Imaging and Data Resource Center (MIDRC) [482] to accelerate insights. By ingesting such assets via extensions to the automated ETL pipelines and data model while reusing the security, governance, and analytics foundations, MINDS could integrate emerging COVID-19 knowledge. More broadly, maintaining interoperable components enables consolidating distributed data silos across domains to advance data-driven medicine beyond just oncology through unified analytics.

While MINDS demonstrates significant benefits in enabling integrated analytics, some core limitations provide fruitful directions for further enhancement. Given infrastructure barriers, a primary constraint centers on directly ingesting raw clinical imagery and video. However, introducing dimension reduction through learned embeddings holds promise for overcoming such hurdles while

preserving semantic representation. Exploring privacy-preserving approaches would also facilitate assimilating regulated data assets beyond public corpora. Additionally, absorbing unstructured physician notes poses non-trivial natural language understanding challenges needing advancement through pre-trained clinical language models. If these addressable constraints are tackled, the potential significance would be immense. MINDS could profoundly transform integrated biomedical investigation paradigms by synergizing heterogeneity and multiplicity across exponentially growing streams. New modalities, data types, and controlled datasets could continually expand the scope. Assimilating free text notes could uncover novel linguistic biomarkers. Exciting enhancements we have highlighted include, incorporating regulated data through privacy-preserving methods, migrating imagery via compact embeddings, absorbing unstructured notes through advanced NLP, expanding across diseases by reusing consolidation components, and scaling across cloud platforms to prevent vendor lock-in.

## **7.6 Conclusions**

The MINDS was designed to address the challenges of integrating and managing large volumes of oncology data from diverse sources. MINDS provides a cost-effective and scalable solution for storing and managing oncology data through its innovative cloud technologies and data mapping techniques. It leverages public datasets to ensure reproducibility and enhance machine learning capabilities while providing a clear pathway for including controlled data in the future. Our results demonstrate that MINDS significantly reduces storage size and associated costs compared to traditional data storage methods. MINDS' compatibility with public datasets ensures no leaks of controlled data while allowing for reproducibility of results. The system also enhances machine learning capabilities by updating patient information as new data is released from clinical trials, providing transparency and reproducibility.

The data integrated into MINDS is sourced from publicly available datasets in the Genomic Data Commons (<https://gdc.cancer.gov>, accessed on 12 January 2024), Proteomics Data Commons (<https://proteomics.cancer.gov>, accessed on 12 January 2024), Imaging Data Commons (<https://imaging.cancer.gov>, accessed on 12 January 2024), and the Cancer Research and Biotechnology Data Commons (<https://cancerresearchandbiotechnology.org>, accessed on 12 January 2024).

[//imaging.cancer.gov](https://imaging.cancer.gov), accessed on 12 January 2024) and cBioPortal for Cancer Genomics (<https://www.cbioportal.org>, accessed on 12 January 2024). Only metadata and identifiers are ingested into MINDS—the datasets remain hosted in their respective repositories. Analyses are performed by querying these sources through their public APIs and computational workbenches. The code implementation for the MINDS platform is available at <https://github.com/lab-rasool/MINDS>, accessed on 11 January 2024.

## **Chapter 8: SeNMo: A Self-Normalizing Deep Learning Model for Enhanced Multi-Omics Data Analysis in Oncology**

### **8.1 Introduction**

Cancer data has multiple modalities, each offering distinct but complementary views of the disease [333]. Radiological images reveal structural and functional anomalies, histopathology slides provide cellular and tissue-level detail, clinical and Electronic Health Records (EHR) data encapsulate patient history and treatment outcomes, and molecular data such as genomics, transcriptomics, proteomics, and metabolomics uncover the underlying biological mechanisms driving cancer progression and response to therapy [77, 383, 494, 124]. Studying cancer through a multimodal perspective is critical for comprehensive understanding and effective treatment strategies [273]. Additionally, multimodal approaches facilitate personalized medicine, enhance our ability to predict disease-related outcomes, and advance our understanding and treatment of cancer [6].

#### **8.1.1 Multimodal and Multiomics Data**

The growth of molecular data has greatly advanced cancer research [124]. The emergence of high-throughput sequencing technologies supported by the development of sophisticated bioinformatics tools and computational algorithms has ushered in an era of “omics” [559]. Multi-omics is a subset of multimodal data that specifically refers to the integrated analysis of various molecular data modalities including genomics, transcriptomics, proteomics, and metabolomics [739]. Multi-omics provides a comprehensive view of the biological processes and molecular mechanisms underlying cancer [808]. By combining these different layers of molecular data, multi-omics transcends the limitations of single-omic studies, which might only provide a partial view of the

disease, by illustrating how various molecular components (DNA mutations, protein expression, and RNA expression, etc) interact within the complex biological network of cancer [278].

### 8.1.2 Pan-cancer Perspective

Researchers have studied individual cancers as well as pan-cancer data. Although studying individual cancers has shown significant benefits in understanding specific pathways and therapeutic responses, the pan-cancer approach offers a broader, more systemic view that can accelerate breakthroughs applicable across multiple types of cancer. Pan-cancer studies have enabled the identification of commonalities and differences across various cancer types, leading to insights that may not be as evident when focusing on a single cancer type [700]. The pan-cancer setting has identified universal cancer vulnerabilities, detailed pathway alterations for cross-cancer diagnostics and treatments, and revealed shared oncogenic pathways and mutation patterns, uncovering new clinically useful features [297, 605, 290, 700]. Similarly, pan-cancer studies have identified key molecular signatures that can predict response to immunotherapy across different tumor types, demonstrating the clinical relevance of pan-cancer approach [684, 421].

### 8.1.3 Existing Landscape of Pan-cancer Multi-omics Analysis

Traditionally, multimodal, multi-omics, and pan-cancer studies are conducted through a variety of techniques and methods that leverage advanced computational tools, bioinformatics, statistical methods, machine learning, and deep learning models to integrate and interpret complex oncology datasets. The data integration techniques in multi-omics are generally categorized into supervised, weakly supervised, and unsupervised methods, which can be further sub-categorized into (1) feature extraction (selection, extraction, and dimensionality reduction), (2) feature engineering (transformation, reducing dimensionality, normalizing and simplifying data, reducing noise, and alignment), (3) network-based methods (patient similarity networks, patient-drug networks, drug-drug networks, etc.), (4) clustering (grouping similar samples, stratification, feature selection, and grouping biological modules), (5) factorization (decompose or factorize features, multiple kernel



learning, Bayesian consensus, similarity network fusion, NMF), and (6) Deep Learning (CNNs, MLPs, RNNs, Transformers, GNNs, etc.) [4, 739, 13, 734]. Deep learning, a subset of machine learning characterized by the use of neural networks with many layers, has dramatically transformed the study of high-width (many features), low-length (fewer samples) molecular data [12, 735]. With its inherent capacity to model complex, non-linear relationships and to handle vast datasets, deep learning has proven adept at uncovering patterns that traditional analysis may overlook. There are elaborate reviews in the existing literature that analyze different pan-cancer, multimodal, multi-omics works [430, 87, 285, 660, 739, 738, 694, 691]. Using multi-omics data to enhance cancer diagnosis, prognosis, and treatment planning offers a transformative opportunity. Recent literature underscores the shift towards computational integration of heterogeneous biological data, revealing critical insights into cancer's multifaceted nature. Here we examine seminal works that encapsulate the paradigm shift in multi-omics cancer studies.

A notable advancement is the utilization of Self-Normalizing Neural Networks (SNNs) for pan-cancer classification, which, as demonstrated by the study on copy number variation (CNV) data from The Cancer Genome Atlas (TCGA) lung adenocarcinoma (LUAD), ovarian (OV), liver hepatocellular carcinoma (LIHC), and breast (BRCA) cancers, underscores the importance of feature selection in managing high-dimensional data for effective disease categorization [410]. The SNN model trained to perform pan-cancer classification yielded superior accuracy and macro F1 scores over traditional algorithms like Random Forest [410]. Complementing the pan-cancer approach, an integrative analysis combining histology-genomic data via multimodal deep learning offered a broad-spectrum understanding of cancer biology [138]. With an extensive dataset from TCGA encompassing 14 cancer types, a deep learning multimodal fusion (MMF) model outperformed attention-based multiple-instance learning (AMIL) model and self-normalizing network (SNN), showcasing the benefits of integrative analytics over singular data type analyses [138]. Further emphasizing multi-omics data integration, DeepProg, an ensemble framework combined deep learning and machine learning for prognosis prediction [557]. By processing RNA-Seq, miRNA

sequencing, and DNA methylation for 32 cancer types from TCGA, DeepProg excelled in survival subtype prediction and risk stratification [557].

Khadirnaikar et al. identified novel subgroups with similar molecular characteristics by combining different models of machine learning and deep learning [352]. By reducing dimensions of multi-omics features (mRNA, miRNA, DNA methylation, protein expression) and applying various classifiers, this approach identified subgroups across 33 tumor types. The authors argued that the number of samples should be commensurate with the number of dimensions for better prediction power of a learning model [352]. Another study used four types of -omics data (gene expression, miRNA expression, protein expression, and DNA methylation) for two datasets (TCGA-BLCA, TCGA-LGG) to predict Progression-Free Interval (PFI) and Overall Survival (OS) through Multiview Factorization AutoEncoder [462]. The identification of pan-cancer prognostic biomarkers through integrated multi-omics data (DNA methylation, gene expression, somatic copy number alteration (CNA), and miRNA expression) across 13 cancers showed the performance of statistical and bioinformatic methods in survival-related gene discovery [804]. The predictive capability of multi-omics data is further evidenced in non-small cell lung cancer (NSCLC) survival prediction, where the combination of five modalities, miRNA, mRNA, DNA methylation, long non-coding RNA (lncRNA) and clinical data, showed superior C-indices compared to individual modalities [205].

The advantage of multimodal data fusion for survival prediction is quantified across various cancer stages and types, with FUSED models exhibiting better average C-index compared to various machine learning and bioinformatics methods [511]. This approach combined clinical features with genomic, transcriptomic, and proteomic data in oncological prognostics across 33 cancer types [511]. Deep learning-based clustering method called MCluster-VAEs predicted subtype discovery using multi-omics data (mRNA, miRNA, DNA methylation, CNA) across 32 cancer types, outperforming traditional methods [592]. Decoupled contrastive learning model, DEDUCE, used multi-head attention decoupled contrastive learning approach for subtype clustering through multi-omics data consisting of gene expression, DNA methylation, and miRNA expression, across

five cancer types (BRCA, GBM, SARC, LUAD, STAD) [535]. The authors of DEDUCE used multi-head attention encoder network for cancer subtype discovery [535].

#### 8.1.4 Challenges and Opportunities

Although good for the task at hand, the above-mentioned methods often struggle to fully capture the complexity and heterogeneity of cancer due to their inherent limitations in handling and interpreting vast, multidimensional datasets. Dimensionality reduction methods such as principal component analysis (PCA) or t-distributed stochastic neighbor embedding (t-SNE) can inadvertently discard subtle yet crucial biological nuances that might be pivotal for understanding disease mechanisms [330]. Learning-based dimensionality reduction methods, such as using deep learning models, lack the discriminating and interpreting ability of extracted features, lack consensus in the balance between the number of deep network layers vs. the number of layer neurons, and cannot handle or recover the missing data [330].

Similarly, feature selection and learning-based feature engineering, despite their effectiveness in identifying key predictors within datasets, can introduce biases and result in models that are suited to specific features of the data used for training [382, 770]. This compromises such methods' ability to perform well across different datasets or in real-world clinical settings [382, 770]. Furthermore, these methods frequently face challenges in terms of generalizability, as they may not perform consistently across diverse patient populations or varying biological conditions, limiting their utility in broader clinical practice. Thus, while these techniques are instrumental in advancing cancer research, their limitations highlight the need for more robust and generalizable framework that can more accurately predict end-points across different cancer types and data modalities.

Recently, a new class of deep learning models called foundation models that comprise LLMs (Large Language Models) and VLMs (Vision-Language Models) have been introduced by training on large multimodal data [733, 276]. These models have demonstrated a strong ability to generalize well across different tasks when provided with ample and diverse training data [733]. Due to their extensive and varied datasets, these models capture a broad range of patterns and nuances,

enabling them to apply learned knowledge flexibly and effectively across different contexts. The key conclusions from the success story of foundation models relevant to this study are as follows.

1. Foundation models are trained on massive datasets that encompass a wide spectrum of information across different domains and modalities. This extensive training helps the models develop a robust understanding of complex patterns and relationships within the data. For instance, models like GPT (from OpenAI) [102] and BERT (developed by Google) [179] have been shown to perform exceptionally well on a variety of natural language processing tasks, from translation to sentiment analysis, precisely because they have been exposed to large amounts of diverse textual data during training [276, 733].
2. VLMs integrate information from both visual and textual sources, allowing them to develop a more comprehensive understanding of the world. For example, models like CLIP (from OpenAI) [563] and ViLBERT (Vision-and-Language BERT) [449] learn to correlate images with text, enabling them to perform tasks such as image captioning or visual question answering with high accuracy. This ability to process and synthesize information across different modalities enhances their flexibility and adaptability to new tasks that may not strictly resemble those they were originally trained on [276].
3. The capability of these models to generalize is further evidenced by their performance across a range of tasks with minimal task-specific tuning [733]. For example, once trained, these models can often switch between tasks such as text classification, summarization, and even complex reasoning without extensive retraining. This adaptability is largely due to their training datasets' comprehensive and diverse nature, which provides a rich background against which the models can evaluate new problems [733, 739, 276].

The establishment of large-scale biological databases and data repositories, such as National Cancer Institute's (NCI) The Cancer Genome Atlas (TCGA) [687] and Clinical Proteomic Tumor Analysis Consortium (CPTAC) [206], hold vast amounts of cancer data, especially multi-omics data, that can be readily used for disease analysis. Despite many efforts in building a foundation

model for the omics data, the existing literature has no foundation model that has been trained on multi-omics pan-cancer data. scGPT is the foundation model trained for the single-cell sequencing data of 33 million cells [162]. SAMMS model has been trained on two cancer types (TCGA's LGG and KIRC) using clinical (age, gender), gene expression, CNV, and miRNA, and WSI data [818]. RNA Foundation Model (RNA-FM) was trained on 23 million non-coding RNA sequences [130]. PATH-GPTOMIC used CNV, genomic mutations, bulk RNA Seq, and WSI data to predict survival outcomes for two datasets (TCGA-GBMLGG, TCGA-KIRC) [723]. The absence of a pan-cancer, multi-omics foundation model can be attributed to the complexity and heterogeneity of such data, lack of comprehensive datasets, specificity of current analytical methods, and large computational and resource constraints. To address these challenges, we consider a multi-omics, pan-cancer framework that involves only essential pre-processing steps. We propose a mini-foundation model called SeNMo (Self-Normalizing Deep Learning Model for Multi-Omics). SeNMo has been trained on six data modalities including clinical, gene expression, miRNA expression, DNA Methylation, DNA Mutations, and reverse-phase protein array (RPPA) expression data across 33 cancer types. We call it a mini-foundation model because the latent representations generated by SeNMo have not been evaluated for properties such as generalization, emergence, expressivity, scalability, and compositionally, which are essential traits for a model to be named as "foundation model" [24, 733]. We evaluated SeNMo's generalization capability to tasks such as OS prediction and primary cancer classification. The rest of the evaluation is beyond the scope of this work and shall be pursued in future endeavors.

Predicting overall survival and accurately classifying cancer types are pivotal endpoints in cancer research and patient care. For patients, these predictions can inform treatment options, influence monitoring strategies, and guide clinical decision-making, thus impacting quality of life and survival prospects. For healthcare systems and researchers, the ability to predict outcomes enhances understanding of the disease, improves the design of clinical trials, and drives the development of new therapeutics. In essence, enhancing the accuracy of survival predictions and cancer classification through advanced computational methods not only stands to revolutionize the

clinical approach to oncology but also embodies the patient-centered ethos that is central to modern medicine. SeNMo offers the following contributions:

1. We provide oncology data analysis using molecular correlates of patient prognosis across 33 cancer types, addressing both disease-wide and individual patient levels.
2. We created a multi-omics, pan-cancer framework with minimal and essential pre-processing steps, eliminating the need for complex, custom-engineered methods. This allows researchers to concentrate on the learning aspect.
3. We developed a mini-foundation model capable of generalizing to different tasks and unseen data through fine-tuning.
4. Our findings indicate that MLP-based networks are highly susceptible to catastrophic forgetting. We show that fine-tuning should involve a fraction of the epochs ( $\leq 30$ ) while adjusting the learning rate, weight decay, and dropout to fractionally update all layers of the trained model.
5. The SeNMo framework represents the first initiative to analyze 33 cancer types using six molecular data modalities: clinical data, gene expression, miRNA expression, DNA methylation, DNA mutations, and protein expression.
6. We present the latent feature vectors learned by SeNMo as an open-access vector database system, HoneyBee. This resource enables researchers to explore further, discover biomarkers, and assess features.
7. The codebase and resultant embeddings are made available as open-source through GitHub and HuggingFace for the research community to use.

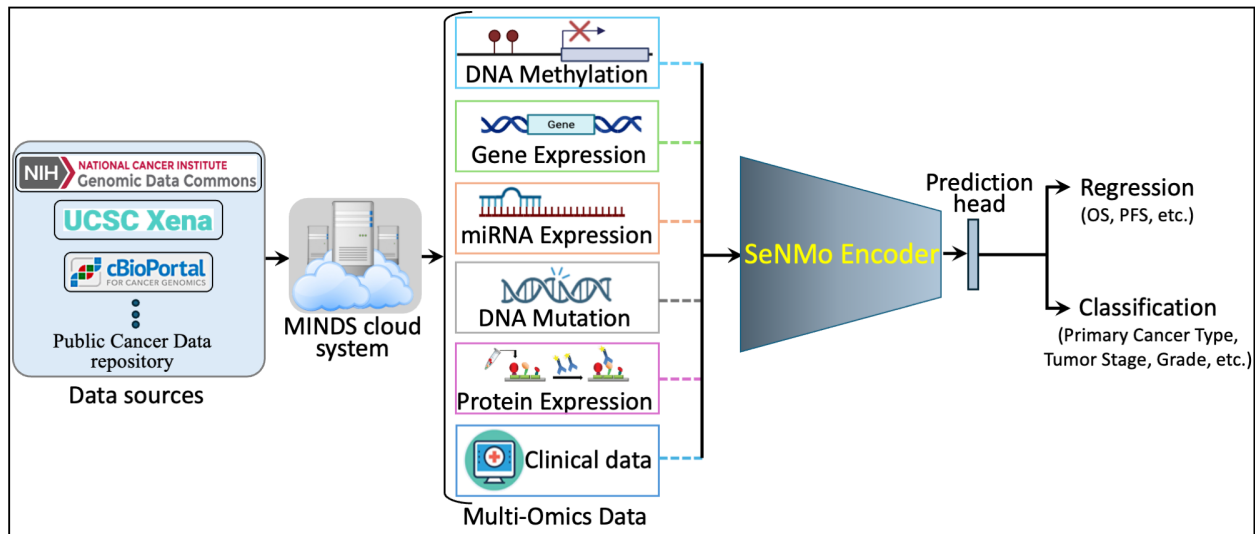


Figure 8.1 Overview of the SeNMmo framework.

## 8.2 Materials and Methods

### 8.2.1 Data Acquisition

TCGA houses one of the largest collections of high-dimensional multi-omics datasets for more than 33 different types of cancer for around 20,500 individual tumor samples [687]. The multi-omics data contains high-throughput RNA-Seq, DNA-Seq, miRNA-Seq, single-nucleotide variant (SNV), CNV, DNA methylation, and RPPA data [687]. Building cohorts for patient data lying across different data formats, modalities, and systems is not a trivial task. To curate the data and build the patient cohorts, we used our previously developed Multimodal Integration of Oncology Data System (MINDS) which is a metadata framework for fusing data from publicly available sources such as the TCGA-GDC and UCSC Xena portal into a machine-learning ready format [687, 248, 691]. MINDS is available as open-source for the cancer research community and we have integrated MINDS into the SeNMmo framework for enhanced outreach and benefit of researchers. For training, validation, and testing our model, we used pan-cancer data from TCGA and Xena comprising 33 cancer types, as shown in Table 8.1. We also fine-tuned the model on the

CPTAC-LSCC [612] and Moffitts' LSCC [659] data to evaluate the model's generalizability and transfer learning capabilities.

### 8.2.2 Data Modalities

Out of the 13 multi-omic modalities contained across each cancer dataset, the ones we chose were gene expression (RNAseq), DNA methylation, miRNA stem-loop expression, RPPA data, DNA mutation, and clinical data. The rationale behind selecting these specific modalities was that these modalities have been frequently preferred over the other data types in studying cancer primarily due to their direct relevance in the fundamental processes of cancer progression, their diagnostic and prognostic capabilities, and their established technologies [611, 421]. These modalities directly influence and reflect key biological processes fundamental to cancer progression, making them extremely valuable for uncovering the molecular mechanisms driving the disease [611]. Moreover, the set of these modalities provides robust predictive and prognostic information, and their integration provides a holistic view of a tumor's multi-omic profile [129, 611, 421]. Lastly, the selected modalities contained a static feature number across each cancer type that helped us in developing the standard data pre-processing pipeline for pan-cancer studies. Below we briefly describe each of the data modalities considered in this study, followed by the preprocessing steps undertaken to select the features used in training the SeNMo model.

1. *DNA Methylation* is a key epigenetic modification where methyl groups are added to the DNA molecule, typically at cytosine bases adjacent to guanine, known as CpG sites [447]. This process plays a crucial role in regulating gene expression without altering the DNA sequence itself [447]. In the context of cancer, DNA methylation patterns are immensely valuable as a data modality. Aberrant methylation of certain genes can lead to their silencing or activation, contributing to oncogenesis and tumor progression [390]. By analyzing methylation profiles across different cancer types, researchers can identify diagnostic markers, predict disease progression, and tailor personalized treatment strategies [390]. Thus, DNA methylation serves as a critical biomarker in oncology, offering insights into the molecular mechanisms



of cancer and enhancing the precision of therapeutic interventions [390]. DNA methylation is quantified through beta values, which range from 0 to 1, with higher values indicating increased methylation [197]. The beta values for DNA methylation data in TCGA-GDC were measured using the Illumina Human Methylation 450 platform, a sophisticated method for detailed methylation profiling [731]. The data consists of 485,576 unique cg and rs methylation sites across several tumor types [731].

2. *Gene Expression (RNAseq) analysis* via RNA sequencing (RNAseq) is a powerful data modality in cancer research, providing deep insights into the transcriptomic landscape of tumors [158]. This technique quantifies the presence and quantity of RNA in a biological sample at a given moment, allowing for a detailed view of transcriptional activity in a cell [158]. RNAseq helps to identify genes that are upregulated or downregulated in cancer cells compared to normal cells, offering clues about oncogenic pathways and potential therapeutic targets [286]. The TCGA-GDC gene expression data is derived from RNAseq, utilizing High-throughput sequence Fragments Per Kilobase of transcript per Million mapped reads (HTseq-FPKM) as a measure for normalization [250]. This approach normalizes raw read counts by gene length and the number of mapped reads. Further processing includes incrementing the FPKM value by one and then applying a log transformation to stabilize variance and enhance statistical analysis [584]. The data spans 60,483 genes, with FPKM values indicating gene expression level. Values over 1000 signify high expression, whereas values between 0.5 and 10 indicate low expression [250, 681].
3. *miRNA Stem Loop Expression* is a pivotal aspect in understanding the intricate regulatory mechanisms that miRNAs, or microRNAs, play in gene expression [551]. These small, non-coding RNA molecules typically function by binding to complementary sequences on target messenger RNA (mRNA) transcripts, leading to their silencing [551]. The expression of miRNAs involves a multi-step process, that ensures specific targeting and effective modulation of gene expression, crucial for both normal cellular function and pathological conditions,

such as cancer [551]. The miRNA expression values for TCGA-GDC were measured using stem-loop expression through Illumina, and values were transformed by adding one and being log transformed [145, 428]. These were mapped across 1880 features representing hsa-miRNA sites, where expressions varied between high and low values.

4. *Protein Expression* is an effective methodology that is similar to western blotting, and is used to quantify protein expression in tissue slides [186]. The technique involves transferring antibodies to a nitrocellulose-coated slide to bind specific proteins, forming quantifiable spots through a DAB calorimetric reaction and tyramide dye deposition, analyzed using "SuperCurve Fitting" software [186, 43]. This process allows for effective comparison of protein expression levels in tumor samples against benign samples, highlighting aberrant protein levels that drive the molecular phenotypes of cancer [186, 132]. Through the quantification of protein expression, RPPA uncovers the functional status of several signaling molecules, phosphorylation molecules, and metabolic molecules [43]. RPPA data has been generated from the profiling of nearly 500 antibody-proteins for a given patient and deposited in The Cancer Proteome Atlas (TCPA) portal [408]. Each data file contains 487 antigen ID (AGID), the peptide target ID, the gene identifier that codes for the protein, various other identifiers, and the antigen's corresponding expression level. Protein expression levels were normalized through log transformation and median centering after being calculated by the SuperCurve fitting software [339].
5. *DNA Mutation* from analysis of DNA sequence identify mutated regions compared to a reference genome, resulting in a variant calling format (VCF) file that details these differences [156, 155]. Aggregating VCF files to exclude low-quality variants and include only somatic mutations produces mutation annotation format (MAF) files [154]. Unlike VCF files, which consider all reference transcripts, MAF files focus on the most affected reference and include detailed characteristics and quantifiable scores that assess the mutation's translational impact and clinical significance[154]. This information is crucial because clinically significant

mutations often cause major defects in protein structure, severely impacting downstream functions. This dysregulation ultimately drives the development of various cancers [479]. SIFT and PolyPhen are features within MAF files that quantify a mutation's effect on its encoded protein and can be further analyzed. MAF files from TCGA-GDC contain 18,090 mutational characteristics [154].

6. *Clinical* data plays a crucial role in cancer research, serving as the foundation for correlating biological data with patient outcomes and demographics [494]. The clinical data encompasses detailed patient information, which is instrumental in understanding the epidemiology of cancer, evaluating treatment responses, and improving prognostic assessments [494]. Integrating clinical data with genomic and proteomic analyses can uncover relationships between molecular profiles and clinical manifestations of cancer [739]. Among the many clinical features and phenotypes, age, gender, race, and cancer stage are particularly emphasized in cancer research due to their significant impact on disease presentation, progression, and treatment efficacy [403, 446, 789, 773]. Age is a critical factor as the incidence and type of cancer often vary significantly with age, influencing both the biological behavior of tumors and the overall prognosis of patients [403]. Gender is another key determinant, as certain cancers are gender-specific, while others may show differences in occurrence and outcome between genders, likely due to biological, hormonal, and social factors [446]. Race has been linked to differences in cancer susceptibility, mortality rates, and treatment outcomes, reflecting underlying genetic, environmental, and socioeconomic factors [789]. Finally, cancer stage at diagnosis is paramount for determining the extent of disease and guiding treatment decisions, directly correlating with survival rates [773].

### 8.2.3 Pre-processing

Multomics data integrates diverse biological data modalities such as genomics, transcriptomics, proteomics, and metabolomics, to understand the complex mechanisms of diseases like cancer. However, before integration, this data requires multiple preprocessing steps to overcome the *big P*,

*small n* problem and other associated challenges of high-throughput molecular data. The *big P, small n* problem refers to a large number of features (P) and a small number of samples (n) in the data [426]. The pan-cancer multi-omics data comes with intra- and inter-dataset correlations, heterogeneous measurement scales, missing values, technical variability, and other background noise. Some of the most significant challenges include, (i) data heterogeneity, where each type of data encompasses unique properties and scales, (ii) volume and complexity, where overwhelming volume of data (often in terabytes), managing, storing, and processing requires substantial computational resources and advanced data management strategies, (iii) quality and variability incurred because of the different platforms resulting in batch effects, differing levels of sensitivity, noise, missingness, and varying error rates, and (iv) lack of standardization in how data is collected and processed across different laboratories and studies. These challenges are further pronounced when selecting the preprocessing steps to make the data machine learning-ready. Some of the tasks to consider while dealing with multi-omic data include:

1. Because of the diverse nature, each type of omics data requires specific normalization techniques to adjust for factors. (i.e. gene length in RNA-seq data or protein abundance in proteomics). Choosing the right normalization method based on the type of data is crucial to ensure that data is comparable [806, 350, 440].
2. Multiomics datasets often contain missing values due to detection limits or experimental errors. At times, as in our case, an entire data modality for a patient is missing. Selecting the robust imputation method for missing data is critical to avoid biased interpretations. Typical imputation methods suggest using mean, median, KNN, Gaussian mixture clustering, Bayesian, and deep learning-based (autoencoders) techniques to handle imputations [650].
3. The high dimensionality of multi-omics data often exceeds the number of samples available, leading to the risk of overfitting. Techniques like PCA, t-SNE, features selection, features engineering, and others are used to reduce dimensionality while preserving the most informative features of the data [48].

4. Proper annotation and comprehensive metadata are essential for the effective preprocessing of multiomics data. Metadata must capture details about sample collection, processing protocols, and experimental conditions, which are crucial for accurate data interpretation and reproducibility [623].
5. Integrating diverse datasets involves sophisticated statistical and computational methods. Techniques such as concatenation, transformation, and advanced modeling (ML/ DL algorithms) are usually used to merge these datasets coherently [399].

Addressing these challenges requires interdisciplinary expertise, including bioinformatics, statistics, and domain-specific knowledge. Here, we describe the preprocessing steps used across molecular data modalities.

- First, we removed the features that had NaNs across all the samples. This reduced the dimension, removed noise, and ensured continuous-numbered features to work with.
- Next, constant/quasi-constant features with a threshold of 0.998 were filtered out using Feature-engine, a Python library for feature engineering and selection [216]. This eliminated features with no expression at all across every sample along with features that were noise, since the expression value was the same across every sample.
- Next, duplicate features between genes were identified that contained the same values across two separate genes, and one of the genes was kept. This may reveal gene-gene relationships between the two genes stemming from an up-regulation pathway or could simply reflect noise.
- Next, we filtered the features having low variance ( $\approx 0.25$ ) because the features having high variance hold the maximum amount of information [93]. We used VarianceThreshold feature selector of scikit learn library that removes low-variance features based on the defined threshold [548]. We chose a threshold for each data modality so that the resulting features have matching dimensions, as shown in Figure 8.2.

- The gene expression data originally contained 60,483 features, with FPKM transformed numbers ranging from 0 to 12. Roughly 30,000 genes remained after the above-mentioned preprocessing steps, which was still a very high number of features. High expression values reveal important biological insights due to an indication that a certain gene product is transcribed in large quantities, revealing that gene features with large expression values within the dataset are highly relevant. Genes containing an expression value greater than 7 (127 FPKM value) were kept, while the rest were discarded. Around 3,000 genes remained after this process, all of which ranged from values between 7 and 12.
- We handled missing features at two levels of data integration. First, for the features within each modality and cancer type, the missing values were imputed with the mean of the samples for that feature. This resulted in the full-length feature vector for each sample. Second, across different cancers and modalities, we padded the missing features with zeros. One may opine that this is equivalent to zero-padding prevalent in the bio-statistics, but we argue that padding zeros across cancers and modalities is not an imputation when integrating very high dimensional, and high-sample-sized data. In deep learning, the zero imputation technique shows the best performance compared to other imputation techniques and deficient data removal techniques [47, 699]. Moreover, there is a line of work that simply used zero padding to minimize the noise in data and achieved state-of-the-art performance on respective datasets [675, 776].

#### 8.2.4 Features Integration

After carrying out the preprocessing steps mentioned above, we integrate the data across cancers and across modalities. We generate two views of the data by combining the features across cancers and across modalities. First view is created by taking the union of features across all cancer patients for each of the six modalities (DNA methylation, gene expression, miRNA expression, protein expression, DNA mutation, and clinical). As a result of the preprocessing explained earlier, the DNA methylation data features were reduced from 485,576 features to  $\approx 4,500$  features for all

cancers. The union of these features from individual cancers resulted in a feature dimension of 52,396. The gene expression data originally had 60,483 features across all cancers, which was reduced to  $\approx 3000$  features. Union of these features resulted in the feature dimension of 8,794. The miRNA expression data originally had 1,880 features across all cancers, which was reduced to  $\approx 1,400$  features. Union of these features resulted in the feature dimension of 1,730. The protein expression data originally had 487 features across all cancers, which was reduced to 472 features unionized to 472 dimensions. The DNA mutation data had 18,090 features across all cancers, pre-processed and unionized to 17,253 features. Lastly, we convert the categorical clinical features to numerical values such as gender, race, and cancer stages. The details of these clinical characteristics are given in Table 8.2. Mathematically, the preprocessing is given below.

Let  $\mathbf{v}$  represent the initial feature having fixed dimension for each cancer. The dimension of each feature set is reduced through a preprocessing step, resulting in the feature vector  $\tilde{\mathbf{v}}$ , which is calculated by a function of  $\mathbf{v}$ , noted as  $f(\mathbf{v})$ , where  $f$  is the dimension reduction function such as those presented in the previous section,  $\tilde{\mathbf{v}} = f(\mathbf{v})$ . For  $n = 33$  cancer types, the reduced dimensional feature vector  $\tilde{\mathbf{v}}$  from each cancer type are then combined through a union operation to generate a feature vector  $V_m$  for each modality  $m$  and  $M = 6$  are the total number of modalities. The feature vector for each modality,  $V_m$ , is defined as:

$$V_m = \begin{cases} \bigcup_{i=1}^n \tilde{\mathbf{v}}_i & \text{if } \tilde{\mathbf{v}}_i \text{ varies by cancer type or modality,} \\ \tilde{\mathbf{v}} & \text{otherwise.} \end{cases} \quad (8.1)$$

Finally, the union of all  $V_m$  across different modalities results in the total pan-cancer, multimodal feature vector  $V_c \in \mathbb{R}^{80,697}$ . The total pan-cancer, multimodal feature vector  $V_c$  can then be expressed as:

$$V_c = \bigcup_{m=1}^M V_m \quad (8.2)$$

Table 8.1 Feature reduction summary of pan-cancer data.

Data	Cases	miRNA Exprn		DNA Methyl		Gene Exprn		Protein Exprn		DNA Mut	
		Before	After	Before	After	Before	After	Before	After	Before	After
TCGA-DLBC	51	1880	1060	485576	4396	60483	850	487	472	18090	17253
TCGA-UCS	61	1880	1101	485576	4632	60483	1231	487	472	18090	17253
TCGA-CHOL	62	1880	967	485576	4479	60483	1261	487	472	18090	17253
TCGA-UVM	80	1880	1162	485576	4019	60483	772	487	472	18090	17253
TCGA-MESO	86	1880	1158	485576	4372	60483	1278	487	472	18090	17253
TCGA-ACC	95	1880	1110	485576	4454	60483	1304	487	472	18090	17253
TCGA-THYM	138	1880	1245	485576	4609	60483	1337	487	472	18090	17253
TCGA-TGCT	139	1880	1290	485576	4762	60483	1343	487	472	18090	17253
TCGA-READ	178	1880	1314	485576	4077	60483	1547	487	472	18090	17253
TCGA-KICH	182	1880	1089	485576	4333	60483	1107	487	472	18090	17253
TCGA-PCPG	189	1880	1251	485576	4550	60483	1216	487	472	18090	17253
TCGA-PAAD	222	1880	1308	485576	4518	60483	1567	487	472	18090	17253
TCGA-ESCA	249	1880	1300	485576	4192	60483	1684	487	472	18090	17253
TCGA-SARC	287	1880	1235	485576	4467	60483	2490	487	472	18090	17253
TCGA-CESC	304	1880	1405	485576	4167	60483	2017	487	472	18090	17253
TCGA-KIRP	376	1880	1297	485576	4078	60483	1798	487	472	18090	17253
TCGA-SKCM	436	1880	1426	485576	4427	60483	2488	487	472	18090	17253
TCGA-BLCA	447	1880	1361	485576	4483	60483	2751	487	472	18090	17253
TCGA-LIHC	463	1880	1336	485576	4023	60483	2017	487	472	18090	17253
TCGA-STAD	499	1880	1397	485576	4196	60483	2354	487	472	18090	17253
TCGA-LGG	533	1880	1287	485576	4193	60483	1560	487	472	18090	17253
TCGA-COAD	539	1880	1460	485576	4671	60483	1931	487	472	18090	17253
TCGA-UCEC	588	1880	1414	485576	4424	60483	2849	487	472	18090	17253
TCGA-HNSC	611	1880	1428	485576	4358	60483	2059	487	472	18090	17253
TCGA-THCA	614	1880	1369	485576	4160	60483	1432	487	472	18090	17253
TCGA-PRAD	623	1880	1334	485576	4006	60483	1635	487	472	18090	17253
TCGA-LAML	626	1880	1140	485576	4415	60483	1032	487	472	18090	17253
TCGA-GBM	649	1880	1023	485576	4076	60483	1206	487	472	18090	17253
TCGA-LUAD	728	1880	1360	485576	4480	60483	2562	487	472	18090	17253
TCGA-OV	731	1880	1430	485576	4254	60483	2116	487	472	18090	17253
TCGA-LUSC	752	1880	1375	485576	4302	60483	2610	487	472	18090	17253
TCGA-KIRC	979	1880	1333	485576	4399	60483	2274	487	472	18090	17253
TCGA-BRCA	1260	1880	1418	485576	4195	60483	3671	487	472	18090	17253

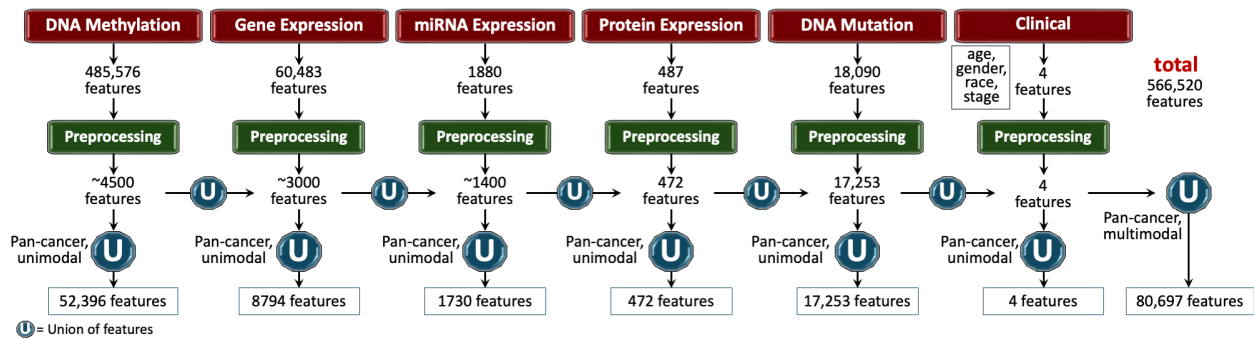


Figure 8.2 Features processing pipeline for pan-cancer data.

### 8.2.5 Clinical End-points

To assess the performance of the SeNMo framework, we chose two end-points that belong to two different categories of machine learning tasks. First is the clinically relevant prediction of



Table 8.2 Summary of patient characteristics for pan-cancer data in this study

Cancer Type	Age (Mean ± SD)	Gender (M/F)	Race (White/Asian/Black/NA/American Indian/Alaska)	Stage (0/I/IA/IB/IC/II/IIA/IIIB/IIIC/III/IIIA/IIIB/IIIC/IV/IVA/IVB/IVC/NA)
TCGA-ACC	47.46 ± 16.20	33/62	79/3/1/12/0	0/9/0/0/0/46/0/0/0/20/0/0/0/17/0/0/0/3
TCGA-BLCA	67.92 ± 10.39	326/121	363/43/23/18/0	0/3/0/0/0/136/0/0/0/159/0/0/0/148/0/0/0/1
TCGA-BRCA	57.94 ± 13.11	13/1247	915/59/198/87/1	0/114/94/7/0/6/404/307/0/2/176/30/74/22/0/0/0/24
TCGA-CESC	48.04 ± 13.70	0/304	211/19/32/30/9	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/304
TCGA-CHOL	64.37 ± 12.21	30/32	55/3/3/1/0	0/30/0/0/0/16/0/0/0/5/0/0/0/2/3/6/0/0
TCGA-COAD	66.93 ± 12.67	288/251	261/11/67/198/2	0/87/1/0/0/46/150/13/2/26/9/69/47/56/18/3/0/12
TCGA-DLBC	56.76 ± 13.68	24/27	32/18/1/0/0	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/51
TCGA-ESCA	64.22 ± 12.11	208/41	162/46/6/35/0	0/14/9/7/0/1/56/43/0/4/1/16/10/9/7/6/0/0/30
TCGA-GBM	57.74 ± 14.32	399/250	547/13/53/36/0	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/649
TCGA-HNSC	61.02 ± 11.92	443/168	522/12/58/17/2	0/29/0/0/0/93/0/0/0/97/0/0/0/302/13/1/76
TCGA-KICH	51.61 ± 14.12	99/83	154/6/19/3/0	0/75/0/0/0/59/0/0/0/34/0/0/0/14/0/0/0/0
TCGA-KIRC	60.67 ± 11.95	641/338	876/16/73/14/0	0/475/0/0/0/102/0/0/0/237/0/0/0/161/0/0/0/4
TCGA-KIRP	61.98 ± 12.20	278/98	275/6/75/16/4	0/219/0/0/0/25/0/0/0/77/0/0/0/21/0/0/0/34
TCGA-LAML	54.82 ± 15.87	345/281	564/8/49/5/0	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/626
TCGA-LGG	42.71 ± 13.32	293/240	492/8/22/10/1	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/533
TCGA-LIHC	60.44 ± 13.71	305/158	255/168/25/14/1	0/211/0/0/0/105/0/0/0/6/78/12/11/2/1/3/0/34
TCGA-LUAD	65.20 ± 10.08	329/399	580/14/84/48/2	0/7/194/195/0/2/67/103/0/0/101/12/0/37/0/0/0/10
TCGA-LUSC	67.28 ± 8.62	548/204	530/12/47/163/0	0/4/127/243/0/4/87/138/0/3/94/33/0/12/0/0/0/7
TCGA-MESO	63.01 ± 9.78	70/16	84/1/1/0/0	0/7/2/1/0/15/0/0/0/45/0/0/0/16/0/0/0/0
TCGA-OV	59.60 ± 11.44	0/731	626/25/43/33/3	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/731
TCGA-PAAD	64.87 ± 11.36	123/99	195/13/8/6/0	0/1/6/15/0/0/36/148/0/6/0/0/0/7/0/0/0/3
TCGA-PCPG	47.02 ± 15.15	84/105	157/7/20/4/1	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/189
TCGA-PRAD	60.93 ± 6.80	623/0	510/13/81/18/1	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/623
TCGA-READ	63.83 ± 11.85	98/80	90/1/7/80/0	0/37/0/0/0/7/40/2/1/6/7/25/14/21/7/0/0/11
TCGA-SARC	60.70 ± 14.38	129/158	253/5/20/9/0	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/287
TCGA-SKCM	57.84 ± 15.41	289/174	441/12/1/9/0	6/30/18/30/0/39/18/28/61/44/16/46/68/23/0/0/0/36
TCGA-STAD	65.44 ± 10.53	320/179	311/108/15/64/0	0/1/21/46/0/37/54/71/0/4/88/67/39/47/0/0/0/24
TCGA-TGCT	31.87 ± 9.19	139/0	124/4/6/5/0	0/69/26/11/0/4/6/1/1/2/1/6/5/0/0/0/0/7
TCGA-THCA	47.17 ± 15.83	166/448	413/59/35/106/1	0/350/0/0/0/64/0/0/0/134/0/0/0/4/52/0/8/2
TCGA-THYM	58.12 ± 13.00	72/66	115/13/8/2/0	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/138
TCGA-UCEC	63.74 ± 11.06	0/588	402/21/120/32/4	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/588
TCGA-UCS	70.07 ± 9.24	0/61	50/1/9/1/0	0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0/61
TCGA-UVM	61.65 ± 13.95	45/35	55/0/0/25/0	0/0/0/0/0/0/12/27/0/0/25/10/1/4/0/0/0/1
Moffitt-LSCC	69.14 ± 8.34	72/36	105/0/3/0/0	0/0/24/25/0/0/31/15/0/0/12/1/0/0/0/0/0

overall survival (OS), which is a regression task. The second is the prediction of the primary cancer type, which is a 33-class classification task.

### 8.2.5.1 Overall Survival (OS)

Cancer prognosis via survival outcome prediction is a standard method used for biomarker discovery, stratification of patients into distinct treatment groups, and therapeutic response prediction [135]. Statistical survival models and a shift towards incorporating deep learning into survival

analysis have improved the task of OS prediction. Previous works have used combinations of molecular data types and employed different statistical and learning-based methods to predict OS in different datasets [545, 462, 205, 511]. The journey towards comprehensive survival analysis continues, aiming to combine different data types to better understand the relationship between molecular features and patient outcomes, ultimately leading to more precise prognostic assessments and tailored treatment approaches. In this study, we utilize clinical, demographic, genomic, and other molecular data types to investigate potential risk factors for cancer patients, specifically examining their correlations with the patients' time-to-event, which in this case is OS. We have implemented prediction of OS as a regression task, i.e., prediction of OS in days. Time-to-event or survival data not only records the occurrence of events like death but also tracks the time from the study's outset to when the event occurs, concludes, or when a patient is no longer followed (known as right censoring). Survival or right-censored times since cancer diagnosis for our pan-cancer data is depicted in Figure 8.3A. Due to censoring, the exact survival time for some patients remains unknown. In such instances, each patient's outcome is defined by two main variables: a censoring indicator, also called the vital status, and the observed time  $T = \min(T_s, T_\delta)$ , where  $T_s$  represents the exact survival time and  $T_\delta$  is the censoring time,  $\{T_s \leq T_\delta\}$  [808]. Survival function to describe the likelihood of a patient surviving beyond a specified time  $t$  is given as:

$$F(t) = P\{T > t\} \tag{8.3}$$

Additionally, the hazard function offers insights into the risk of the event occurring at a given time, assuming survival up to that point. The hazard function represents the instantaneous rate at which events (such as death) occur at a specific time, given that the individual has survived up to that time. It provides insights into the risk of experiencing the event at any given moment, conditional on having survived up to that point in time. Mathematically, the hazard function  $h(t)$  is defined as the ratio of the probability of the event occurring in a short time interval around  $t$  to the probability of surviving beyond  $t$ :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (8.4)$$

where,  $h(t)$  is the hazard function at time  $t$ ,  $T$  is the survival time,  $P(t \leq T < t + \Delta t | T \geq t)$  is the conditional probability that the event occurs in the time interval  $[t, t + \Delta t)$  given that survival time is greater than or equal to  $t$ , and  $\Delta t$  represents an infinitesimally small time interval.

Based on survival data, the hazard function describes the instantaneous risk of experiencing the event of interest at any given time. In our pan-cancer data, the (right) censoring was defined as censor  $\delta = 1$  in case of an event (e.g., death), and 0 otherwise.

### 8.2.5.2 Primary Cancer Type

The primary cancer type prediction task involves classifying a cancer sample into one of the 33 possible types based on various biological and clinical features. This task is fundamentally important and clinically relevant because accurate identification of the primary cancer type is critical for choosing the most effective treatment strategy, improving patient outcomes, and personalizing therapy approaches [484]. Cancer treatments and prognostics can vary dramatically between different cancer types, often requiring specific interventions that are tailored to the unique biological characteristics of each type. Correctly predicting the primary cancer type helps in planning follow-up care and surveillance, enhancing the likelihood of early detection of recurrence. Thus, achieving high accuracy in this classification not only supports better clinical decision-making but also significantly impacts patient survival and quality of life. The pan-cancer data comprising 33 cancers and the distribution of patients (cases) across these datasets are depicted in Figure 8.3B.

### 8.2.6 SeNMo Deep Learning Model

In learning scenarios with hundreds to thousands of features and relatively few training samples, feedforward networks are susceptible to overfitting [135]. Unlike Convolutional Neural Networks (CNNs), the weights in feedforward networks are shared, making them more prone to training instabilities caused by perturbations and regularization techniques like stochastic gradient descent

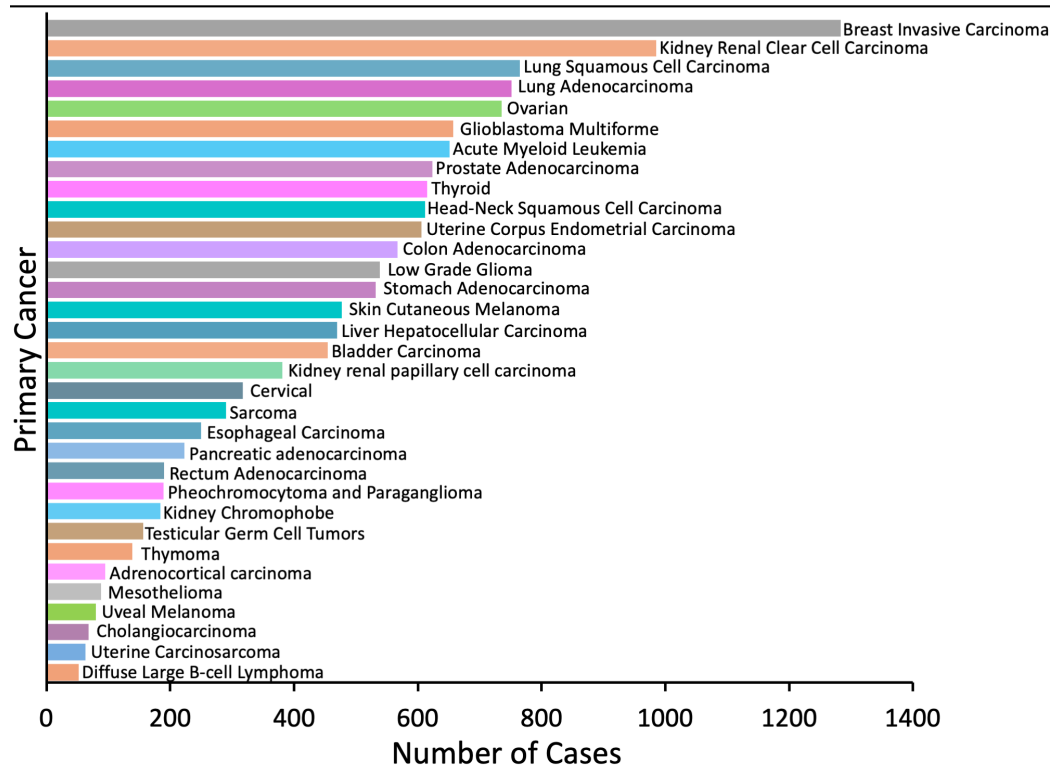


Figure 8.3 Summary of the number of cases in the pan-cancer data.

and dropout. CNNs do not perform as well on high-dimensional, low-sample data for several reasons, including the spatial invariance assumption, fixed input size, and parameter efficiency viz-a-viz multi-omics data sparsity. Transformers-based models are also not inherently optimized for high-dimensional, low-sample data, such as in genomics or other multi-omics datasets, because these models use attention mechanism to predict the next token in case of language tasks and meaningful patterns in non-language tasks, which fails in case of highly sparse molecular data. To address overfitting in high-dimensional, low-sample-size multi-omics data and employ more robust regularization techniques during training, we draw inspiration from Self-Normalizing Networks (SNN) introduced by Klambauer et al. [377]. SNNs have been extensively used for their ability to effectively handle high-dimensional data with limited samples, making them particularly relevant in multi-omics data analysis. Our learning framework is based on the stacked layers of SNNs, as described below.

As shown in Figure 8.4, SeNMo comprises of stacked blocks of SNN layers, where each block is composed of a linear unit, SELU activation, followed by Alpha-dropout. Combined, these blocks enable high-level abstract representations by keeping neuron activations converged towards zero mean and unit variance [377]. Linear unit is essentially equivalent to what is commonly referred to as a “fully connected” (FC) or a “multilayer perceptron” (MLP) layer in traditional neural network architectures. SELU activations are an alternative to the traditional rectified linear unit (ReLU) activations commonly used in neural networks. Klambauer et al. demonstrated through the Banach fixed-point theorem that activations with close proximity to zero mean and unit variance, propagating through numerous network layers, will ultimately converge to zero mean and unit variance [377]. SELU activations have the unique property of self-normalization, meaning that the activations tend to converge to a mean of zero and a standard deviation of one, regardless of the input distribution. The mathematical equation for the SELU activation function is:

$$\text{SELU}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x \leq 0 \end{cases} \quad (8.5)$$

where,  $\lambda$  is a scaling factor (typically set to 1.05071) and  $\alpha$  is the negative scale factor (typically set to 1.6733).

Dropout is a regularization technique that randomly sets a fraction of input units to zero during training to prevent overfitting. Alpha-dropout is a modified version of traditional dropout regularization, specifically designed to preserve the self-normalizing property of SELU activations. Alpha-dropout applies a dropout mask to the activations during training, scaled by a factor that ensures the mean and variance of the activations remain unchanged. This scaling factor is computed based on the dropout rate and the SELU parameters ( $\lambda$  and  $\alpha$ ). Mathematically,

$$\text{Alpha-dropout}(x) = \frac{x - \mu(x)}{\text{std}(x)} \times \text{mask} + \mu(x) \quad (8.6)$$

where,  $x$  is the input activation,  $\mu(x)$ ,  $\text{std}(x)$  are mean and standard deviation of the input activation, respectively, and  $\text{mask}$  is a binary mask generated with the specified dropout rate.

Together, SELU activations and Alpha-dropout enable SeNMo blocks to maintain a stable mean and variance of activations throughout the network layers, facilitating more stable training and better generalization performance. Similarly, SELU activation and Alpha-dropout help mitigate training instabilities caused by vanishing or exploding gradients in the feedforward networks. Our network architecture consists of seven fully connected hidden layers followed by SELU activation and Alpha-dropout, where the number of neurons in each block is shown in the inset of Figure 8.4. The final fully connected layer is utilized to learn a latent representation of the sample, which we call the patient embedding  $\mathbf{x} \in \mathbb{R}^{48}$ .

### 8.2.7 Training and Evaluation

We evaluate SeNMo's performance with the quantitative and statistical metrics common for survival outcome prediction and grade classification. For survival analysis, we evaluated the model using the Concordance Index (C-Index). For the primary cancer type classification, we generate the classification report comprising average accuracy, average precision, recall, F1-score, confusion

matrix, and scatter plot. For testing the statistical significance of the predictions, we use the LogRank Test. Below, we explain the loss, evaluation metrics, and statistical tests in detail.

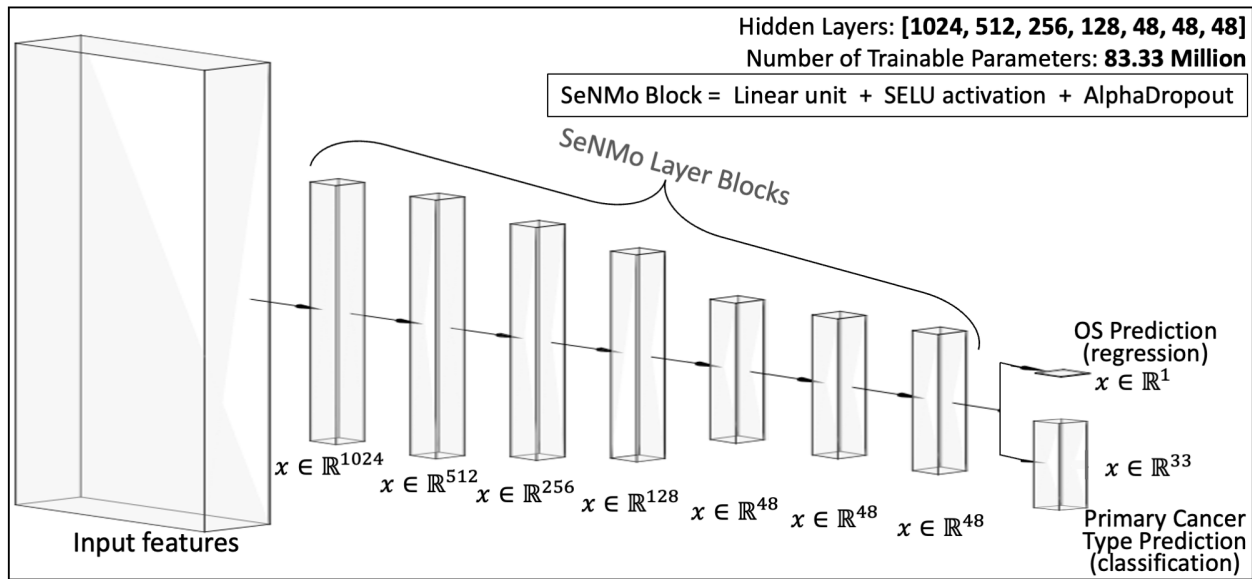


Figure 8.4 Architecture of the SeNMo encoder network.

### 8.2.7.1 Loss Function

The loss being used for backpropagation in the model is a combination of three components: Cox loss, cross-entropy loss, and regularization loss. This combined loss function aims to simultaneously optimize the model's ability to predict survival outcomes (Cox loss), encourage model-simplicity or sparsity (regularization loss), and model the likelihood of cancer types (cross-entropy loss). The overall loss is a weighted sum of these three components, where each component is multiplied by a corresponding regularization hyperparameter ( $\lambda_c$ ,  $\lambda_{ce}$ ,  $\lambda_r$ ). This weighted sum allows for balancing the influence of each loss component on the optimization process. Mathematically, the overall loss can be expressed as:

$$L = \lambda_c L_{cox} + \lambda_{ce} L_{ce} + \lambda_r L_{reg} \quad (8.7)$$

- *Cox proportional hazards loss* ( $L_{\text{cox}}$ ) is a measure of dissimilarity between the predicted hazard scores and the true event times in survival analysis. It is calculated using the Cox proportional hazards model and penalizes deviations between predicted and observed survival outcomes of all individuals who are at risk at time  $t_i$ , weighted by the censoring indicator [141]. The function takes a vector of survival times for each individual in the batch, censoring status for each individual (1 if the event occurred, 0 if censored), and the predicted log hazard ratio for each individual from the neural network, and returns the Cox loss for the batch, which is used to train the neural network via backpropagation. This backpropagation encourages the model to assign higher hazards to high-risk individuals and lower predicted hazards to censored individuals or those who experience the event later. Mathematically, the Cox loss is expressed as:

$$L_{\text{cox}} = -\frac{1}{N} \sum_{i=1}^N \left( \theta_i - \log \sum_{j=1}^N e^{\theta_j} \cdot R_{ij} \right) \cdot \delta_i \quad (8.8)$$

where,  $N$  is the batch size (number of samples),  $\theta_i$  is the predicted hazard for sample  $i$ ,  $R_{ij}$  is the indicator function that equals 1 if the survival time of sample  $j$  is greater than or equal to the survival time of sample  $i$ , and 0 otherwise, and  $\delta_i$  is the censoring indicator for sample  $i$ , which equals 1 if the event is observed for sample  $i$  and 0 otherwise.

- *Cross-entropy loss* ( $L_{\text{ce}}$ ) is a common loss function used for multi-class classification problems, particularly when each sample belongs to one of the  $C$  classes. When combined with a LogSoftmax layer, the function measures how well a model's predicted log probabilities match the true distribution across various classes. For a multi-class classification problem having  $C$  classes, the model's outputs (raw class scores or logits) are transformed into log probabilities using a LogSoftmax layer. The cross-entropy loss compares these log probabilities to the true distribution, which is usually represented in a one-hot encoded format. The loss is calculated by negating the log probability of the true class across all samples in a batch and then averaging these values. For the given output of LogSoftmax,  $\log(p_{n,c})$  for



each class  $c$  in each sample  $n$ , the cross-entropy loss for a multi-class problem can be defined as:

$$L_{ce} = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_{n,c} \log(p_{n,c}) \quad (8.9)$$

where,  $N$  is the total number of samples,  $C$  are the total classes, and  $y_{n,c}$  is the target label for sample  $n$  and class  $c$ , typically 1 for the true class and 0 otherwise.

- *Regularization loss* ( $L_{reg}$ ) encourages the model's weights to remain small or sparse, thus preventing overfitting and improving generalization. We used  $L1$  regularization to the SeNMo's parameters, which penalizes the absolute values of the weights.

#### 8.2.7.2 Concordance Index (C-Index)

The C-Index is a frequently used evaluation metric in survival analysis to assess the predictive accuracy of a model for the time-to-event outcomes [808]. It measures the degree to which the model's predictions correlate with the actual survival times observed in the data. It quantifies the model's ability to correctly rank pairs of subjects based on their predicted survival times. The C-Index evaluates the probability that, in a randomly selected pair of individuals, the one who experienced the event (like death or failure) first also had a higher risk score predicted by the model. Risk score is the output of the survival model and represents the expected order of the events; the higher the score, the higher the risk of experiencing the event sooner [808]. We used the `concordance_index` Lifelines function to calculate the C-Index [167]. This function takes the predicted hazard scores for each individual, the true event indicator (e.g., 1 if an event occurred, 0 if censored) for each individual, and the survival times (time to event or censoring) for each individual. The C-Index function computes the fraction of all pairs of subjects whose predicted event times are correctly ordered among all pairs where one subject experienced an event and the other did not. C-Index ranges between 0 and 1 where 0.5 is the expected result from random predictions, 1.0 is a perfect concordance, and 0.0 is perfect anti-concordance [808]. Mathematically,

$$\text{C-Index} = \frac{(\text{Number of concordant pairs} + 0.5 \times \text{tied pairs})}{\text{Total number of evaluable pairs}}, \quad (8.10)$$

$$\text{C-Index} = \Pr(\hat{S}_i < \hat{S}_j | T_i < T_j, \delta_i = 1)$$

where, concordant pairs are pairs of individuals where the predicted survival times are correctly ordered relative to the observed survival times. Tied pairs are number of pairs where the predictions are equal or survival times are the same. Total number of evaluable pairs are the total pairs considered, excluding pairs with censoring issues or other exclusions.  $\hat{S}_i$  and  $\hat{S}_j$  represent the predicted risks or survival probabilities for individuals  $i$  and  $j$ , respectively.  $T_i < T_j$  implies that individual  $i$  experienced the event before individual  $j$ .  $\delta_i = 1$  indicates that the event for individual  $i$  was observed (not censored).

### 8.2.7.3 Cox Log-rank Function

The Cox Log-rank function calculates the p-value using the log-rank test based on predicted hazard scores, censor values, and the true overall survival times. The log-rank test is a statistical method to compare the survival distributions of two groups, where the null hypothesis is that there is no difference between the groups. It is commonly used in survival analysis to compare the observed number of events in each group to the number of events expected under the null hypothesis. For the hazard ratio  $h_i(t)$  of group  $i$  at time  $t$ , the hypotheses are given by,

$$\begin{aligned} H_0 : h_1(t) &= h_2(t) \\ H_A : h_1(t) &= \delta h_2(t), \quad \delta \neq 1 \end{aligned} \quad (8.11)$$

The test statistic for the log-rank test is calculated as the sum of the differences between the observed and expected number of events squared, divided by the expected number of events, summed over all observed time points. The p-value obtained from the log-rank test indicates the

significance of the difference in survival distributions between the two groups. The test statistic is chi-squared under the null hypothesis [167].

$$\chi^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i} \quad (8.12)$$

where,  $O_i$  is the observed number of events at time point  $i$  in the sample,  $E_i$  is the expected number of events at time point  $i$  under the null hypothesis, and  $N$  is the total number of observed time points.

Table 8.3 Hyperparameters for training and finetuning.

<b>Hyperparams</b>	<b>Training (range)</b>	<b>Finetuning (range)</b>
Learning Rate	[1e-6, 1e-1]	[1e-8, 1e-3]
Weight Decay	[1e-6, 1e-1]	[1e-4, 1e-2]
Dropout	[0.1, 0.65]	[0.05, 0.45]
Batch Size	[64, 128, 256, 512]	[8, 16, 32, 48]
Epochs	[50, 100]	[8, 10, 15, 20, 30]
Hidden Layers	[1, 2, 3, 4, 5, 6, 7, 8, 9]	-
Hidden Neurons	[2048, 1024, 512, 256, 128, 48, 32]	-
Optimizer	[adam, sgd, rmsprop, adamw]	[adam, adamw]
Learning Rate Policy	[linear, exp, step, plateau, cosine]	[linear, exp, plateau]
Frozen Layers	-	[7, 6, 5, 4, 3, 2]
Additional Layers	-	[1, 2, 3]

#### 8.2.7.4 Hyperparameters Search

Hyperparameters are non-learnable parameters of a deep learning model and are crucial as they govern the learning process and model architecture. Hyperparameter tuning involves selecting the optimal combination of parameters that results in the best model performance. Common hyperparameters include learning rate and policy, batch size, number of epochs, weight decay, dropout type and probability, and architecture specifics such as the number of hidden layers and

neurons in each layer. Methods for hyperparameter search range from grid search, where all possible combinations of parameters are evaluated; to random search, which randomly samples parameter combinations within predefined bounds. More sophisticated techniques like Bayesian optimization or using automated machine learning (AutoML) tools can dynamically adjust parameters based on previous results to find the best solutions more efficiently. We employed weights and biases [81] utility to carry out random and Bayesian methods of hyperparameters search. The list of hyperparameters we searched for training and fine-tuning is given in Table 8.3. For model training, we conducted around 400 simulations to find the current hyperparameters. For fine-tuning the SeNMo model on CPTAC-LUSC and Moffitt’s Lung SCC data, we ran around 1000 and 450 simulations, respectively. The plots for these simulations are given in the appendix figures 8.9, 8.10, and 8.11.

#### 8.2.7.5 *Frameworks, Compute Resources, Wall-clock Times*

We trained SeNMo model using the Moffitt Cancer Center’s HPC machine using one Tesla V100 32GB GPU running Ubuntu 22.04.4 and CUDA 12.2. The entire code was developed in Python and PyTorch frameworks. The software frameworks and corresponding packages used in our codebase are given in Table 8.4. Training time for our current 83.33 Million parameter SeNMo encoder is approximately 11 hours. We conducted the hyperparameters search of the pan-cancer model for approximately 20 days using multiple GPUs in parallel. Finetuning the trained model on a given data having around 150 patients approximately takes 15 minutes. It took us one week to find the fine-tuning hyperparameters for the CPTAC-LSCC data.

Table 8.4 Frameworks and packages used in our codebase.

	<b>Package name</b>	<b>Version</b>
<b>Operating systems</b>	Ubuntu	20.04.4
<b>Programming languages</b>	Python	3.10.13
<b>Deep learning framework</b>	Pytorch	2.2.0
	torchvision	0.17.0
<b>Miscellaneous</b>	feature-engine	1.6.2
	imbalanced-learn	0.12.0
	scipy	1.12.0
	scikit-learn	1.4.0
	numpy	1.26.3
	PyYaml	6.0.1
	jupyter	1.0.0
	pandas	2.2.0
	pickle5	0.0.11
	protobuf	4.25.2
wandb	0.16.3	

## 8.3 Results

### 8.3.1 Pan-Cancer Multimodal Analysis - Overall Survival

For the overall survival (OS) task, the pan-cancer data was randomly divided into the training-validation set (80%) and the held-out test set (20%) for each cancer type. The pan-cancer training was carried out by combining the training-validation cohort of all 33 cancer types and adopting the 10-fold cross-validation with the 80 – 20% division of samples. The training-validation cohort has 11, 050 patients, each having  $\mathbb{R}^{80,697}$  features, comprising the six multi-omics modalities, gene expression, DNA methylation, miRNA expression, protein expression, DNA mutation, and the four clinical features (age, gender, race, stage). The SeNMo encoder model was trained on the training-validation cohort for the regression task of predicting the OS. C-Index was used as the evaluation metric of the hazard score predicted by the model. We used weights and biases to find the optimal set of hyperparameters for our deep learning model [81]. Figure 8.9 shows the visualization of the parallel sweeps across all hyperparameters, resulting in training around 400 unique models. The optimal model had a learning rate of 0.00058, a weight decay of 0.00598,

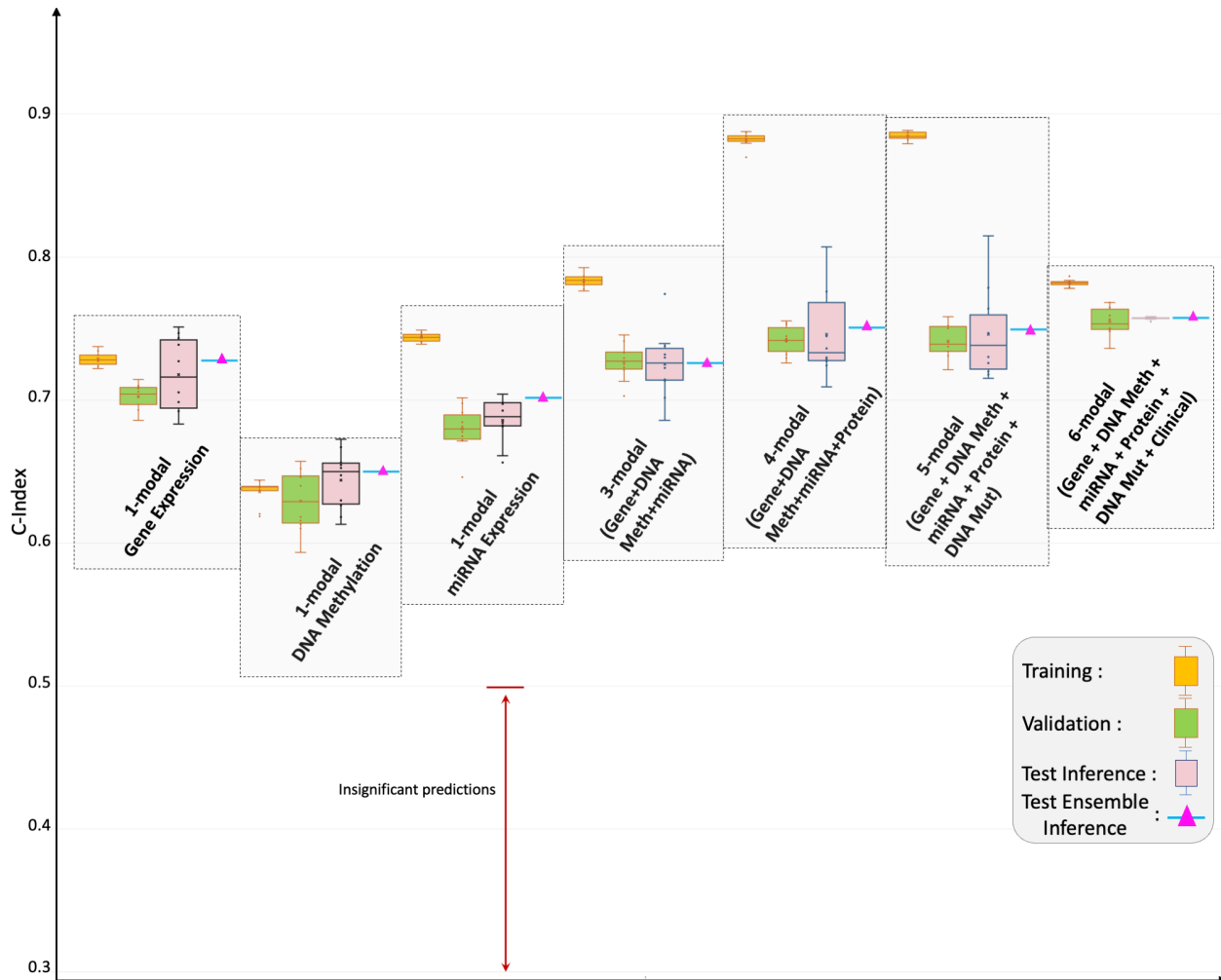


Figure 8.5 Pan-cancer C-Index results for OS prediction.

0.1058 dropout, 256 batch size, 100 epochs, and seven hidden layers with neurons in these layers as [1024, 512, 256, 128, 48, 48, 48]. The trained model contained 83.33 million trainable parameters. Checkpoints were saved for this model for each of the 10 folds. The model’s training resulted in the average training C-Index of 0.78 and average validation C-Index of 0.76 across the 10 folds.

For the evaluation/testing of the trained model, the inference data was created by combining the held-out test set from all 33 cancer types, resulting in 2,754 patients, each having  $\mathbb{R}^{80,697}$  features. The inference on the test set showed the C-Index of 0.757, the average of the C-Indices from the 10 checkpoints. To further validate our findings, we created an ensemble of the 10 checkpoints by averaging the prediction vectors from all the models and then evaluating the final averaged prediction vector for C-Index. For the pan-cancer, multi-omics SeNMo model, an ensemble C-

Index of 0.758 was achieved on the held-out test set. The significance level in all these analyses is 95%, i.e.,  $p < 0.05$ , indicating statistically significant values. These results are depicted in the Figure 8.5.

We further tested the optimal hyperparameters of our trained model to train different combinations of the pan-cancer data modalities. We call these 1-modal, 3-modal (gene expression, DNA methylation, miRNA expression), 4-modal (3+protein expression), 5-modal (4+DNA mutation), and 6-modal (all modalities) cohorts. Although our initial model was trained on all 6 modalities, these experiments aim to see how the model performs on each of these pan-cancer cohorts where one or more of the data modalities is missing. As depicted in Figure 8.5, the SeNMo model trained on the pan-cancer 1-modal (Gene expression) cohort showed a C-Index for training, validation, testing, and ensemble inference as 0.729, 0.702, 0.718, and 0.728, respectively. For the pan-cancer 1-modal (DNA methylation) cohort, the model's training, validation, testing, and ensemble inference C-indices are 0.636, 0.629, 0.644, and 0.65, respectively. For the pan-cancer 1-modal (miRNA expression) cohort, the model's training, validation, testing, and ensemble inference C-indices are 0.744, 0.68, 0.686, and 0.702, respectively. We did not analyze the model individually on the rest of the three modalities because clinical and protein expression features are too small for an 83 million-parameter model, whereas the DNA mutation data comprised the binarized features of mutations. Evaluating the model on the 3-modal cohort showed the training, validation, testing, and ensemble inference C-indices of 0.783, 0.727, 0.725, and 0.726, respectively. Further adding the protein expression to the 3-modal data, we trained and evaluated the model on the 4-modal cohort and got the C-Indices of 0.88, 0.742, 0.746, and 0.751 for training, validation, testing, and ensemble inference, respectively. Lastly, the model's performance on the 5-modal cohort showed the training, validation, testing, and ensemble inference C-indices of 0.885, 0.741, 0.746, and 0.749, respectively. Next, we analyze how the model trained on pan-cancer, 6-modal data fared on individual cancer patients' data.

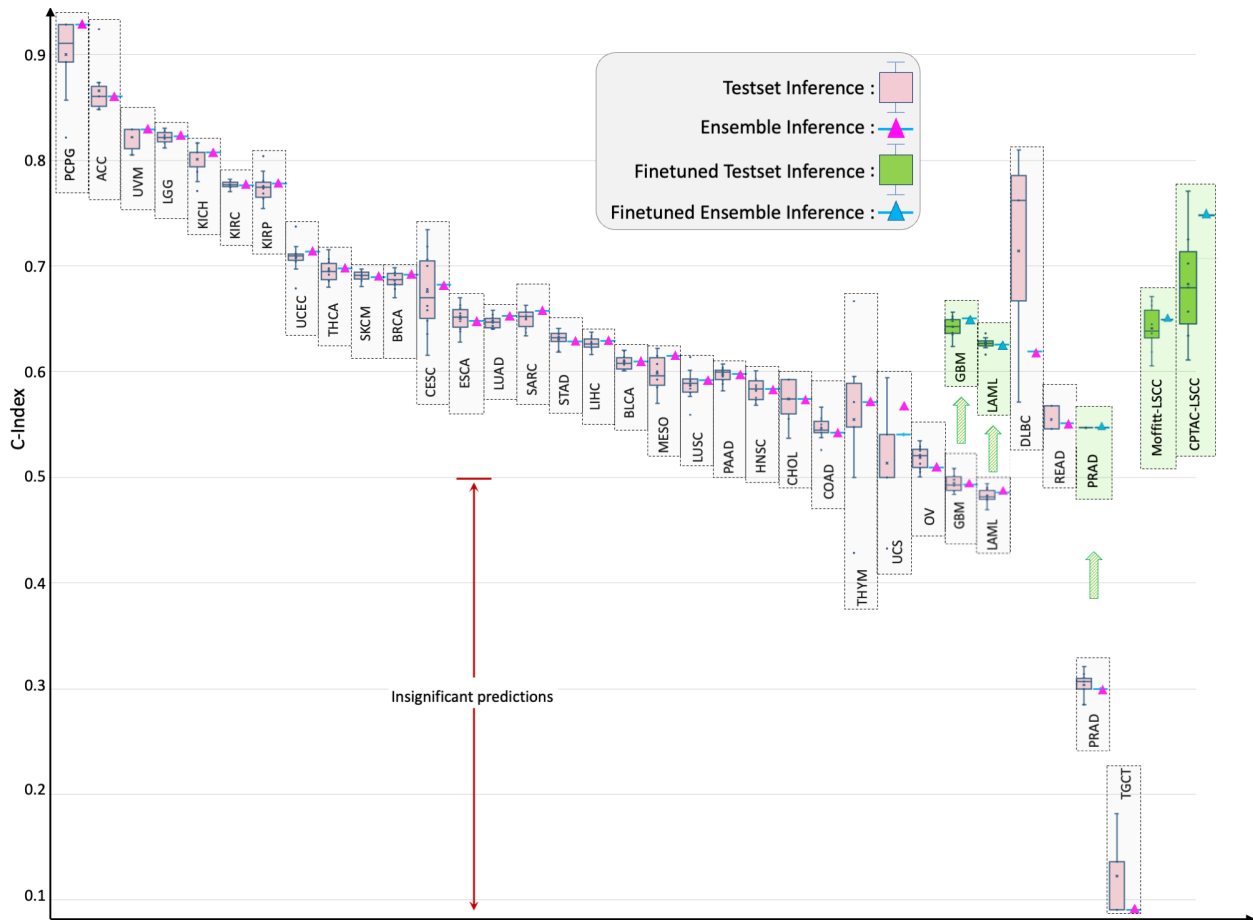


Figure 8.6 Cancer-wise C-Index results for OS prediction.

### 8.3.2 Cancer-wise Multimodal Analysis - Overall Survival

We evaluated the model trained on the 6-modal pan-cancer cohort on the held-out individual cancer data from an individual cancer-wise perspective. The number of patients in these cancer cohorts was a randomly selected subset of the cases shown in Figure 8.3 and Tables 8.1, 8.2, which accounts for the 20% of the total samples. The trained model was evaluated on each of the 33 individual cancer data using simple inference and the ensemble of the 10-fold checkpoints. Figure 8.6 shows the evaluation performance of the model on 33 cancer types. The model showed the best predictive performance on TCGA-PCPG data with an average C-Index on the test set of 0.9 and ensemble inference of 0.929. SeNMo’s performance on the other cancer types in format  $\{Test Infer-$



ence, Ensemble Inference} is shown in Table 8.5, where 29 cancer types have significant C-Indices. We noticed that the results for TCGA-GBM, TCGA-LAML, TCGA-PRAD, and TCGA-TGCT were not statistically significant, i.e.,  $p > 0.05$ . So, we fine-tuned the model for these datasets by reducing the learning rate, increasing the weight decay and dropout, and letting the model fine-tune for 10 epochs. Resultantly, the model’s performance increased for TCGA-GBM= {0.642, 0.650}, TCGA-LAML= {0.627, 0.626}, and TCGA-PRAD= {0.541, 0.542}. These improvements are depicted with the green arrows and green boxes in Figure 8.6. However, the model failed to converge for TCGA-TGCT data and consistently gave predictions that were not significant,  $p > 0.05$ .

Table 8.5 C-Index for test and ensemble inference across cancer types

<b>Cancer Type</b>	<b>C-Index</b> {Test, Ensemble}	<b>Cancer Type</b>	<b>C-Index</b> {Test, Ensemble}
TCGA-PCPG	{0.900, 0.929}	TCGA-BLCA	{0.609, 0.609}
TCGA-ACC	{0.866, 0.861}	TCGA-MESO	{0.599, 0.615}
TCGA-UVM	{0.822, 0.829}	TCGA-LUSC	{0.588, 0.592}
TCGA-LGG	{0.821, 0.823}	TCGA-PAAD	{0.597, 0.598}
TCGA-KICH	{0.801, 0.807}	TCGA-HNSC	{0.583, 0.583}
TCGA-KIRC	{0.777, 0.776}	TCGA-CHOL	{0.574, 0.574}
TCGA-KIRP	{0.775, 0.778}	TCGA-COAD	{0.546, 0.542}
TCGA-UCEC	{0.708, 0.713}	TCGA-THYM	{0.555, 0.571}
TCGA-THCA	{0.696, 0.698}	TCGA-UCS	{0.514, 0.541}
TCGA-SKCM	{0.691, 0.689}	TCGA-OV	{0.518, 0.509}
TCGA-BRCA	{0.687, 0.692}	TCGA-GBM	{0.495, 0.493}
TCGA-CESC	{0.676, 0.682}	TCGA-LAML	{0.482, 0.485}
TCGA-ESCA	{0.650, 0.648}	TCGA-DLBC	{0.714, 0.619}
TCGA-LUAD	{0.647, 0.653}	TCGA-READ	{0.550, 0.551}
TCGA-SARC	{0.650, 0.658}	TCGA-PRAD	{0.304, 0.300}
TCGA-STAD	{0.631, 0.628}	TCGA-TGCT	{0.123, 0.091}
TCGA-LIHC	{0.627, 0.629}		

### 8.3.3 Out-of-distribution Evaluation and Fine-tuning

To further verify the performance of our model, we evaluated the model with the off-the-shelf datasets CPTAC-LSCC [612] and Moffitt's LSCC [659]. Since these datasets were not part of the training data, using these datasets represented out-of-distribution evaluation for the model. Evaluating the model without fine-tuning showed the  $\{Test\ Inference, Ensemble\ Inference\}$  of CPTAC-LSCC =  $\{0.48, 0.50\}$ , and Moffit-LSCC =  $\{0.581, 0.59\}$ . Fine-tuning the model for 10 epochs, with reduced learning rate, and increased weight decay and dropout resulted in the improvement of C-Indices as CPTAC-LSCC =  $\{0.677, 0.73\}$ , and Moffit-LSCC =  $\{0.647, 0.656\}$ . These fine-tuning results are depicted in Figure 8.6 as the green box plots.

### 8.3.4 Patient Stratification

Risk stratification of patients allows clinicians and researchers to identify patients who might need more intensive care or monitoring and those who may have a better prognosis, facilitating more personalized treatment approaches. We further investigated the SeNMo's ability to stratify the patients based on low, intermediate, and high risk conditions. We plot Kaplan-Meier (KM) curves of our model on the pan-cancer, multi-omics held-out test set, as shown in Figure 8.7. We select the low/ intermediate/ high risk stratification distribution as the 33-66-100 percentile of hazard predictions [135, 422]. The hazard scores predicted by SeNMo are used to evaluate the model's stratification ability. The KM comparative analysis shows that SeNMo distinguished the patients across the three groups. The low-risk group (green) exhibited the highest survival probability, maintaining close to 100% survival up to approximately 5 years, and gradually declining to about 60% by the 25-year mark. The intermediate-risk group (blue) showed a significantly lower survival probability, starting to diverge from the low-risk group early on and reaching around 40% by the 15-year mark of the study period. The high-risk group (orange) displayed the most pronounced decline in survival probability, with a steep drop to approximately 20% survival within the first 10 years, and further reducing to below 10% after 10 years. The logrank test to evaluate the significance of this stratification shows that the p-value of low vs. intermediate curves is  $1.66e - 05$ , low vs. high

is  $1.156e - 46$ , and intermediate vs. high is  $1.92e - 22$ , showing significant results, i.e.,  $p < 0.05$ . The 95% confidence intervals around each curve show the reliability of these estimates.

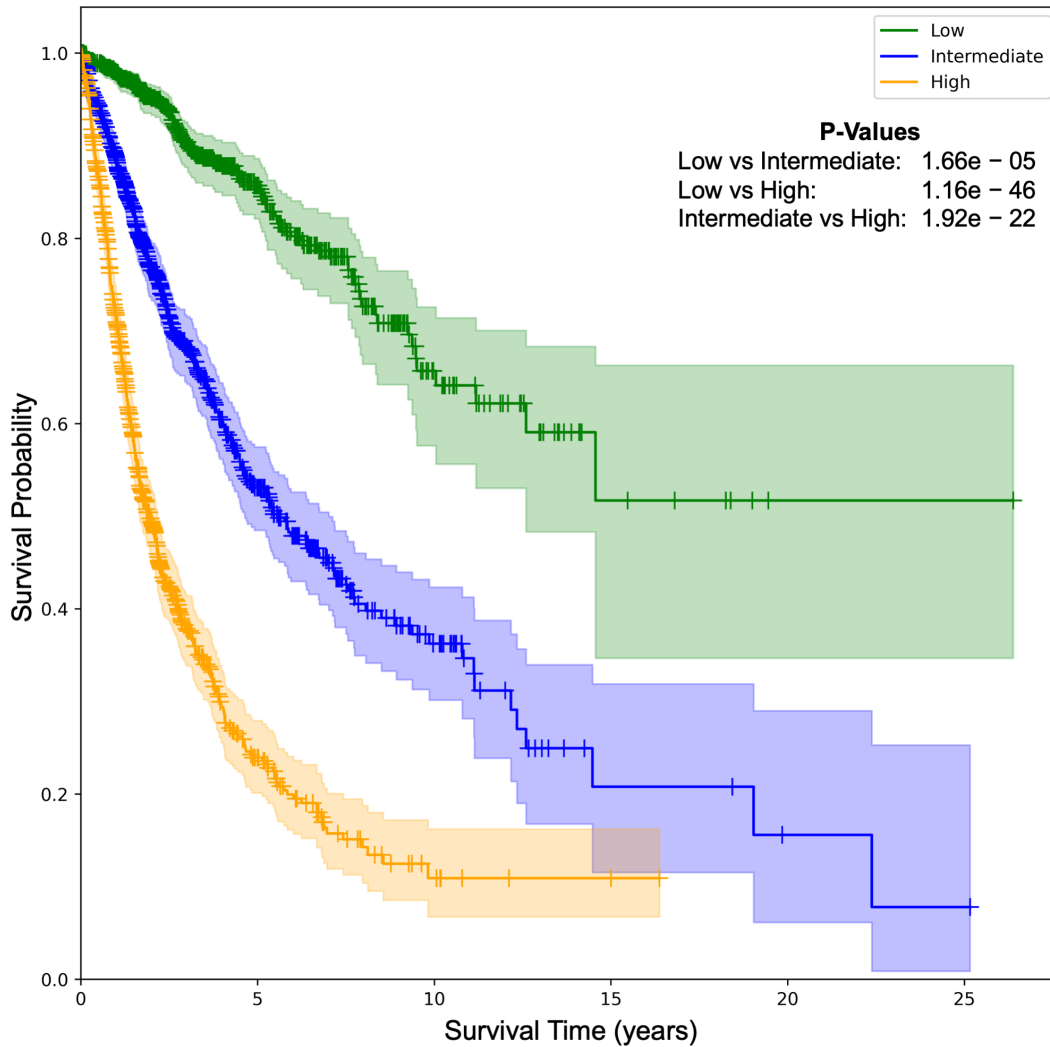


Figure 8.7 KM comparative analysis of using SeNMo in stratifying patient outcomes.

### 8.3.5 Primary Cancer Type Prediction

To test the generalizability of SeNMo across different tasks, we carried out the prediction of primary cancer type from pan-cancer, multi-omics data. We set the problem as a classification problem, where the multi-omics data is used to predict the type of cancer for the given patient data

among the 33 classes. It is imperative to mention here that the four clinical features in the initial data contained the cancer stage, as shown in Figure 8.2 and Table 8.2. When considering a cancer type classification problem, the stage adds a bias in the data because of the staging distribution among different cancers. Therefore, for the cancer classification simulations, we excluded the “stage” feature in the clinical data. As shown in Figure 8.8, the model achieves near-perfect accuracy levels, with 99.9% average accuracy in training, 99.8% in validation, and consistent performance in both simple and ensemble inference approaches. The confusion matrix depicts a clear concentration of values along the diagonal, indicating a high rate of correct predictions across all cancer types. The scatter plot shows an alignment of predicted labels with true labels along the diagonal line, highlighting the model’s robust predictive accuracy. The classification report across various cancer types reveals that the model consistently maintains high precision, recall, and F1-scores, approaching a value of 1 for almost all categories. The robust predictive power of our model emphasizes the fact that each cancer has a unique molecular landscape, highlighted through differences in gene, protein, and miRNA expression, DNA methylation, and types of somatic mutations seen in our data.

## **8.4 Discussion and Conclusion**

We analyzed pan-cancer dataset of 33 cancer types comprising five molecular data modalities (with varying amount of features) and four clinical data features using our SeNMo encoder-based framework. Public databases such as CPTAC and TCGA contain common identifiers within their data that connect data from the same patient. Therefore, molecular data, such as gene expression, miRNA expression, DNA methylation, somatic mutations, and protein expression can be consolidated to represent a singular patient. However, such high-dimensional data has intra- and inter-dataset correlations, heterogeneous measurement scales, missing values, technical variations, and other forms of noise [808]. This necessitates the need for a variety of preprocessing techniques such as the removal of low variance features and the imputing of missing features among others prior to training. Training such a large dataset having high-dimensional heterogeneous data required

proper computational resources and a precise pipeline for training, testing, and validation. After extensive training-evaluation runs, we found, through optimal parameters searching, a model that performs very well across the different data types and tasks (refer to Figures 8.5 and 8.9). The model has been shown to outperform the existing works in OS prediction when considering the six data modalities included in our data [511]. Moreover, we observed that adding more data and types of modalities increased the model’s performance.

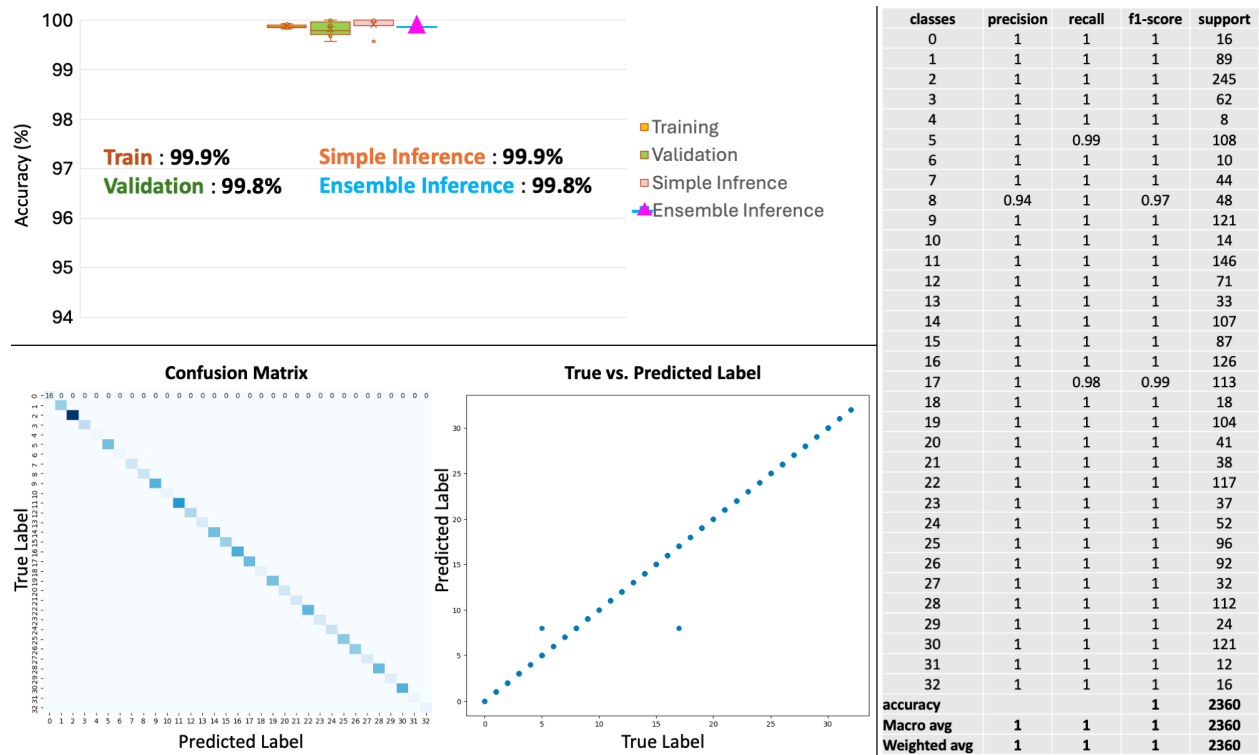


Figure 8.8 Pan-cancer primary cancer type prediction.

The model’s performance was evaluated on individual cancers at test-time through simple inference and ensembling methods. We observed that the model’s predictive power improved when an ensemble of the checkpoints was employed, (refer to Figure 8.6). However, for four cancer types, TCGA-GBM, TCGA-LAML, TCGA-PRAD, and TCGA-TGCT, the model did not show significant predictive power. During the investigation, we observed that these datasets had non-admissible pairs in some of the data folds, i.e., all samples had censor value  $\delta = 0$  in Equation 8.10. In the case

of TCGA-PRAD and TCGA-TGCT, the number of samples having  $\delta = 1$  in the training/validation cohort was 12 and 3, respectively. To address the lack of predictive power, we fine-tuned the model for these datasets by using the stratified k-folds to offset the class-representation problem in the data folds. After searching for the optimal hyperparameters for fine-tuning, the model's performance became significant ( $p < 0.05$ ) for three out of four datasets, (refer to green box plots in Figure 8.6). The model failed to converge for the TCGA-TGCT data despite numerous efforts at fine-tuning. A possible rationale behind this is that because of the more pronounced class imbalance in this particular dataset, i.e., 3 out of 111 samples in the training cohort have censor value  $\delta = 1$ , the model cannot learn to distinguish the samples based on the OS.

It is imperative to mention here that MLPs-based networks are very sensitive to catastrophic forgetting when presented with out-of-distribution data or when subjected to a different task [443]. We fine-tuned the SeNMo encoder for one public data (CPTAC-LSCC) and one internal data (Moffitt's LSCC) [612, 659]. In our simulations to fine-tune the model, we encountered the catastrophic forgetting phenomenon in SeNMo, where the model would fail to converge on both new datasets. This was more pronounced when a certain number of hidden layers were frozen, and the rest were trained with lower learning rates, as shown in Figures 8.10 and 8.11. We resorted to the option of unfreezing all the layers of the encoder and fine-tuning the model with a very small learning rate ( $4e - 5$ ), high weight decay and dropout (0.35), and just 10 epochs. This method worked and the model showed significant performance on the out-of-distribution datasets.

The KM survival curves depicted in Figure 8.7 demonstrate a clear stratification of survival probabilities among three risk-defined patient groups. These results underscore the effectiveness of the risk stratification model in predicting long-term outcomes and highlight the critical need for targeted therapeutic strategies based on individual risk assessments. This stratification allows for more personalized patient management and could potentially guide clinical decision-making toward improving OS rates across diverse patient populations.

The classification results in Figure 8.8 illustrate the superior generalizability of the model's predictive power to classify primary cancer types through the SeNMo encoder, despite it being

primarily trained for predicting OS. Additionally, the detailed classification report across various cancer types reveals that the model consistently maintains high precision, recall, and F1-scores for almost all cancer types. Such metrics not only confirm the model's effectiveness in accurately identifying the correct cancer class but also its reliability in replicating these results across different samples. This level of performance suggests the capability of the model to successfully learn high level representations from heterogenous, high-dimension, multivariate data stemming from complex molecular modalities such as gene expression, miRNA expression, somatic mutations, DNA methylation, and protein expression.

We made the entire codebase of SeNMo publicly available on GitHub (<https://github.com/lab-rasool/SeNMo>). We have made the latent representations of patient data generated from SeNMo available to the research community through our HoneyBee system [693]. HoneyBee stores these representations, also known as patient embeddings, in a structured format using Hugging Face datasets, effectively creating a vector database. HoneyBee has demonstrated the effectiveness of using patient embeddings, offering a significant advantage over the traditional approach of using raw data and extensive pre-processing [693]. The molecular data, overall survival information, and other phenotypes from the TCGA and corresponding labels are available from NIH Genomic Data Commons (<https://portal.gdc.cancer.gov/>). The gene expression, miRNA expression, and DNA Methylation data was obtained from UCSC XENA (<https://xena.ucsc.edu/>). The CPTAC-LSCC and Moffitt LSCC data are available at [248, 659]. The codebase for the project are available at <https://github.com/lab-rasool/SeNMo>.

## 8.5 Hyperparameters Search - Pan-cancer Training

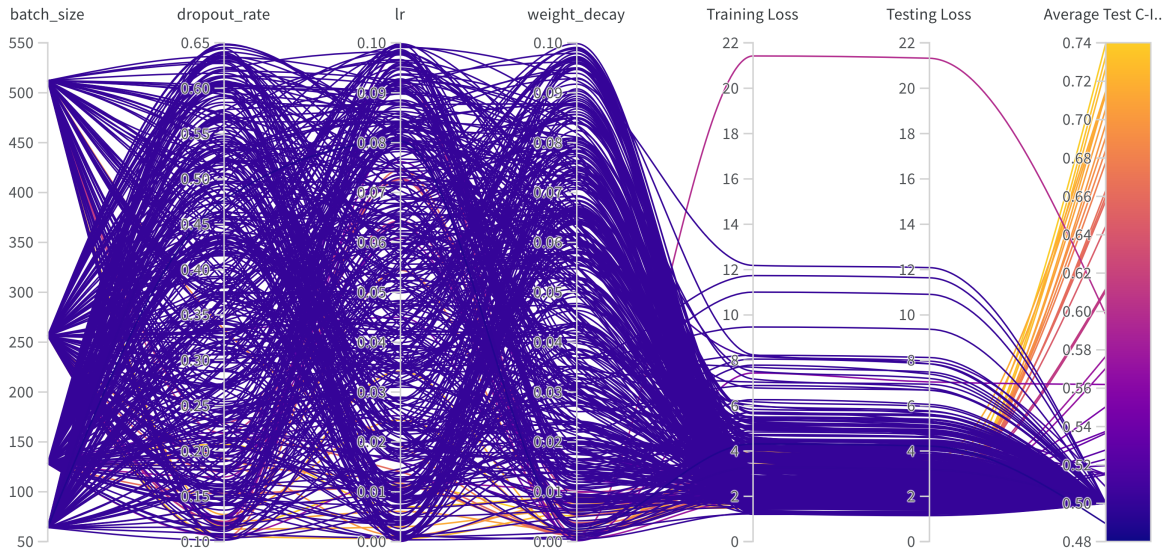


Figure 8.9 Hyperparameters search for training on Pan-cancer multiomics data.

## 8.6 Hyperparameters Search - Finetuning CPTAC-LSCC

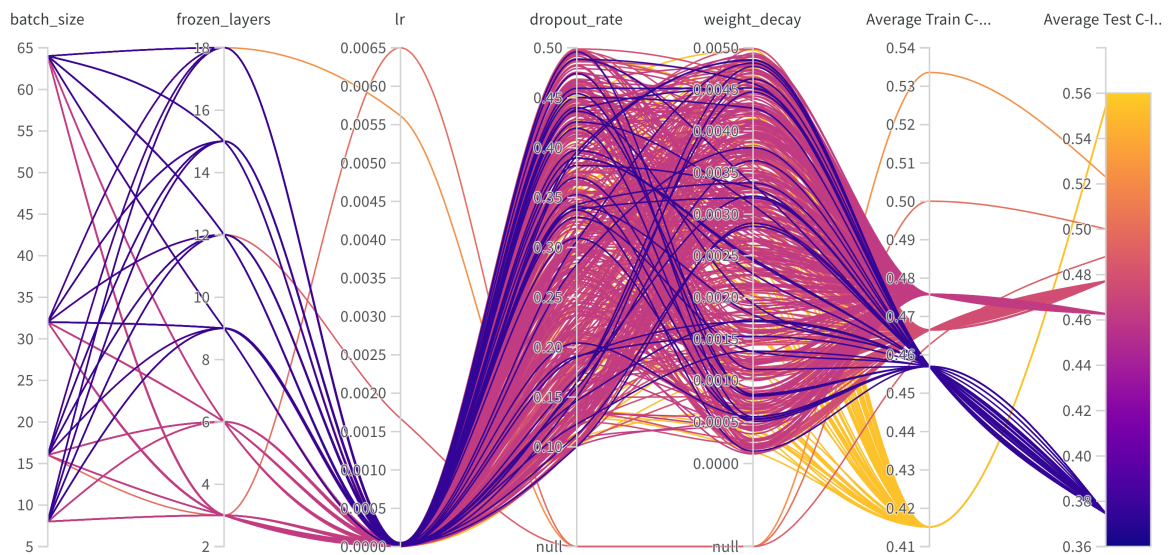


Figure 8.10 Hyperparameters search for finetuning on CPTAC-LSCC multiomics data.



## 8.7 Hyperparameters Search - Finetuning Moffitt-SCC

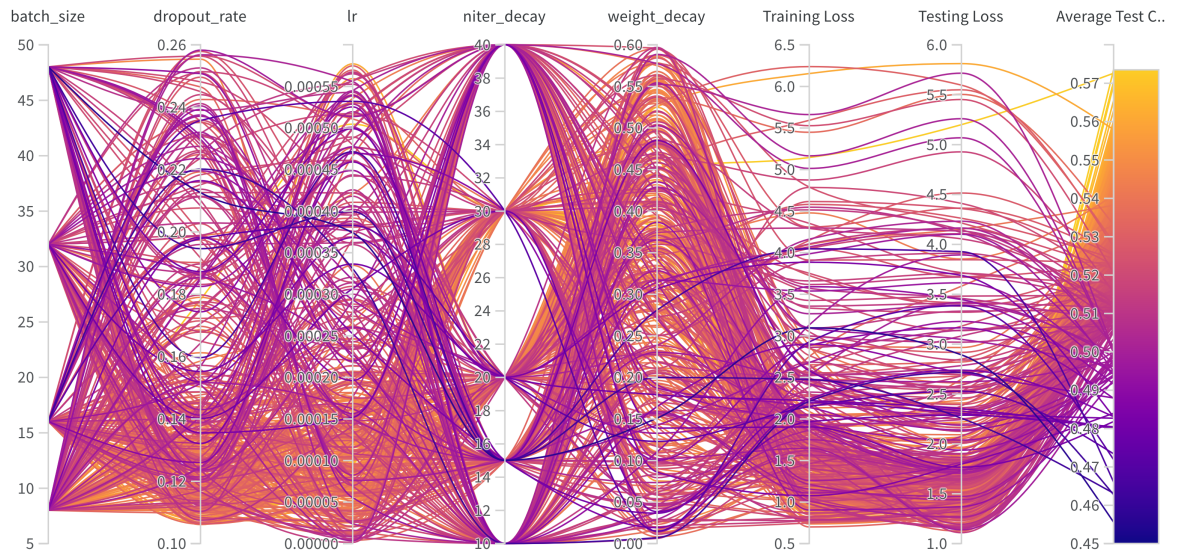


Figure 8.11 Hyperparameters search for fine-tuning on Moffitt's Lung SCC multiomics data.

## **Chapter 9: Embedding-based Multimodal Learning on Pan-Squamous Cell Carcinomas for Improved Survival Outcomes**

### **9.1 Introduction**

Clinical information about cancerous tumors is routinely recorded at different scales and resolutions throughout the progression of the disease, treatment, and survivorship [136, 703, 90]. The resulting data may include multiple diverse modalities [740], including (1) molecular and -omics information recorded from genome, proteome, transcriptome, epigenome, and microbiome, (2) diagnostic radiological imaging, e.g., ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) [213], (3) histopathology, immunohistochemistry (IHC), and immunofluorescence (IF) images and data, e.g., whole slide images (WSI) recorded from stained tumor tissue samples [268, 468], and (4) Clinical data including Electronic Health Records (EHR) that consist of structured and unstructured data about the patient, their disease, clinical notes from routine visits, labs tests and vitals, and clinical reports from radiology, pathology or biopsy [52, 756]. Jointly learning from such multimodal, multiscale, heterogeneous information with the possibility of out-of-distribution (OOD) inputs (e.g., unknown disease or cancer type), incomplete, noisy, and missing data is challenging but crucial for tackling complex diseases such as cancer. The state-of-the-art multimodal artificial intelligence/machine learning (AI/ML) techniques use data fusion methods and various flavors of deep neural networks, including Transformers, convolutions neural networks (CNNs), multilayer perceptions (MLPs), etc. [703, 135, 89, 458, 138, 430, 448, 763, 238]. None of these models are intrinsically designed to handle heterogeneous multimodal datasets with noisy, incomplete, and missing data during training and after deployment. Given the enormous growth in the volume and variety of medical data and the inherent noise in the data recording procedures/instruments, there is a need to develop robust

AI/ML models that can learn simultaneously from multimodal heterogeneous datasets to answer clinical questions.

### 9.1.1 Multiscale, Heterogeneous Data in Cancer

The challenge of understanding cancer exists at different scales, including (1) genetic and molecular aspects of cancer and its micro-environment, (2) pathological information about tissue, (3) radiological information about the organ, (4) physiology and health of the patient, and (5) their lifestyle. We must also understand the dynamic changes happening over time during the development and progression of the disease [703, 6]. The advanced MedTech hardware/software allows us to take snapshots of cancer development from a normal cell to a pre-malignant lesion to a malignant tumor in many different ways and at different scales and resolutions. The challenge is to coherently ingest, process, denoise, and learn from these datasets [204, 660]. Integrating data from heterogeneous modalities is vital for creating a unified view of cancer, which can be more insightful and predictive than a view created by a single data modality.

Cancer patients undergo various diagnostic imaging scans, lab tests, medical procedures, biopsies, and treatment regimens, including surgical resection, chemotherapy, radiation therapy, immunotherapy, or targeted therapies. During these activities, healthcare facilities collect and store various forms or modalities of data, which are later analyzed by medical professionals for treatment planning, disease monitoring, surveillance, and post-treatment survivorship. In this work, we focus on three types of data modalities: (1) digitized histopathology slides saved in the WSI file format, (3) -omics data that includes genomics, proteomics, and transcriptomics, (4) clinical data, which consists of patients' demographic information, clinical notes (including pathology reports), and lab results/vitals. Evidently, we are confronted with diverse data types that capture different yet complementary views of the underlying disease at different scales. That is, from molecular scale (captured using -omics data) to tissue (quantified with histopathology data) to demographic data (from EHR) and finally, a mixture of everything in semi-structured or unstructured text format (in the form of clinical notes and medical reports captured in the EHR).

### 9.1.2 Squamous Cell Carcinoma

Squamous cell carcinoma (SCC) is a type of cancer that can arise in various organs and tissues beyond the skin, including the lungs, bladder, cervix, esophagus, and head and neck region. It arises from squamous cells, which are flat cells that line many surfaces in the body. Lung SCC accounts for approximately 15-20% of all lung cancers, and head and neck SCC is the seventh most common cancer worldwide [600, 69]. These cancers can be aggressive and have significant mortality rates, highlighting the need for better understanding and treatment strategies. SCCs, particularly in the head and neck, can severely impact the quality of life due to their effects on essential functions like breathing, eating, and speaking. The survival rate for SCC is very high when detected early, with a 5-year survival rate of 99% [120]. The management of advanced SCCs often involves a multidisciplinary approach combining surgery, radiation therapy, chemotherapy, and targeted or immunotherapies. Ongoing research aims to optimize these multimodal treatment strategies and identify the most effective combinations for different SCC types and stages. Researchers are actively exploring potential biomarkers for early detection, prognosis, and treatment response prediction in various SCCs [110, 651]. In this work, we studied pan-squamous cell carcinoma (PanSCC) comprising lung, bladder, cervix, esophagus, and head and neck subtypes.

### 9.1.3 Multimodal Learning (MML)

MML techniques combine information from various modalities to improve the accuracy and reliability of a given ML task [6, 378, 430, 696, 647, 660, 549]. We can define five stages of multimodal learning, including preprocessing, feature extraction, data fusion, primary learning, and final classification [645]. The “data fusion” combines raw/extracted features or class prediction vectors from multiple modalities to create a single data representation. Data fusion can be performed in different ways: (1) *early fusion* involves merging the modality-specific features (or embeddings) into a single feature vector before training AI/ML model, (2) *intermediate fusion* involves training separate models for each data modality and combining model outputs for prediction, and (3) in *late fusion*, the output of each model is used to produce a separate decision, which are then combined

to make a final decision. The choice of fusion technique depends on the characteristics of the data and the specific problem being addressed [6, 332, 378].

#### 9.1.4 State-of-the-Art and Challenges in Oncology MML

Modern and classical AI/ML models have been used to fuse various modalities of oncology data [439, 277, 703, 89, 458, 138, 430, 448, 763, 740, 736]. However, these techniques are generally ad-hoc and limit their analysis to a selected set of modalities instead of using all available data.

The state-of-the-art MML models use Transformer-based architectures to fuse and jointly learn from image and text data. Using these models out-of-the-box or their variants for fusing medical data that includes various types of radiology images, gigapixel histopathology/IHC/IF images, a variety of -omics data, and semi-structured EHR data is neither straightforward nor optimal.

Most MML models targeted for oncology applications are ad-hoc by design, use various types of AI/ML models and fusion methods subjectively, and involve a significant amount of manual feature/model engineering focused on a specific cancer type (or sub-type) and an organ [439, 277, 703, 89].

Oncology MML models generally target a limited number of data modalities. For example, radiology and genomics are fused for radio-genomics analysis, pathology and -omics are mixed and referred to as pathomics, or any one modality (e.g., radiology or pathology imaging) is fused with demographic data (e.g., age or smoking status) from the EHR [703, 416, 98, 799].

The current state-of-the-art models use varying methods to address the challenges of missing and noisy data, which is pervasive in medical datasets. These models are not designed and trained to be robust to routine changes in the data in medical settings, e.g., changes in the data recording protocols, machine hardware/software/firmware updates, changes in patients' demographics owing to changes in the algorithm deployment site or a new mutation or disease variant.

Answering cancer-related clinical questions using multimodal oncology datasets has its unique challenges, nuances, and subtleties and therefore warrants its own unified data integration frame-

work that can (1) handle data heterogeneity, missingness, and noise, (2) learn relations between different modalities of a patient and between patients, and (3) make accurate predictions.

Our proposed framework has six distinct components. First, we introduce an embedding-based flexible and robust approach to tackle multi-modality cancer data. Second, we introduce graphs on the generated embeddings and perform graph structure learning to identify intra-/inter-cancer relationships/patterns using GNNs. Third, graph-based and learning-based methods with supervised techniques are introduced to fuse embeddings and handle missing/incomplete data. Fourth, the proposed graph structure enables multiscale learning across modalities. Fifth, we introduce a fusion mechanism that keeps the maximum amount of information intact while capturing relational patterns about the disease across different views of data. Sixth, we use the self-normalizing weights initialization on graph convolutional layers as well as use exponential linear unit (ELU) activation that ensures the self-normalizing property of GNNs. The evaluation metric for our framework is predicting accuracy for the given tasks of predicting OS quantified using the concordance index (C-index) [701]. Our work will potentially play a transformative role in the area of learning from multimodal, heterogeneous data in general and oncology data. The overview of our multimodal embeddings-based learning framework is illustrated in Figure 9.1.

## 9.2 Methods

We propose PARADIGM (Pan-Squamous Cell Carcinoma Representation using Advanced Multimodal learning with Graph-based Modeling). Our proposed framework benefits from the state-of-the-art AI/ML models for learning modality-specific embeddings and combine these embeddings hierarchically using GNNs for learning inter- and intra-cancer patterns.

### 9.2.1 PARADIGM Architecture

We use standard mathematical notations, i.e., lowercase letters ( $e$ ) for scalar values, lowercase bold ( $\mathbf{e}$ ) for column vectors, and uppercase letters ( $E$ ) for matrices. The graphs are depicted by  $G_{\text{sub}}=(V, C)$  having node-set  $V=\{v_1, v_2, \dots, v_n\}$ , where node  $v$  has feature vector  $\mathbf{x}_v$ , edge set

$C = \{(v_i, v_j) \mid v_i, v_j \in V\}$ , and subscript  $_{\text{sub}}$  represents the purpose or task associated with the graph  $G$ . Fig. 9.1 presents the main components of PARADIGM, (a) modality-specific sample embeddings, (b) aggregation and fusion into patient embeddings, (c) patient-specific graphs, and (d) joint graph learning. We describe these components in detail in the following:

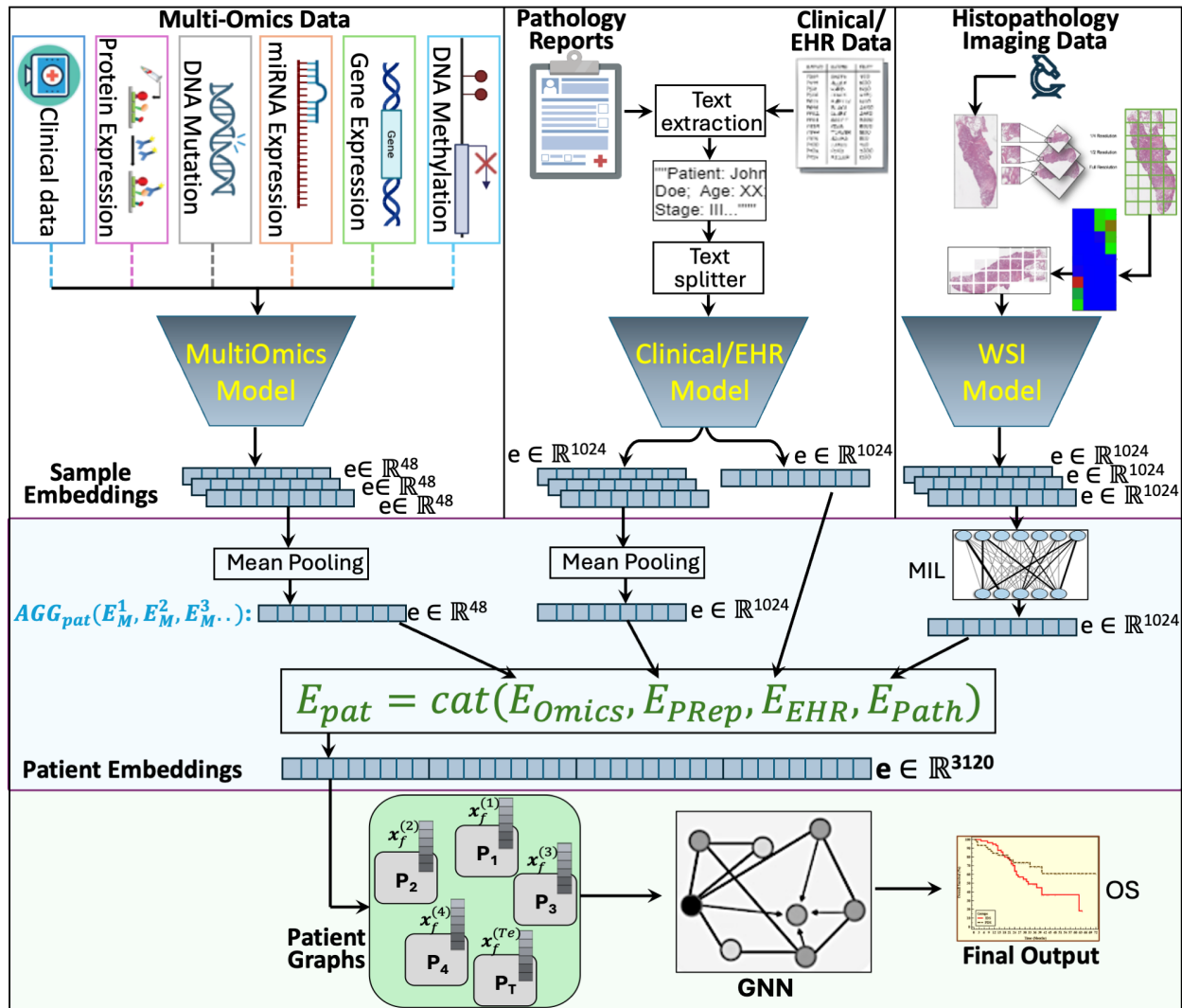


Figure 9.1 The schematic layout of the multi-modal data integration framework.

### 9.2.1.1 Modality-Specific Embeddings

The first component of the PARADIGM architecture consists of a set of pre-trained, locally fine-tuned modality-specific AI/ML models, e.g., various types of Transformers, CNNs, SNNs, ViTs, etc. [193, 705, 772, 736]. Our motivation is to leverage the state-of-the-art AI/ML pre-trained modality-specific models to find the most representative embeddings for all four data modalities. We prefer the models that have been pre-trained using related medical datasets [772, 56, 134, 733, 741, 560, 751]. After the first data processing step of modality-specific operations, we get:  $E_{\text{Path}} = [\mathbf{e}_p^{(1)}, \mathbf{e}_p^{(2)}, \dots, \mathbf{e}_p^{(T_p)}]$ ,  $E_{\text{Omics}} = [\mathbf{e}_o^{(1)}, \mathbf{e}_o^{(2)}, \dots, \mathbf{e}_o^{(T_o)}]$ ,  $E_{\text{EHR}} = [\mathbf{e}_e^{(1)}, \mathbf{e}_e^{(2)}, \dots, \mathbf{e}_e^{(T_e)}]$ , and  $E_{\text{PRep}} = [\mathbf{e}_{pr}^{(1)}, \mathbf{e}_{pr}^{(2)}, \dots, \mathbf{e}_{pr}^{(T_{pr})}]$  for WSI patches, -omics, EHR, and pathology report datasets, respectively. We have  $E_{\text{Path}} \in \mathbb{R}^{D_p \times T_p}$ ,  $E_{\text{Omics}} \in \mathbb{R}^{D_o \times T_o}$ ,  $E_{\text{EHR}} \in \mathbb{R}^{D_e \times T_e}$ , and  $E_{\text{PRep}} \in \mathbb{R}^{D_{pr} \times T_{pr}}$ , which gives  $\mathbf{e}_p \in \mathbb{R}^{D_p}$ ,  $\mathbf{e}_o \in \mathbb{R}^{D_o}$ ,  $\mathbf{e}_e \in \mathbb{R}^{D_e}$ , and  $\mathbf{e}_{pr} \in \mathbb{R}^{D_{pr}}$ . Here,  $T_p$ ,  $T_o$ ,  $T_e$ , and  $T_{pr}$  represent the total number of patients, and  $D_p$ ,  $D_o$ ,  $D_e$ , and  $D_{pr}$  represent the size of embedding vector for the WSI patches, -omics, EHR, and pathology reports datasets, respectively.  $D_p$ ,  $D_o$ ,  $D_e$ , and  $D_{pr}$  can have different values based on the complexity of the data modality. The EHR data is considered the baseline and is always available for all patients. Other data modalities may be missing, i.e.,  $T_p$ ,  $T_o$ ,  $T_{pr} \leq T_e$ .

### 9.2.1.2 Concatenation and Aggregation

We used a two-step process (1) aggregate the sample embeddings into patient-level embeddings and (2) concatenate resulting embeddings across all modalities. In the first step, we aggregate the different samples' embedding vectors for the same patient into a single patient embedding for each modality. We use a learning-based aggregation function to combine all available WSI sample embeddings for each patient to produce  $E_M = \text{AGG}_{\text{avail}}(E_M^1, E_M^2, E_M^3, \dots, E_M^u)$ , where  $\text{AGG}_{\text{avail}}$  represents an aggregator function implemented with a multiple-instance learning network for modality  $M$  having total samples  $u$  for each patient.  $\text{AGG}_{\text{avail}}$  is trained using the ground-truth data of 'age at index' and MSE loss function. For the -omics and pathology reports data modalities, we simply aggregate the sample embeddings by mean pooling to generate patient embeddings for the



respective modalities. As the result of this aggregation stage, we get aggregated modality embeddings  $\mathbf{e}_o \in \mathbb{R}^{48}$ ,  $\mathbf{e}_p \in \mathbb{R}^{1024}$ ,  $\mathbf{e}_e \in \mathbb{R}^{1024}$ , and  $\mathbf{e}_{pr} \in \mathbb{R}^{1024}$  for -omics, WSIs, EHR, and pathology reports, respectively. The second step involves concatenating the embedding vectors from different modalities into a single latent representation for each patient. The aggregated and concatenated patient embedding is represented by  $E_{\text{pat}} = \text{cat}(E_{\text{Omics}}^u, E_{\text{PRep}}^u, E_{\text{EHR}}^u, E_{\text{Path}}^u)$ , where  $\text{cat}$  represents the concatenation operation such that  $E_{\text{pat}} \in \mathbb{R}^{3120 \times T_e}$ . Thus, the new embedding vector for each patient now has information from the initial modality embeddings,  $E_M$ .

### 9.2.1.3 Patient-Specific Graphs

We construct patient graph  $G_F$  where each node of  $G_F$  represents a patient and the node embeddings are given by  $E_f = [\mathbf{x}_f^{(1)}, \mathbf{x}_f^{(2)}, \dots, \mathbf{x}_f^{(T_e)}]$ . The (weighted) adjacency matrix for the patient graph is calculated using the Euclidean distance between embedding vectors [656]. For the patient graph  $G_F$ , the adjacency matrix can be found using:  $A_{\text{pat}} = \text{Euc}_{ij}(E_M) = \sqrt{\sum_{k=1}^{D_e} (E_{M_{ki}} - E_{M_{kj}})^2}$ . The patient graph has the number of nodes equal to the number of patients in the clinical (EHR) data  $|V_{\text{pat}}| = T_e$ . We also generated the adjacency matrix using cosine distance between patient embeddings for comparing the effect of selecting different distance metrics on the downstream analysis.

### 9.2.1.4 Joint Graph Learning

Finally, the joint patient graph is used to train GNN using supervised learning and the MSE loss function. As the result of the graph learning, the node feature vector for each patient is updated from its neighbors in the message passing mechanism.

## 9.2.2 Learning Strategy

Having presented the architecture of the PARADIGM framework, we now present the learning strategy of our framework as follows.

### 9.2.2.1 Modality-Specific Learning

The modality-specific learning involves three tasks, (1) dataset development, (2) evaluating and selecting the modality-specific publicly available (foundation) models, and (3) transfer learning and fine-tuning of modality-specific models, as enumerated below.

1. Training multimodal models requires multimodal cancer datasets with EHR data as the base modality and one or more of the following: radiology, pathology, or -omics. We used public databases developed and shared by the National Institute of Health (NIH) and National Cancer Institute (NCI) and Moffitt's local datasets to train and validate our framework. Generally, the publicly available datasets may have only one or two data modalities. In case multiple modalities are present, they are scattered across different databases, and it is difficult to identify and link patients across databases. We use the already developed multimodal database, MINDS, to curate and build patient cohorts across five squamous cell carcinoma dataset (TCGA-LUSC, TCGA-HNSC, TCGA-CESC, TCGA-ESCA, and TCGA-BLCA) [691]. We acquired four different types of data modalities. The multi-omics data includes protein expression, DNA mutation, miRNA expression, gene expression, and DNA methylation, along with clinical data. We used the preprocessing steps for multiomics data adopted by SeNMo [736]. Pathology reports are processed through text extraction and text splitting techniques to obtain meaningful segments of text data. Clinical/EHR, which includes structured clinical records, and histopathology data comprising WSIs have been curated from the MINDS system [691]. We also evaluated the framework by training on the data collected at Moffitt ( $T = 103$ ) and consisted of EHR data (including age at diagnosis, gender, ethnicity, race, smoking status, year of diagnosis, vital status (alive/dead), and tumor cellularity), pathology images, and -omics data (included RNA-Seq expression and protein expression) [546].
2. We leverage pre-trained models for initial modality-specific data processing. Based on our analysis, we have selected pre-trained models for obtaining initial embeddings after

supervised and self-supervised fine-tuning of these models individually and jointly [772, 736]. For the WSIs, we evaluated various models and selected UNI for generating embeddings [134]. For the clinical data such as EHR and pathology reports, we evaluated GatorTron and ClinicalT5 models [772, 455]. Based on our experiments, we selected GatorTron as our embedding model because of its superior performance on medical language tasks and text interpretation from pathology reports [772]. For the multi-omics model, we selected the only model that has been trained on the five molecular data types across 33 cancer sites and is publicly available, SeNMo [736]. All of these embeddings are publicly available through HoneyBee framework and hosted on Hugging Face [693].

3. The preprocessed data was then used to fine-tune the modality-specific models on different tasks such as OS, or self-supervised contrastive loss. After fine-tuning, we generated the embeddings for each modality model for each data sample. It is pertinent to mention here that these are sample-level vectors in the case of molecular and pathology report data and patch-level embeddings in the case of WSIs.

#### 9.2.2.2 Concatenation and Aggregation of Embeddings

As a result of the processing presented above, we get the initial embeddings for each data sample in each category of modality. Next, we combine these for each modality to build a single feature vector for each patient, referred to as patient-level embedding  $E_M = AGG_{\text{avail}}(E_M^1, E_M^2, E_M^3, \dots, E_M^u)$ . In the case of patch-level embeddings for WSIs generated from the GatorTron model, we employed multiple-instance learning (MIL) where a neural network is trained in a weakly-supervised fashion to generate the slide-level embeddings [234]. We experimented with attention-based MIL (ABMIL), mean-pooling, and max-pooling learning strategies and selected ABMIL for its superior performance in predicting the patient’s age at diagnosis [309]. The molecular and pathology reports embeddings had only a few patients with more than one sample, so we employed a simple pooling using the mean of the sample vectors. As the result of aggregation, we get the uniform-dimensional embedding vectors for each patient in each modality type. After aggregation,

we concatenated the patient’s modality embeddings into a single embedding vector per patient,  $E_{\text{pat}} = \text{cat}(E_{\text{Omics}}^u, E_{\text{PRep}}^u, E_{\text{EHR}}^u, E_{\text{Path}}^u)$ . It is important to note here that the concatenation operation caters for the missing modalities by padding zeros to the missing feature vectors. This way, the framework takes the union of patients across different modalities to ensure that patients with missing data modalities, as in the real-world scenario, are not excluded from the analysis. The concatenation approach allows the retention of all the available information for the given patient. The embedding aggregation and concatenation operations are needed for three reasons, (1) to reduce the size of modality-specific embeddings to represent patient-level embeddings, (2) to combine embeddings from different data modalities in a joint space for pathology, -omics, and EHR data, and (3) to produce a joint embedding of the same size despite the possibility of missing some data modalities for some patients.

### 9.2.2.3 Patient Graphs and Joint Graph Learning

The final set of embeddings are used as nodal features in the joint graph, which are trained to predict the selected endpoints of OS. It is important to highlight that the choice of the endpoint (i.e., OS) does not limit the PARADIGM framework. Based on the datasets, cancer type, and organ/site, our framework can support various other endpoints: tumor response rate, tumor shrinkage, disease-free survival, time to treatment discontinuation, toxicity, and time to next treatment. We use joint embedding for each patient to build a graph that represents one disease, i.e., one graph for each of the five cancer types. We then combine these subgraphs into a joint PanSCC graph data and use it to train the GNN. We used various measures to quantify the change in the embeddings, e.g., Euclidean or cosine distance. For the prediction of OS, we used convolution-based GNNs as they aggregate neighboring nodes’ embeddings through a stack of multiple layers [758]. We tested commonly used spatial and spectral convolutional GNNs such as GraphSAGE [271], Graph Attention Network (GAT) [708], Graph Convolutional Network (GCN) [375], and Graph Isomorphism Network (GIN) [761]. Based on our preliminary experiments, we selected the GCN model as our downstream model [375].

### 9.2.3 Model Evaluation

We have five trained models based on our selection of five cancers (cervical, bladder, head and neck, lung, and esophageal) and one model for PanSCC. We also compare the performance of PARADIGM models with the single-modality, two-modality, and three-modality models to highlight the benefits of using the multimodal approach. We also compare the performance of our models with the state-of-the-art multimodal models. We evaluate the models' performance using the C-Index [427], which compares the predicted outcome probabilities from the model with the ground-truth data. C-Index ranges from 0 to 1, where a value of 1 indicates the model can perfectly distinguish between individuals who experience an event earlier and those who experience it later. A value of 0.5 suggests that the model performs no better than random chance, while a value below 0.5 indicates that the model's predictions are inversely related to the observed outcomes. We use logrank statistical test to quantify the performance of the models to establish statistical significance.

## 9.3 Results

### 9.3.1 Multimodal Integration of Oncology Database System

Accessing, storing, pre-processing, harmonizing, clearing, and feeding multimodal data, consisting of various images, text, real numbers, and categories, to our models for training is a significant challenge. Given that no such system existed in oncology, we developed an in-house database system called MINDS and presented it in Fig. 9.2 [691]. MINDS can handle all types of data, including -omics (e.g., genomics, transcriptomics, and proteomics), diagnostic radiological imaging, histopathology/IHC/IF imaging, and EHR. MINDS employs data harmonization techniques to integrate all data types seamlessly and enables training PARADIGM models with up to petabytes of data. MINDS allows building cohorts for various cancer sites and types/sub-types. We followed the following scheme to build MINDS:

1. We collected (semi-)structured data from Genetic Data Commons (GDC), Imaging Data Commons (IDC), and Proteomic Data Commons (PDC) [217, 259, 682]. AWS Glue was used to pre-process and aggregate data [36, 38].
2. The pre-processed semi-structured data were uploaded to an Amazon S3 ingest buckets [35]. The structured data was stored in an AWS RDS instance [33].
3. We used AWS Data Lake Formation tool [37], which automated transforming the semi-structured data stored in the S3 bucket into a query-able data lake using AWS Glue crawlers and AWS Athena [30].
4. The structured data stored in AWS RDS was integrated into a data warehouse using Amazon Redshift [34].
5. We used AWS step functions to coordinate and manage the data export process for the selected cohort. We also provide manifest files to interact with GDC, IDC, and PDC and download all the unstructured data using their APIs.

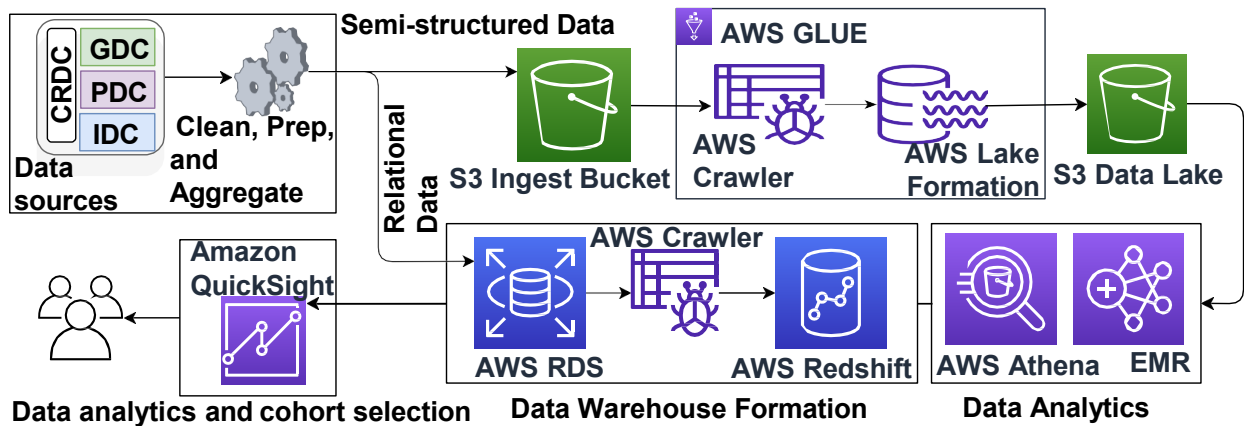


Figure 9.2 A schematic layout of MINDS multimodal data for PARADIGM models training.

### 9.3.2 Modality-Specific Models, Fine-tuning, and Transfer Learning

We investigated various modality-specific models for our data modalities, pathology, -omics, pathology reports, and EHR. Our criterion for evaluating the representativeness or optimality of these modality-specific models includes their predictive performance against the ground-truth data of OS. As a result of preliminary analysis, we used: (1) UNI for histopathology images [134]. UNI is a vision transformer model pre-trained on whole slide images using the DINOv2 self-supervised framework [134]. We selected UNI for its ability to capture intricate details and patterns in histopathology images through self-supervised pre-training [134]. (2) GatorTron for the EHR data (including clinical notes and pathology reports) [772]. GatorTron is a large language model specifically designed for clinical NLP. It was trained on a massive corpus of 277 billion words, including 82 billion words from de-identified clinical notes and 195 billion words from various English texts. (3) Self-Normalizing Networks (SeNMo) for -omics data [736], which is a mini-foundation model that has been trained on five molecular data types for more than 13,000 patients across 33 cancer types. Our experiments showed that UNI and GatorTron worked well with fine-tuning using small datasets [772, 56, 546]; however, SeNMo always required transfer learning with larger datasets, potentially linked to the complexity and variability of the -omics datasets [736, 740].

### 9.3.3 Prediction of OS in PanSCC Cohort

We predicted OS for pan-squamous cell carcinoma (PanSCC) patients using our framework, and a host of multimodal and uni-modal MLPs, Transformers, and classical ML algorithms such as eXtreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR).

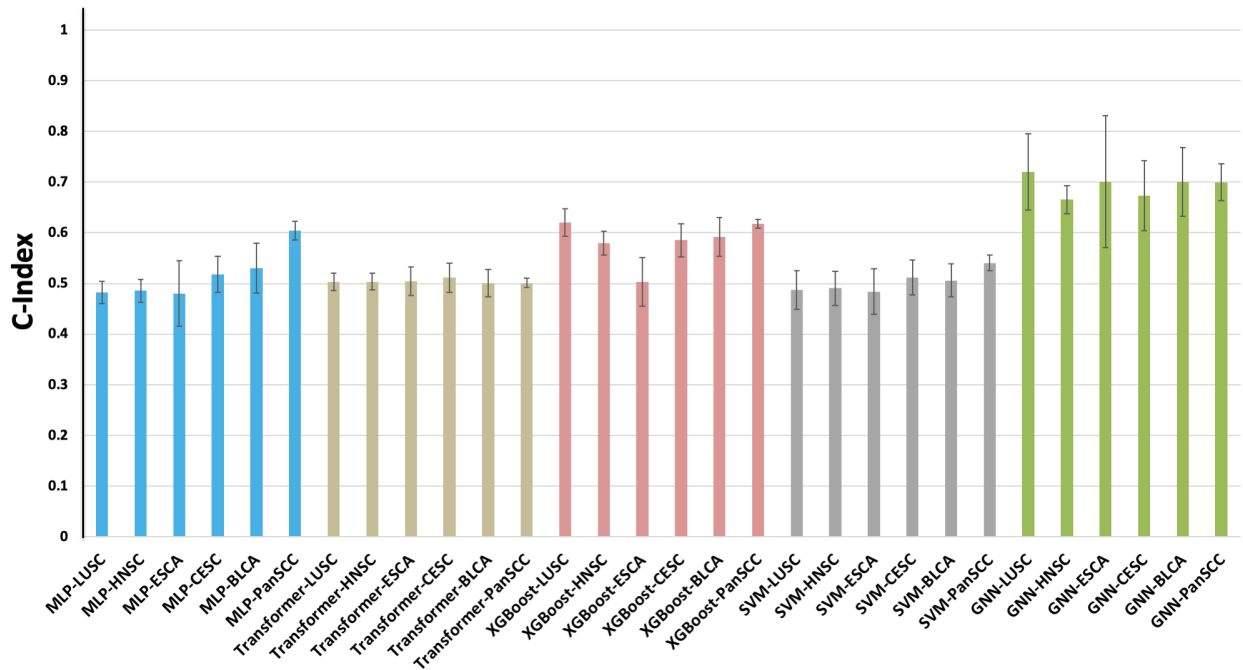


Figure 9.3 C-Index on training on clinical, pathology reports, WSIs, and molecular data.

### 9.3.3.1 Four-modality Analysis

Figure 9.3 shows the performance of different ML and DL models trained on multimodal datasets for SCC comprising four modalities; clinical, pathology reports, whole slide images (WSIs), and molecular data. The models include Multilayer Perceptrons (MLPs), Transformers, XGBoost, Support Vector Machines (SVMs), and GNNs. Each model was tested across various types of squamous cell carcinoma, lung squamous cell carcinoma (LUSC), head and neck squamous cell carcinoma (HNSC), esophageal carcinoma (ESCA), bladder carcinoma (BLCA), and pan-squamous cell carcinoma (PanSCC) comprising all five cancer types. The performance metric used is the C-Index for OS predictions, shown on the vertical axis. The MLP models, represented in blue, exhibit moderate performance with C-Index values around 0.5 to 0.6, with the best performance on the panSCC cohort. Transformer models, shown in beige, demonstrate similar performance to



MLPs. The XGBoost models, depicted in red, generally outperform the MLP and Transformer models, achieving higher C-Index values, indicating better prediction accuracy. The SVM models, represented in gray, show consistent but slightly lower performance compared to XGBoost models. The GNN models, shown in green, achieve the highest C-Index values across all cancer types, suggesting that GNNs are particularly effective in leveraging multimodal data for accurate survival predictions. The error bars indicate the variability in the performance across different folds, providing insight into the reliability of the predictions.

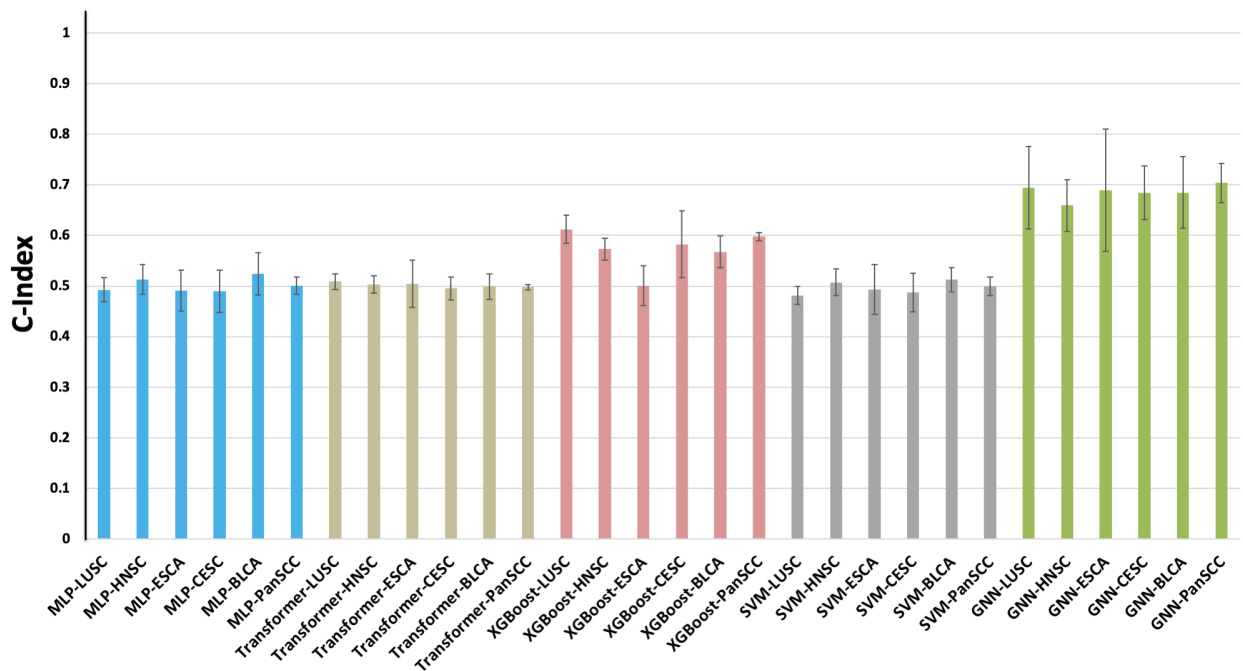


Figure 9.4 C-Index for OS predictions on clinical, pathology reports, and WSIs SCC data.

### 9.3.3.2 Three-modality Analysis

Figure 9.4 illustrates the performance of different ML and DL models on multimodal datasets for SCC, encompassing clinical data, pathology reports, and WSIs. The MLP models, show moderate performance with C-Index values around 0.5, indicating the random predictions. Transformer

models, depicted in beige, demonstrate similar performance to MLPs, with C-Index values clustering around the same range. The XGBoost models outperform both MLP and Transformer models, achieving higher C-Index values, which suggests better survival prediction accuracy. The SVM models, represented in gray, exhibit comparable performance to MLPs and Transformers. The GNN models achieve the highest C-Index values across all cancer types, significantly outperforming the other models.

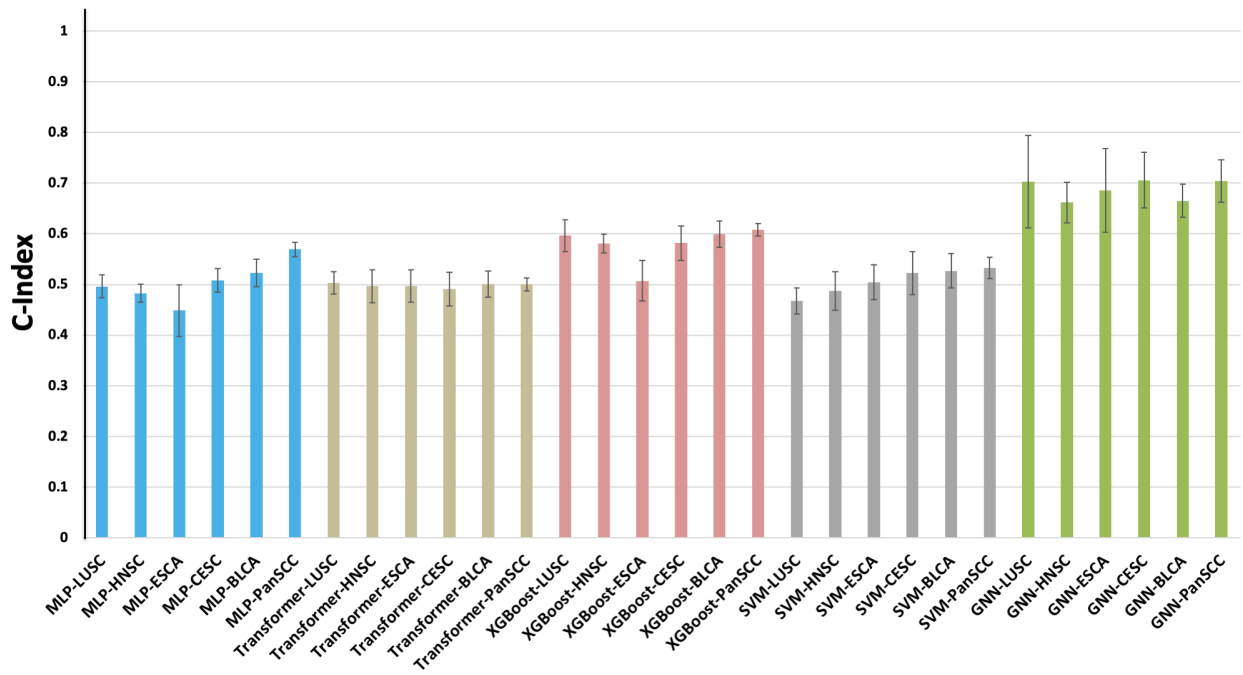


Figure 9.5 C-Index for OS predictions on clinical, molecular, and WSIs SCC data.

Figure 9.5 illustrates the performance of MLPs, Transformers, XGBoost, SVMs, and GNNs SCC data incorporating clinical data, molecular data, and WSIs. The MLP models show random performance for LUSC, HNSC, and ESCA with C-Index values  $<0.5$ , while a gradual improvement as we move from CESC, BLCA to panSCC data cohorts with C-Index  $<0.6$ . Transformer models do not perform well with all cohorts having C-Index  $<0.5$ . The XGBoost models consistently outperform the MLPs, Transformers, and SVM models. GNNs achieved the highest C-Index values across all cancer types, significantly outperforming the other models.

Figure 9.6 shows the MLPs, Transformers, XGBoost, SVMs, and GNNs trained on SCC data incorporating clinical data, pathology reports, and molecular data. The MLP models performed better than all previous three-modality cohorts having C-Index  $>0.5$  in five out of six datasets. Transformer models have similar ranges in C-Indices as before. The XGBoost models generally outperform the MLP and Transformer models, achieving higher C-Index values. The SVM models have shown better performance in this cohort with comparable performance to the XGBoost models. The GNN models achieved the highest C-Indices across all cancer types, significantly outperforming the other models.

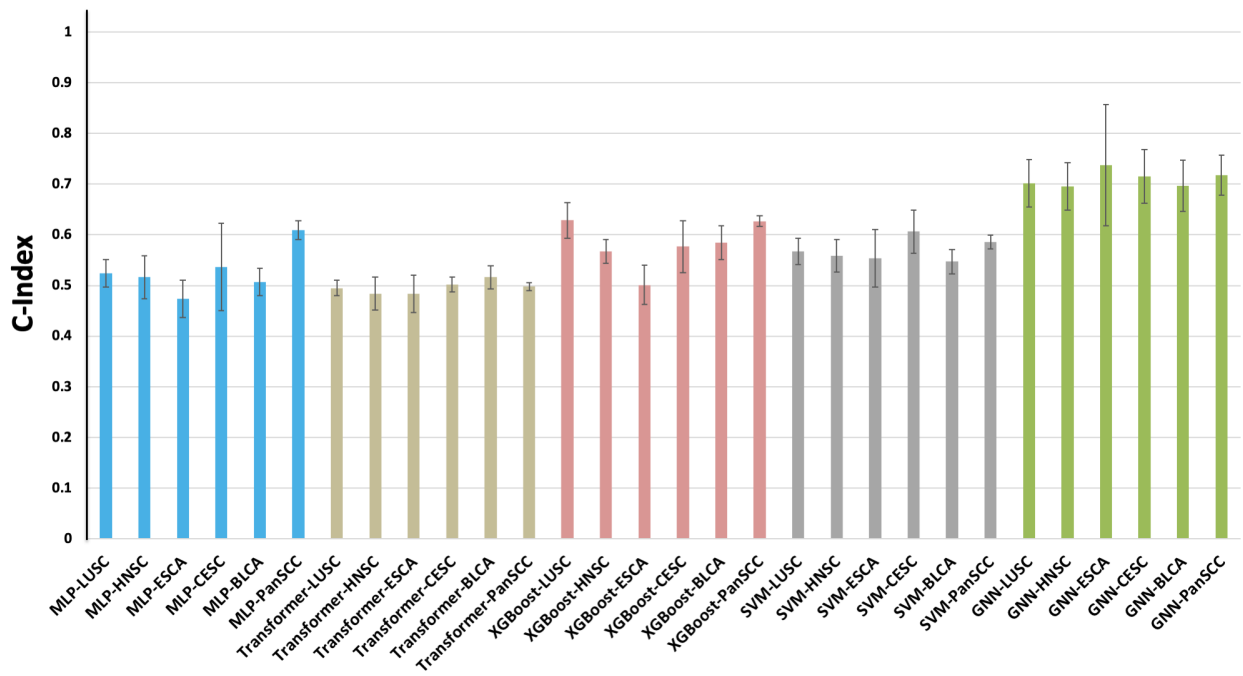


Figure 9.6 C-Index for OS predictions on clinical, pathology reports, and molecular data.

### 9.3.3.3 Two-modality Analysis

Figure 9.7 shows the performance of MLPs, Transformers, XGBoost, SVMs, and GNNs applied to multimodal datasets for SCC, integrating clinical data and WSIs. The MLPs, Transformers, and

SVM models exhibit random performance with C-Index  $<0.5$  for most of the cohorts. The XGBoost models outperform the MLPs, Transformers, and SVMs, achieving C-Index values from 0.5 to 0.6. GNNs achieve the highest C-Index values across all cancer types, ranging from 0.65 to 0.72.

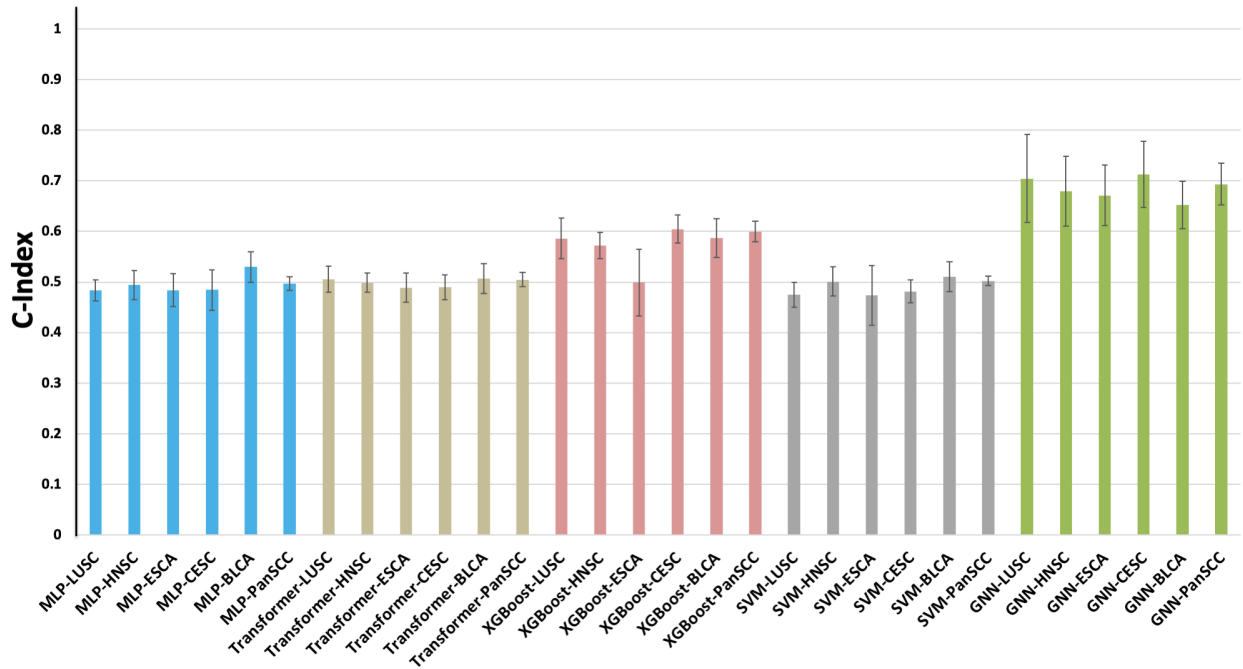


Figure 9.7 Performance comparison of OS prediction on clinical and WSIs SCC data.

Figure 9.8 illustrates that MLPs achieved the predictive performance for C-Index between 0.5 and 0.59, while XGBoost achieved values between 0.49 for ESCA to 0.63 for LUSC. Transformers fail to perform well, while SVMs have moderate performance range of 0.49-0.56. GNN models, illustrated in green, achieve the highest C-Index values across all cancer types, with all values around 0.7.

Figure 9.9 presents the performance of various models applied to SCC data with clinical and molecular data types. As with the previous results, the order of performance is Transformers, SVMs, MLPs, XGBoost, and GNNs, in increasing C-Index on predicting OS.

### 9.3.3.4 *Internal Data Analysis*

Figure 9.10 shows the results from analyzing Lung SCC data collected at Moffitt Cancer Center [659]. The data consists of four modalities, including pathology images, clinical data, and two subtypes of molecular data, RNA-Seq expression, and protein expressions [659]. We generated pathology embeddings using REMEDIS [56] and EHR embeddings using GatorTron [772], without fine-tuning the models. We trained SeNMo for -omics data to generate RNA-Seq expression and protein expression embeddings. Model evaluation was done using C-index and 10-fold cross-validation, as shown in Figure 9.10. The models selected for evaluation are MLP, Transformer, self-normalizing network (SNN), and GNN. For two-modality combinations, MLP models show varying performance, with C-Index values ranging from approximately 0.5 to 0.7. Transformer models with two modalities exhibit poor performance compared to MLPs. For the three-modality combinations, the models generally show improved performance, especially the Transformers models exhibit better performance compared to two modalities. The four-modality combinations exhibit the highest performance. The MLP model trained on EHR, RNA, Protein, and WSIs achieves a C-Index of around 0.65, indicating strong predictive accuracy. The Transformer model with the same four modalities achieves a high C-Index, 0.91. The GNN model further improves the C-Index to 0.93, demonstrating the benefit of integrating all four data modalities.

## 9.4 Discussion

Disease-related information for oncology resides at varying scales and resolutions of data. Clinicians routinely fuse such diverse information mentally for decision-making, but there are limits to human processing capabilities. On the other hand, AI/ML models struggle when ingesting such heterogeneous, multiscale information for critical decision-making. Graphs and GNNs have been shown to perform well in contextual learning as well as represent the robust architectures of deep neural networks [735, 740]. Jointly learning from such multimodal, multiscale, heterogeneous information with the possibility of incomplete and missing data modalities is challenging but crucial for understanding and tackling complex diseases like cancer. Our proposed framework

for multimodal learning from multiscale, heterogeneous oncology datasets, has been trained and validated for accurately and robustly predicting overall survival (OS) for different tumor types (e.g., head and neck, lung, esophageal, cervical, and bladder cancers). We have also evaluated the framework on the Moffitt Cancer Center & Research Institute’s internal lung squamous cell carcinoma cohort comprising 103 patients. Our framework is built on pre-trained AI/ML models, Graph Neural Networks (GNNs), and self-supervised and supervised techniques to learn from multimodal, heterogeneous datasets. Our multimodal framework’s potential applications go beyond cancer and healthcare settings, such as the deployment of AI/ML in many mission-critical application areas, including aviation safety, autonomous control of vehicles, automated decision-making in human-machine integrative environments, data fusion applications, and financial systems. Overall, the results indicate that models trained with more modalities tend to achieve higher C-Index values, reflecting better predictive accuracy for OS panSCC data. The highest performance is observed with the four-modality combinations, particularly for the GNNs-based PARADIGM model. All models and processed data have been made available via Hugging Face and GitHub.

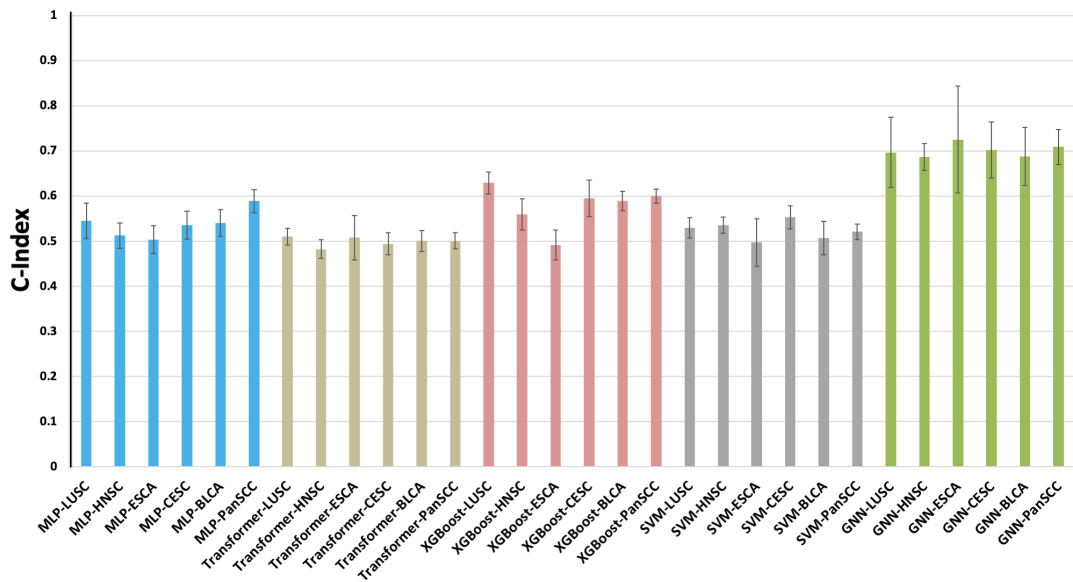


Figure 9.8 Performance comparison of OS prediction on clinical and pathology reports.

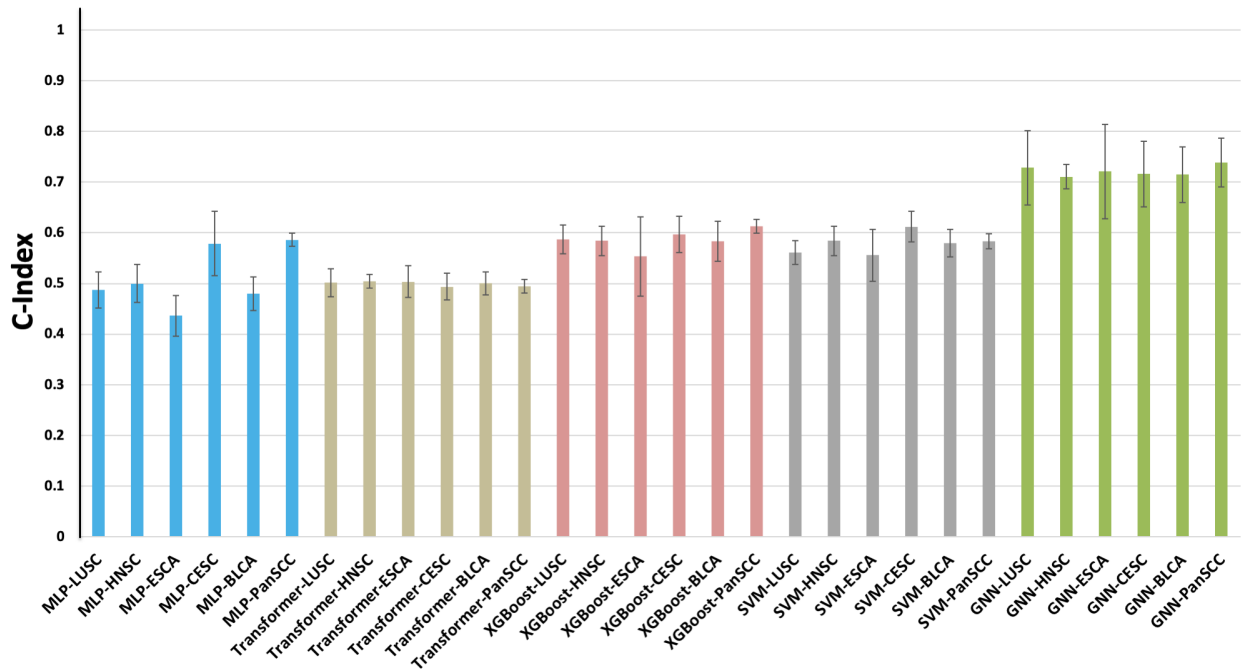


Figure 9.9 Performance comparison of OS prediction on clinical and molecular SCC data.

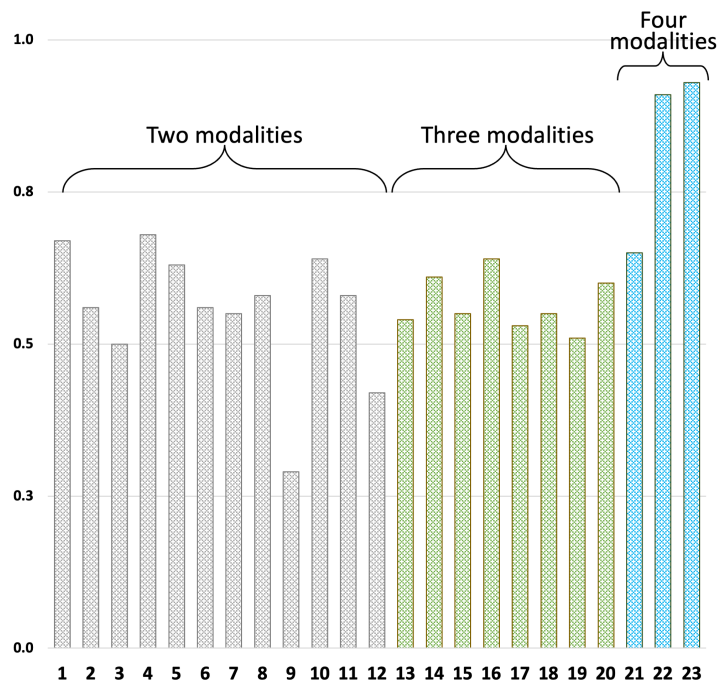


Figure 9.10 C-indices from different models trained on combinations of modalities.

#### 9.4.1 Embedding-based Approach

The proposed framework introduced an embedding-based flexible and robust approach for multimodal learning. It benefits from the open-source models, pre-trained with histopathology, -omics, and EHR data [56, 772, 736, 134], later fine-tuning these models to use them as the feature (or embeddings) extractors. This helped with rich feature extraction from all modalities at a significantly reduced computational cost.

#### 9.4.2 Graph Structure and Contextual Learning

Our framework introduced graph structure and GNNs on these embeddings to perform intra-/inter-cancer learning from heterogeneous datasets. The GNN-based approach enabled the representation of patient embeddings in the form of graph-structured data, where nodes represented patients and weighted edges between nodes represented inter-patient similarities. The process of updating the features of the current node based on its neighbors and the strength of the edges between these nodes helped in intra-/inter-cancer learning even though neighboring nodes may represent different cancers.

#### 9.4.3 Effect of Graph Sparsity

In analyzing the graph structures and their effects on the model's performance, we varied the sparsity of the input graphs. Sparsity is controlled by thresholding the weights on the edges quantified by the distance metrics, such as Euclidean or cosine distance. Figure 9.11 illustrates the effect of changing the graph sparsity on the prediction of OS by the same GNN model. Across all combinations of modalities and cancer data types, the dense graph configuration generally achieves significant C-Index values. When we increase the graph sparsity, there is a notable increase in the model's performance, with C-Index values increasing to  $>0.8$  in almost all the cases. This indicates that sparse graphs carry high signal-to-noise ratio and better predictive power for the GNN models, compared to dense graphs. Moreover, the inclusion of more data modalities generally enhances predictive accuracy, but the benefit is maximized when the graph is sparsely connected.



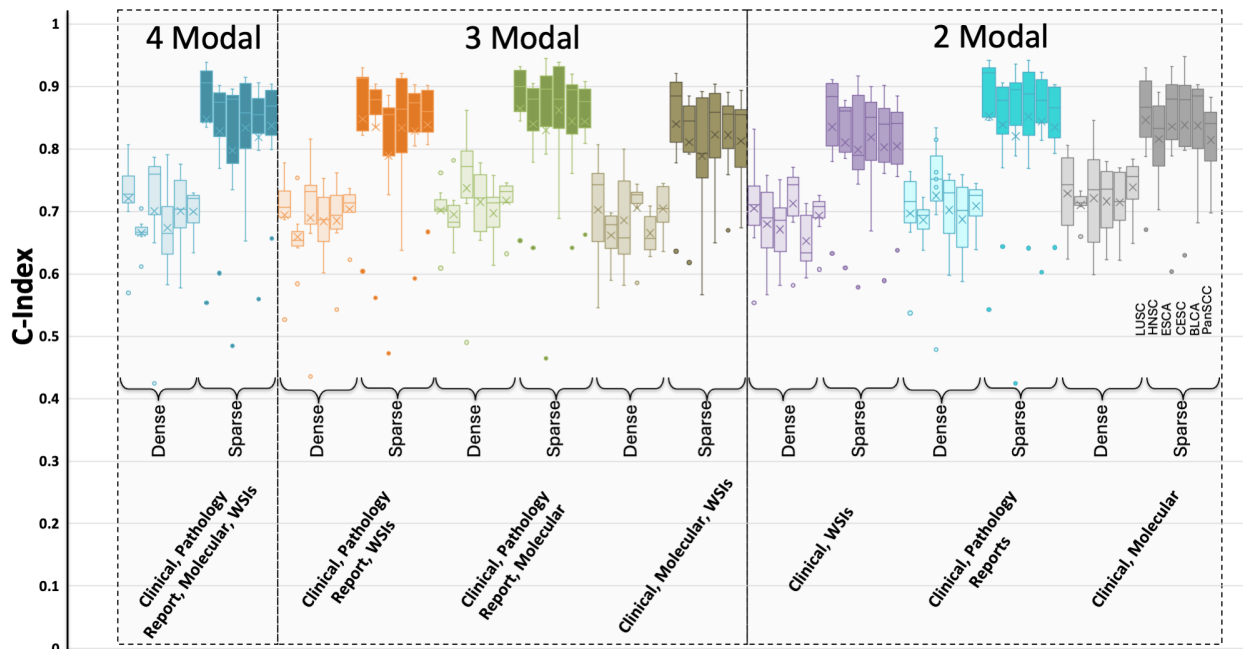


Figure 9.11 Impact of input graph sparsity on the predictive performance of GNNs.

#### 9.4.4 Handling Missing Modalities

The challenge of missing modalities is handled in our framework. Some patients may not have all the modalities; however, our framework allows learning with the available data and updating nodal features.

#### 9.4.5 Hierarchical Learning

The framework's hierarchical nature enables it to handle sub-modalities seamlessly. We can add another data modality to fuse with the existing data types. For example, radiological scans such as MRI, CT, PET scans can be processed through REMEDIS foundation model to generate sample embeddings, or different stains of histopathology images such as IHC or IF slides can be added to the existing pipeline seamlessly. This innovative methodology of hierarchically fusing

heterogeneous information based on the data modalities and sub-modalities is vital for learning from multiscale datasets that capture various aspects of underlying phenomena in space and time.

Our proposed framework, PARADIGM, demonstrates a significant advancement in the field of multimodal, heterogeneous data integration for cancer outcome prediction. By leveraging graph neural networks and modality-specific foundation models, we successfully generated comprehensive patient-level representations that capture the intricate relationships across various data modalities, including EHR, whole slide images, pathology reports, and molecular data. Our approach not only improved the predictive accuracy for survival analysis in pan-Squamous Cell Carcinomas (SCC) across multiple cancer types but also showcased the robustness and scalability of GNNs in handling complex, multimodal datasets. The superior performance of PARADIGM compared to traditional machine learning models underscores the importance of integrating diverse data sources to achieve a holistic understanding of cancer progression. Our findings suggest that the convergence of genetic, physiological, and psychosocial data into a unified framework can provide deeper insights and more accurate predictions, ultimately contributing to personalized and precision oncology. This research paves the way for future studies to explore the full potential of GNNs and multimodal data integration in various clinical applications, offering a promising direction for improving patient outcomes in cancer and beyond.

## Chapter 10: Conclusion and Future Work

This potential of integrating graph-theoretic measures, advanced AI/ML models, and multimodal data fusion techniques in enhancing the robustness and predictive capabilities of deep artificial neural networks is significant. We present the domain of graphs, graph learning, and their relation with the robust deep neural networks and multimodal learning in oncology. We demonstrated that graph structural properties such as topological entropy and Olivier-Ricci curvature can predict the robustness of DANNs before training, providing valuable insights for designing resilient neural network architectures. The exploration of brain tumor segmentation using advanced DNN architectures like U-Net highlighted the transformative impact of machine learning on medical imaging, improving the precision and efficiency of treatment planning and tumor surveillance. The rise of foundation models and generative AI offers a promising avenue for overcoming the limitations of task-specific AI/ML models. By leveraging large-scale training data and fine-tuning capabilities, these models provide robust performance across diverse applications, including digital pathology. In cancer research, the integration of multimodal data through advanced models like Graph Neural Networks (GNNs) and Transformers has shown remarkable potential in enhancing the accuracy of diagnostic and prognostic models. Systems like MINDS exemplify the future of oncology data integration, enabling comprehensive analysis and personalized cancer care.

There are many open questions related to the presented topics that are of interest for the research community. In future we will focus on expanding the scope of these models to encompass a wider range of cancer types, data modalities, and clinical applications. Efforts will be directed towards refining the integration frameworks to handle the ever-growing volume of biomedical data, ensuring scalability, security, and efficiency. Additionally, the development of new algorithms and techniques for improving the interpretability and transparency of AI/ML models will be a

critical area of research, aiming to foster greater trust and adoption in clinical practice. We will also focus on implementing these frameworks in the clinical setting with specialist-in-the-loop evaluation and assistance. The convergence of deep learning, graph theory, and multimodal data integration holds immense promise for revolutionizing cancer research and clinical practice. Continued innovation and collaboration across these fields will be essential in realizing the full potential of these technologies, ultimately leading to more effective, efficient, and personalized approaches to cancer care.

## References

- [1] Martin Abadi et al. TensorFlow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pages 265–283, 2016. (Savannah, GA, USA).
- [2] Esther Abels, Liron Pantanowitz, Famke Aeffner, Mark D Zarella, Jeroen van der Laak, Marilyn M Bui, Venkata NP Vemuri, Anil V Parwani, Jeff Gibbs, Emmanuel Agosto-Arroyo, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *The Journal of pathology*, 249(3):286–294, 2019.
- [3] Insight Ace. Ai-based digital pathology solutions market size, share & trends analysis report. @ONLINE, February 2024.
- [4] Debabrata Acharya and Anirban Mukhopadhyay. A comprehensive review of machine learning techniques for multi-omics data integration: challenges and applications in precision oncology. *Briefings in Functional Genomics*, page elae013, 2024.
- [5] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [6] Julián N Acosta, Guido J Falcone, Pranav Rajpurkar, and Eric J Topol. Multimodal biomedical ai. *Nature Medicine*, 28(9):1773–1784, 2022.

- [7] George Adam, Ladislav Rampásek, Zhaleh Safikhani, Petr Smirnov, Benjamin Haibe-Kains, and Anna Goldenberg. Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precision Oncology*, 4(1):19, 2020.
- [8] Mohammed Adnan, Shivam Kalra, and Hamid R Tizhoosh. Representation learning of histopathology images using graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 988–989, 2020.
- [9] Famke Aeffner, Mark D Zarella, Nathan Buchbinder, Marilyn M Bui, Matthew R Goodman, Douglas J Hartman, Giovanni M Lujan, Mariam A Molani, Anil V Parwani, Kate Lillard, et al. Introduction to digital image analysis in whole-slide imaging: A white paper from the Digital Pathology Association. *Journal of pathology informatics*, 10(1):9, 2019.
- [10] Raag Agrawal and Sudhakaran Prabakaran. Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity*, 124(4):525–534, 2020.
- [11] Sabeen Ahmed, Dimah Dera, Muhammad Saud Ul Hassan, Nidhal Carla Bouaynaya, and Ghulam Rasool. Failure detection in deep neural networks for medical imaging. *Frontiers in Medical Technology*, page 41, 2022.
- [12] Sabeen Ahmed, Dimah Dera, Saud Ul Hassan, Nidhal Bouaynaya, and Ghulam Rasool. Failure detection in deep neural networks for medical imaging. *Frontiers in Medical Technology*, 4, 2022.
- [13] Sabeen Ahmed, Ian E Nielsen, Aakash Tripathi, Shamooun Siddiqui, Ravi P Ramachandran, and Ghulam Rasool. Transformers in time-series analysis: A tutorial. *Circuits, Systems, and Signal Processing*, 42(12):7433–7466, 2023.
- [14] Sabeen Ahmed, Ian E Nielsen, Aakash Tripathi, Shamooun Siddiqui, Ghulam Rasool, and Ravi P Ramachandran. Transformers in time-series analysis: A tutorial. *arXiv preprint arXiv:2205.01138*, 2022. <https://arxiv.org/abs/2205.01138>.

- [15] David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, and Lars Petersson. A survey on graph-based deep learning for computational histopathology. *Computerized Medical Imaging and Graphics*, 95:102027, 2022.
- [16] Euijoon Ahn, Ashnil Kumar, Dagan Feng, Michael Fulham, and Jinman Kim. Unsupervised deep transfer feature learning for medical image classification. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, pages 1915–1918. IEEE, 2019.
- [17] Aiforia Team. AI for Image Analysis (AIFORIA), 2023. available at: <https://www.aiforia.com/>. Last accessed on Mar 31, 2023.
- [18] Muhammad Joan Ailia, Nishant Thakur, Jamshid Abdul-Ghafar, Chan Kwon Jung, Kwangil Yim, and Yosep Chong. Current trend of artificial intelligence patents in digital pathology: a systematic evaluation of the patent landscape. *Cancers*, 14(10):2400, 2022.
- [19] Khalid K. Al-jabery, Tayo Obafemi-Ajayi, Gayla R. Olbricht, and Donald C. Wunsch II. Data preprocessing. In Khalid K. Al-jabery, Tayo Obafemi-Ajayi, Gayla R. Olbricht, and Donald C. Wunsch II, editors, *Computational Learning Approaches to Data Analytics in Biomedical Applications*, pages 7–27. Academic Press, 2020.
- [20] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, et al. Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35:23716–23736, 2022.

- [21] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob Menick, Sebastian Borgeaud, Andrew Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikolaj Binkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karen Simonyan. Flamingo: a Visual Language Model for Few-Shot Learning. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.
- [22] Samer Albahra, Tom Gorbett, Scott Robertson, Giana D'Aleo, Sushasree Vasudevan Suseel Kumar, Samuel Ockunzzi, Daniel Lallo, Bo Hu, and Hooman H Rashidi. Artificial Intelligence and Machine Learning Overview in Pathology & Laboratory Medicine: A General Review of Data Preprocessing and Basic Supervised Concepts. In *Seminars in Diagnostic Pathology*. Elsevier, 2023.
- [23] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [24] Saghir Alfasly, Peyman Nejat, Sobhan Hemati, Jibrán Khan, Isaiah Lahr, Areej Alsaafin, Abubakr Shafique, Nneka Comfere, Dennis Murphree, Chady Meroueh, et al. When is a foundation model a foundation model. *arXiv preprint arXiv:2309.11510*, 2023.
- [25] Hussam Alkaissi and Samy I McFarlane. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*, 15(2):e35179, February 2023.
- [26] Areej Alsaafin, Amir Safarpour, Milad Sikaroudi, Jason D Hipp, and HR Tizhoosh. Learning to predict RNA sequence expressions from whole slide images with applications for search and classification. *Communications Biology*, 6(1):304, 2023.



- [27] Charlems Alvarez-Jimenez, Alvaro A Sandino, Prateek Prasanna, Amit Gupta, Satish E Viswanath, and Eduardo Romero. Identifying cross-scale associations between radiomic and pathomic signatures of non-small cell lung cancer subtypes: preliminary results. *Cancers*, 12(12):3663, 2020.
- [28] Ahmad Alwosheel, Sander van Cranenburgh, and Caspar G Chorus. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, 28:167–182, 2018.
- [29] Ashraf A Aly, Safaai Bin Deris, and Nazar Zaki. Research review for digital image segmentation techniques. *International Journal of Computer Science & Information Technology*, 3(5):99, 2011.
- [30] Amazon Web Services. Amazon Athena. Available online: <https://aws.amazon.com/athena/>. (accessed on 1 March 2023).
- [31] Amazon Web Services. Amazon CloudWatch. Available online: <https://aws.amazon.com/cloudwatch/>. (accessed on 7 August 2023).
- [32] Amazon Web Services. Amazon QuickSight. Available online: <https://aws.amazon.com/quicksight/>. (accessed on 1 March 2023).
- [33] Amazon Web Services. Amazon RDS. Available online: <https://aws.amazon.com/rds/>. (accessed on 1 March 2023).
- [34] Amazon Web Services. Amazon Redshift. Available online: <https://aws.amazon.com/redshift/>. (accessed on 1 March 2023).
- [35] Amazon Web Services. Amazon S3. Available online: <https://aws.amazon.com/s3/>. (accessed on 1 March 2023).
- [36] Amazon Web Services. AWS Glue. Available online: <https://aws.amazon.com/glue/>. (accessed on 1 March 2023).

- [37] Amazon Web Services. AWS Lake Formation. Online. Available online: <https://aws.amazon.com/lake-formation/>. (accessed on 1 March 2023).
- [38] Amazon Web Services. Data Catalog and crawlers in AWS Glue. Available online: <https://docs.aws.amazon.com/glue/latest/dg/catalog-and-crawler.html>. (accessed on 1 March 2023).
- [39] Amazon Web Services. Encryption at rest. Available online: <https://docs.aws.amazon.com/redshift/latest/mgmt/security-server-side-encryption.html>. (accessed on 7 August 2023).
- [40] Amazon Web Services. Security in AWS Glue. Available online: <https://docs.aws.amazon.com/glue/latest/dg/security.html>. (accessed on 7 August 2023).
- [41] Amazon Web Services. Serverless Computing - AWS Lambda - Amazon Web Services. Available online: <https://aws.amazon.com/lambda/>. (accessed on 7 August 2023).
- [42] Deepak Anand, Shrey Gadiya, and Amit Sethi. Histograms: graphs in histopathology. In *Medical Imaging 2020: Digital Pathology*, volume 11320, pages 150–155. SPIE, 2020.
- [43] MD Anderson. Rppa description. [https://www.mdanderson.org/documents/core-facilities/FunctionalProteomicsRPPACoreFacility/RPPADescription\\_2016.pdf](https://www.mdanderson.org/documents/core-facilities/FunctionalProteomicsRPPACoreFacility/RPPADescription_2016.pdf), 2024. Accessed: 2024-05-13.
- [44] Ove Andrén, Katja Fall, Lennart Franzén, Swen-Olof Andersson, Jan-Erik Johansson, and Mark A Rubin. How well does the gleason score predict prostate cancer death? a 20-year followup of a population based cohort in sweden. *The Journal of urology*, 175(4):1337–1340, 2006.
- [45] Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, and Guang-Zhong Yang. Big data for health. *IEEE Journal of Biomedical and Health Informatics*, 19(4):1193–1208, 2015.

- [46] Christof Angermueller, Heather J Lee, Wolf Reik, and Oliver Stegle. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome biology*, 18(1):1–13, 2017.
- [47] Fetty Tri Anggraeny, Intan Yuniar Purbasari, M Syahrul Munir, Faisal Muttaqin, Eka Prakarsa Mandiyarta, and Fawwaz Ali Akbar. Analysis of simple data imputation in disease dataset. In *International Conference on Science and Technology (ICST 2018)*, pages 471–475. Atlantis Press, 2018.
- [48] F Anowar, S Sadaoui, and B Selim. Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne), *comput. sci. rev.*, 40, 100378. *ISI*, 2021.
- [49] Lauren G Aoude, Bernadette ZY Wong, Vanessa F Bonazzi, Sandra Brosda, Shaun B Walters, Lambros T Koufariotis, Marjan M Naeini, John V Pearson, Harald Oey, Kalpana Patel, et al. Radiomics biomarkers correlate with cd8 expression and predict immune signatures in melanoma patients. *Molecular Cancer Research*, 19(6):950–956, 2021.
- [50] Giovanni Apruzzese, Mauro Andreolini, Luca Ferretti, Mirco Marchetti, and Michele Colajanni. Modeling realistic adversarial attacks against network intrusion detection systems. *Digital Threats: Research and Practice*, 2021.
- [51] Onur Asan, Ann B Nattinger, Ayse P Gurses, Jeanne T Tyszka, and Tina WF Yen. Oncologists’ views regarding the role of electronic health records in care coordination. *JCO clinical cancer informatics*, 2:1–12, 2018.
- [52] Briseis Aschebrook-Kilfoy, Paul Zakin, Andrew Craver, Sameep Shah, Muhammad G Kibriya, Elizabeth Stepniak, Andrea Ramirez, Cheryl Clark, Elizabeth Cohn, Lucila Ohno-Machado, et al. An overview of cancer in the first 315,000 all of us participants. *PloS one*, 17(9):e0272522, 2022.

- [53] Rebecca Asiimwe, Stephanie Lam, Samuel Leung, Shanzhao Wang, Rachel Wan, Anna Tinker, Jessica N McAlpine, Michelle MM Woo, David G Huntsman, and Aline Talhouk. From biobank and data silos into a data commons: convergence to support translational medicine. *Journal of Translational Medicine*, 19:1–13, 2021.
- [54] Marc Aubreville, Nikolas Stathonikos, Christof A Bertram, Robert Klopffleisch, Natalie Ter Hoeve, Francesco Ciompi, Frauke Wilm, Christian Marzahl, Taryn A Donovan, Andreas Maier, et al. Mitosis domain generalization in histopathology images—The MIDOG challenge. *Medical Image Analysis*, 84:102699, 2023.
- [55] Reza Azad, Amirhossein Kazerouni, Moein Heidari, Ehsan Khodapanah Aghdam, Amirali Molaei, Yiwei Jia, Abin Jose, Rijo Roy, and Dorit Merhof. Advances in medical image analysis with vision transformers: A comprehensive review. *arXiv preprint arXiv:2301.03505*, 2023.
- [56] Shekoofeh Azizi, Laura Culp, Jan Freyberg, Basil Mustafa, Sebastien Baur, Simon Kornblith, Ting Chen, Patricia MacWilliams, S Sara Mahdavi, Ellery Wulczyn, et al. Robust and efficient medical imaging with self-supervision. *arXiv preprint arXiv:2205.09723*, 2022. <https://arxiv.org/abs/2205.09723>.
- [57] Deven Kishor Babre. Clinical data interchange standards consortium: a bridge to overcome data standardisation, 2013.
- [58] Maroua Bahri, Flavia Salutari, Andrian Putina, and Mauro Sozio. Automl: state of the art with a focus on anomaly detection, challenges, and research directions. *International Journal of Data Science and Analytics*, pages 1–14, 2022.
- [59] Bijie Bai, Xilin Yang, Yuzhu Li, Yijie Zhang, Nir Pillar, and Aydogan Ozcan. Deep learning-enabled virtual histological staining of biological samples. *Light: Science & Applications*, 12(1):57, 2023.

- [60] Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021.
- [61] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-igg collection. *The cancer imaging archive*, 286, 2017.
- [62] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13, 2017.
- [63] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. *arXiv preprint arXiv:1811.02629*, 2018.
- [64] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443, 2018.
- [65] Peter Bankhead, Maurice B Loughrey, José A Fernández, Yvonne Dombrowski, Darragh G McArt, Philip D Dunne, Stephen McQuaid, Ronan T Gray, Liam J Murray, Helen G Coleman, et al. Qupath: Open source software for digital pathology image analysis. *Scientific reports*, 7(1):1–7, 2017.

- [66] Oshrat Bar, Amnon Drory, and Raja Giryes. A spectral perspective of DNN robustness to label noise. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *International Conference on Artificial Intelligence and Statistics, AISTATS 2022, 28-30 March 2022, Virtual Event*, volume 151 of *Proceedings of Machine Learning Research*, pages 3732–3752. PMLR, 2022.
- [67] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.
- [68] Yassine Barhoumi, Nidhal C Bouaynaya, and Ghulam Rasool. Efficient Scopeformer: Towards Scalable and Rich Feature Extraction for Intracranial Hemorrhage Detection. *arXiv preprint arXiv:2302.00220*, 2023.
- [69] Adam Barsouk, John Sukumar Aluru, Prashanth Rawla, Kalyan Saginala, and Alexander Barsouk. Epidemiology, risk factors, and prevention of head and neck squamous cell carcinoma. *Medical Sciences*, 11(2):42, 2023.
- [70] Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, 20(3):353–364, 2017.
- [71] Danielle Smith Bassett and ED Bullmore. Small-world brain networks. *The neuroscientist*, 12(6):512–523, 2006.
- [72] Michèle Basseville, Igor V Nikiforov, et al. *Detection of abrupt changes: theory and application*, volume 104. prentice Hall Englewood Cliffs, 1993.
- [73] Stefan Bauer, Roland Wiest, Lutz-P Nolte, and Mauricio Reyes. A survey of MRI-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology*, 58(13):R97–R129, 2013.
- [74] 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.

- [75] Norman J Beauchamp, R Nick Bryan, Marilyn M Bui, Gabriel P Krestin, Geraldine B McGinty, Carolyn C Meltzer, and Michael Neumaier. Integrative diagnostics: the time is now—a report from the international society for strategic studies in radiology. *Insights into Imaging*, 14(1):54, 2023.
- [76] Corey M Benedum, Arjun Sondhi, Erin Fidyk, Aaron B Cohen, Sheila Nemeth, Blythe Adamson, Melissa Estévez, and Selen Bozkurt. Replication of real-world evidence in oncology using electronic health record data extracted by machine learning. *Cancers*, 15(6):1853, 2023.
- [77] Kaustav Bera, Nathaniel Braman, Amit Gupta, Vamsidhar Velcheti, and Anant Madabhushi. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nature reviews Clinical oncology*, 19(2):132–146, 2022.
- [78] Kaustav Bera, Kurt A Schalper, David L Rimm, Vamsidhar Velcheti, and Anant Madabhushi. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nature reviews Clinical oncology*, 16(11):703–715, 2019.
- [79] Berkeley and Change Healthcare and Duke Health and Google and JHU and Mayo Clinic and Microsoft and MITRE and SAS and Stanford Medicine and UCSF and Vanderbilt University Medical Center. Coalition for Health AI, 2023. available at: <https://coalitionforhealthai.org/>. Last accessed on Jul 31, 2023.
- [80] Seth J Berkowitz, David Kwan, Toby C Cornish, Elliot L Silver, Karen S Thullner, Alex Aisen, Marilyn M Bui, Shawn D Clark, David A Clunie, Monief Eid, et al. Interactive multimedia reporting technical considerations: Himss-siim collaborative white paper. *Journal of Digital Imaging*, 35(4):817–833, 2022.
- [81] Lukas Biewald. Experiment Tracking with Weights and Biases, 2020. Software available from wandb.com.

- [82] Mohsin Bilal, Mohammed Nimir, David Snead, Graham S Taylor, and Nasir Rajpoot. Role of AI and Digital Pathology for Colorectal Immuno-oncology. *British Journal of Cancer*, 128(1):3–11, 2023.
- [83] Karl Y Bilimoria, Andrew K Stewart, David P Winchester, and Clifford Y Ko. The national cancer data base: a powerful initiative to improve cancer care in the united states. *Annals of surgical oncology*, 15(3):683–690, 2008.
- [84] M. Bishop. *Pattern Recognition & Machine Learning*. Springer, New York, August 2006.
- [85] Andrew Blake and Michael Isard. *Active contours: the application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion*. Springer Science & Business Media, 2012.
- [86] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.
- [87] Kevin Boehm, Pegah Khosravi, Rami Vanguri, Jianjiong Gao, and Sohrab Shah. Harnessing multimodal data integration to advance precision oncology. *Nature Reviews Cancer*, 22:1–13, 10 2021.
- [88] Kevin M Boehm, Emily A Aherne, Lora Ellenson, Ines Nikolovski, Mohammed Alghamdi, Ignacio Vázquez-García, Dmitriy Zamarin, Kara Long Roche, Ying Liu, Druv Patel, et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nature cancer*, 3(6):723–733, 2022.
- [89] Kevin M Boehm, Emily A Aherne, Lora Ellenson, Ines Nikolovski, Mohammed Alghamdi, Ignacio Vázquez-García, Dmitriy Zamarin, Kara Long Roche, Ying Liu, Druv Patel, et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nature cancer*, 3(6):723–733, 2022.



- [90] 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.
- [91] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models, 2022.

- [92] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021. <https://arxiv.org/abs/2108.07258>.
- [93] Andrea Bommert, Thomas Welchowski, Matthias Schmid, and Jörg Rahnenführer. Benchmark of filter methods for feature selection in high-dimensional gene expression survival data. *Briefings in Bioinformatics*, 23(1):bbab354, 2022.
- [94] Vadim Borisov, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk, and Gjergji Kasneci. Deep Neural Networks and Tabular Data: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21, 2022.
- [95] Luis Bote-Curiel, Sergio Muñoz-Romero, Alicia Gerrero-Curienes, and José Luis Rojo-Álvarez. Deep learning and big data in healthcare: A double review for critical beginners. *Applied Sciences*, 9(11), 2019.
- [96] Abdelmajid Bousselham, Omar Bouattane, Mohamed Youssfi, and Abdelhadi Raihani. Towards reinforced brain tumor segmentation on mri images based on temperature changes on pathologic area. *International journal of biomedical imaging*, 2019, 2019.
- [97] Nathaniel Braman, Jacob WH Gordon, Emery T Goossens, Caleb Willis, Martin C Stumpe, and Jagadish Venkataraman. Deep orthogonal fusion: multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pages 667–677. Springer, 2021.

- [98] Valentina Brancato, Carlo Cavaliere, Nunzia Garbino, Francesco Isgrò, Marco Salvatore, and Marco Aiello. The relationship between radiomics and pathomics in glioblastoma patients: Preliminary results from a cross-scale association study. *Frontiers in Oncology*, 12:1005805, 2022.
- [99] Claude R Brice and Claude L Fennema. Scene analysis using regions. *Artificial intelligence*, 1(3-4):205–226, 1970.
- [100] E Brodsky and Boris S Darkhovsky. *Nonparametric methods in change point problems*, volume 243. Springer Science & Business Media, 2013.
- [101] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901, 2020.
- [102] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
- [103] Jan C Buckner. Factors influencing survival in high-grade gliomas. In *Seminars in oncology*, volume 30, pages 10–14. Elsevier, 2003.
- [104] 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.

- [105] Henriett Butz and Attila Patócs. Brief summary of the most important molecular genetic methods (pcr, qpcr, microarray, next-generation sequencing, etc.). *Genetics of Endocrine diseases and syndromes*, pages 33–52, 2019.
- [106] Dmitrii Bychkov, Riku Turkki, Caj Haglund, Nina Linder, and Johan Lundin. Deep learning for tissue microarray image-based outcome prediction in patients with colorectal cancer. In *Medical Imaging 2016: Digital Pathology*, volume 9791, pages 298–303. SPIE, 2016.
- [107] Daniel E Cahall, Ghulam Rasool, Nidhal C Bouaynaya, and Hassan M Fathallah-Shaykh. Inception modules enhance brain tumor segmentation. *Frontiers in computational neuroscience*, 13:44, 2019.
- [108] Zhaowei Cai, Gukyeong Kwon, Avinash Ravichandran, Erhan Bas, Zhuowen Tu, Rahul Bhotika, and Stefano Soatto. X-DETR: A Versatile Architecture For Instance-Wise Vision-Language Tasks. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVI*, page 290–308, Berlin, Heidelberg, 2022. Springer-Verlag.
- [109] Minal Çalışkan and Koichi Tazaki. Ai/ml advances in non-small cell lung cancer biomarker discovery. *Frontiers in Oncology*, 13, 2023.
- [110] Joshua D Campbell, Christina Yau, Reanne Bowlby, Yuexin Liu, Kevin Brennan, Huihui Fan, Alison M Taylor, Chen Wang, Vonn Walter, Rehan Akbani, et al. Genomic, pathway network, and immunologic features distinguishing squamous carcinomas. *Cell reports*, 23(1):194–212, 2018.
- [111] Rui Cao, Fan Yang, Si-Cong Ma, Li Liu, Yu Zhao, Yan Li, De-Hua Wu, Tongxin Wang, Wei-Jia Lu, Wei-Jing Cai, et al. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer. *Theranostics*, 10(24):11080, 2020.

- [112] Zhi-Jie Cao and Ge Gao. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature Biotechnology*, 40(10):1458–1466, 2022.
- [113] Giuseppina Carannante, Dimah Dera, Nidhal C Bouaynaya, Hassan M Fathallah-Shaykh, and Ghulam Rasool. Trustworthy Medical Segmentation with Uncertainty Estimation. *arXiv preprint arXiv:2111.05978*, 2021. <https://arxiv.org/abs/2111.05978>.
- [114] Giuseppina Carannante, Dimah Dera, Ghulam Rasool, and Nidhal C Bouaynaya. Self-Compression in Bayesian Neural Networks. In *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2020.
- [115] Nicholas Carlini, Anish Athalye, Nicolas Papernot, Wieland Brendel, Jonas Rauber, Dimitris Tsipras, Ian Goodfellow, Aleksander Madry, and Alexey Kurakin. On evaluating adversarial robustness. *arXiv preprint arXiv:1902.06705*, 2019.
- [116] Nicholas Carlini and David A. Wagner. Towards evaluating the robustness of neural networks. In IEEE, editor, *2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA, May 22-26, 2017*, pages 39–57, 2017.
- [117] CBTR-US. Central brain tumor registry of the us. <https://cbtrus.org/>, 2020.
- [118] CDC. Ncpr. <https://www.cdc.gov/cancer/npcr/index.htm>, 2020.
- [119] CDISC. Clinical Data Interchange Standards Consortium. Available online: <https://www.cdisc.org/>. (accessed on 1 December 2023).
- [120] Moffitt Cancer Center. Squamous cell carcinoma survival rate, June 2024. Available at: <https://www.moffitt.org/cancers/squamous-cell-carcinoma/survival-rate/>. Last accessed on Jun 2, 2024.

- [121] Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J Byrne, Michael L Heuer, Erik Larsson, et al. The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*, 2(5):401–404, 2012.
- [122] Heang-Ping Chan, Lubomir M. Hadjiiski, and Ravi K. Samala. Computer-aided diagnosis in the era of deep learning. *Medical Physics*, 47(5):e218–e227, 2020.
- [123] Chun-Hung Chao, Zhuotun Zhu, Dazhou Guo, Ke Yan, Tsung-Ying Ho, Jinzheng Cai, Adam P Harrison, Xianghua Ye, Jing Xiao, Alan Yuille, et al. Lymph node gross tumor volume detection in oncology imaging via relationship learning using graph neural network. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23*, pages 772–782. Springer, 2020.
- [124] Kasit Chatsirisupachai, Tom Lesluyes, Luminita Paraoan, Peter Van Loo, and João Pedro De Magalhães. An integrative analysis of the age-associated multi-omic landscape across cancers. *Nature communications*, 12(1):2345, 2021.
- [125] Michail Chatzianastasis, Michalis Vazirgiannis, and Zijun Zhang. Explainable multilayer graph neural network for cancer gene prediction. *arXiv preprint arXiv:2301.08831*, 2023.
- [126] Anika Cheerla and Olivier Gevaert. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics*, 35(14):i446–i454, 2019.
- [127] Eduard Chelebian, Christophe Avenel, Kimmo Kartasalo, Maja Marklund, Anna Tanoglidi, Tuomas Mirtti, Richard Colling, Andrew Erickson, Alastair D Lamb, Joakim Lundeberg, et al. Morphological features extracted by ai associated with spatial transcriptomics in prostate cancer. *Cancers*, 13(19):4837, 2021.

- [128] Borui Chen, Jing Jin, Haichao Liu, Zhengyu Yang, Haoming Zhu, Yu Wang, Jianping Lin, Shizhong Wang, and Shaoqing Chen. Trends and hotspots in research on medical images with deep learning: a bibliometric analysis from 2013 to 2023. *Frontiers in Artificial Intelligence*, 6:1289669, 2023.
- [129] Feng Chen, Michael C Wendl, Matthew A Wyczalkowski, Matthew H Bailey, Yize Li, and Li Ding. Moving pan-cancer studies from basic research toward the clinic. *Nature cancer*, 2(9):879–890, 2021.
- [130] Jiayang Chen, Zhihang Hu, Siqi Sun, Qingxiong Tan, Yixuan Wang, Qinze Yu, Licheng Zong, Liang Hong, Jin Xiao, Tao Shen, et al. Interpretable rna foundation model from unannotated data for highly accurate rna structure and function predictions. *arXiv preprint arXiv:2204.00300*, 2022.
- [131] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.
- [132] Mei-Ju May Chen, Jun Li, Yumeng Wang, Rehan Akbani, Yiling Lu, Gordon B Mills, and Han Liang. Tcpa v3. 0: an integrative platform to explore the pan-cancer analysis of functional proteomic data. *Molecular & Cellular Proteomics*, 18(8):S15–S25, 2019.
- [133] Richard J Chen, Chengkuan Chen, Yicong Li, Tiffany Y Chen, Andrew D Trister, Rahul G Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16144–16155, 2022.

- [134] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862, 2024.
- [135] Richard J Chen, Ming Y Lu, Jingwen Wang, Drew FK Williamson, Scott J Rodig, Neal I Lindeman, and Faisal Mahmood. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Transactions on Medical Imaging*, 2020.
- [136] Richard J Chen, Ming Y Lu, Wei-Hung Weng, Tiffany Y Chen, Drew FK Williamson, Trevor Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4025, 2021.
- [137] Richard J. Chen, Ming Y. Lu, Wei-Hung Weng, Tiffany Y. Chen, Drew FK. Williamson, Trevor Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3995–4005, 2021.
- [138] Richard J Chen, Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Jana Lipkova, Zahra Noor, Muhammad Shaban, Maha Shady, Mane Williams, Bumjin Joo, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell*, 40(8):865–878, 2022.
- [139] Yongxin Chen, Tryphon Georgiou, Michele Pavon, and Allen Tannenbaum. Robust transport over networks. *IEEE transactions on automatic control*, 62(9):4675–4682, 2016.



- [140] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask Transformer for universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1290–1299, 2022.
- [141] Travers Ching. Cox regression. <http://traversc.github.io/cox-nnet/docs/>, 2024. Accessed: 2024-05-13.
- [142] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, and Jimeng Sun. GRAM: graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 787–795, 2017.
- [143] François Chollet et al. Keras. <https://keras.io>, 2015.
- [144] Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2021.
- [145] Andy Chu, Gordon Robertson, Denise Brooks, Andrew J Mungall, Inanc Birol, Robin Coope, Yussanne Ma, Steven Jones, and Marco A Marra. Large-scale profiling of micrnas for the cancer genome atlas. *Nucleic acids research*, 44(1):e3–e3, 2016.
- [146] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
- [147] Didem Cifci, Sebastian Foersch, and Jakob Nikolas Kather. Artificial intelligence to identify genetic alterations in conventional histopathology. *The Journal of Pathology*, 257(4):430–444, 2022.

- [148] Didem Cifci, Gregory P Veldhuizen, Sebastian Foersch, and Jakob Nikolas Kather. AI in Computational Pathology of Cancer: Improving Diagnostic Workflows and Clinical Outcomes? *Annual Review of Cancer Biology*, 7, 2023.
- [149] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging*, 26:1045–1057, 2013.
- [150] David A Clunie, Don K Dennison, Dawn Cram, Kenneth R Persons, Mark D Bronkalla, and Henri “Rik” Primo. Technical challenges of enterprise imaging: HIMSS-SIIM collaborative white paper. *Journal of digital imaging*, 29:583–614, 2016.
- [151] Coalition for Health AI. Blueprint for Trustworthy AI Implementation Guidance and Assurance for Healthcare, 2023. available at: [https://coalitionforhealthai.org/papers/blueprint-for-trustworthy-ai\\_V1.0.pdf](https://coalitionforhealthai.org/papers/blueprint-for-trustworthy-ai_V1.0.pdf). Last accessed on Jul 31, 2023.
- [152] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.
- [153] Common Crawl. Common crawl, 2023. Available online: <https://commoncrawl.org/>. (accessed on 18 September 2023).
- [154] Genomic Data Commons. File format - maf. [https://docs.gdc.cancer.gov/Data/File\\_Formats/MAF\\_Format/](https://docs.gdc.cancer.gov/Data/File_Formats/MAF_Format/), 2024. Accessed: 2024-05-13.
- [155] Genomic Data Commons. File format - vcf. [https://docs.gdc.cancer.gov/Data/File\\_Formats/VCF\\_Format/](https://docs.gdc.cancer.gov/Data/File_Formats/VCF_Format/), 2024. Accessed: 2024-05-13.
- [156] Genomic Data Commons. Mutation annotation format. [https://docs.gdc.cancer.gov/Encyclopedia/pages/Mutation\\_Annotation\\_Format/](https://docs.gdc.cancer.gov/Encyclopedia/pages/Mutation_Annotation_Format/), 2024. Accessed: 2024-05-13.

- [157] United States Congress. Health insurance portability and accountability act of 1996. <https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf>, 1996.
- [158] Luis A Corchete, Elizabeta A Rojas, Diego Alonso-López, Javier De Las Rivas, Norma C Gutiérrez, and Francisco J Burguillo. Systematic comparison and assessment of rna-seq procedures for gene expression quantitative analysis. *Scientific reports*, 10(1):19737, 2020.
- [159] David J Cote, Quinn T Ostrom, Haley Gittleman, Kelsey R Duncan, Travis S Creve-Coeur, Carol Kruchko, Timothy R Smith, Meir J Stampfer, and Jill S Barnholtz-Sloan. Glioma incidence and survival variations by county-level socioeconomic measures. *Cancer*, 125(19):3390–3400, 2019.
- [160] Daniel Cremers, Mikael Rousson, and Rachid Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International journal of computer vision*, 72(2):195–215, 2007.
- [161] Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, Yang Zhou, Kaizhao Liang, Jintai Chen, Juanwu Lu, Zichong Yang, Kuei-Da Liao, et al. A survey on multimodal large language models for autonomous driving. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 958–979, 2024.
- [162] Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods*, pages 1–11, 2024.
- [163] Hui Cui, Ping Xuan, Qiangguo Jin, Mingjun Ding, Butuo Li, Bing Zou, Yiyue Xu, Bingjie Fan, Wanlong Li, Jinming Yu, et al. Co-graph attention reasoning based imaging and clinical features integration for lymph node metastasis prediction. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pages 657–666. Springer, 2021.

- [164] Miao Cui and David Y Zhang. Artificial intelligence and computational pathology. *Laboratory Investigation*, 101(4):412–422, 2021.
- [165] Yufei CUI, Ziquan Liu, Xiangyu Liu, Xue Liu, Cong Wang, Tei-Wei Kuo, Chun Jason Xue, and Antoni B. Chan. Bayes-MIL: A New Probabilistic Perspective on Attention-based Multiple Instance Learning for Whole Slide Images. In *The Eleventh International Conference on Learning Representations*, 2023.
- [166] Suresh Dara and Priyanka Tumma. Feature extraction by using deep learning: A survey. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2018.
- [167] Cameron Davidson-Pilon. lifelines, survival analysis in python. <https://doi.org/10.5281/zenodo.10456828>, Jan 2024. Accessed: 2024-05-13.
- [168] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.
- [169] Demetra Demetriou, Rodney Hull, Mmamolella Kgoebane-Maseko, Zarina Lockhat, and Zodwa Dlamini. AI-Enhanced Digital Pathology and Radiogenomics in Precision Oncology. In *Artificial Intelligence and Precision Oncology: Bridging Cancer Research and Clinical Decision Support*, pages 93–113. Springer, 2023.
- [170] Lloyd Demetrius, Volker Matthias Gundlach, and Gunter Ochs. Complexity and demographic stability in population models. *Theoretical population biology*, 65(3):211–225, 2004.
- [171] Lloyd A Demetrius. Boltzmann, darwin and directionality theory. *Physics reports*, 530(1):1–85, 2013.

- [172] Lloyd A Demetrius and Christian Wolf. Directionality theory and the second law of thermodynamics. *Physica A: Statistical Mechanics and its Applications*, 598:127325, 2022.
- [173] D. Dera, G. Rasool, and N. Bouaynaya. Extended variational inference for propagating uncertainty in convolutional neural networks. In *IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, Oct 2019.
- [174] Dimah Dera, Sabeen Ahmed, Nidhal C Bouaynaya, and Ghulam Rasool. Trustworthy uncertainty propagation for sequential time-series analysis in rnns. *IEEE Transactions on Knowledge and Data Engineering*, 2023.
- [175] Dimah Dera, Nidhal Bouaynaya, Ghulam Rasool, R Shterenberg, and Hassan Fathallah-Shaykh. PremiUm-CNN: Propagating Uncertainty Towards Robust Convolutional Neural Networks. *IEEE Transactions on Signal Processing*, 2021.
- [176] Dimah Dera, Ghulam Rasool, and Nidhal Bouaynaya. Extended Variational Inference for Propagating Uncertainty in Convolutional Neural Networks. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2019.
- [177] Dimah Dera, Ghulam Rasool, Nidhal C Bouaynaya, Adam Eichen, Stephen Shanko, Jeff Cammerata, and Sanipa Arnold. Bayes-SAR net: Robust SAR image classification with uncertainty estimation using bayesian convolutional neural network. In *2020 IEEE International Radar Conference (RADAR)*, pages 362–367. IEEE, 2020.
- [178] Austin Derrow-Pinion, Jennifer She, David Wong, Oliver Lange, Todd Hester, Luis Perez, Marc Nunkesser, Seongjae Lee, Xueying Guo, Brett Wiltshire, et al. ETA prediction with graph neural networks in google maps. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3767–3776, 2021.

- [179] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics, June 2019.
- [180] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics*, 2019.
- [181] Stephanie Dick. Artificial Intelligence. *Harvard Data Science Review*, 1(1), jul 1 2019.
- [182] DICOM. Digital imaging and communications in medicine. <https://www.dicomstandard.org/about-home>.
- [183] Yadong Ding, Yu Wu, Chengyue Huang, Siliang Tang, Fei Wu, Yi Yang, Wenwu Zhu, and Yueting Zhuang. Nap: Neural architecture search with pruning. *Neurocomputing*, 477:85–95, 2022.
- [184] Alara Dirik and Sayak Paul. A Dive into Vision-Language Models, 2023.
- [185] Manfredo Perdigao Do Carmo and J Flaherty Francis. *Riemannian geometry*, volume 6. Springer, 1992.
- [186] GDC Documentation. Reverse phase protein array. [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/RPPA\\_intro/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/RPPA_intro/), 2024. Accessed: 2024-05-13.
- [187] Robert H Dolin, Liora Alschuler, Calvin Beebe, Paul V Biron, Sandra Lee Boyer, Daniel Essin, Elliot Kimber, Tom Lincoln, and John E Mattison. The hl7 clinical document architecture. *Journal of the American Medical Informatics Association*, 8(6):552–569, 2001.

- [188] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *annual conference on medical image understanding and analysis*, pages 506–517. Springer, 2017.
- [189] Xuanyi Dong and Yi Yang. Nas-bench-201: Extending the scope of reproducible neural architecture search. *arXiv preprint arXiv:2001.00326*, 2020.
- [190] David A Dorr, Laura Adams, and Peter Embí. Harnessing the promise of artificial intelligence responsibly. *JAMA*, 329(16):1347–1348, 2023.
- [191] 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.
- [192] 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.
- [193] 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*, 2021.
- [194] S Ben Driss, Mahmoud Soua, Rostom Kachouri, and Mohamed Akil. A comparison study between mlp and convolutional neural network models for character recognition. In SPIE, editor, *Real-Time Image and Video Processing 2017*, volume 10223, page 1022306. International Society for Optics and Photonics, 2017.

- [195] Jojanneke Drogdt, Megan Milota, Shoko Vos, Annelien Bredenoord, and Karin Jongsma. Integrating artificial intelligence in pathology: a qualitative interview study of users' experiences and expectations. *Modern Pathology*, 35(11):1540–1550, 2022.
- [196] Hao Du, Jiashi Feng, and Mengling Feng. Zoom in to where it matters: a hierarchical graph based model for mammogram analysis. *arXiv preprint arXiv:1912.07517*, 2019.
- [197] Pan Du, Xiao Zhang, Chiang-Ching Huang, Nadereh Jafari, Warren A Kibbe, Lifang Hou, and Simon M Lin. Comparison of beta-value and m-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics*, 11:1–9, 2010.
- [198] Vincent Dumoulin and Francesco Visin. Convolution arithmetic. [https://www.github.com/vdumoulin/conv\\_arithmetic](https://www.github.com/vdumoulin/conv_arithmetic).
- [199] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016.
- [200] Amelie Echle, Niklas Timon Rindtorff, Titus Josef Brinker, Tom Luedde, Alexander Thomas Pearson, and Jakob Nikolas Kather. Deep learning in cancer pathology: a new generation of clinical biomarkers. *British Journal of Cancer*, 124(4):686–696, 2021.
- [201] Deepak Chowdary Edara, Lakshmi Prasanna Vanukuri, Venkatramaphanikumar Sistla, and Venkata Krishna Kishore Kolli. Sentiment analysis and text categorization of cancer medical records with LSTM. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–17, 2019.
- [202] Yasha Ektefaie, George Dasoulas, Ayush Noori, Maha Farhat, and Marinka Zitnik. Geometric multimodal representation learning. *arXiv preprint arXiv:2209.03299*, 2023.
- [203] Yasha Ektefaie, George Dasoulas, Ayush Noori, Maha Farhat, and Marinka Zitnik. Multimodal learning with graphs. *Nature Machine Intelligence*, 5(4):340–350, 2023.



- [204] Issam El Naqa and Dana Rollison. Moffitt Cancer Center: Why we are building the first machine learning department in oncology. *The Cancer Letter*, 47(7):5–10, 2021.
- [205] Jacob G Ellen, Etai Jacob, Nikos Nikolaou, and Natasha Markuzon. Autoencoder-based multimodal prediction of non-small cell lung cancer survival. *Scientific Reports*, 13(1):15761, 2023.
- [206] Matthew J. Ellis, Michael Gillette, Steven A. Carr, Amanda G. Paulovich, Richard D. Smith, Karin K. Rodland, R. Reid Townsend, Christopher Kinsinger, Mehdi Mesri, Henry Rodriguez, and Daniel C. Liebler. Connecting Genomic Alterations to Cancer Biology with Proteomics: The NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discovery*, 3(10):1108–1112, 10 2013.
- [207] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural architecture search: A survey. *The Journal of Machine Learning Research*, 20(1):1997–2017, 2019.
- [208] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.
- [209] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639):115–118, 2017.
- [210] Kianoush Falahkheirkhah, Saumya Tiwari, Kevin Yeh, Sounak Gupta, Loren Herrera-Hernandez, Michael R McCarthy, Rafael E Jimenez, John C Cheville, and Rohit Bhargava. Deepfake Histologic Images for Enhancing Digital Pathology. *Laboratory Investigation*, 103(1), 2023.
- [211] Hamza Farooq, Yongxin Chen, Tryphon T Georgiou, Allen Tannenbaum, and Christophe Lenglet. Network curvature as a hallmark of brain structural connectivity. *Nature communications*, 10(1):1–11, 2019.

- [212] Hamza Farooq, Christophe Lenglet, and Flavia Nelson. Robustness of brain structural networks is affected in cognitively impaired ms patients. *Frontiers in neurology*, 11:1542, 2020.
- [213] Leonard Fass. Imaging and cancer: a review. *Molecular oncology*, 2(2):115–152, 2008.
- [214] Hassan M. Fathallah-Shaykh, Andrew DeAtkine, Elizabeth Coffee, Elias Khayat, Asim K. Bag, Xiaosi Han, Paula Province Warren, Markus Bredel, John Fiveash, James Markert, Nidhal Bouaynaya, and Louis B. Nabors. Diagnosing growth in low-grade gliomas with and without longitudinal volume measurements: A retrospective observational study. *PLOS Medicine*, 16(5):1–16, 05 2019.
- [215] FDA. FDA Approved (AI/ML)-Enabled Medical Devices, 2022. Available at: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>. Last accessed on Jul 20, 2023.
- [216] Feature-engine. Feature-engine, a python library for feature engineering and selection. <https://feature-engine.trainindata.com/en/latest/index.html>.
- [217] A. Fedorov, W. Longabaugh, D. Pot, D. Clunie, S. Pieper, R. Lewis, H. Aerts, A. Homeyer, M. Herrmann, U. Wagner, T. Pihl, K. Farahani, and R. Kikinis. Nci imaging data commons. *International Journal of Radiation Oncology\*Biophysics*, 111(3, Supplement):e101, 2021. 2021 Proceedings of the ASTRO 63rd Annual Meeting.
- [218] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2):167–181, 2004.

- [219] Lili Feng, Zhenyu Liu, Chaofeng Li, Zhenhui Li, Xiaoying Lou, Lizhi Shao, Yunlong Wang, Yan Huang, Haiyang Chen, Xiaolin Pang, et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *The Lancet Digital Health*, 4(1):e8–e17, 2022.
- [220] Yue-Hua Feng, Shao-Wu Zhang, and Jian-Yu Shi. DPDDI: a deep predictor for drug-drug interactions. *BMC bioinformatics*, 21:1–15, 2020.
- [221] Kristen L. Fessele. The rise of big data in oncology. *Seminars in Oncology Nursing*, 34(2):168–176, 2018. Technology in Cancer Care.
- [222] Samuel G Finlayson, John D Bowers, Joichi Ito, Jonathan L Zittrain, Andrew L Beam, and Isaac S Kohane. Adversarial attacks on medical machine learning. *Science*, 363(6433):1287–1289, 2019.
- [223] David Flatow and Daniel Penner. On the robustness of convnets to training on noisy labels. *Technical report, Stanford University*, 2017.
- [224] National Institute for Health and Care Excellence (Great Britain). *Brain tumours (primary) and brain metastases in adults*. National Institute for Health and Care Excellence, 2018.
- [225] Dave Fornell. FDA Cleared AI Algorithms. <https://healthexec.com/topics/artificial-intelligence/fda-has-now-cleared-more-500-healthcare-ai-algorithms>, 2023.
- [226] Alex Fout, Jonathon Byrd, Basir Shariat, and Asa Ben-Hur. Protein interface prediction using graph convolutional networks. *Advances in neural information processing systems*, 30, 2017.
- [227] Scott Freitas, Diyi Yang, Srijan Kumar, Hanghang Tong, and Duen Horng Chau. Graph vulnerability and robustness: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

- [228] Cornelius Fritz, Emilio Dorigatti, and David Rügamer. Combining graph neural networks and spatio-temporal disease models to improve the prediction of weekly covid-19 cases in germany. *Scientific Reports*, 12(1):3930, 2022.
- [229] Xiao Fu, Erik Sahai, and Anna Wilkins. Application of digital pathology-based advanced analytics of tumour microenvironment organisation to predict prognosis and therapeutic response. *The Journal of Pathology*, 260(5):578–591, 2023.
- [230] Yu Fu, Alexander W Jung, Ramon Viñas Torne, Santiago Gonzalez, Harald Vöhringer, Artem Shmatko, Lucy R Yates, Mercedes Jimenez-Linan, Luiza Moore, and Moritz Gerstung. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nature Cancer*, 1(8):800–810, 2020.
- [231] Kunihiko Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pages 267–285. Springer, 1982.
- [232] William K Funkhouser. Pathology: the clinical description of human disease. In *Essential Concepts in Molecular Pathology*, pages 177–190. Elsevier, 2020.
- [233] Andreas Futschik, Thomas Hotz, Axel Munk, and Hannes Sieling. Multiscale dna partitioning: statistical evidence for segments. *Bioinformatics*, 30(16):2255–2262, 2014.
- [234] Michael Gadermayr and Maximilian Tschuchnig. Multiple instance learning for digital pathology: A review of the state-of-the-art, limitations & future potential. *Computerized Medical Imaging and Graphics*, page 102337, 2024.
- [235] Andrea Galassi, Marco Lippi, and Paolo Torrioni. Attention in Natural Language Processing. *IEEE Transactions on Neural Networks and Learning Systems*, 32(10):4291–4308, 2021.

- [236] Zhe Gan, Linjie Li, Chunyuan Li, Lijuan Wang, Zicheng Liu, Jianfeng Gao, et al. Vision-language pre-training: Basics, recent advances, and future trends. *Foundations and Trends in Computer Graphics and Vision*, 14(3–4):163–352, 2022.
- [237] Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, Erik Larsson, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cbioportal. *Science signaling*, 6(269):p11–p11, 2013.
- [238] Jianliang Gao, Tengfei Lyu, Fan Xiong, Jianxin Wang, Weimao Ke, and Zhao Li. MGNN: A multimodal graph neural network for predicting the survival of cancer patients. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1697–1700, 2020.
- [239] Yang Gao, Jeff M Phillips, Yan Zheng, Renqiang Min, P Thomas Fletcher, and Guido Gerig. Fully convolutional structured LSTM networks for joint 4D medical image segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, pages 1104–1108. IEEE, 2018.
- [240] Narmin Ghaffari Laleh, Marta Ligeró, Raquel Perez-Lopez, and Jakob Nikolas Kather. Facts and hopes on the use of artificial intelligence for predictive immunotherapy biomarkers in cancer. *Clinical Cancer Research*, 29(2):316–323, 2023.
- [241] Narmin Ghaffari Laleh, Daniel Truhn, Gregory Patrick Veldhuizen, Tianyu Han, Marko van Treeck, Roman D Buelow, Rupert Langer, Bastian Dislich, Peter Boor, Volkmar Schulz, et al. Adversarial attacks and adversarial robustness in computational pathology. *Nature Communications*, 13(1):5711, 2022.

- [242] Golnaz Ghiasi, Xiuye Gu, Yin Cui, and Tsung-Yi Lin. Scaling Open-Vocabulary Image Segmentation With Image-Level Labels. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVI*, page 540–557, Berlin, Heidelberg, 2022. Springer-Verlag.
- [243] Blake A Gibson, Elizabeth McKinnon, Rex C Bentley, Jeffrey Mohlman, Benjamin L Witt, Eric J Yang, Daniel Geisler, and Marie DeFrances. Communicating Certainty in Pathology Reports Interpretation Differences Among Staff Pathologists, Clinicians, and Residents in a Multicenter Study. *Archives of Pathology & Laboratory Medicine*, 146(7):886–893, 2022.
- [244] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.
- [245] Fantin Girard, Conrad Kavalec, and Farida Cheriet. Joint segmentation and classification of retinal arteries/veins from fundus images. *Artificial intelligence in medicine*, 94:96–109, 2019.
- [246] Karan Goel, Nazneen Fatema Rajani, Jesse Vig, Zachary Taschdjian, Mohit Bansal, and Christopher Ré. Robustness gym: Unifying the NLP evaluation landscape. In Avi Sil and Xi Victoria Lin, editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 42–55. Association for Computational Linguistics, 2021.
- [247] S Larry Goldenberg, Guy Nir, and Septimiu E Salcudean. A new era: artificial intelligence and machine learning in prostate cancer. *Nature Reviews Urology*, 16(7):391–403, 2019.

- [248] Mary Goldman, Brian Craft, Mim Hastie, Kristupas Repečka, Fran McDade, Akhil Kamath, Ayan Banerjee, Yunhai Luo, Dave Rogers, Angela N Brooks, et al. The ucsc xena platform for public and private cancer genomics data visualization and interpretation.  *biorxiv*, page 326470, 2018.
- [249] Mary J Goldman, Brian Craft, Mim Hastie, Kristupas Repečka, Fran McDade, Akhil Kamath, Ayan Banerjee, Yunhai Luo, Dave Rogers, Angela N Brooks, et al. Visualizing and interpreting cancer genomics data via the xena platform.  *Nature biotechnology*, 38(6):675–678, 2020.
- [250] Augusto Gonzalez, Dario A Leon, Yasser Perera, and Rolando Perez. On the gene expression landscape of cancer.  *Plos one*, 18(2):e0277786, 2023.
- [251] Carlos Vladimiro Gonzalez Zelaya. Towards explaining the effects of data preprocessing on machine learning. In  *IEEE 35th International Conference on Data Engineering (ICDE)*, pages 2086–2090, 2019.
- [252] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning book.  *MIT Press*, 521(7553):800, 2016.
- [253] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In  *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [254] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples.  *arXiv preprint arXiv:1412.6572*, 2014.
- [255] Google. Gpus on compute engine. <https://cloud.google.com/compute/docs/gpus>.
- [256] Stephen Gould, Tianshi Gao, and Daphne Koller. Region-based segmentation and object detection. In  *Advances in neural information processing systems*, pages 655–663, 2009.

- [257] Roberto Gozalo-Brizuela and Eduardo C Garrido-Merchán. A survey of generative ai applications. *arXiv preprint arXiv:2306.02781*, 2023.
- [258] Antonietta Gerarda Gravina, Raffaele Pellegrino, Giovanna Palladino, Giuseppe Imperio, Andrea Ventura, and Alessandro Federico. Charting new ai education in gastroenterology: cross-sectional evaluation of chatgpt and perplexity ai in medical residency exam. *Digestive and Liver Disease*, 2024.
- [259] Robert L. Grossman, Allison P. Heath, Vincent Ferretti, Harold E. Varmus, Douglas R. Lowy, Warren A. Kibbe, and Louis M. Staudt. Toward a shared vision for cancer genomic data. *New England Journal of Medicine*, 375(12):1109–1112, 2016.
- [260] DEKON Group. Medical imaging computing and computerassisted intervention. <https://miccai2020.org/en/>.
- [261] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377, 2018.
- [262] Xiuye Gu, Tsung-Yi Lin, Weicheng Kuo, and Yin Cui. Open-vocabulary Object Detection via Vision and Language Knowledge Distillation. In *International Conference on Learning Representations*, 2022.
- [263] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.
- [264] Ruoyu Guo, Kunzi Xie, Maurice Pagnucco, and Yang Song. SAC-Net: Learning with weak and noisy labels in histopathology image segmentation. *Medical Image Analysis*, page 102790, 2023.



- [265] Rajarsi Gupta, Tahsin Kurc, Ashish Sharma, Jonas S Almeida, and Joel Saltz. The emergence of pathomics. *Current Pathobiology Reports*, 7:73–84, 2019.
- [266] 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.
- [267] Tatyana S Gurina and Lary Simms. Histology, staining. In *StatPearls*. StatPearls Publishing, 2022.
- [268] David A Gutman, Mohammed Khalilia, Sanghoon Lee, Michael Nalisnik, Zach Mullen, Jonathan Beezley, Deepak R Chittajallu, David Manthey, and Lee AD Cooper. The digital slide archive: A software platform for management, integration, and analysis of histology for cancer research. *Cancer research*, 77(21):e75–e78, 2017.
- [269] Miriam Hägele, Philipp Seegerer, Sebastian Lapuschkin, Michael Bockmayr, Wojciech Samek, Frederick Klauschen, Klaus-Robert Müller, and Alexander Binder. Resolving challenges in deep learning-based analyses of histopathological images using explanation methods. *Scientific reports*, 10(1):6423, 2020.
- [270] Benjamin Haibe-Kains, George Alexandru Adam, Ahmed Hosny, Farnoosh Khodakarami, Thakkar Shraddha, Rebecca Kusko, Susanna-Assunta Sansone, Weida Tong, Russ D. Wolfinger, Christopher E. Mason, Wendell Jones, Joaquin Dopazo, Cesare Furlanello, Levi Waldron, Bo Wang, Chris McIntosh, Anna Goldenberg, Anshul Kundaje, Casey S. Greene, Tamara Broderick, Michael M. Hoffman, Jeffrey T. Leek, Keegan Korthauer, Wolfgang Huber, Alvis Brazma, Joelle Pineau, Robert Tibshirani, Trevor Hastie, John P. A. Ioannidis, John Quackenbush, and Hugo J. W. L. Aerts and. Transparency and reproducibility in artificial intelligence. *Nature*, 586(7829):E14–E16, 2020.
- [271] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

- [272] Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, Zhaohui Yang, Yiman Zhang, and Dacheng Tao. A Survey on Vision Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):87–110, 2023.
- [273] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011.
- [274] Matthew G Hanna, Niels H Olson, Mark Zarella, Rajesh C Dash, Markus D Herrmann, Larissa V Furtado, Michelle N Stram, Patricia M Raciti, Lewis Hassell, Alex Mays, et al. Recommendations for performance evaluation of machine learning in pathology: A concept paper from the college of american pathologists. *Archives of Pathology & Laboratory Medicine*, 2023.
- [275] Robert M Haralick and Linda G Shapiro. Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1):100–132, 1985.
- [276] Iryna Hartsock and Ghulam Rasool. Vision-language models for medical report generation and visual question answering: A review. *arXiv preprint arXiv:2403.02469*, 2024.
- [277] Arman Hasanzadeh, Ehsan Hajiramezanali, Nick Duffield, and Xiaoning Qian. MoReL: Multi-omics Relational Learning. *arXiv preprint arXiv:2203.08149*, 2022.
- [278] Yehudit Hasin, Marcus Seldin, and Aldons Lusic. Multi-omics approaches to disease. *Genome biology*, 18:1–15, 2017.
- [279] Lewis A Hassell, Syeda Fatima Absar, Chhavi Chauhan, Suzanne Dintzis, Carol F Farver, Samreen Fathima, Eric F Glassy, Jeffery A Goldstein, Rama Gullapalli, Jonhan Ho, et al. Pathology education powered by virtual and digital transformation. *Arch Pathol Lab Med*, 2022.

- [280] Mohammad Havaei et al. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017.
- [281] HDR-UK. The health data research hub for cancer. <https://www.hdruk.ac.uk/help-with-your-data/our-hubs-across-the-uk/data-can/>, 2020.
- [282] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015. (Santiago, Chile).
- [283] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In IEEE, editor, *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778, 2016.
- [284] 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.
- [285] Xiujing He, Xiaowei Liu, Fengli Zuo, Hubing Shi, and Jing Jing. Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. In *Seminars in Cancer Biology*, volume 88, pages 187–200. Elsevier, 2023.
- [286] Sara Hijazo-Pechero, Ania Alay, Raúl Marín, Noelia Vilariño, Cristina Muñoz-Pinedo, Alberto Villanueva, David Santamaría, Ernest Nadal, and Xavier Solé. Gene expression profiling as a potential tool for precision oncology in non-small cell lung cancer. *Cancers*, 13(19):4734, 2021.
- [287] Khan Hikmat, Rasool Ghulam, B Nidhal, CT Tyler, and Thompson Lacey. Deep ensemble for rotorcraft attitude prediction. In *Vertical Flight Society's 77th Annual Forum Technology Display*, 2021.

- [288] Izumi V. Hinkson, Tanja M. Davidsen, Juli D. Klemm, Ishwar Chandramouliswaran, Anthony R. Kerlavage, and Warren A. Kibbe. A comprehensive infrastructure for big data in cancer research: Accelerating cancer research and precision medicine. *Frontiers in Cell and Developmental Biology*, 5, 2017.
- [289] HL7. HL7 FHIR. Available online: <https://www.hl7.org/fhir/>. (accessed on 1 December 2023).
- [290] Katherine A Hoadley, Christina Yau, Toshinori Hinoue, Denise M Wolf, Alexander J Lazar, Esther Drill, Ronglai Shen, Alison M Taylor, Andrew D Cherniack, Vésteinn Thorsson, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell*, 173(2):291–304, 2018.
- [291] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [292] Eric C Holland. Progenitor cells and glioma formation. *Current opinion in neurology*, 14(6):683–688, 2001.
- [293] 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.
- [294] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification, 2018.
- [295] Arkar Htet, Sui Reng Liana, Theingi Aung, and Amiya Bhaumik. Chatgpt in content creation: Techniques, applications, and ethical implications. In *Advanced Applications of Generative AI and Natural Language Processing Models*, pages 43–68. IGI Global, 2024.

- [296] Ronghang Hu and Amanpreet Singh. Unit: Multimodal multitask learning with a unified transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1439–1449, 2021.
- [297] Zheng Hu, Zan Li, Zhicheng Ma, and Christina Curtis. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nature genetics*, 52(7):701–708, 2020.
- [298] 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.
- [299] Shih-Cheng Huang, Anuj Pareek, Malte Jensen, Matthew P Lungren, Serena Yeung, and Akshay S Chaudhari. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. *NPJ Digital Medicine*, 6(1):74, 2023.
- [300] Yongxiang Huang and Albert CS Chung. Edge-variational graph convolutional networks for uncertainty-aware disease prediction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VII 23*, pages 562–572. Springer, 2020.
- [301] Yu Huang, Chenzhuang Du, Zihui Xue, Xuanyao Chen, Hang Zhao, and Longbo Huang. What makes multi-modal learning better than single (provably). In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.
- [302] Yu Huang, Junyang Lin, Chang Zhou, Hongxia Yang, and Longbo Huang. Modality competition: What makes joint training of multi-modal network fail in deep learning?(provably). In *International Conference on Machine Learning*, pages 9226–9259. PMLR, 2022.

- [303] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. Ccnet: Criss-cross attention for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 603–612, 2019.
- [304] Drew A. Hudson and Christopher D. Manning. GQA: A New Dataset for Real-World Visual Reasoning and Compositional Question Answering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [305] Drew A Hudson and Larry Zitnick. Generative adversarial transformers. In *International conference on machine learning*, pages 4487–4499. PMLR, 2021.
- [306] Ahmad Ibrahim, Hoda K. Mohamed, Ali Maher, and Baochang Zhang. A survey on human cancer categorization based on deep learning. *Frontiers in Artificial Intelligence*, 5, June 2022.
- [307] Asmaa Ibrahim, Paul Gamble, Ronnachai Jaroensri, Mohammed M Abdelsamea, Craig H Mermel, Po-Hsuan Cameron Chen, and Emad A Rakha. Artificial intelligence in digital breast pathology: techniques and applications. *The Breast*, 49:267–273, 2020.
- [308] Nabil Ibtehaz and M Sohel Rahman. MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. *arXiv [Preprint]*, 2019.
- [309] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018.
- [310] National Cancer Institute. Prostate-specific antigen. <https://www.cancer.gov/types/prostate/psa-fact-sheet/>.
- [311] National Cancer Institute. surveillance. <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/surveillance>.
- [312] National Cancer Institute. Seer program. [https://seer.cancer.gov/registries/cancer\\_registry/index.html](https://seer.cancer.gov/registries/cancer_registry/index.html), 2020.

- [313] Institution Of Analysts & Programmers. Trustworthy Software Foundation, 2023. available at: <https://www.tsfdn.org/>. Last accessed on Jul 31, 2023.
- [314] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [315] 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.
- [316] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge. In *International MICCAI Brainlesion Workshop*, pages 287–297. Springer, 2017.
- [317] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. No New-Net. In *International MICCAI Brainlesion Workshop*, pages 234–244. Springer, 2018. (Granada, Spain).
- [318] Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, et al. Perceiver io: A general architecture for structured inputs & outputs. *arXiv preprint arXiv:2107.14795*, 2021.
- [319] Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, et al. Perceiver IO: A General Architecture for Structured Inputs & Outputs. *International Conference on Learning Representations*, 2022.
- [320] Anil K Jain and Richard C Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.

- [321] Anil K Jain, Robert PW Duin, and Jianchang Mao. Statistical pattern recognition: A review, *IEEE transactions on pattern analysis and machine intelligence*. vol, 22, 2000.
- [322] Anil K Jain, Alexander Topchy, Martin HC Law, and Joachim M Buhmann. Landscape of clustering algorithms. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 1, pages 260–263. IEEE, 2004.
- [323] Jitesh Jain, Jiachen Li, Mang Tik Chiu, Ali Hassani, Nikita Orlov, and Humphrey Shi. Oneformer: One transformer to rule universal image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2989–2998, 2023.
- [324] Camden Jansen, Ricardo N Ramirez, Nicole C El-Ali, David Gomez-Cabrero, Jesper Tegner, Matthias Merckenschlager, Ana Conesa, and Ali Mortazavi. Building gene regulatory networks from scATAC-seq and scRNA-seq using linked self organizing maps. *PLoS computational biology*, 15(11):e1006555, 2019.
- [325] Bruna V Jardim-Perassi, Wei Mu, Suning Huang, Michal R Tomaszewski, Jan Poleszczuk, Mahmoud A Abdalah, Mikalai M Budzevich, William Dominguez-Viqueira, Damon R Reed, Marilyn M Bui, et al. Deep-learning and MR images to target hypoxic habitats with evofosfamide in preclinical models of sarcoma. *Theranostics*, 11(11):5313, 2021.
- [326] Adrián Javaloy, Maryam Meghdadi, and Isabel Valera. Mitigating Modality Collapse in Multimodal VAEs via Impartial Optimization. In *International Conference on Machine Learning*, pages 9938–9964. PMLR, 2022.
- [327] Yanrong Ji, Zhihan Zhou, Han Liu, and Ramana V Davuluri. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. *bioRxiv*, 2020.



- [328] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 4904–4916. PMLR, 18–24 Jul 2021.
- [329] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *International Conference on Machine Learning*, pages 4904–4916. PMLR, 2021.
- [330] Weikuan Jia, Meili Sun, Jian Lian, and Sujuan Hou. Feature dimensionality reduction: a review. *Complex & Intelligent Systems*, 8(3):2663–2693, 2022.
- [331] Hao Jiang, Yanning Zhou, Yi Lin, Ronald C.K. Chan, Jiang Liu, and Hao Chen. Deep learning for computational cytology: A survey. *Medical Image Analysis*, 84:102691, February 2023.
- [332] Peng Jiang, Sanju Sinha, Kenneth Aldape, Sridhar Hannenhalli, Cenk Sahinalp, and Eytan Ruppin. Big Data in basic and translational cancer research. *Nature Reviews Cancer*, 22(11):625–639, 2022.
- [333] Peng Jiang, Sanju Sinha, Kenneth Aldape, Sridhar Hannenhalli, Cenk Sahinalp, and Eytan Ruppin. Big data in basic and translational cancer research. *Nature Reviews Cancer*, 22(11):625–639, 2022.
- [334] Licheng Jiao, Jie Chen, Fang Liu, Shuyuan Yang, Chao You, Xu Liu, Lingling Li, and Biao Hou. Graph representation learning meets computer vision: A survey. *IEEE Transactions on Artificial Intelligence*, 2022.
- [335] Di Jin, Cuiying Huo, Jianwu Dang, Peican Zhu, Weixiong Zhang, Witold Pedrycz, and Lingfei Wu. Heterogeneous Graph Neural Networks using Self-supervised Reciprocally Contrastive Learning. *arXiv preprint arXiv:2205.00256*, 2022.

- [336] Wei Jin, Yaxing Li, Han Xu, Yiqi Wang, Shuiwang Ji, Charu Aggarwal, and Jiliang Tang. Adversarial attacks and defenses on graphs. *ACM SIGKDD Explorations Newsletter*, 22(2):19–34, 2021.
- [337] Weina Jin, Xiaoxiao Li, Mostafa Fatehi, and Ghassan Hamarneh. Guidelines and evaluation of clinical explainable AI in medical image analysis. *Medical Image Analysis*, 84:102684, 2023.
- [338] S Joo, ES Ko, S Kwon, E Jeon, H Jung, JY Kim, MJ Chung, and YH Im. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. *Sci Rep*, 11(1):18800, 2021.
- [339] Zhenlin Ju, Wenbin Liu, Paul L Roebuck, Doris R Siwak, Nianxiang Zhang, Yiling Lu, Michael A Davies, Rehan Akbani, John N Weinstein, Gordon B Mills, et al. Development of a robust classifier for quality control of reverse-phase protein arrays. *Bioinformatics*, 31(6):912–918, 2015.
- [340] Emily Kaczmarek, Amoon Jamzad, Tashifa Imtiaz, Jina Nanayakkara, Neil Renwick, and Parvin Mousavi. Multi-omic graph transformers for cancer classification and interpretation. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022*, pages 373–384. World Scientific, 2021.
- [341] Kaggle. Tiny imagenet. <https://www.kaggle.com/c/tiny-imagenet/overview>, 2021. Accessed: 02-21-2022.
- [342] M Esat Kalfaoglu, Sinan Kalkan, and A Aydin Alatan. Late Temporal Modeling in 3D CNN Architectures with BERT for Action Recognition. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 731–747. Springer, 2020.

- [343] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.
- [344] Mihyun Kang and Zdenek Petrášek. Random graphs: Theory and applications from nature to society to the brain. *Internationale Mathematische Nachrichten*, 227:1–24, 2014.
- [345] Wendi Kang, Xiang Qiu, Yingen Luo, Jianwei Luo, Yang Liu, Junqing Xi, Xiao Li, and Zhengqiang Yang. Application of radiomics-based multiomics combinations in the tumor microenvironment and cancer prognosis. *Journal of Translational Medicine*, 21(1):598, 2023.
- [346] Po-Yu Kao, Thuyen Ngo, Angela Zhang, Jefferson W Chen, and BS Manjunath. Brain Tumor Segmentation and Tractographic Feature Extraction from Structural MR Images for Overall Survival Prediction. In *International MICCAI Brainlesion Workshop*, pages 128–141. Springer, 2018. (Granada, Spain).
- [347] Mohamed Esmail Karar, Denis R Merk, Volkmar Falk, and Oliver Burgert. A simple and accurate method for computer-aided transapical aortic valve replacement. *Computerized Medical Imaging and Graphics*, 50:31–41, 2016.
- [348] Jakob Nikolas Kather, Narmin Ghaffari Laleh, Sebastian Foersch, and Daniel Truhn. Medical domain knowledge in domain-agnostic generative AI. *NPJ Digital Medicine*, 5(1):90, 2022.
- [349] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.
- [350] Poorvi Kaushik, Evan J Molinelli, Martin L Miller, Weiqing Wang, Anil Korkut, Wenbin Liu, Zhenlin Ju, Yiling Lu, Gordon Mills, and Chris Sander. Spatial normalization of reverse phase protein array data. *PloS one*, 9(12):e97213, 2014.

- [351] Auguste Kerckhoffs. La cryptographic militaire. *Journal des sciences militaires*, pages 5–38, 1883.
- [352] Seema Khadirnaikar, Sudhanshu Shukla, and SRM Prasanna. Integration of pan-cancer multi-omics data for novel mixed subgroup identification using machine learning methods. *Plos one*, 18(10):e0287176, 2023.
- [353] Hikmat Khan. Brain-inspired continual learning: Rethinking the role of features in the stability-plasticity dilemma. *Rowan University*, 2024.
- [354] Hikmat Khan. *Brain-Inspired Continual Learning: Rethinking the Role of Features in the Stability-Plasticity Dilemma*. Phd thesis, Rowan University, 2024. Theses and Dissertations. 3212.
- [355] Hikmat Khan, Syed Farhan Alam Zaidi, Asad Safi, and Shahab Ud Din. A comprehensive analysis of mri based brain tumor segmentation using conventional and deep learning methods. In *Intelligent Computing Systems: Third International Symposium, ISICS 2020, Sharjah, United Arab Emirates, March 18–19, 2020, Proceedings 3*, pages 92–104. Springer, 2020.
- [356] Hikmat Khan, Nidhal C. Bouaynaya, and Ghulam Rasool. The importance of robust features in mitigating catastrophic forgetting. In *2023 IEEE Symposium on Computers and Communications (ISCC)*, pages 752–757, 2023.
- [357] Hikmat Khan, Nidhal C Bouaynaya, and Ghulam Rasool. The importance of robust features in mitigating catastrophic forgetting. In *2023 IEEE Symposium on Computers and Communications (ISCC)*, pages 752–757. IEEE, 2023.
- [358] Hikmat Khan, Nidhal Carla Bouaynaya, and Ghulam Rasool. Adversarially Robust Continual Learning. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2022.

- [359] Hikmat Khan, Nidhal Carla Bouaynaya, and Ghulam Rasool. Adversarially robust continual learning. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2022.
- [360] Hikmat Khan, Nidhal Carla Bouaynaya, and Ghulam Rasool. Brain-inspired continual learning: Robust feature distillation and re-consolidation for class incremental learning. *IEEE Access*, 12:34054–34073, 2024.
- [361] Hikmat Khan, Ghulam Rasool, Nidhal Carla Bouaynaya, and Charles C Johnson. Rotorcraft flight information inference from cockpit videos using deep learning. In *Vertical Flight Society's 75th Annual Forum and Technology Display*, 2019.
- [362] Hikmat Khan, Ghulam Rasool, Nidhal Carla Bouaynaya, Tyler Travis, Lacey Thompson, and Charles C Johnson. Explainable ai: Rotorcraft attitude prediction. In *Vertical Flight Society's 76th Annual Forum and Technology Display*, 2020.
- [363] Hikmat Khan, Pir Masoom Shah, Munam Ali Shah, Saif ul Islam, and Joel JPC Rodrigues. Cascading handcrafted features and convolutional neural network for iot-enabled brain tumor segmentation. *Computer communications*, 153:196–207, 2020.
- [364] Hikmat Khan, Pir Masoom Shah, Syed Farhan Alam Zaidi, et al. Susceptibility of continual learning against adversarial attacks. *arXiv preprint arXiv:2207.05225*, 2022.
- [365] MA Khan, I Ashraf, M Alhaisoni, R Damaševičius, R Scherer, A Rehman, and SAC Bukhari. Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists. *Diagnostics (Basel)*, 10(8):565, 2020.
- [366] Muhammad Ashfaq Khan, Md. Rezaul Karim, and Yangwoo Kim. A two-stage big data analytics framework with real world applications using spark machine learning and long short-term memory network. *Symmetry*, 10(10), 2018.

- [367] Sulaiman Khan, Hazrat Ali, and Zubair Shah. Identifying the role of vision transformer for skin cancer—a scoping review. *Frontiers in Artificial Intelligence*, 6, July 2023.
- [368] Navid Khoshavi, Saman Sargolzaei, Yu Bi, and Arman Roohi. Entropy-based modeling for estimating adversarial bit-flip attack impact on binarized neural network. In n.a., editor, *ASPDAC '21: 26th Asia and South Pacific Design Automation Conference, Tokyo, Japan*, pages 493–498. ACM, 2021.
- [369] Pegah Khosravi, Maria Lysandrou, Mahmoud Eljalby, Qianzi Li, Ehsan Kazemi, Pantelis Zisimopoulos, Alexandros Sigaras, Matthew Brendel, Josue Barnes, Camir Ricketts, et al. A deep learning approach to diagnostic classification of prostate cancer using pathology–radiology fusion. *Journal of Magnetic Resonance Imaging*, 54(2):462–471, 2021.
- [370] Tim-Rasmus Kiehl. Digital and computational pathology: A specialty reimaged. In *The Future Circle of Healthcare: AI, 3D Printing, Longevity, Ethics, and Uncertainty Mitigation*, pages 227–250. Springer, 2022.
- [371] Chang-Jin Kim, James C Morley, and Charles R Nelson. The structural break in the equity premium. *Journal of Business & Economic Statistics*, 23(2):181–191, 2005.
- [372] David Kim, Kaitlin E Sundling, Renu Virk, Michael J Thrall, Susan Alperstein, Marilyn M Bui, Heather Chen-Yost, Amber D Donnelly, Oscar Lin, Xiaoying Liu, et al. Digital cytology part 2: Artificial intelligence in cytology a concept paper with review and recommendations from the american society of cytopathology digital cytology task force. *Journal of the American Society of Cytopathology*, 2023.
- [373] Inho Kim, Kyungmin Kang, Youngjae Song, and Tae-Jung Kim. Application of Artificial Intelligence in Pathology: Trends and Challenges. *Diagnostics*, 12(11):2794, 2022.
- [374] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

- [375] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- [376] Nfn Kiran, FNU Sapna, FNU Kiran, Deepak Kumar, FNU Raja, Sheena Shiwlani, Antonella Paladini, FNU Sonam, Ahmed Bendari, Raja Sandeep Perakash, et al. Digital pathology: Transforming diagnosis in the digital age. *Cureus*, 15(9), 2023.
- [377] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. *Advances in neural information processing systems*, 30, 2017.
- [378] Adrienne Kline, Hanyin Wang, Yikuan Li, Saya Dennis, Meghan Hutch, Zhenxing Xu, Fei Wang, Feixiong Cheng, and Yuan Luo. Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine*, 5(1):171, 2022.
- [379] Nikita Klyuchnikov, Ilya Trofimov, Ekaterina Artemova, Mikhail Salnikov, Maxim Fedorov, Alexander Filippov, and Evgeny Burnaev. Nas-bench-nlp: neural architecture search benchmark for natural language processing. *IEEE Access*, 10:45736–45747, 2022.
- [380] Ching-Yun Ko, Zhaoyang Lyu, Lily Weng, Luca Daniel, Ngai Wong, and Dahua Lin. POPQORN: quantifying robustness of recurrent neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 3468–3477. PMLR, 2019.
- [381] Joost N Kok, Egbert J Boers, Walter A Kusters, Peter Van der Putten, and Mannes Poel. Artificial intelligence: definition, trends, techniques, and cases. *Artificial intelligence*, 1:270–299, 2009.
- [382] Jerzy Krawczuk and Tomasz Łukaszuk. The feature selection bias problem in relation to high-dimensional gene data. *Artificial intelligence in medicine*, 66:63–71, 2016.

- [383] R Krithiga and P Geetha. Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review. *Archives of Computational Methods in Engineering*, 28(4):2607–2619, 2021.
- [384] Alex Krizhevsky. Learning multiple layers of features from tiny images. university of toronto (2012). <http://www.cs.toronto.edu/kriz/cifar.html>, last accessed, 5:13, 2022.
- [385] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [386] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [387] Carol Kruchko, Quinn T Ostrom, Haley Gittleman, and Jill S Barnholtz-Sloan. The cbtrus story: providing accurate population-based statistics on brain and other central nervous system tumors for everyone, 2018.
- [388] K Kuhn et al. The cancer biomedical informatics grid (cabig™): Infrastructure and applications for a worldwide research community. *Medinfo*, 1:330, 2007.
- [389] Alvina G Lai, Laura Pasea, Amitava Banerjee, Spiros Denaxas, Michail Katsoulis, Wai Hoong Chang, Bryan Williams, Deenan Pillay, Mahdad Noursadeghi, David Linch, et al. Estimating excess mortality in people with cancer and multimorbidity in the covid-19 emergency. *medRxiv*, 2020.
- [390] Ranjani Lakshminarasimhan and Gangning Liang. The role of dna methylation in cancer. *DNA Methyltransferases-Role and Function*, pages 151–172, 2016.
- [391] Michele Larobina and Loredana Murino. Medical image file formats. *Journal of digital imaging*, 27(2):200–206, 2014.



- [392] Juan Alberto Lecaros. Biobanks for biomedical research: Evolution and future. In *Handbook of Bioethical Decisions. Volume I: Decisions at the Bench*, pages 295–323. Springer, 2023.
- [393] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.
- [394] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep Learning. *Nature*, 521(7553):436, 2015.
- [395] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [396] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, sep 2019.
- [397] Peter Lee, Sebastien Bubeck, and Joseph Petro. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *New England Journal of Medicine*, 388(13):1233–1239, 2023.
- [398] Joffrey L Leevy and Taghi M Khoshgoftaar. A short survey of LSTM models for de-identification of medical free text. In *2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC)*, pages 117–124. IEEE, 2020.
- [399] Brian Lei, Xinyin Jiang, and Anjana Saxena. Tcga expression analyses of 10 carcinoma types reveal clinically significant racial differences. *Cancers*, 15(10):2695, 2023.
- [400] Johannes Lemke, Jan Scheele, Thomas Kapapa, Silvia Von Karstedt, Christian Rainer Wirtz, Doris Henne-Bruns, and Marko Kornmann. Brain metastases in gastrointestinal cancers: is there a role for surgery? *International journal of molecular sciences*, 15(9):16816–16830, 2014.

- [401] Dongjin Leng, Linyi Zheng, Yuqi Wen, Yunhao Zhang, Lianlian Wu, Jing Wang, Meihong Wang, Zhongnan Zhang, Song He, and Xiaochen Bo. A benchmark study of deep learning-based multi-omics data fusion methods for cancer. *Genome Biology*, 23(1):1–32, 2022.
- [402] Brian Lester, Rami Al-Rfou, and Noah Constant. The Power of Scale for Parameter-Efficient Prompt Tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3045–3059. Association for Computational Linguistics, November 2021.
- [403] Anna Lewandowska, Grzegorz Rudzki, Tomasz Lewandowski, Aleksandra Strykowska-Gora, and Sławomir Rudzki. Risk factors for the diagnosis of colorectal cancer. *Cancer Control*, 29:10732748211056692, 2022.
- [404] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Annual Meeting of the Association for Computational Linguistics*, 2019.
- [405] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, 2020.
- [406] Boyi Li, Kilian Q Weinberger, Serge Belongie, Vladlen Koltun, and Rene Ranftl. Language-driven Semantic Segmentation. In *International Conference on Learning Representations*, 2022.
- [407] Haichun Li, Ao Li, and Minghui Wang. A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. *Computers in biology and medicine*, 108:150–160, 2019.

- [408] Jun Li, Yiling Lu, Rehan Akbani, Zhenlin Ju, Paul L Roebuck, Wenbin Liu, Ji-Yeon Yang, Bradley M Broom, Roeland GW Verhaak, David W Kane, et al. Tcpa: a resource for cancer functional proteomics data. *Nature methods*, 10(11):1046–1047, 2013.
- [409] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. *arXiv preprint arXiv:2301.12597*, 2023. <https://arxiv.org/abs/2301.12597>.
- [410] Junyi Li, Qingzhe Xu, Mingxiao Wu, Tao Huang, and Yadong Wang. Pan-cancer classification based on self-normalizing neural networks and feature selection. *Frontiers in Bioengineering and Biotechnology*, 8:766, 2020.
- [411] L. Li, P. Zhang, H. Zhang, J. Yang, C. Li, Y. Zhong, L. Wang, L. Yuan, L. Zhang, J. Hwang, K. Chang, and J. Gao. Grounded language-image pre-training. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10955–10965, Los Alamitos, CA, USA, jun 2022. IEEE Computer Society.
- [412] Michelle M Li, Kexin Huang, and Marinka Zitnik. Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering*, pages 1–17, 2022.
- [413] Peibo Li, Yixing Yang, Maurice Pagnucco, and Yang Song. Explainability in graph neural networks: An experimental survey. *arXiv preprint arXiv:2203.09258*, 2022.
- [414] Peizhao Li, Jiuxiang Gu, Jason Kuen, Vlad I Morariu, Handong Zhao, Rajiv Jain, Varun Manjunatha, and Hongfu Liu. Selfdoc: Self-supervised document representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5652–5660, 2021.
- [415] Pengyong Li, Jun Wang, Yixuan Qiao, Hao Chen, Yihuan Yu, Xiaojun Yao, Peng Gao, Guotong Xie, and Sen Song. An effective self-supervised framework for learning expressive molecular global representations to drug discovery. *Briefings in Bioinformatics*, 22(6):bbab109, 2021.

- [416] Simin Li and Baosen Zhou. A review of radiomics and genomics applications in cancers: the way towards precision medicine. *Radiation Oncology*, 17(1):1–10, 2022.
- [417] Song Li and Geoffrey Kwok Fai Tso. Bottleneck supervised u-net for pixel-wise liver and tumor segmentation. *arXiv preprint arXiv:1810.10331*, 2018.
- [418] Xia Li, Zhisheng Zhong, Jianlong Wu, Yibo Yang, Zhouchen Lin, and Hong Liu. Expectation-maximization attention networks for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9167–9176, 2019.
- [419] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. Hdenseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging*, 37(12):2663–2674, 2018.
- [420] Yi Li, Haozhi Qi, Jifeng Dai, Xiangyang Ji, and Yichen Wei. Fully convolutional instance-aware semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2359–2367, 2017.
- [421] Yize Li, Eduard Porta-Pardo, Collin Tokheim, Matthew H Bailey, Tomer M Yaron, Vasileios Stathias, Yifat Geffen, Kathleen J Imbach, Song Cao, Shankara Anand, et al. Pan-cancer proteogenomics connects oncogenic drivers to functional states. *Cell*, 186(18):3921–3944, 2023.
- [422] Zhe Li, Yuming Jiang, Mengkang Lu, Ruijiang Li, and Yong Xia. Survival prediction via hierarchical multimodal co-attention transformer: A computational histology-radiology solution. *IEEE Transactions on Medical Imaging*, 2023.
- [423] Jie Lian, Jiajun Deng, Edward S Hui, Mohamad Koochi-Moghadam, Yunlang She, Chang Chen, and Varut Vardhanabhuti. Early stage NSCLS patients’ prognostic prediction with multi-information using transformer and graph neural network model. *eLife*, 11:e80547, oct 2022.

- [424] Junjie Liang, Cihui Yang, Mengjie Zeng, and Xixi Wang. TransConver: transformer and convolution parallel network for developing automatic brain tumor segmentation in MRI images. *Quantitative Imaging in Medicine and Surgery*, 12(4), 2022.
- [425] Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. Foundations and recent trends in multimodal machine learning: Principles, challenges, and open questions. *arXiv preprint arXiv:2209.03430*, 2022.
- [426] JG Liao and Khew-Voon Chin. Logistic regression for disease classification using microarray data: model selection in a large p and small n case. *Bioinformatics*, 23(15):1945–1951, 2007.
- [427] Lifelines. Concordance Index for Survival Analysis, 2023.
- [428] Shuting Lin, Jie Zhou, Yiqiong Xiao, Bridget Neary, Yong Teng, and Peng Qiu. Integrative analysis of tcga data identifies mirnas as drug-specific survival biomarkers. *Scientific Reports*, 12(1):6785, 2022.
- [429] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.
- [430] Jana Lipkova, Richard J Chen, Bowen Chen, Ming Y Lu, Matteo Barbieri, Daniel Shao, Anurag J Vaidya, Chengkuan Chen, Luoting Zhuang, Drew FK Williamson, et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell*, 40(10):1095–1110, 2022.
- [431] Jana Lipkova, Richard J. Chen, Bowen Chen, Ming Y. Lu, Matteo Barbieri, Daniel Shao, Anurag J. Vaidya, Chengkuan Chen, Luoting Zhuang, Drew F.K. Williamson, Muhammad Shaban, Tiffany Y. Chen, and Faisal Mahmood. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell*, 40:1095–1110, 2022.

- [432] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.
- [433] Hanxiao Liu, Karen Simonyan, and Yiming Yang. Darts: Differentiable architecture search. *arXiv preprint arXiv:1806.09055*, 2018.
- [434] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *arXiv preprint arXiv:2304.08485*, 2023. <https://arxiv.org/abs/2304.08485>.
- [435] Jinsha Liu, Priyanka Pandya, and Sepideh Afshar. Therapeutic advances in oncology. *International Journal of Molecular Sciences*, 22(4), 2021.
- [436] Nianfeng Liu, Haiqing Li, Man Zhang, Jing Liu, Zhenan Sun, and Tieniu Tan. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In *2016 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2016.
- [437] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35, 2023.
- [438] Sicen Liu, Xiaolong Wang, Yang Xiang, Hui Xu, Hui Wang, and Buzhou Tang. Multi-channel fusion LSTM for medical event prediction using EHRs. *Journal of Biomedical Informatics*, 127:104011, 2022.
- [439] T Liu, J Huang, T Liao, R Pu, S Liu, and Y Peng. A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data. *Irbm*, 43(1):62–74, 2022.
- [440] Wenbin Liu, Zhenlin Ju, Yiling Lu, Gordon B Mills, and Rehan Akbani. A comprehensive comparison of normalization methods for loading control and variance stabilization of reverse-phase protein array data. *Cancer informatics*, 13:CIN–S13329, 2014.

- [441] Xiaolong Liu, Zhidong Deng, and Yuhan Yang. Recent progress in semantic image segmentation. *Artificial Intelligence Review*, 52(2):1089–1106, 2019.
- [442] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach, 2020.
- [443] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*, 2024.
- [444] Michelle Livne, Jana Rieger, Orhun Utku Aydin, Abdel Aziz Taha, Ela Marie Akay, Tabea Kossen, Jan Sobesky, John D Kelleher, Kristian Hildebrand, Dietmar Frey, et al. A u-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. *Frontiers in neuroscience*, 13:97, 2019.
- [445] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [446] Camila M Lopes-Ramos, John Quackenbush, and Dawn L DeMeo. Genome-wide sex and gender differences in cancer. *Frontiers in oncology*, 10:597788, 2020.
- [447] Netanel Loyfer, Judith Magenheimer, Ayelet Peretz, Gordon Cann, Joerg Bredno, Agnes Klochandler, Ilana Fox-Fisher, Sapir Shabi-Porat, Merav Hecht, Tsuria Pelet, et al. A dna methylation atlas of normal human cell types. *Nature*, 613(7943):355–364, 2023.
- [448] Haoyu Lu, Qiongyi Zhou, Nanyi Fei, Zhiwu Lu, Mingyu Ding, Jingyuan Wen, Changde Du, Xin Zhao, Hao Sun, Huiguang He, et al. Multimodal foundation models are better simulators of the human brain. *arXiv preprint arXiv:2208.08263*, 2022. <https://arxiv.org/abs/2208.08263>.

- [449] Jiasen Lu, Dhruv Batra, Devi Parikh, and Stefan Lee. Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. *Advances in neural information processing systems*, 32, 2019.
- [450] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Kenji Ikamura, Georg Gerber, Ivy Liang, Long Phi Le, Tong Ding, Anil V Parwani, et al. A foundational multimodal vision language ai assistant for human pathology. *arXiv preprint arXiv:2312.07814*, 2023.
- [451] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, 2024.
- [452] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Andrew Zhang, Long Phi Le, et al. Towards a visual-language foundation model for computational pathology. *arXiv preprint arXiv:2307.12914*, 2023. <https://arxiv.org/abs/2307.12914>.
- [453] Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Andrew Zhang, Long Phi Le, et al. Towards a visual-language foundation model for computational pathology. *arXiv preprint arXiv:2307.12914*, 2023.
- [454] Ming Y Lu, Richard J Chen, Dehan Kong, Jana Lipkova, Rajendra Singh, Drew FK Williamson, Tiffany Y Chen, and Faisal Mahmood. Federated learning for computational pathology on gigapixel whole slide images. *Medical image analysis*, 76:102298, 2022.
- [455] Qiu hao Lu, Dejing Dou, and Thien Nguyen. Clinicalt5: A generative language model for clinical text. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5436–5443, 2022.



- [456] Zhichao Lu, Gautam Sreekumar, Erik Goodman, Wolfgang Banzhaf, Kalyanmoy Deb, and Vishnu Naresh Boddeti. Neural architecture transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(9):2971–2989, 2021.
- [457] Timo Lüddecke and Alexander Ecker. Image Segmentation Using Text and Image Prompts. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7076–7086. IEEE, 2022.
- [458] Jia Luo, Rami S Vanguri, Andrew T Aukerman, Jacklynn V Egger, Christopher J Fong, Natally Horvat, Andrew Pagano, Jose Araujo-Filho, Luke Geneslaw, Hira Rizvi, et al. Multimodal integration of radiology, pathology, and genomics for prediction of response to pd-1 blockade in patients with non–small cell lung cancer., 2022.
- [459] Jie Ma, Jun Liu, Qika Lin, Bei Wu, Yaxian Wang, and Yang You. Multitask learning for visual question answering. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15, 2021.
- [460] Lingxiao Ma, Zhi Yang, Youshan Miao, Jilong Xue, Ming Wu, Lidong Zhou, and Yafei Dai. NeuGraph: Parallel Deep Neural Network Computation on Large Graphs. In *USENIX Annual Technical Conference*, pages 443–458, 2019.
- [461] Ruibin Ma, Po-Hsuan Cameron Chen, Gang Li, Wei-Hung Weng, Angela Lin, Krishna Gadepalli, and Yuannan Cai. Human-centric metric for accelerating pathology reports annotation. *arXiv preprint arXiv:1911.01226*, 2019. <https://arxiv.org/abs/1911.01226>.
- [462] Tianle Ma and Aidong Zhang. Integrate multi-omics data with biological interaction networks using multi-view factorization autoencoder (mae). *BMC genomics*, 20(Suppl 11):944, 2019.

- [463] Xiao Ma and Fucang Jia. Brain tumor classification with multimodal MR and pathology images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part II 5*, pages 343–352. Springer, 2020.
- [464] Yao Ma, Xiaorui Liu, Tong Zhao, Yozen Liu, Jiliang Tang, and Neil Shah. A unified view on graph neural networks as graph signal denoising. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 1202–1211, 2021.
- [465] Yao Ma and Jiliang Tang. *Deep learning on graphs*. Cambridge University Press, 2021.
- [466] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.
- [467] Shino Magaki, Seyed A Hojat, Bowen Wei, Alexandra So, and William H Yong. An introduction to the performance of immunohistochemistry. *Biobanking: Methods and Protocols*, pages 289–298, 2019.
- [468] Cristina Magi-Galluzzi. Prostate cancer: diagnostic criteria and role of immunohistochemistry. *Modern Pathology*, 31:12–21, 2018.
- [469] Faisal Mahmood, Daniel Borders, Richard J Chen, Gregory N McKay, Kevan J Salimian, Alexander Baras, and Nicholas J Durr. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Transactions on Medical Imaging*, 39(11):3257–3267, 2019.
- [470] Eliana Marostica, Rebecca Barber, Thomas Denize, Isaac S Kohane, Sabina Signoretti, Jeffrey A Golden, and Kun-Hsing Yu. Development of a histopathology informatics pipeline for classification and prediction of clinical outcomes in subtypes of renal cell carcinoma. *Clinical Cancer Research*, 27(10):2868–2878, 2021.

- [471] Anne L Martel, Dan Hosseinzadeh, Caglar Senaras, Yu Zhou, Azadeh Yazdanpanah, Rushin Shojaii, Emily S Patterson, Anant Madabhushi, and Metin N Gurcan. An image analysis resource for cancer research: Piip—pathology image informatics platform for visualization, analysis, and management. *Cancer research*, 77(21):e83–e86, 2017.
- [472] Marvel. Marvel movies. <https://www.marvel.com/movies>.
- [473] Naoki Masuda, Michiko Sakaki, Takahiro Ezaki, and Takamitsu Watanabe. Clustering coefficients for correlation networks. *Frontiers Neuroinformatics*, 12:7, 2018.
- [474] mayoclinic. Braintumor. <https://www.mayoclinic.org/diseases-conditions/brain-tumor/symptoms-causes/syc-20350084>.
- [475] MayoFoundation. Biomedical imaging resource. <https://analyzedirect.com/>.
- [476] Clare McGenity, Emily L Clarke, Charlotte Jennings, Gillian Matthews, Caroline Cartlidge, Henschel Freduah-Agyemang, Deborah D Stocken, and Darren Treanor. Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy. *npj Digital Medicine*, 7(1):114, 2024.
- [477] Scott Mayer McKinney, Alan Karthikesalingam, Daniel Tse, Christopher J Kelly, Yun Liu, Greg S Corrado, and Shravya Shetty. Reply to: Transparency and reproducibility in artificial intelligence. *Nature*, 586(7829):E17–E18, 2020.
- [478] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg S Corrado, Ara Darzi, et al. International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788):89–94, 2020.
- [479] Gaurav Mendiratta, Eugene Ke, Meraj Aziz, David Liarakos, Melinda Tong, and Edward C Stites. Cancer gene mutation frequencies for the us population. *Nature communications*, 12(1):5961, 2021.

- [480] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.
- [481] Christina Messiou, Richard Lee, and Manuel Salto-Tellez. Multimodal analysis and the oncology patient: Creating a hospital system for integrated diagnostics and discovery. *Computational and Structural Biotechnology Journal*, 21:4536–4539, 2023.
- [482] MIDRC. Medical Imaging and Data Resource Center (MIDRC). Available online: <https://www.midrc.org/>. (accessed on 28 November 2023).
- [483] Mite Mijalkov, Ehsan Kakaei, Joana B Pereira, Eric Westman, Giovanni Volpe, and Alzheimer’s Disease Neuroimaging Initiative. Braph: a graph theory software for the analysis of brain connectivity. *PloS one*, 12(8):e0178798, 2017.
- [484] Kimberly D Miller, Leticia Nogueira, Angela B Mariotto, Julia H Rowland, K Robin Yabroff, Catherine M Alfano, Ahmedin Jemal, Joan L Kramer, and Rebecca L Siegel. Cancer treatment and survivorship statistics, 2019. *CA: a cancer journal for clinicians*, 69(5):363–385, 2019.
- [485] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.
- [486] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *arXiv preprint arXiv:2001.05566*, 2020.

- [487] Matthias Minderer, Alexey Gritsenko, Austin Stone, Maxim Neumann, Dirk Weissenborn, Alexey Dosovitskiy, Aravindh Mahendran, Anurag Arnab, Mostafa Dehghani, Zhuoran Shen, Xiao Wang, Xiaohua Zhai, Thomas Kipf, and Neil Houlsby. Simple open-vocabulary object detection with vision transformers. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part X*, page 728–755, Berlin, Heidelberg, 2022. Springer-Verlag.
- [488] Robert C Miner. Image-guided neurosurgery. *Journal of medical imaging and radiation sciences*, 48(4):328–335, 2017.
- [489] Riccardo Miotto, Li Li, Brian A Kidd, and Joel T Dudley. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, 6(1):1–10, 2016.
- [490] Bilal Mirza, Wei Wang, Jie Wang, Howard Choi, Neo Christopher Chung, and Peipei Ping. Machine learning and integrative analysis of biomedical big data. *Genes*, 10(2):87, 2019.
- [491] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019.
- [492] Shaocong Mo, Ming Cai, Lanfen Lin, Ruofeng Tong, Qingqing Chen, Fang Wang, Hongjie Hu, Yutaro Iwamoto, Xian-Hua Han, and Yen-Wei Chen. Multimodal priors guided segmentation of liver lesions in MRI using mutual information based graph co-attention networks. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*, pages 429–438. Springer, 2020.

- [493] Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M. Krumholz, Jure Leskovec, Eric J. Topol, and Pranav Rajpurkar. Foundation Models for Generalist Medical Artificial Intelligence. *Nature*, 616(7956):259–265, Apr 2023.
- [494] Olivier Morin, Martin Vallières, Steve Braunstein, Jorge Barrios Ginart, Taman Upadhaya, Henry C Woodruff, Alex Zwanenburg, Avishek Chatterjee, Javier E Villanueva-Meyer, Gilmer Valdes, et al. An artificial intelligence framework integrating longitudinal electronic health records with real-world data enables continuous pan-cancer prognostication. *Nature Cancer*, 2(7):709–722, 2021.
- [495] Wei Mu, Matthew B Schabath, and Robert J Gillies. Images are data: challenges and opportunities in the clinical translation of radiomics. *Cancer Research*, 82(11):2066–2068, 2022.
- [496] L. J. Muhammad and Alessandro Bria. Editorial: Ai applications for diagnosis of breast cancer. *Frontiers in Artificial Intelligence*, 6, October 2023.
- [497] Wazir Muhammad, Saad-bin-Saeed Ahmed, Shahzaib Naeem, Agha Agha Muhammad Hamad Khan, Bilal Mazhar Qureshi, Amjad Hussain, and Bulent Aydogan. Artificial neural network-assisted prediction of radiobiological indices in head and neck cancer. *Frontiers in Artificial Intelligence*, 7:1329737, 2024.
- [498] Shawn N Murphy, Griffin Weber, Michael Mendis, Vivian Gainer, Henry C Chueh, Susanne Churchill, and Isaac Kohane. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *Journal of the American Medical Informatics Association*, 17(2):124–130, 2010.
- [499] Keisuke Nakagawa, Lama Moukheiber, Leo Celi, Malhar Patel, Faisal Mahmood, Dibson Gondim, Michael Hogarth, and Richard Levenson. Ai in pathology: What could possibly go wrong? In *Seminars in Diagnostic Pathology*,. Elsevier, 2023.

- [500] Athira Nambiar and Divyansh Mundra. An overview of data warehouse and data lake in modern enterprise data management. *Big Data and Cognitive Computing*, 6(4), 2022.
- [501] National Cancer Institute. CCG’s Genome Characterization Pipeline. Available online: <https://www.cancer.gov/ccg/research/genome-characterization-pipeline>. (accessed on 18 June 2023).
- [502] NCI. NCI Thesaurus. Available online: <https://ncit.nci.nih.gov/ncitbrowser/>. (accessed on 1 December 2023).
- [503] Thin Nguyen, Hang Le, Thomas P Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh. GraphDTA: predicting drug-target binding affinity with graph neural networks. *Bioinformatics*, 37(8):1140–1147, 2021.
- [504] Jan Moritz Niehues, Philip Quirke, Nicholas P West, Heike I Grabsch, Marko van Treeck, Yoni Schirris, Gregory P Veldhuizen, Gordon GA Hutchins, Susan D Richman, Sebastian Foersch, et al. Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: A retrospective multi-centric study. *Cell Reports Medicine*, 2023.
- [505] Ian E Nielsen, Dimah Dera, Ghulam Rasool, Ravi P Ramachandran, and Nidhal Carla Bouaynaya. Robust explainability: A tutorial on gradient-based attribution methods for deep neural networks. *IEEE Signal Processing Magazine*, 39(4):73–84, 2022.
- [506] Ian E Nielsen, Ravi P Ramachandran, Nidhal Bouaynaya, Hassan M Fathallah-Shaykh, and Ghulam Rasool. EvalAttAI: A Holistic Approach to Evaluating Attribution Maps in Robust and Non-Robust Models. *arXiv preprint arXiv:2303.08866*, 2023. <https://arxiv.org/abs/2303.08866>.
- [507] Michael Nielsen. Ch-4: A visual proof that neural nets can compute any function. <http://neuralnetworksanddeeplearning.com/chap4.html>.

- [508] NIH. Cancer Data Aggregator. Available online: <https://datacommons.cancer.gov/cancer-data-aggregator>. (accessed on 15 June 2023).
- [509] NIH. Neuroimaging informatics technology initiative. <https://nifti.nimh.nih.gov/> and <https://brainder.org/2012/09/23/the-nifti-file-format/>.
- [510] NIH. Overview of SNOMED CT. National Library of Medicine. Available online: [https://www.nlm.nih.gov/healthit/snomedct/snomed\\_overview.html](https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html). (accessed on 1 December 2023).
- [511] Nikolaos Nikolaou, Domingo Salazar, Harish RaviPrakash, Miguel Goncalves, Rob Mulla, Nikolay Burlutskiy, Natasha Markuzon, and Etai Jacob. Quantifying the advantage of multimodal data fusion for survival prediction in cancer patients. *bioRxiv*, pages 2024–01, 2024.
- [512] NRRD. Nearly raw raster data. <http://teem.sourceforge.net/nrrd/>.
- [513] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*, 2018.
- [514] College of American Pathologists. Digital pathology codes. *College of American Pathologists*, June 2023.
- [515] Perelman School of Medicine University of Pennsylvania. Brain tumor segmentation. <http://braintumorsegmentation.org/>.
- [516] American Association of Neurological Surgeons. Braintumortypes. <https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Brain-Tumors>.
- [517] Se-Ra Oh, Young-Duk Seo, Euijong Lee, and Young-Gab Kim. A comprehensive survey on security and privacy for electronic health data. *International Journal of Environmental Research and Public Health*, 18(18):9668, 2021.



- [518] Ron Ohlander, Keith Price, and D Raj Reddy. Picture segmentation using a recursive region splitting method. *Computer graphics and image processing*, 8(3):313–333, 1978.
- [519] Yann Ollivier. Ricci curvature of metric spaces. *Comptes Rendus Mathematique*, 345(11):643–646, 2007.
- [520] Yann Ollivier. Ricci curvature of markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009.
- [521] Adam B Olshen, ES Venkatraman, Robert Lucito, and Michael Wigler. Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics*, 5(4):557–572, 2004.
- [522] Marwan Omar, Soohyeon Choi, DaeHun Nyang, and David Mohaisen. Robust natural language processing: Recent advances, challenges, and future directions. *arXiv preprint arXiv:2201.00768*, 2022.
- [523] OpenAI. Introducing ChatGPT, 2022. <https://openai.com/blog/chatgpt>. Last accessed on: March 10, 2023.
- [524] OpenAI. GPT-4, 2023. Available at: <https://openai.com/research/gpt-4>. Last accessed on April 5, 2023.
- [525] OpenAI. Hello gpt-4o. @ONLINE, May 2024.
- [526] José Ignacio Orlando, Philipp Seeböck, Hrvoje Bogunović, Sophie Klimscha, Christoph Grechenig, Sebastian Waldstein, Bianca S Gerendas, and Ursula Schmidt-Erfurth. U2-net: A bayesian u-net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological oct scans. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1441–1445. IEEE, 2019.

- [527] Quinn T Ostrom, David J Cote, Mustafa Ascha, Carol Kruchko, and Jill S Barnholtz-Sloan. Adult glioma incidence and survival by race or ethnicity in the united states from 2000 to 2014. *JAMA oncology*, 4(9):1254–1262, 2018.
- [528] Quinn T Ostrom, Haley Gittleman, et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011–2015. *Neuro-Oncology*, 20, 10 2018.
- [529] Daniel W. Otter, Julian R. Medina, and Jugal K. Kalita. A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, 32:604–624, 2021.
- [530] Maria Otth and Katrin Scheinemann. Surveillance imaging for high-grade childhood brain tumors: What to do 10 years after completion of treatment? *Pediatric blood & cancer*, 65(11):e27311, 2018.
- [531] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Gray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.
- [532] Ezi Ozoani, Marissa Gerchick, and Margaret Mitchell. Model Cards, 2022. Available at: <https://huggingface.co/blog/model-cards>.
- [533] Paige Team. Pathology Artificial Intelligence Guidance Engine (PAIGE), 2023. available at: <https://paige.ai/>. Last accessed on Mar 31, 2023.
- [534] Susmita Palmal, Nikhilanand Arya, Sriparna Saha, and Somanath Tripathy. Breast cancer survival prognosis using the graph convolutional network with choquet fuzzy integral. *Scientific Reports*, 13(1):14757, 2023.

- [535] Liangrui Pan, Dazhen Liu, Yutao Dou, Lian Wang, Zhichao Feng, Pengfei Rong, Liwen Xu, and Shaoliang Peng. Multi-head attention mechanism learning for cancer new subtypes and treatment based on cancer multi-omics data. *arXiv preprint arXiv:2307.04075*, 2023.
- [536] Liron Pantanowitz, Douglas Hartman, Yan Qi, Eun Yoon Cho, Beomseok Suh, Kyunghyun Paeng, Rajiv Dhir, Pamela Michelow, Scott Hazelhurst, Sang Yong Song, et al. Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses. *Diagnostic Pathology*, 15:1–10, 2020.
- [537] Liron Pantanowitz, Ashish Sharma, Alexis B Carter, Tahsin Kurc, Alan Sussman, and Joel Saltz. Twenty years of digital pathology: an overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives. *Journal of pathology informatics*, 9(1):40, 2018.
- [538] Liron Pantanowitz, Paul N. Valenstein, Andrew J. Evans, Keith J. Kaplan, John D. Pfeifer, David C. Wilbur, Laura C. Collins, and Terence J. Colgan. Review of the current state of whole slide imaging in pathology. *Journal of Pathology Informatics*, 2(1):36, 2011.
- [539] A. Papoulis and S. U. Pillai. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Higher Education, 4 edition, 2002.
- [540] Jiwoong Park, Junho Cho, Hyung Jin Chang, and Jin Young Choi. Unsupervised hyperbolic representation learning via message passing auto-encoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5516–5526, 2021.
- [541] Min-Koo Park, Jin-Muk Lim, Jinwoo Jeong, Yeongjae Jang, Ji-Won Lee, Jeong-Chan Lee, Hyungyu Kim, Euiyul Koh, Sung-Joo Hwang, Hong-Gee Kim, et al. Deep-Learning Algorithm and Concomitant Biomarker Identification for NSCLC Prediction Using Multi-Omics Data Integration. *Biomolecules*, 12:1839, 2022.

- [542] Ankush U Patel, Nada Shaker, Sambit Mohanty, Shivani Sharma, Shivam Gangal, Catarina Eloy, and Anil V Parwani. Cultivating Clinical Clarity through Computer Vision: A Current Perspective on Whole Slide Imaging and Artificial Intelligence. *Diagnostics*, 12(8):1778, 2022.
- [543] Pushpak Pati, Guillaume Jaume, Lauren Alisha Fernandes, Antonio Foncubierta-Rodríguez, Florinda Feroce, Anna Maria Anniciello, Giosue Scognamiglio, Nadia Brancati, Daniel Riccio, Maurizio Di Bonito, et al. HACT-Net: A Hierarchical Cell-to-Tissue Graph Neural Network for Histopathological Image Classification. In *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis: Second International Workshop, UNSURE 2020, and Third International Workshop, GRAIL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2*, pages 208–219. Springer, 2020.
- [544] Sarthak Pati, Ujjwal Baid, Brandon Edwards, Micah Sheller, Shih-Han Wang, G Anthony Reina, Patrick Foley, Alexey Gruzdev, Deepthi Karkada, Christos Davatzikos, et al. Federated learning enables big data for rare cancer boundary detection. *Nature communications*, 13(1):7346, 2022.
- [545] Kedar A Patwardhan, Harish RaviPrakash, Nikos Nikolaou, Ignacio Gonzalez-Garcia, Jose Domingo Salazar, Paul Metcalfe, and Joachim Reischl. Towards a survival risk prediction model for metastatic nslc patients on durvalumab using whole-lung ct radiomics. *bioRxiv*, pages 2024–02, 2024.
- [546] Doru Paul and Natalia L Komarova. Multi-scale network targeting: A holistic systems-biology approach to cancer treatment. *Progress in biophysics and molecular biology*, 165:72–79, 2021.
- [547] T Pavlidis. *Structural pattern recognition*. Springer, 1977.

- [548] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [549] Xiangdong Pei, Ke Zuo, Yuan Li, and Zhengbin Pang. A review of the application of multi-modal deep learning in medicine: Bibliometrics and future directions. *International Journal of Computational Intelligence Systems*, 16(1):44, 2023.
- [550] Hao Peng, Casey Moore, Yuanyuan Zhang, Debabrata Saha, Steve Jiang, and Robert Timmerman. An ai-based approach for modeling the synergy between radiotherapy and immunotherapy. *Scientific Reports*, 14(1):8250, 2024.
- [551] Yong Peng and Carlo M Croce. The role of micrnas in human cancer. *Signal transduction and targeted therapy*, 1(1):1–9, 2016.
- [552] Kathryn L Penney, Meir J Stampfer, Jaquelyn L Jahn, Jennifer A Sinnott, Richard Flavin, Jennifer R Rider, Stephen Finn, Edward Giovannucci, Howard D Sesso, Massimo Loda, et al. Gleason grade progression is uncommon. *Cancer research*, 73(16):5163–5168, 2013.
- [553] PennMedicine. Common types of brain tumors. <https://www.pennmedicine.org/updates/blogs/neuroscience-blog/2018/november/what-are-the-most-common-types-of-brain-tumors>.
- [554] Sérgio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5):1240–1251, 2016.
- [555] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations, 2018.

- [556] Fabio Pierazzi, Feargus Pendlebury, Jacopo Cortellazzi, and Lorenzo Cavallaro. Intriguing properties of adversarial ml attacks in the problem space. In n.a., editor, *2020 IEEE Symposium on Security and Privacy (SP)*, pages 1308–1325. IEEE Computer Society, 2020.
- [557] Olivier B Poirion, Zheng Jing, Kumardeep Chaudhary, Sijia Huang, and Lana X Garmire. DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome medicine*, 13(1):1–15, 2021.
- [558] Yixuan Qiao, Lianhe Zhao, Chunlong Luo, Yufan Luo, Yang Wu, Shengtong Li, Dechao Bu, and Yi Zhao. Multi-modality artificial intelligence in digital pathology. *Briefings in Bioinformatics*, 23(6):bbac367, 2022.
- [559] Dahui Qin. Next-generation sequencing and its clinical application. *Cancer biology & medicine*, 16(1):4, 2019.
- [560] Jianing Qiu, Lin Li, Jiankai Sun, Jiachuan Peng, Peilun Shi, Ruiyang Zhang, Yinzhaodong, Kyle Lam, Frank P. W. Lo, Bo Xiao, Wu Yuan, Dong Xu, and Benny Lo. Large AI Models in Health Informatics: Applications, Challenges, and the Future, 2023. <https://arxiv.org/abs/2303.11568> and <https://github.com/Jianing-Qiu/Awesome-Healthcare-Foundation-Models>.
- [561] Yixuan Qiu, Feng Lin, Weitong Chen, and Miao Xu. Pre-training in Medical Data: A Survey. *Machine Intelligence Research*, pages 1–33, 2023.
- [562] Martha Quinn, Jane Forman, Molly Harrod, Suzanne Winter, Karen E. Fowler, Sarah L. Krein, Ashwin Gupta, Sanjay Saint, Hardeep Singh, and Vineet Chopra. Electronic health records, communication, and data sharing: challenges and opportunities for improving the diagnostic process. *Diagnosis*, 6(3):241–248, 2019.
- [563] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

- [564] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.
- [565] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 18–24 Jul 2021.
- [566] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training, 2018. Available at <https://openai.com/research/language-unsupervised>. Last accessed: April 5, 2023.
- [567] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. *OpenAI*, 2018.
- [568] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Better language models and their implications, 2019. available at: <https://openai.com/research/better-language-models>. Last accessed on April 5, 2023.
- [569] Ilija Radosavovic, Justin Johnson, Saining Xie, Wan-Yen Lo, and Piotr Dollár. On network design spaces for visual recognition. In n.a., editor, *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 1882–1890. IEEE, 2019.

- [570] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text Transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551, 2020.
- [571] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *The Journal of Machine Learning Research*, 21(1), jan 2020.
- [572] Nishant Rajadhyaksha and Aarushi Chitkara. Graph contrastive learning for multi-omics data. *arXiv preprint arXiv:2301.02242*, 2023.
- [573] R Rajeev, J Abdul Samath, and NK Karthikeyan. An intelligent recurrent neural network with long short-term memory (LSTM) BASED batch normalization for medical image denoising. *Journal of Medical Systems*, 43:1–10, 2019.
- [574] Pranav Rajpurkar and Matthew P Lungren. The current and future state of ai interpretation of medical images. *New England Journal of Medicine*, 388(21):1981–1990, 2023.
- [575] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical Text-Conditional Image Generation with CLIP Latents, 2022. <https://arxiv.org/abs/2204.06125>.
- [576] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, pages 8821–8831. PMLR, 2021.
- [577] Jiahua Rao, Xiang Zhou, Yutong Lu, Huiying Zhao, and Yuedong Yang. Imputing single-cell RNA-seq data by combining graph convolution and autoencoder neural networks. *Iscience*, 24(5):102393, 2021.



- [578] Y. Rao, W. Zhao, G. Chen, Y. Tang, Z. Zhu, G. Huang, J. Zhou, and J. Lu. DenseCLIP: Language-Guided Dense Prediction with Context-Aware Prompting. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18061–18070, Los Alamitos, CA, USA, jun 2022. IEEE Computer Society.
- [579] Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, and Degui Zhi. Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *npj Digital Medicine*, 4(1), 2021.
- [580] Ghulam Rasool. HOME: Hands-On Machine Learning Course for Medical Professionals, 2023. Available at: <https://lab.moffitt.org/rasool/home/>.
- [581] Faisal Altaf Rathore, Hafiz Saad Khan, Hafiz Mudassar Ali, Marwa Obayya, Saim Rasheed, Lal Hussain, Zaki Hassan Kazmi, Mohamed K Nour, Abdullah Mohamed, and Abdelwahed Motwakel. Survival prediction of glioma patients from integrated radiology and pathology images using machine learning ensemble regression methods. *Applied Sciences*, 12(20):10357, 2022.
- [582] Saima Rathore, Tamim Niazi, Muhammad Aksam Iftikhar, and Ahmad Chaddad. Glioma grading via analysis of digital pathology images using machine learning. *Cancers*, 12(3):578, 2020.
- [583] Prasan Ratnayake, Sugandima Weragoda, Janaka Wansapura, Dharshana Kasthurirathna, and Mahendra Piraveenan. Quantifying the robustness of complex networks with heterogeneous nodes. *Mathematics*, 9(21), 2021.
- [584] Andrea Rau, Michael Flister, Hallgeir Rui, and Paul L Auer. Exploring drivers of gene expression in the cancer genome atlas. *Bioinformatics*, 35(1):62–68, 2019.
- [585] RCCA. Active surveillance: Its role in low-risk cancer. <https://www.regionalcancercare.org/services/active-surveillance/>.

- [586] Jaxk Reeves, Jien Chen, Xiaolan L Wang, Robert Lund, and Qi Qi Lu. A review and comparison of changepoint detection techniques for climate data. *Journal of applied meteorology and climatology*, 46(6):900–915, 2007.
- [587] Elliott Remmer. Explainability methods for transformer-based artificial neural networks:: a comparative analysis, 2022.
- [588] Stephanie Rendón de la Torre, Jaan Kalda, Robert Kitt, and Jüri Engelbrecht. Fractal and multifractal analysis of complex networks: Estonian network of payments. *The European Physical Journal B*, 90(12):1–8, 2017.
- [589] Edward M Riseman and Michael A Arbib. Computational techniques in the visual segmentation of static scenes. *Computer Graphics and Image Processing*, 6(3):221–276, 1977.
- [590] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022.
- [591] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- [592] Zhiwei Rong, Zhilin Liu, Jiali Song, Lei Cao, Yipe Yu, Mantang Qiu, and Yan Hou. Mcluster-vaes: an end-to-end variational deep learning-based clustering method for subtype discovery using multi-omics data. *Computers in Biology and Medicine*, 150:106085, 2022.
- [593] 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.

- [594] Azriel Rosenfeld and Larry S Davis. Image segmentation and image models. *Proceedings of the IEEE*, 67(5):764–772, 1979.
- [595] Steven P. Rowe and Martin G. Pomper. Molecular imaging in oncology: Current impact and future directions. *CA: A Cancer Journal for Clinicians*, 72(4):333–352, 2022.
- [596] Benedek Rozemberczki, Anna Gogleva, Sebastian Nilsson, Gavin Edwards, Andriy Nikolov, and Eliseo Papa. MOOMIN: Deep Molecular Omics Network for Anti-Cancer Drug Combination Therapy. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 3472–3483, 2022.
- [597] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.
- [598] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.*, 115(3):211–252, 2015.
- [599] Khaled Saab, Tao Tu, Wei-Hung Weng, Ryutaro Tanno, David Stutz, Ellery Wulczyn, Fan Zhang, Tim Strother, Chunjong Park, Elahe Vedadi, et al. Capabilities of gemini models in medicine. *arXiv preprint arXiv:2404.18416*, 2024.
- [600] Bhanusivakumar R Sabbula, David P Gasalberti, and Fatima Anjum. Squamous cell lung cancer. In *StatPearls [Internet]*. StatPearls Publishing, 2023.
- [601] Taro Sakamoto, Tomoi Furukawa, Kris Lami, Hoa Hoang Ngoc Pham, Wataru Uegami, Kishio Kuroda, Masataka Kawai, Hidenori Sakanashi, Lee Alex Donald Cooper, Andrey Bychkov, et al. A narrative review of digital pathology and artificial intelligence: focusing on lung cancer. *Translational Lung Cancer Research*, 9(5):2255, 2020.

- [602] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Auto-Context Convolutional Neural Network (Auto-Net) for Brain Extraction in Magnetic Resonance Imaging. *IEEE Transactions on Medical Imaging*, 36(11):2319–2330, 2017.
- [603] Joel Saltz, Ashish Sharma, Ganesh Iyer, Erich Bremer, Feiqiao Wang, Alina Jasniewski, Tammy DiPrima, Jonas S Almeida, Yi Gao, Tianhao Zhao, et al. A containerized software system for generation, management, and exploration of features from whole slide tissue images. *Cancer research*, 77(21):e79–e82, 2017.
- [604] Stephen-John Sammut, Mireia Crispin-Ortuzar, Suet-Feung Chin, Elena Provenzano, Helen A Bardwell, Wenxin Ma, Wei Cope, Ali Dariush, Sarah-Jane Dawson, Jean E Abraham, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature*, 601(7894):623–629, 2022.
- [605] Francisco Sanchez-Vega, Marco Mina, Joshua Armenia, Walid K Chatila, Augustin Luna, Konnor C La, Sofia Dimitriadoy, David L Liu, Havish S Kantheti, Sadegh Saghafinia, et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell*, 173(2):321–337, 2018.
- [606] Romeil Sandhu, Tryphon Georgiou, Ed Reznik, Liangjia Zhu, Ivan Kolesov, Yasin Senbabaoglu, and Allen Tannenbaum. Graph curvature for differentiating cancer networks. *Scientific reports*, 5(1):1–13, 2015.
- [607] Romeil S Sandhu, Tryphon T Georgiou, and Allen R Tannenbaum. Ricci curvature: An economic indicator for market fragility and systemic risk. *Science advances*, 2(5):e1501495, 2016.
- [608] Poornachandra Sandur, C Naveena, VN Manjunath Aradhya, and KB Nagasundara. Segmentation of brain tumor tissues in hgg and lgg mr images using 3d u-net convolutional neural network. *International Journal of Natural Computing Research (IJNCR)*, 7(2):18–30, 2018.
- [609] Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. Dynamic graph representation learning via self-attention networks. *arXiv preprint arXiv:1812.09430*, 2018.

- [610] Aida Santaolalla, Tim Hulsen, Jenson Davis, Hashim U Ahmed, Caroline M Moore, Shonit Punwani, Gert Attard, Neil McCartan, Mark Emberton, Anthony Coolen, et al. The reimagine multimodal warehouse: Using artificial intelligence for accurate risk stratification of prostate cancer. *Frontiers in Artificial Intelligence*, 4:769582, 2021.
- [611] Virinder Kaur Sarhadi and Gemma Armengol. Molecular biomarkers in cancer. *Biomolecules*, 12(8):1021, 2022.
- [612] Shankha Satpathy, Karsten Krug, Pierre M Jean Beltran, Sara R Savage, Francesca Petralia, Chandan Kumar-Sinha, Yongchao Dou, Boris Reva, M Harry Kane, Shayan C Avanesian, et al. A proteogenomic portrait of lung squamous cell carcinoma. *Cell*, 184(16):4348–4371, 2021.
- [613] Camillo Saueressig, Adam Berkley, Elliot Kang, Reshma Munbodh, and Ritambhara Singh. Exploring graph-based neural networks for automatic brain tumor segmentation. In *From Data to Models and Back: 9th International Symposium, DataMod 2020, Virtual Event, October 20, 2020, Revised Selected Papers 9*, pages 18–37. Springer, 2021.
- [614] Andrew M. Saxe, Pang Wei Koh, Zhenghao Chen, Maneesh Bhand, Bipin Suresh, and Andrew Y. Ng. On random weights and unsupervised feature learning. In Lise Getoor and Tobias Scheffer, editors, *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pages 1089–1096. Omnipress, 2011.
- [615] Rebeca Scalco, Yamah Hamsafar, Charles L White III, Julie A Schneider, Robert Ross Reichard, Stefan Prokop, Richard J Perrin, Peter T Nelson, Sean Mooney, Andrew P Lieberman, et al. The status of digital pathology and associated infrastructure within alzheimer’s disease centers. *Journal of Neuropathology & Experimental Neurology*, 82(3):202–211, 2023.
- [616] Philips Scanners. parrec. <https://nipy.org/nibabel/reference/nibabel.parrec.html>.

- [617] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. BLOOM: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*, 2022.
- [618] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. *arXiv preprint arXiv:2211.05100*, 2022.
- [619] Elisabeth Scheufele, Dina Aronzon, Robert Coopersmith, Michael T McDuffie, Manish Kapoor, Christopher A Uhrich, Jean E Avitabile, Jinlei Liu, Dan Housman, and Matvey B Palchuk. transmart: an open source knowledge management and high content data analytics platform. *AMIA Summits on Translational Science Proceedings*, 2014:96, 2014.
- [620] Carissa Schoenick, Peter Clark, Oyvind Tafjord, Peter Turney, and Oren Etzioni. Moving beyond the turing test with the Allen AI science challenge. *Communications of the ACM*, 60(9):60–64, 2017.
- [621] Stefan Schulz, Ann-Christin Woerl, Florian Jungmann, Christina Glasner, Philipp Stenzel, Stephanie Strobl, Aurélie Fernandez, Daniel-Christoph Wagner, Axel Haferkamp, Peter Mildenerger, Wilfried Roth, and Sebastian Foersch. Multimodal deep learning for prognosis prediction in renal cancer. *Frontiers in Oncology*, 11, 2021.
- [622] Scopio Team. Scopio Labs, 2023. available at: <https://scopiolabs.com/>. Last accessed on Mar 31, 2023.
- [623] Marzia Settino and Mario Cannataro. Survey of main tools for querying and analyzing tcga data. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1711–1718. IEEE, 2018.

- [624] Pir Masoom Shah, Hikmat Khan, Uferah Shafi, Saif ul Islam, Mohsin Raza, Tran The Son, and Hoa Le-Minh. 2d-cnn based segmentation of ischemic stroke lesions in mri scans. In *Advances in Computational Collective Intelligence: 12th International Conference, ICCCI 2020, Da Nang, Vietnam, November 30–December 3, 2020, Proceedings 12*, pages 276–286. Springer, 2020.
- [625] Murray Shanahan. Talking About Large Language Models. *arXiv preprint arXiv:2212.03551*, 2023. <https://arxiv.org/abs/2212.03551>.
- [626] Junyuan Shang, Tengfei Ma, Cao Xiao, and Jimeng Sun. Pre-training of graph augmented transformers for medication recommendation. *arXiv preprint arXiv:1906.00346*, 2019.
- [627] Sharath M. Shankaranarayana, Keerthi Ram, Kaushik Mitra, and Mohanasankar Sivaprakasam. Fully Convolutional Networks for Monocular Retinal Depth Estimation and Optic Disc-Cup Segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2019.
- [628] 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, 2021.
- [629] Neeraj Sharma and Lalit M Aggarwal. Automated medical image segmentation techniques. *Journal of medical physics/Association of Medical Physicists of India*, 35(1):3, 2010.
- [630] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19:221–248, 2017.
- [631] Jun Shi, Ruoyu Wang, Yushan Zheng, Zhiguo Jiang, and Lanlan Yu. Graph convolutional networks for cervical cell classification. In *MICCAI 2019 Computational Pathology Workshop COMPAY*, 2019.

- [632] Wuzhen Shi, Feng Jiang, and Debin Zhao. Single image super-resolution with dilated convolution based multi-scale information learning inception module. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 977–981. IEEE, 2017.
- [633] Zhouxing Shi, Huan Zhang, Kai-Wei Chang, Minlie Huang, and Cho-Jui Hsieh. Robustness verification for transformers. In n.a., editor, *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.
- [634] Artem Shmatko, Narmin Ghaffari Laleh, Moritz Gerstung, and Jakob Nikolas Kather. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nature Cancer*, 3(9):1026–1038, 2022.
- [635] Jayson Sia, Edmond Jonckheere, and Paul Bogdan. Ollivier-ricci curvature-based method to community detection in complex networks. *Scientific reports*, 9(1):1–12, 2019.
- [636] Aisha Siam, Abdel Rahman Alsaify, Bushra Mohammad, Md Rafiul Biswas, Hazrat Ali, and Zubair Shah. Multimodal deep learning for liver cancer applications: a scoping review. *Frontiers in Artificial Intelligence*, 6, 2023.
- [637] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1):7–34, 2019.
- [638] Dan Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 2006.
- [639] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [640] Cedars Sinai. Brain tumors and brain cancer. <https://www.cedars-sinai.org/health-library/diseases-and-conditions/b/brain-tumors-and-brain-cancer.html>.



- [641] Amanpreet Singh, Ronghang Hu, Vedanuj Goswami, Guillaume Couairon, Wojciech Galuba, Marcus Rohrbach, and Douwe Kiela. Flava: A foundational language and vision alignment model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15638–15650, 2022.
- [642] Amanpreet Singh, Ronghang Hu, Vedanuj Goswami, Guillaume Couairon, Wojciech Galuba, Marcus Rohrbach, and Douwe Kiela. FLAVA: A foundational language and vision alignment model. In *CVPR*, 2022.
- [643] Gagandeep Singh, Sunil Manjila, Nicole Sakla, Alan True, Amr H Wardeh, Niha Beig, Anatoliy Vaysberg, John Matthews, Prateek Prasanna, and Vadim Spektor. Radiomics and radiogenomics in gliomas: a contemporary update. *British Journal of Cancer*, 125(5):641–657, 2021.
- [644] Ranwir K Sinha, Asitava Deb Roy, Nikhil Kumar, Himel Mondal, and Ranwir Sinha. Applicability of ChatGPT in Assisting to Solve Higher Order Problems in Pathology. *Cureus*, 15(2), 2023.
- [645] William C Sleeman IV, Rishabh Kapoor, and Preetam Ghosh. Multimodal classification: Current landscape, taxonomy and future directions. *ACM Computing Surveys*, 55(7):1–31, 2022.
- [646] Shaden Smith, Mostofa Patwary, Brandon Norick, Patrick LeGresley, Samyam Rajbhandari, Jared Casper, Zhun Liu, Shrimai Prabhumoye, George Zerveas, Vijay Korthikanti, et al. Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model. *arXiv preprint arXiv:2201.11990*, 2022. <https://arxiv.org/abs/2201.11990>.

- [647] Luis R Soenksen, Yu Ma, Cynthia Zeng, Leonard Boussioux, Kimberly Villalobos Carballo, Liangyuan Na, Holly M Wiberg, Michael L Li, Ignacio Fuentes, and Dimitris Bertsimas. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ digital medicine*, 5(1):149, 2022.
- [648] Andrew H Song, Guillaume Jaume, Drew FK Williamson, Ming Y Lu, Anurag Vaidya, Tiffany R Miller, and Faisal Mahmood. Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering*, pages 1–20, 2023.
- [649] Jingqi Song, Yuanjie Zheng, Muhammad Zakir Ullah, Junxia Wang, Yanyun Jiang, Chenxi Xu, Zhenxing Zou, and Guocheng Ding. Multiview multimodal network for breast cancer diagnosis in contrast-enhanced spectral mammography images. *International Journal of Computer Assisted Radiology and Surgery*, 16(6):979–988, 2021.
- [650] Meng Song, Jonathan Greenbaum, Joseph Luttrell IV, Weihua Zhou, Chong Wu, Hui Shen, Ping Gong, Chaoyang Zhang, and Hong-Wen Deng. A review of integrative imputation for multi-omics datasets. *Frontiers in Genetics*, 11:570255, 2020.
- [651] Qi Song, Ye Yang, Dongxian Jiang, Zhaoyu Qin, Chen Xu, Haixing Wang, Jie Huang, Lingli Chen, Rongkui Luo, Xiaolei Zhang, et al. Proteomic analysis reveals key differences between squamous cell carcinomas and adenocarcinomas across multiple tissues. *Nature Communications*, 13(1):4167, 2022.
- [652] Qianqian Song, Jing Su, and Wei Zhang. scGCN is a graph convolutional networks algorithm for knowledge transfer in single cell omics. *Nature communications*, 12(1):3826, 2021.
- [653] David S Specht, Asim Waqas, Ghulam Rasool, Charles Clifford, and Nidhal Bouaynaya. Intelligent helipad detection and (grad-cam) estimation using satellite imagery. *Transportation Research Board*, 2021.
- [654] Olaf Sporns. Graph theory methods for the analysis of neural connectivity patterns. In Rolf Kötter, editor, *Neuroscience databases*, pages 171–185. Springer, 2003.

- [655] CJ Stam. Connected brains: introduction to graph theory. *Department of Clinical Neurophysiology, VU University Medical Centre*, 2013.
- [656] Claudio Stamile, Aldo Marzullo, and Enrico Deusebio. *Graph Machine Learning: Take graph data to the next level by applying machine learning techniques and algorithms*. Packt Publishing Ltd, 2021.
- [657] Stanford. Tutorial 3: Image segmentation. <https://ai.stanford.edu/~syyeung/cvweb/tutorial3.html>.
- [658] Stefan G Stark, Joanna Ficek, Francesco Locatello, Ximena Bonilla, Stéphane Chevrier, Franziska Singer, Gunnar Rätsch, and Kjong-Van Lehmann. SCIM: universal single-cell matching with unpaired feature sets. *Bioinformatics*, 36(Supplement\_2):i919–i927, 2020.
- [659] Paul A Stewart, Eric A Welsh, Robbert JC Slebos, Bin Fang, Victoria Izumi, Matthew Chambers, Guolin Zhang, Ling Cen, Fredrik Pettersson, Yonghong Zhang, et al. Proteogenomic landscape of squamous cell lung cancer. *Nature communications*, 10(1):1–17, 2019.
- [660] Sandra Steyaert, Marija Pizurica, Divya Nagaraj, Priya Khandelwal, Tina Hernandez-Boussard, Andrew J Gentles, and Olivier Gevaert. Multimodal data fusion for cancer biomarker discovery with deep learning. *Nature machine intelligence*, 5(4):351–362, 2023.
- [661] Linda Studer, Jannis Wallau, Heather Dawson, Inti Zlobec, and Andreas Fischer. Classification of intestinal gland cell-graphs using graph neural networks. In *2020 25th International conference on pattern recognition (ICPR)*, pages 3636–3643. IEEE, 2021.
- [662] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision*, pages 843–852, 2017.

- [663] Hao Sun, Jiaqing Liu, Shurong Chai, Zhaolin Qiu, Lanfen Lin, Xinyin Huang, and Yenwei Chen. Multi-Modal Adaptive Fusion Transformer Network for the estimation of depression level. *Sensors*, 21(14):4764, 2021.
- [664] Li Sun, Songtao Zhang, and Lin Luo. Tumor segmentation and survival prediction in glioma with deep learning. In *International MICCAI Brainlesion Workshop*, pages 83–93. Springer, 2018.
- [665] Mookund Sureka, Abhijeet Patil, Deepak Anand, and Amit Sethi. Visualization for histopathology images using graph convolutional neural networks. In *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 331–335. IEEE, 2020.
- [666] Dídac Surís, Sachit Menon, and Carl Vondrick. Vipergpt: Visual inference via python execution for reasoning. *arXiv preprint arXiv:2303.08128*, 2023. <https://arxiv.org/abs/2303.08128>.
- [667] Kyle Swanson, Eric Wu, Angela Zhang, Ash A Alizadeh, and James Zou. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell*, 2023. In Press, Corrected Proof available at: <https://doi.org/10.1016/j.cell.2023.01.035>.
- [668] Khajamoinuddin Syed, William C Sleeman IV, Michael Hagan, Jatinder Palta, Rishabh Kapoor, and Preetam Ghosh. Multi-view data integration methods for radiotherapy structure name standardization. *Cancers*, 13(8):1796, 2021.
- [669] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*, 2016.

- [670] 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.
- [671] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In IEEE, editor, *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 1–9, 2015.
- [672] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian J. Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.
- [673] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
- [674] Raheleh Talebi, Carlos A Celis-Morales, Abolfazl Akbari, Atefeh Talebi, Nasrin Borumandnia, and Mohamad Amin Pourhoseingholi. Machine learning-based classifiers to predict metastasis in colorectal cancer patients. *Frontiers in Artificial Intelligence*, 7, 2024.
- [675] Divyanshu Talwar, Aanchal Mongia, Debarka Sengupta, and Angshul Majumdar. Autoimpute: Autoencoder based imputation of single-cell rna-seq data. *Scientific reports*, 8(1):16329, 2018.
- [676] Jiajia Tang, Kang Li, Ming Hou, Xuanyu Jin, Wanzeng Kong, Yu Ding, and Qibin Zhao. MMT: Multi-way Multi-modal Transformer for Multimodal Learning. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 3458–3465. International Joint Conferences on Artificial Intelligence Organization, 2022.

- [677] Peng Tang, Xintong Yan, Yang Nan, Xiaobin Hu, Bjoern H Menzee Krammer, Tobias Lasser, et al. Joint-individual fusion structure with fusion attention module for multi-modal skin cancer classification. *arXiv preprint arXiv:2312.04189*, 2023.
- [678] Allen Tannenbaum, Chris Sander, Liangjia Zhu, Romeil Sandhu, Ivan Kolesov, Eduard Reznik, Yasin Senbabaoglu, and Tryphon Georgiou. Ricci curvature and robustness of cancer networks. *arXiv preprint arXiv:1502.04512*, 2015.
- [679] Shuchang Tao, Huawei Shen, Qi Cao, Liang Hou, and Xueqi Cheng. Adversarial immunization for certifiable robustness on graphs. In Liane Lewin-Eytan, David Carmel, Elad Yom-Tov, Eugene Agichtein, and Evgeniy Gabrilovich, editors, *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, pages 698–706. ACM, 2021.
- [680] Thomas E Tavolara, Ziyu Su, Metin N Gurcan, and M Khalid Khan Niazi. One label is all you need: Interpretable ai-enhanced histopathology for oncology. In *Seminars in Cancer Biology*. Elsevier, 2023.
- [681] EBI Gene Expression Team. Expression atlas. Software available from <https://www.ebi.ac.uk>.
- [682] Ratna Rajesh Thangudu, Paul A. Rudnick, Michael Holck, Deepak Singhal, Michael J. MacCoss, Nathan J. Edwards, Karen A. Ketchum, Christopher R. Kinsinger, Erika Kim, and Anand Basu. Abstract LB-242: Proteomic Data Commons: A resource for proteogenomic analysis. *Cancer Research*, 80(16\_Supplement):LB–242, 08 2020.
- [683] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. YFCC100M: The new data in multimedia research. *Communications of the ACM*, 59:64–73, 2016.

- [684] Vésteinn Thorsson, David L Gibbs, Scott D Brown, Denise Wolf, Dante S Bortone, Tai-Hsien Ou Yang, Eduard Porta-Pardo, Galen F Gao, Christopher L Plaisier, James A Eddy, et al. The immune landscape of cancer. *Immunity*, 48(4):812–830, 2018.
- [685] Zhiqiang Tian, Xiaojian Li, Yaoyue Zheng, Zhang Chen, Zhong Shi, Lizhi Liu, and Baowei Fei. Graph-convolutional-network-based interactive prostate segmentation in MR images. *Medical physics*, 47(9):4164–4176, 2020.
- [686] Hamid Reza Tizhoosh and Liron Pantanowitz. Artificial intelligence and digital pathology: challenges and opportunities. *Journal of Pathology Informatics*, 9(1):38, 2018.
- [687] Katarzyna Tomczak, Patrycja Czerwińska, and Maciej Wiznerowicz. Review The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemporary Oncology*, 2015(1):68–77, 2015.
- [688] Li Tong, Hang Wu, and May D Wang. Integrating multi-omics data by learning modality invariant representations for improved prediction of overall survival of cancer. *Methods*, 189:74–85, 2021.
- [689] Matteo Tortora, Ermanno Cordelli, Rosa Sicilia, Lorenzo Nibid, Edy Ippolito, Giuseppe Perrone, Sara Ramella, and Paolo Soda. Radiopathomics: Multimodal learning in non-small cell lung cancer for adaptive radiotherapy. *IEEE Access*, 2023.
- [690] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. LLaMA: Open and Efficient Foundation Language Models, 2023. <https://arxiv.org/abs/2302.13971>.
- [691] Aakash Tripathi, Asim Waqas, Kavya Venkatesan, Yasin Yilmaz, and Ghulam Rasool. Building flexible, scalable, and machine learning-ready multimodal oncology datasets. *Sensors*, 24(5):1634, 2024.

- [692] Aakash Tripathi, Asim Waqas, Kavya Venkatesan, Yasin Yilmaz, and Ghulam Rasool. Building flexible, scalable, and machine learning-ready multimodal oncology datasets. *Sensors*, 24(5):1634, 2024.
- [693] Aakash Tripathi, Asim Waqas, Yasin Yilmaz, and Ghulam Rasool. Honeybee: A scalable modular framework for creating multimodal oncology datasets with foundational embedding models. *arXiv preprint arXiv:2405.07460*, 2024.
- [694] Aakash Tripathi, Asim Waqas, Yasin Yilmaz, and Ghulam Rasool. Multimodal transformer model improves survival prediction in lung cancer compared to unimodal approaches. *Cancer Research*, 84(6\_Supplement):4905–4905, 2024.
- [695] Satvik Tripathi, Ethan Jacob Moyer, Alisha Isabelle Augustin, Alex Zavalny, Suhani Dheer, Rithvik Sukumaran, Daniel Schwartz, Brandon Gorski, Farouk Dako, and Edward Kim. RadGenNets: Deep learning-based radiogenomics model for gene mutation prediction in lung cancer. *Informatics in Medicine Unlocked*, 33:101062, 2022.
- [696] Satvik Tripathi, Ethan Jacob Moyer, Alisha Isabelle Augustin, Alex Zavalny, Suhani Dheer, Rithvik Sukumaran, Daniel Schwartz, Brandon Gorski, Farouk Dako, and Edward Kim. RadGenNets: Deep learning-based radiogenomics model for gene mutation prediction in lung cancer. *Informatics in Medicine Unlocked*, 33:101062, 2022.
- [697] Pei-Chen Tsai, Tsung-Hua Lee, Kun-Chi Kuo, Fang-Yi Su, Tsung-Lu Michael Lee, Eliana Marostica, Tomotaka Ugai, Melissa Zhao, Mai Chan Lau, Juha P Väyrynen, et al. Histopathology images predict multi-omics aberrations and prognoses in colorectal cancer patients. *Nature communications*, 14(1):2102, 2023.
- [698] Joseph Turian, Lev Ratinov, and Yoshua Bengio. Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 384–394, 2010.



- [699] Tomas Rakvåg Ulriksborg. Imputation of missing time series values using statistical and mathematical strategies. *Department of Informatics*, 2022.
- [700] Timothy Underwood. Pan-cancer analysis of whole genomes. *Nature*, 578(7793):82–93, 2020.
- [701] Hajime Uno, Tianxi Cai, Michael J Pencina, Ralph B D’Agostino, and Lee-Jen Wei. On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*, 30(10):1105–1117, 2011.
- [702] Diego Valsesia, Giulia Fracastoro, and Enrico Magli. RAN-GNNs: Breaking the Capacity Limits of Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [703] Rami S Vanguri, Jia Luo, Andrew T Aukerman, Jacklynn V Egger, Christopher J Fong, Nattally Horvat, Andrew Pagano, Jose de Arimateia Batista Araujo-Filho, Luke Geneslaw, Hira Rizvi, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L) 1 blockade in patients with non-small cell lung cancer. *Nature Cancer*, 3(10):1151–1164, 2022.
- [704] Elena V Varlamova, Maria A Butakova, Vlada V Semyonova, Sergey A Soldatov, Artem V Poltavskiy, Oleg I Kit, and Alexander V Soldatov. Machine learning meets cancer. *Cancers*, 16(6):1100, 2024.
- [705] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [706] 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.

- [707] Drew A Vecchio, Mark D Hammig, Xiongye Xiao, Anwesha Saha, Paul Bogdan, and Nicholas A Kotov. Spanning network gels from nanoparticles and graph theoretical analysis of their structure and properties. *Advanced Materials*, page 2201313, 2022.
- [708] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
- [709] Charles Vesteghem, Rasmus Froberg Brøndum, Mads Sønderkær, Mia Sommer, Alexander Schmitz, Julie Støve Bødker, Karen Dybkær, Tarec Christoffer El-Galaly, and Martin Bøgsted. Implementing the fair data principles in precision oncology: review of supporting initiatives. *Briefings in bioinformatics*, 21(3):936–945, 2020.
- [710] Mitko Veta, Paul J Van Diest, Stefan M Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, Anders BL Larsen, Jacob S Vestergaard, Anders B Dahl, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical Image Analysis*, 20(1):237–248, 2015.
- [711] Robert D. Vincent, Peter Neelin, Najmeh Khalili-Mahani, Andrew L. Janke, Vladimir S. Fonov, Steven M. Robbins, Leila Baghdadi, Jason Lerch, John G. Sled, Reza Adalat, David MacDonald, Alex P. Zijdenbos, D. Louis Collins, and Alan C. Evans. Minc 2.0: A flexible format for multi-modal images. *Frontiers in Neuroinformatics*, 10:35, 2016.
- [712] Francesco Visin, Marco Ciccone, Adriana Romero, Kyle Kastner, Kyunghyun Cho, Yoshua Bengio, Matteo Matteucci, and Aaron Courville. Reseg: A recurrent neural network-based model for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 41–48, 2016.
- [713] Francesco Visin, Kyle Kastner, Kyunghyun Cho, Matteo Matteucci, Aaron Courville, and Yoshua Bengio. Renet: A recurrent neural network based alternative to convolutional networks. *arXiv preprint arXiv:1505.00393*, 2015.

- [714] Minh Vu and My T Thai. PGM-Explainer: Probabilistic Graphical Model Explanations for Graph Neural Networks. *Advances in neural information processing systems*, 33:12225–12235, 2020.
- [715] Quoc Dang Vu, Kashif Rajpoot, Shan E Ahmed Raza, and Nasir Rajpoot. Handcrafted Histological Transformer (H2T): Unsupervised representation of whole slide images. *Medical Image Analysis*, page 102743, 2023.
- [716] Alex Waibel, Toshiyuki Hanazawa, Geoffrey Hinton, Kiyohiro Shikano, and Kevin J Lang. Phoneme recognition using time-delay neural networks. *IEEE transactions on acoustics, speech, and signal processing*, 37(3):328–339, 1989.
- [717] Lilapati Waikhom and Ripon Patgiri. A survey of graph neural networks in various learning paradigms: methods, applications, and challenges. *Artificial Intelligence Review*, pages 1–70, 2022.
- [718] Lijuan Wan, Zhuo Sun, Wenjing Peng, Sicong Wang, Jiangtao Li, Qing Zhao, Shuhao Wang, Han Ouyang, Xinming Zhao, Shuangmei Zou, et al. Selecting candidates for organ-preserving strategies after neoadjuvant chemoradiotherapy for rectal cancer: Development and validation of a model integrating mri radiomics and pathomics. *Journal of Magnetic Resonance Imaging*, 56(4):1130–1142, 2022.
- [719] C Wang, Martin Rajchl, ADC Chan, and Eranga Ukwatta. An ensemble of U-Net architecture variants for left atrial segmentation. In *Medical Imaging 2019: Computer-Aided Diagnosis*, volume 10950, page 109500M. International Society for Optics and Photonics, 2019. (San Diego, CA, USA).
- [720] Dongyang Wang, Junli Su, and Hongbin Yu. Feature extraction and analysis of natural language processing for deep learning english language. *IEEE Access*, 8:46335–46345, 2020.

- [721] Fei Wang, Liren Chen, Cheng Li, Shiyao Huang, Yanjie Chen, Chen Qian, and Chen Change Loy. The devil of face recognition is in the noise. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part IX*, volume 11213 of *Lecture Notes in Computer Science*, pages 780–795. Springer, 2018.
- [722] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *International MICCAI brainlesion workshop*, pages 178–190. Springer, 2017.
- [723] Hongxiao Wang, Yang Yang, Zhuo Zhao, Pengfei Gu, Nishchal Sapkota, and Danny Z Chen. Path-gptomic: A balanced multi-modal learning framework for survival outcome prediction. *arXiv preprint arXiv:2403.11375*, 2024.
- [724] Jingwen Wang, Richard J Chen, Ming Y Lu, Alexander Baras, and Faisal Mahmood. Weakly supervised prostate TMA classification via graph convolutional networks. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 239–243. IEEE, 2020.
- [725] Juexin Wang, Anjun Ma, Yuzhou Chang, Jianting Gong, Yuexu Jiang, Ren Qi, Cankun Wang, Hongjun Fu, Qin Ma, and Dong Xu. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nature communications*, 12(1):1882, 2021.
- [726] Renjie Wang, Weixing Dai, Jing Gong, Mingzhu Huang, Tingdan Hu, Hang Li, Kailin Lin, Cong Tan, Hong Hu, Tong Tong, et al. Development of a novel combined nomogram model integrating deep learning-pathomics, radiomics and immunoscore to predict postoperative outcome of colorectal cancer lung metastasis patients. *Journal of hematology & oncology*, 15(1):11, 2022.
- [727] Sheng Wang, Siqi Sun, Zhen Li, Renyu Zhang, and Jinbo Xu. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS computational biology*, 13(1):e1005324, 2017.

- [728] Tongxin Wang, Wei Shao, Zhi Huang, Haixu Tang, Jie Zhang, Zhengming Ding, and Kun Huang. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nature communications*, 12(1):3445, 2021.
- [729] Xiyue Wang, Jinxi Xiang, Jun Zhang, Sen Yang, Zhongyi Yang, Ming-Hui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. SCL-WC: Cross-slide contrastive learning for weakly-supervised whole-slide image classification. *Advances in Neural Information Processing Systems*, 35:18009–18021, 2022.
- [730] Yanan Wang, Yu Guang Wang, Changyuan Hu, Ming Li, Yanan Fan, Nina Otter, Ikuang Sam, Hongquan Gou, Yiqun Hu, Terry Kwok, et al. Cell graph neural networks enable the precise prediction of patient survival in gastric cancer. *NPJ precision oncology*, 6(1):45, 2022.
- [731] Zhenxing Wang, XiaoLiang Wu, and Yadong Wang. A framework for analyzing dna methylation data from illumina infinium humanmethylation450 beadchip. *BMC bioinformatics*, 19:15–22, 2018.
- [732] Zirui Wang, Jiahui Yu, Adams Wei Yu, Zihang Dai, Yulia Tsvetkov, and Yuan Cao. SimVLM: Simple visual language model pretraining with weak supervision. In *International Conference on Learning Representations*, 2022.
- [733] Asim Waqas, Marilyn M Bui, Eric F Glassy, Issam El Naqa, Piotr Borkowski, Andrew A Borkowski, and Ghulam Rasool. Revolutionizing digital pathology with the power of generative artificial intelligence and foundation models. *Laboratory Investigation*, page 100255, 2023.
- [734] Asim Waqas, Dimah Dera, Ghulam Rasool, Nidhal Carla Bouaynaya, and Hassan M Fathallah-Shaykh. Brain Tumor Segmentation and Surveillance with Deep Artificial Neural Networks. *Deep Learning for Biomedical Data Analysis*, pages 311–350, 2021.

- [735] Asim Waqas, Hamza Farooq, Nidhal C Bouaynaya, and Ghulam Rasool. Exploring Robust Architectures for Deep Artificial Neural Networks. *Communications Engineering*, 1(1):46, 2022.
- [736] Asim Waqas, Aakash Tripathi, Sabeen Ahmed, Ashwin Mukund, Hamza Farooq, Matthew B Schabath, Paul Stewart, Mia Naeini, and Ghulam Rasool. SeNMo: A Self-Normalizing Deep Learning Model for Enhanced Multi-Omics Data Analysis in Oncology. *arXiv e-prints*, pages arXiv–2405, 2024.
- [737] Asim Waqas, Aakash Tripathi, Sabeen Ahmed, Ashwin Mukund, Paul Stewart, Mia Naeini, Hamza Farooq, and Ghulam Rasool. Senmo: A self-normalizing deep learning model for enhanced multi-omics data analysis in oncology. *Cancer Research*, 84(6\_Supplement):908–908, 2024.
- [738] Asim Waqas, Aakash Tripathi, Ashwin Mukund, Paul Stewart, Mia Naeini, and Ghulam Rasool. Bio24-031: Hierarchical multimodal learning on pan-squamous cell carcinomas for improved survival outcomes. *Journal of the National Comprehensive Cancer Network*, 22(2.5), 2024.
- [739] Asim Waqas, Aakash Tripathi, Ravi P Ramachandran, Paul Stewart, and Ghulam Rasool. Multimodal Data Integration for Oncology in the Era of Deep Neural Networks: A Review. *arXiv preprint arXiv:2303.06471*, 2023. <https://arxiv.org/abs/2303.06471>.
- [740] Asim Waqas, Aakash Tripathi, Ravi P Ramachandran, Paul Stewart, and Ghulam Rasool. Multimodal Data Integration for Oncology in the Era of Deep Neural Networks: A Review. *arXiv preprint arXiv:2303.06471*, 2024. <https://arxiv.org/abs/2303.06471>.
- [741] Jakob Wasserthal, Hanns-Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, Daniel Boll, Joshy Cyriac, Shan Yang, Michael Bach, and Martin Segeroth. TotalSegmentator: robust segmentation of 104 anatomical structures in CT images, 2023. <https://arxiv.org/abs/2303.12961>.

- [742] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442, 1998.
- [743] Hongxin Wei, Lei Feng, Xiangyu Chen, and Bo An. Combating noisy labels by agreement: A joint training method with co-regularization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13726–13735, 2020.
- [744] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research*, 2022. Survey Certification.
- [745] Yinwei Wei, Xiang Wang, Liqiang Nie, Xiangnan He, Richang Hong, and Tat-Seng Chua. MMGCN: Multi-modal graph convolution network for personalized recommendation of micro-video. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1437–1445, 2019.
- [746] Amy Wells, Shaan Patel, Jason B Lee, and Kiran Motaparathi. Artificial intelligence in dermatopathology: Diagnosis, education, and research. *Journal of Cutaneous Pathology*, 48(8):1061–1068, 2021.
- [747] Hongzhi Wen, Jiayuan Ding, Wei Jin, Yiqi Wang, Yuying Xie, and Jiliang Tang. Graph neural networks for multimodal single-cell data integration. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4153–4163, 2022.
- [748] Zhuoyu Wen, Shidan Wang, Donghan M Yang, Yang Xie, Mingyi Chen, Justin Bishop, and Guanghua Xiao. Deep learning in digital pathology for personalized treatment plans of cancer patients. In *Seminars in Diagnostic Pathology*, volume 40, pages 109–119. Elsevier, 2023.

- [749] Martin J Willeminck, Holger R Roth, and Veit Sandfort. Toward Foundational Deep Learning Models for Medical Imaging in the New Era of Transformer Networks. *Radiology: Artificial Intelligence*, 4(6):e210284, 2022.
- [750] Stefan M. Willems, Sanne Abeln, K. Anton Feenstra, Remco de Bree, Egge F. van der Poel, Robert J. Baatenburg de Jong, Jaap Heringa, and Michiel W.M. van den Brekel. The potential use of big data in oncology. *Oral Oncology*, 98:8–12, 2019.
- [751] Michael Wornow, Yizhe Xu, Rahul Thapa, Birju Patel, Ethan Steinberg, Scott Fleming, Michael A. Pfeffer, Jason Fries, and Nigam H. Shah. The Shaky Foundations of Clinical Foundation Models: A Survey of Large Language Models and Foundation Models for EMRs, 2023. <https://arxiv.org/abs/2303.12961>.
- [752] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32:4–24, 2020.
- [753] Kun Xia and Jinzhuo Wang. Recent advances of transformers in medical image analysis: a comprehensive review. *MedComm–Future Medicine*, 2(1):e38, 2023.
- [754] Yu Xiang and Dieter Fox. Da-rnn: Semantic mapping with data associated recurrent neural networks. *arXiv preprint arXiv:1703.03098*, 2017.
- [755] Xiongye Xiao, Hanlong Chen, and Paul Bogdan. Deciphering the generating rules and functionalities of complex networks. *Scientific reports*, 11(1):1–15, 2021.
- [756] Feng Xie, Han Yuan, Yilin Ning, Marcus Eng Hock Ong, Mengling Feng, Wynne Hsu, Bibhas Chakraborty, and Nan Liu. Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. *Journal of Biomedical Informatics*, 126:103980, 2022.



- [757] Saining Xie, Alexander Kirillov, Ross B. Girshick, and Kaiming He. Exploring randomly wired neural networks for image recognition. In IEEE, editor, *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 1284–1293, 2019.
- [758] Yaochen Xie, Zhao Xu, Jingtun Zhang, Zhengyang Wang, and Shuiwang Ji. Self-supervised learning of graph neural networks: A unified review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2412–2429, 2022.
- [759] Yutong Xie, Jianpeng Zhang, Chunhua Shen, and Yong Xia. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 171–180. Springer, 2021.
- [760] Jiarui Xu, Shalini De Mello, Sifei Liu, Wonmin Byeon, Thomas Breuel, Jan Kautz, and Xiaolong Wang. GroupViT: Semantic Segmentation Emerges from Text Supervision. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18113–18123, 2022.
- [761] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.
- [762] P. Xu, X. Zhu, and D. A. Clifton. Multimodal learning with transformers: A survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 45(10):12113–12132, oct 2023.
- [763] Peng Xu, Xiatian Zhu, and David A. Clifton. Multimodal learning with transformers: A survey, 2023.
- [764] Xinying Xu, Guiqing Li, Gang Xie, Jinchang Ren, and Xinlin Xie. Weakly supervised deep semantic segmentation using cnn and elm with semantic candidate regions. *Complexity*, 2019, 2019.

- [765] Yang Xu, Priyojit Das, and Rachel Patton McCord. SMILE: mutual information learning for integration of single-cell omics data. *Bioinformatics*, 38(2):476–486, 2022.
- [766] Bin Yan and Mingtao Pei. Clinical-BERT: Vision-Language Pre-training for Radiograph Diagnosis and Reports Generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2982–2990, 2022.
- [767] Jianwei Yang, Chunyuan Li, Pengchuan Zhang, Bin Xiao, Ce Liu, Lu Yuan, and Jianfeng Gao. Unified Contrastive Learning in Image-Text-Label Space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19163–19173, June 2022.
- [768] Karren Dai Yang, Anastasiya Belyaeva, Saradha Venkatachalapathy, Karthik Damodaran, Abigail Katcoff, Adityanarayanan Radhakrishnan, GV Shivashankar, and Caroline Uhler. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nature communications*, 12(1):31, 2021.
- [769] Linyi Yang, Tin Lok James Ng, Barry Smyth, and Riu Hai Dong. HTML: Hierarchical Transformer-Based Multi-Task Learning for Volatility Prediction. In *Proceedings of The Web Conference 2020, WWW '20*, page 441–451, New York, NY, USA, 2020. Association for Computing Machinery.
- [770] Shuai Yang, Xianjie Guo, Kui Yu, Xiaoling Huang, Tingting Jiang, Jin He, and Lichuan Gu. Causal feature selection in the presence of sample selection bias. *ACM Transactions on Intelligent Systems and Technology*, 14(5):1–18, 2023.
- [771] Tianchi Yang, Linmei Hu, Chuan Shi, Houye Ji, Xiaoli Li, and Liqiang Nie. HGAT: Heterogeneous graph attention networks for semi-supervised short text classification. *ACM Transactions on Information Systems (TOIS)*, 39(3):1–29, 2021.

- [772] Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Anthony B Costa, Mona G Flores, et al. A large language model for electronic health records. *npj Digital Medicine*, 5(1):194, 2022.
- [773] Xinyu Yang, Dongmei Mu, Hao Peng, Hua Li, Ying Wang, Ping Wang, Yue Wang, and Siqi Han. Research and application of artificial intelligence based on electronic health records of patients with cancer: systematic review. *JMIR Medical Informatics*, 10(4):e33799, 2022.
- [774] Jordan Yap, William Yolland, and Philipp Tschandl. Multimodal skin lesion classification using deep learning. *Experimental dermatology*, 27(11):1261–1267, 2018.
- [775] Hai-Cheng Yi, Zhu-Hong You, De-Shuang Huang, and Chee Keong Kwoh. Graph representation learning in bioinformatics: trends, methods and applications. *Briefings in Bioinformatics*, 23(1):bbab340, 2022.
- [776] Joonyoung Yi, Juhyuk Lee, Kwang Joon Kim, Sung Ju Hwang, and Eunho Yang. Why not to use zero imputation? correcting sparsity bias in training neural networks. *arXiv preprint arXiv:1906.00150*, 2019.
- [777] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. GN-Explainer: Generating Explanations for Graph Neural Networks. *Advances in neural information processing systems*, 32, 2019.
- [778] Jiaxuan You, Jure Leskovec, Kaiming He, and Saining Xie. Graph structure of neural networks. In PMLR, editor, *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 10881–10891, 2020.
- [779] Safoora Yousefi, Fatemeh Amrollahi, Mohamed Amgad, Chengliang Dong, Joshua E Lewis, Congzheng Song, David A Gutman, Sameer H Halani, Jose Enrique Velazquez Vega, Daniel J Brat, et al. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. *Scientific reports*, 7(1):1–11, 2017.

- [780] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.
- [781] Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. CoCa: Contrastive Captioners are Image-Text Foundation Models. *Transactions on Machine Learning Research*, Aug 2022, 2022.
- [782] Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. Coca: Contrastive captioners are image-text foundation models. *Transactions on Machine Learning Research*, 2022.
- [783] Hao Yuan, Haiyang Yu, Jie Wang, Kang Li, and Shuiwang Ji. On explainability of graph neural networks via subgraph explorations. In *International Conference on Machine Learning*, pages 12241–12252. PMLR, 2021.
- [784] Yading Yuan, Ming Chao, and Yeh-Chi Lo. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE transactions on medical imaging*, 36(9):1876–1886, 2017.
- [785] Yuhui Yuan and Jingdong Wang. Ocnet: Object context network for scene parsing. *arXiv preprint arXiv:1809.00916*, 2018.
- [786] Song Yuheng and Yan Hao. Image segmentation algorithms overview. *arXiv preprint arXiv:1707.02051*, 2017.
- [787] Yuhang Zang, Wei Li, Kaiyang Zhou, Chen Huang, and Chen Change Loy. Open-Vocabulary DETR With Conditional Matching. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part IX*, page 106–122, Berlin, Heidelberg, 2022. Springer-Verlag.

- [788] Mark D Zarella, Douglas Bowman, Famke Aeffner, Navid Farahani, Albert Xthona, Syeda Fatima Absar, Anil Parwani, Marilyn Bui, and Douglas J Hartman. A practical guide to whole slide imaging: a white paper from the digital pathology association. *Archives of pathology & laboratory medicine*, 143(2):222–234, 2019.
- [789] Valentina A Zavala, Paige M Bracci, John M Carethers, Luis Carvajal-Carmona, Nicole B Coggins, Marcia R Cruz-Correa, Melissa Davis, Adam J de Smith, Julie Dutil, Jane C Figueiredo, et al. Cancer health disparities in racial/ethnic minorities in the united states. *British journal of cancer*, 124(2):315–332, 2021.
- [790] Yuansong Zeng, Xiang Zhou, Jiahua Rao, Yutong Lu, and Yuedong Yang. Accurately clustering single-cell RNA-seq data by capturing structural relations between cells through graph convolutional network. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020.
- [791] Zitao Zeng, Weihao Xie, Yunzhe Zhang, and Yao Lu. Ric-unet: An improved neural network based on unet for nuclei segmentation in histology images. *Ieee Access*, 7:21420–21428, 2019.
- [792] He Zhang, Bang Wu, Xingliang Yuan, Shirui Pan, Hanghang Tong, and Jian Pei. Trustworthy graph neural networks: Aspects, methods and trends. *arXiv preprint arXiv:2205.07424*, 2022.
- [793] Jianxin Zhang, Xiaogang Lv, Qiule Sun, Qiang Zhang, Xiaopeng Wei, and Bin Liu. Sdresunet: Separable and dilated residual u-net for mri brain tumor segmentation. *Current Medical Imaging Formerly Current Medical Imaging Reviews*, 15, 08 2019.
- [794] Jieqiu Zhang, Qi Wu, Wei Yin, Lu Yang, Bo Xiao, Jianmei Wang, and Xiaopeng Yao. Development and validation of a radiopathomic model for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer patients. *BMC cancer*, 23(1):1–12, 2023.

- [795] Ning Zhang. Learning adversarial transformer for symbolic music generation. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–10, 2020.
- [796] Richard Zhang. Making convolutional networks shift-invariant again. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 7324–7334. PMLR, 2019.
- [797] Wentao Zhang, Yu Shen, Zheyu Lin, Yang Li, Xiaosen Li, Wen Ouyang, Yangyu Tao, Zhi Yang, and Bin Cui. Pasca: A graph neural architecture search system under the scalable paradigm. In Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini, editors, *Proceedings of the ACM Web Conference 2022*, pages 1817–1828, 2022.
- [798] Yu-Dong Zhang, Suresh Chandra Satapathy, David S Guttery, Juan Manuel Górriz, and Shui-Hua Wang. Improved breast cancer classification through combining graph convolutional network and convolutional neural network. *Information Processing & Management*, 58(2):102439, 2021.
- [799] Zhenhua Zhang, Enqi Chen, Xueyuan Zhang, Jing Yang, Xianqi Wang, PinZhen Chen, Meng Zeng, Mingshan Du, Senlin Xu, Zhiqing Yang, Fei Ren, and Wei Chen. Multi-modal fusion of radiomics and pathomics to predict the survival of pancreatic cancer patients based on asymmetric twinning information interaction network. *SSRN Electronic Journal*, 2022. <http://dx.doi.org/10.2139/ssrn.4260135>.
- [800] Ziqi Zhang, Chengkai Yang, and Xiuwei Zhang. scDART: integrating unmatched scRNA-seq and scATAC-seq data and learning cross-modality relationship simultaneously. *Genome Biology*, 23(1):139, 2022.
- [801] Bin Zhao, Maoguo Gong, and Xuelong Li. Hierarchical multimodal transformer to summarize videos. *Neurocomputing*, 468:360–369, 2022.

- [802] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.
- [803] Mingyang Zhao, Xiaoshui Huang, Jingen Jiang, Luntian Mou, Dong-Ming Yan, and Lei Ma. Accurate Registration of Cross-Modality Geometry via Consistent Clustering. *IEEE Transactions on Visualization and Computer Graphics*, 2023.
- [804] Ning Zhao, Maozu Guo, Kuanquan Wang, Chunlong Zhang, and Xiaoyan Liu. Identification of pan-cancer prognostic biomarkers through integration of multi-omics data. *Frontiers in Bioengineering and Biotechnology*, 8:268, 2020.
- [805] Xin Zhao, Zeru Zhang, Zijie Zhang, Lingfei Wu, Jiayin Jin, Yang Zhou, Ruoming Jin, Dejing Dou, and Da Yan. Expressive 1-lipschitz neural networks for robust multiple graph learning against adversarial attacks. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 12719–12735. PMLR, 2021.
- [806] Yingdong Zhao, Ming-Chung Li, Mariam M Konaté, Li Chen, Biswajit Das, Chris Karlovich, P Mickey Williams, Yvonne A Evrard, James H Doroshov, and Lisa M McShane. Tpm, fpkm, or normalized counts? a comparative study of quantification measures for the analysis of rna-seq data from the nci patient-derived models repository. *Journal of translational medicine*, 19(1):269, 2021.
- [807] Yuanshen Zhao, Longsong Li, Ke Han, Tao Li, Jingxian Duan, Qiuchang Sun, Chaofan Zhu, Dong Liang, Ningli Chai, and Zhi-Cheng Li. A radio-pathologic integrated model for prediction of lymph node metastasis stage in patients with gastric cancer. *Abdominal Radiology*, 48(11):3332–3342, 2023.

- [808] Zhi Zhao, John Zobolas, Manuela Zucknick, and Tero Aittokallio. Tutorial on survival modeling with applications to omics data. *Bioinformatics*, page btae132, 2024.
- [809] Y. Zhong, J. Yang, P. Zhang, C. Li, N. Codella, L. Li, L. Zhou, X. Dai, L. Yuan, Y. Li, and J. Gao. RegionCLIP: Region-based Language-Image Pretraining. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16772–16782, Los Alamitos, CA, USA, jun 2022. IEEE Computer Society.
- [810] Zeyun Zhong, David Schneider, Michael Voit, Rainer Stiefelhagen, and Jürgen Beyerer. Anticipative Feature Fusion Transformer for Multi-Modal Action Anticipation. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6057–6066, 2023.
- [811] Ce Zhou, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, Hao Peng, Jianxin Li, Jia Wu, Ziwei Liu, Pengtao Xie, Caiming Xiong, Jian Pei, Philip S. Yu, and Lichao Sun. A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT, 2023. <https://arxiv.org/abs/2302.09419>.
- [812] Chong Zhou, Chen Change Loy, and Bo Dai. Extract Free Dense Labels From CLIP. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVIII*, page 696–712, Berlin, Heidelberg, 2022. Springer-Verlag.
- [813] Daquan Zhou, Zhiding Yu, Enze Xie, Chaowei Xiao, Anima Anandkumar, Jiashi Feng, and Jose M Alvarez. Understanding the robustness in vision transformers. *arXiv preprint arXiv:2204.12451*, 2022.
- [814] Kaixiong Zhou, Zirui Liu, Keyu Duan, and Xia Hu. *Graph Neural Networks: AutoML*, pages 371–389. Springer Nature Singapore, Singapore, 2022.



- [815] Xingyi Zhou, Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, and Ishan Misra. Detecting twenty-thousand classes using image-level supervision. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part IX*, page 350–368, Berlin, Heidelberg, 2022. Springer-Verlag.
- [816] Yanning Zhou, Simon Graham, Navid Alemi Koohbanani, Muhammad Shaban, Pheng-Ann Heng, and Nasir Rajpoot. CGC-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.
- [817] Heqin Zhu, Xu Sun, Yuexiang Li, Kai Ma, S Kevin Zhou, and Yefeng Zheng. Dftr: Depth-supervised fusion transformer for salient object detection. *arXiv preprint arXiv:2203.06429*, 2022.
- [818] Wen Zhu, Yiwen Chen, Shanling Nie, and Hai Yang. Samms: Multi-modality deep learning with the foundation model for the prediction of cancer patient survival. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 3662–3668. IEEE, 2023.
- [819] Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. A robustly optimized BERT pre-training approach with post-training. In Sheng Li, Maosong Sun, Yang Liu, Hua Wu, Kang Liu, Wanxiang Che, Shizhu He, and Gaoqi Rao, editors, *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–1227, Huhhot, China, August 2021. Chinese Information Processing Society of China.
- [820] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, 2018.
- [821] Daniel Zügner, Oliver Borchert, Amir Akbarnejad, and Stephan Günnemann. Adversarial attacks on graph neural networks: Perturbations and their patterns. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(5):1–31, 2020.

## Appendix A: Copyright Permissions

The permission below is for the reproduction of material in Chapter 2.

Permissions requests | Nature Portfolio

<https://www.nature.com/nature-portfolio/reprints-and-permissions/perm...>

training materials (including continuing medical education), promotional materials, and web sites. Some permission requests can be granted free of charge, others carry a fee.

Springer Nature does not allow PDFs of full papers to be reproduced online, however e-print PDFs can be [purchased as commercial reprints](#). If you wish to purchase multiple stand-alone copies of a Nature Portfolio paper, which is then printed and shipped to you, please go to [commercial reprints](#).

[Top of page ↑](#)

### **Get permission to reuse Springer Nature content online**

#### **Permission requests from authors**

The author of articles published by Springer Nature do not usually need to seek permission for re-use of their material as long as the journal is credited with initial publication.

Ownership of copyright in original research articles remains with the Author, and provided that, when reproducing the contribution or extracts from it or from the Supplementary Information, the Author acknowledges first and reference publication in the Journal, the Author retains the following non-exclusive rights:

To reproduce the contribution in whole or in part in any printed volume (book or thesis) of which they are the author(s).

The author and any academic institution where they work at the time may reproduce the contribution for the purpose of course

6/1/24, 10:07 PM

## The permission below is for the reproduction of material in Chapter 3.

How-to: Get permission to reuse Springer content online | Springer — I... <https://www.springer.com/us/rights-permissions/obtaining-permissions/882>

at Copyright Clearance Center via phone +1-855-239-3415 or +1-978-646-2777 or email [springernaturesupport@copyright.com](mailto:springernaturesupport@copyright.com).

### How to obtain permission to reuse Springer Nature content not available online on SpringerLink

Requests for permission to reuse content (e.g. figure or table, abstract, text excerpts) from Springer Nature publications currently not available online must be submitted in writing. Please be as detailed and specific as possible about what, where, how much, and why you wish to reuse the content.

#### Your contacts to obtain permission for the reuse of material from:

- books: [bookpermissions@springernature.com](mailto:bookpermissions@springernature.com)
- journals: [journalpermissions@springernature.com](mailto:journalpermissions@springernature.com)

### Author reuse

Please check the Copyright Transfer Statement (CTS) or Licence to Publish (LTP) that you have signed with Springer Nature to find further information about the reuse of your content.

Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution. Authors must properly cite the published article in their thesis according to current citation standards.

Material from: 'AUTHOR, TITLE, JOURNAL TITLE, published [YEAR], [publisher - as it appears on our copyright page]'

If you are any doubt about whether your intended re-use is covered, please contact [journalpermissions@springernature.com](mailto:journalpermissions@springernature.com) for confirmation.

6/1/24, 10:13 PM

The permission below is for the reproduction of material in Chapter 4.

Copyright | Elsevier policy

<https://www.elsevier.com/about/policies-and-standards/copyright>

other works.		
<p>Use and share their works for scholarly purposes (with full acknowledgement of the original article):</p> <ol style="list-style-type: none"> <li>1. In their own classroom teaching. Electronic and physical distribution of copies is permitted</li> <li>2. If an author is speaking at a conference, they can present the article and distribute copies to the attendees</li> <li>3. Distribute the article, including by email, to their students and to research colleagues who they know for their personal use</li> <li>4. Share and publicize the article via Share Links, which offers 50 days' free access for anyone, without signup or registration</li> <li>5. Include in a thesis or dissertation (provided this is not published commercially)</li> <li>6. Share copies of their article privately as part of an invitation-only work group on commercial sites with which the publisher has a hosting agreement</li> </ol>	√	√
Publicly share the preprint on any website or repository at any time.	√	√
Publicly share the accepted manuscript on non-commercial sites	√	√ using a CC BY-NC-ND license and usually only after an embargo

6/1/24, 10:09 PM

## The permission below is for the reproduction of material in Chapter 5.

BIR Publications

<https://www.birpublications.org/page/permissions#thesis>

- **Creative commons attribution licence (CC BY)**

If the article is open access under a creative commons attribution licence (CC BY) you are free to reuse, change and build upon the work, even commercially. Your new work must acknowledge the original source.

---

### Are you the author of the original article?

- ***BJR*, *DMFR* and *Imaging***

Authors of *BJR*, *DMFR* and *Imaging*, articles are free to reuse figures, tables, charts and short sections of text (up to 350 words) from their article in other publications, free of charge providing they cite the original source article.

- ***BJR* | *case reports* and *BJR* | *Open***

*BJR* | *case reports* and *BJR* | *Open* articles are open access under a creative commons attribution licence (CC BY), so authors (and third parties) are free to reuse, change and build upon the work, even commercially. The new work must acknowledge the original source.

---

### Do you want to reproduce content from the article in a PhD thesis?

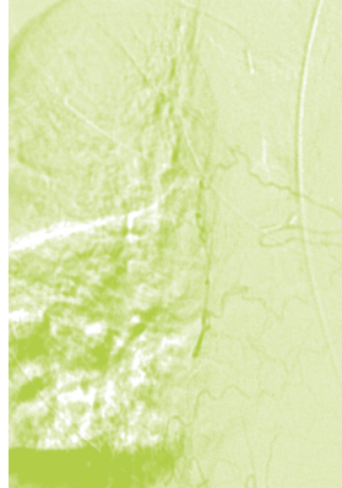
If you are the author of the original article and wish to reproduce the article in full or in part in your PhD thesis you are free to do so providing you cite the original source article.

If you are not the author of the original article but wish reproduce 1-2 figures, tables, charts and short sections of text (up to 350 words) from a *BJR*, *DMFR* or *Imaging* article, you are free to do so providing you cite the original source article. Requests to reproduce

Advertising policy

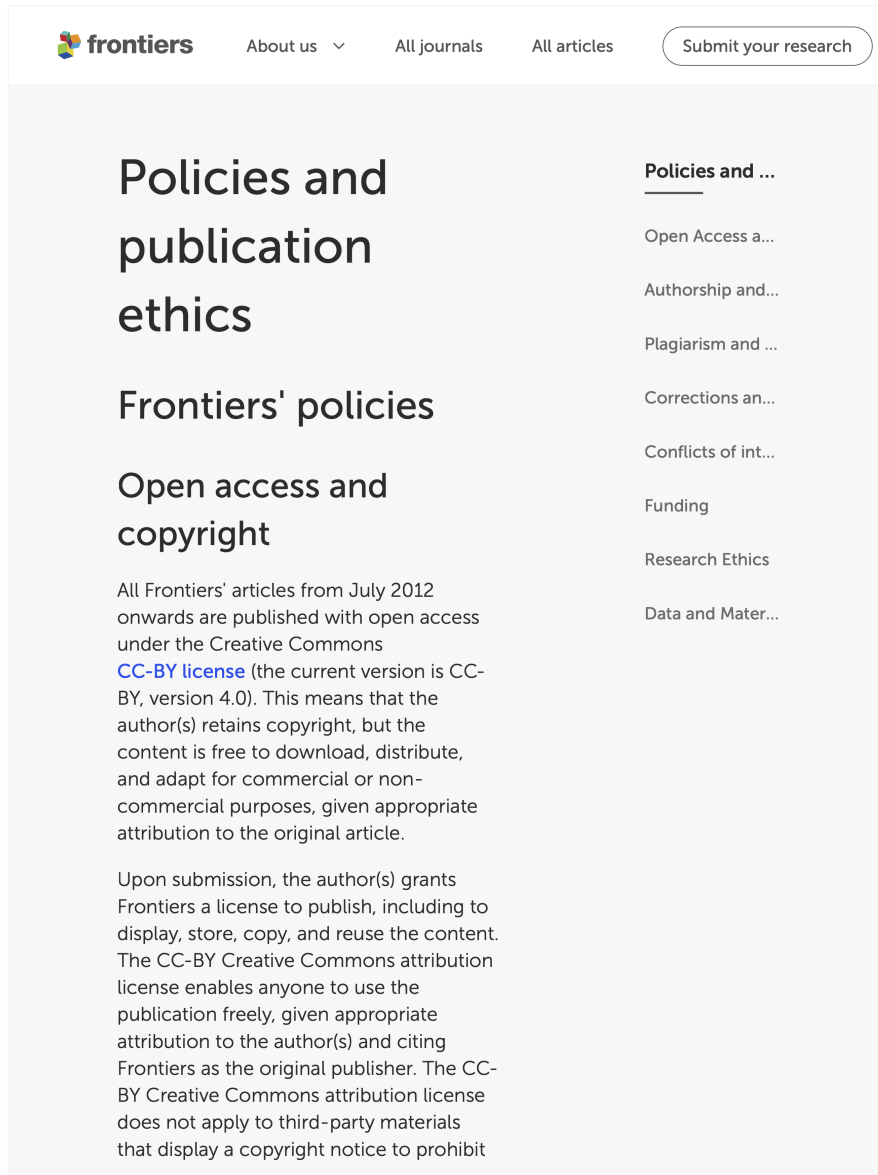
Equality and Diversity

Permissions



6/1/24, 10:55 PM

The permission below is for the reproduction of material in Chapter 6.



The screenshot shows the top navigation bar of the Frontiers website. It includes the Frontiers logo, a dropdown menu for 'About us', links for 'All journals' and 'All articles', and a 'Submit your research' button. The main content area features a large heading 'Policies and publication ethics' and a sub-heading 'Frontiers' policies'. Below this, there is a section titled 'Open access and copyright' with two paragraphs of text. A right-hand sidebar contains a list of links: 'Policies and ...', 'Open Access a...', 'Authorship and...', 'Plagiarism and ...', 'Corrections an...', 'Conflicts of int...', 'Funding', 'Research Ethics', and 'Data and Mater...'. The date and time '6/1/24, 10:19 PM' are displayed in the bottom right corner of the screenshot.

**frontiers** About us ▾ All journals All articles [Submit your research](#)

# Policies and publication ethics

## Frontiers' policies

### Open access and copyright

All Frontiers' articles from July 2012 onwards are published with open access under the Creative Commons [CC-BY license](#) (the current version is CC-BY, version 4.0). This means that the author(s) retains copyright, but the content is free to download, distribute, and adapt for commercial or non-commercial purposes, given appropriate attribution to the original article.

Upon submission, the author(s) grants Frontiers a license to publish, including to display, store, copy, and reuse the content. The CC-BY Creative Commons attribution license enables anyone to use the publication freely, given appropriate attribution to the author(s) and citing Frontiers as the original publisher. The CC-BY Creative Commons attribution license does not apply to third-party materials that display a copyright notice to prohibit

**Policies and ...**  
Open Access a...  
Authorship and...  
Plagiarism and ...  
Corrections an...  
Conflicts of int...  
Funding  
Research Ethics  
Data and Mater...

6/1/24, 10:19 PM

The permission below is for the reproduction of material in Chapter 7.

## Manuscript Submission & Instructions for Authors

Sensors ▼

Instructions for Authors   Aims & Scope   Submit Manuscript

### Authors and Readers Benefit from MDPI's Pledges to:

- **publish thoroughly peer-reviewed journals** of high scholarly impact
- **maintain quick publication** — manuscripts are published within 5-7 weeks of submission (provided no major revisions are required)
- **publish full open access journals** — readers can access all content published on this platform for free
- **publish citation-tracked journals** — MDPI continuously works towards quick coverage and citation-tracking of all of its journals in the top databases of **Scopus**, **Web of Science**, **PMC**, **PubMed**, and **MEDLINE**, along with various other databases

### For Authors and Readers Open Access Means:

Read the full open access information here.

- **free availability** of the literature without any subscription or price barriers
- **immediate access** once an article is published (no embargo period)
- **authors retain all copyrights** - authors will not be forced to sign any copyright transfer agreements
- **permission of re-use** of the published material if proper accreditation is given (**Creative Commons Attribution License** )

## **About the Author**

Asim Waqas is a computer engineer by education. With a focus on Machine Learning, he has contributed significantly to the advancement of deep learning, and cancer research.

Asim holds a Ph.D. in Electrical Engineering from the University of South Florida. Over the years, his research has been published in numerous prestigious journals and conferences, including Springer, Nature Communications Engineering, MDPI, Elsevier, Frontiers, BJR, NCCN, and AACR.

His groundbreaking work in graphs-theoretic assessment of robustness of deep neural networks and using graph neural networks for multimodal learning on oncology data has not only garnered significant academic recognition but also led to increased interest by the research community.

Outside of academia, Asim enjoys hiking, reading, and traveling, reflecting his well-rounded approach to life and work.