
Deep Learning and Single-cell Omics in Colorectal Cancer: A Survey

www.surveyx.cn

Abstract

Colorectal cancer (CRC) presents a significant global health challenge due to its genetic and molecular complexity, necessitating advancements in precision medicine. This survey explores the convergence of deep learning and single-cell omics in CRC research, highlighting their transformative potential in diagnostics and therapeutics. The integration of deep learning models, such as convolutional neural networks and attention-based architectures, with single-cell omics data provides a granular understanding of tumor heterogeneity, facilitating the identification of novel biomarkers and therapeutic targets. Key advancements include improved diagnostic accuracy through enhanced image-based classification and the development of predictive models leveraging multi-omics data. Despite challenges in data integration and computational complexity, innovative approaches like multi-modal fusion and advanced feature selection demonstrate the potential to overcome these barriers. Future directions emphasize the need for robust frameworks to integrate diverse datasets, improve model interpretability, and ensure generalizability across clinical scenarios. By harnessing interdisciplinary methodologies, this integration promises to advance precision medicine, offering personalized treatment strategies and improving patient outcomes in CRC research.

1 Introduction

1.1 Colorectal Cancer: A Major Health Challenge

Colorectal cancer (CRC) ranks among the most common cancers worldwide, significantly contributing to cancer-related mortality [1]. Its complexity arises from genetic, molecular, and histopathological heterogeneity, complicating treatment and prognosis [2]. This situation necessitates a shift towards precision medicine to effectively address CRC's multifaceted nature [2].

Despite available screening methods like colonoscopy, their underutilization exacerbates the health challenge posed by CRC [3]. Early detection through timely screening and polyp removal can significantly lower mortality rates. The integration of advanced computational techniques is essential to enhance the precision of these diagnostic methods [1].

Investigating glycolysis-related genes and their influence on the immune microenvironment is vital for understanding CRC pathophysiology and developing targeted therapies [4]. The persistent issue of CRC relapse post-surgery highlights the limitations of traditional predictive methods, underscoring the urgent need for innovative, interdisciplinary approaches that incorporate advanced computational technologies to mitigate its impact [2].

1.2 Structure of the Survey

This survey provides a comprehensive exploration of the intersection between deep learning and single-cell omics in colorectal cancer research. The initial sections establish the significance of CRC as a major health challenge and the potential of integrating advanced computational techniques to

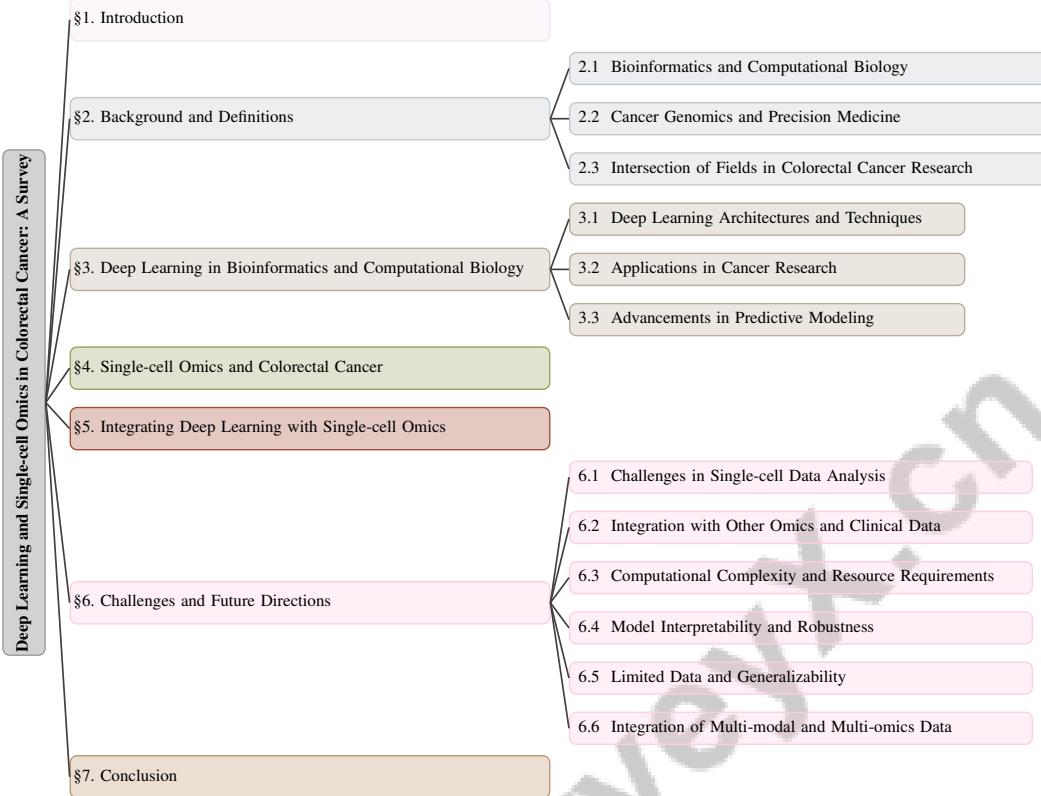


Figure 1: chapter structure

address its complexities. Following the introduction, Section 2 presents background information and definitions, offering an overview of key concepts such as deep learning, single-cell omics, bioinformatics, cancer genomics, computational biology, and precision medicine.

Section 3 examines the role of deep learning in bioinformatics and computational biology, focusing on its application in analyzing large-scale omics data. It discusses various deep learning architectures, techniques, and recent advancements in cancer research, particularly in CRC. Section 4 highlights the importance of single-cell omics for understanding CRC heterogeneity and identifying novel biomarkers and therapeutic targets.

The integration of deep learning with single-cell omics is the focus of Section 5, which explores interdisciplinary approaches to enhance biomarker prediction and classification. This section also emphasizes the integration of omics data with radiomic and histopathology data for comprehensive analysis. Section 6 identifies current challenges, such as computational complexity and resource demands, while discussing potential future directions and innovations to address these issues. The survey concludes with a summary of key points, highlighting the transformative potential of combining deep learning with single-cell omics in advancing colorectal cancer research.

This structured approach aims to provide a clear roadmap for readers, facilitating a deeper understanding of how these interdisciplinary fields converge to enhance precision medicine strategies in colorectal cancer research [5]. The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Bioinformatics and Computational Biology

Bioinformatics and computational biology are pivotal in advancing cancer research by providing sophisticated tools for analyzing complex biological data, thereby enhancing our understanding of cancer genomics and gene properties, including genomic characteristics and regulatory functions [6]. These fields support the development of computational tools for organoid data analysis, crucial for

elucidating disease mechanisms [7]. In colorectal cancer (CRC), bioinformatics facilitates automated diagnostics, such as texture classification in histological images [8], and integrates genetic algorithms with deep learning for innovative drug discovery approaches [9].

Advanced multiplexed imaging technologies, like CODEX, enable detailed tissue microenvironment analysis at the single-cell level, underscoring bioinformatics' role in managing large-scale imaging data [10]. Techniques such as scRank XMBD demonstrate bioinformatics' application in analyzing single-cell RNA-seq data to identify prognostic gene signatures and cell subpopulations linked to CRC recurrence risk [11]. Furthermore, bioinformatics addresses inadequate risk stratification in CRC by incorporating diverse histopathological features into predictive models [12]. Tools like BetterNet, a convolutional neural network for polyp segmentation, exemplify how advanced computational techniques enhance diagnostic accuracy through residual learning and attention mechanisms [13].

Frameworks such as Bio-Bi-LSTM-CRF illustrate deep learning's integration into bioinformatics, utilizing Bidirectional Long Short-Term Memory networks with Conditional Random Fields for extracting clinical entities from unstructured text [14]. Additionally, methods like Directed Information-based Motif Discovery (DTI-MD) analyze gene expression data, emphasizing bioinformatics' critical role in cancer research [15]. The categorization of research based on invasiveness, sensitivity, and specificity further highlights bioinformatics and computational biology's relevance in CRC research [3]. As these fields evolve, their contributions to cancer research are expected to expand, offering new insights and solutions that propel the pursuit of precision medicine.

2.2 Cancer Genomics and Precision Medicine

Cancer genomics and precision medicine are essential for advancing colorectal cancer (CRC) research by elucidating the genetic and molecular complexities of the disease, thus enabling personalized treatment strategies. Comprehensive analyses of genetic alterations in cancer cells are vital for understanding tumorigenesis and disease progression, with the classification of CRC into microsatellite instability (MSI) and microsatellite stable (MSS) subtypes guiding treatment decisions [16]. Transformer-based methodologies have improved MSI status prediction, addressing limitations of traditional convolutional neural networks (CNNs) [17].

Precision medicine leverages genomic insights to tailor medical interventions to individual genetic and molecular profiles, integrating diverse datasets, including genomic, histopathological, and lifestyle factors, to enhance diagnostic accuracy and therapeutic efficacy [18]. The analysis of whole-slide images (WSIs) for CRC survival prediction highlights precision medicine's role in managing complex datasets to inform clinical decision-making [19]. Incorporating immune cell heterogeneity within the tumor microenvironment (TME) into diagnostic models exemplifies precision medicine's expanding scope [20].

Identifying candidate genes and chromatin regulators that contribute to CRC is crucial for advancing precision medicine, aiding in targeted therapy development and enhancing understanding of disease mechanisms [21]. The integration of AI technologies in CRC diagnosis underscores the relevance of cancer genomics and precision medicine in improving diagnostic workflows and treatment outcomes [22]. AI-driven models analyzing histopathological datasets, such as the NCT-CRC-HE, address data quality and biases, enhancing CRC management precision [23].

Precision medicine also emphasizes effective biomarkers for guiding immune checkpoint inhibitors (ICIs) application in CRC, as only a small subset of patients benefit from existing treatments [24]. Integrating high-dimensional omics data with established clinical covariates improves cancer prognosis models, particularly in CRC, by enhancing predictive power and interpretability [25]. As the field evolves, the integration of diverse data types and advanced computational methods promises to enhance CRC management precision, ultimately improving patient outcomes through personalized therapeutic interventions.

2.3 Intersection of Fields in Colorectal Cancer Research

The integration of deep learning, single-cell omics, and bioinformatics has significantly advanced understanding and treatment of colorectal cancer (CRC). Deep learning models, such as convolutional neural networks (CNNs), enhance diagnostic accuracy by enabling automatic classification of colon cancer from confocal laser microscopy images [26]. These models are optimized through genomic

variations, leading to 'biologically-primed' CNNs that improve CRC subtype classification from histopathological images [16].

Single-cell omics provides a detailed view of cellular heterogeneity and interactions within the tumor microenvironment, crucial for identifying distinct cellular subpopulations and their roles in CRC progression. Current methods for predicting recurrence risk often fall short due to diverse cellular compositions within tumors, leading to inconsistent prognostic outcomes [11]. The categorization of feature selection methods into filters, wrappers, embedded, ensemble, and hybrid approaches exemplifies their utility in analyzing single-cell data, facilitating meaningful insights extraction from complex datasets [27].

Bioinformatics serves as the computational backbone that integrates heterogeneous omics datasets, including genomic, transcriptomic, and microbiome data, to provide a comprehensive understanding of CRC. Methods like DTI-MD, which employs directed information for feature selection in tissue-specific gene expression analysis, underscore the intersection of these fields [15]. Innovative applications of bioinformatics in understanding microbial aspects of CRC, exemplified by extreme value analysis (EVA), identify major microbial signatures associated with CRC progression [28].

This intersection is further enriched by machine learning paradigms, such as MLP, addressing the spatio-temporal complexity of cancer cell responses, highlighting the synergy between deep learning, single-cell analysis, and phenotypic investigation [29]. Advanced models like BetterNet, designed for improved segmentation accuracy in medical imaging, capture complex features of polyps, demonstrating the intersection of these fields [13].

This interdisciplinary approach not only advances CRC understanding but also drives innovative diagnostic and therapeutic strategies. By harnessing deep learning's capabilities for feature extraction, insights from single-cell omics, and robust computational resources from bioinformatics, researchers significantly enhance the field of precision medicine. This integration facilitates personalized treatment strategies tailored to individual patient profiles, ultimately improving health outcomes. Recent studies indicate that interpretable deep learning models can effectively analyze complex single-cell data, enhancing understanding of molecular regulators and improving prediction transparency. Additionally, natural language processing (NLP) models in digital pathology have shown superior performance in biomarker prediction compared to traditional CNNs, particularly in scenarios with limited training data, indicating a promising shift towards more effective and personalized diagnostic approaches in clinical settings [30, 31].

3 Deep Learning in Bioinformatics and Computational Biology

3.1 Deep Learning Architectures and Techniques

Method Name	Architectural Variants	Application Domains	Enhanced Capabilities
ASP[32]	Deep Learning Classifier	Liver Segmentation Tasks	Improved Segmentation Accuracy
BBL-CRF[14]	Bi-LSTM-CRF	Clinical Entity Extraction	Improved Contextual Understanding
eSPA[33]	Lstm	Single-cell Rna	Cross-validated Label

Table 1: Table 1 presents a comparative analysis of various deep learning methods employed in bioinformatics and computational biology. The table details the architectural variants, application domains, and enhanced capabilities of each method, illustrating their contributions to advancing precision in biological data analysis.

Deep learning has become a pivotal tool in bioinformatics and computational biology, providing advanced architectures for analyzing complex biological datasets. Convolutional Neural Networks (CNNs) are widely used for image-based tasks, such as classifying colorectal cancer (CRC) pathologies from CT images. The Adaptive Segmentation Pipeline (ASP) exemplifies this by classifying CT images into distinct pathologies with tailored segmentation models, enhancing diagnostic imaging precision essential for treatment planning [32].

Various architectures, including deep neural networks (DNNs), recurrent neural networks (RNNs), and CNNs, offer unique advantages for bioinformatics applications [34]. DNNs efficiently manage large-scale genomic data, while RNNs excel in processing sequential data, suitable for gene expression analysis and sequence prediction. The Bio-Bi-LSTM-CRF model, integrating Bidirectional Long

Short-Term Memory networks with Conditional Random Fields, improves clinical feature extraction from colonoscopy reports, enhancing diagnostic accuracy [14].

Attention mechanisms and graph neural networks (GNNs) further expand deep learning's capabilities in bioinformatics, addressing high-dimensional biological data challenges and enabling more accurate, interpretable models. A categorization of deep learning architectures provides insights into their applications and effectiveness in healthcare [35].

These diverse architectures enhance data analysis in bioinformatics, driving innovative solutions for complex biological problems. AI-augmented histopathologic review systems and transformer-based natural language processing models significantly boost precision medicine capabilities, facilitating accurate tumor characterization and improving precancerous lesion classification [36, 37, 38, 31]. Integrating single-cell multiomics data with advanced deep learning techniques supports the identification of prognostic biomarkers and therapeutic targets, making treatment strategies more effective and efficient [35].

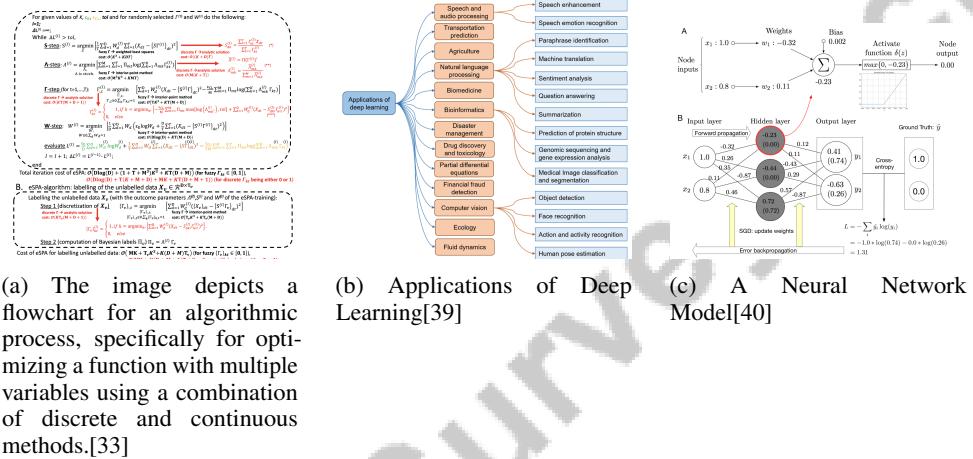


Figure 2: Examples of Deep Learning Architectures and Techniques

As illustrated in Figure 2, deep learning's integration into bioinformatics and computational biology has led to significant advancements through innovative architectures and techniques. The flowchart shows an algorithmic process for optimizing functions with multiple variables, crucial for complex computational tasks. The second image highlights deep learning's expansive applications, while the neural network model exemplifies the architecture of deep learning systems, enhancing predictive accuracy. Together, these visuals capture deep learning's pivotal role in advancing bioinformatics and computational biology. Table 1 provides a detailed overview of selected deep learning methods, highlighting their architectural designs, specific application domains, and the enhancements they bring to bioinformatics and computational biology.

3.2 Applications in Cancer Research

Deep learning's integration into colorectal cancer (CRC) research has significantly enhanced diagnostic and prognostic capabilities [35]. Recent advancements have led to robust models improving CRC diagnosis and treatment outcomes, such as increased polyp detection rates in CRC screening for early diagnosis [1]. Machine learning frameworks like the Glycolysis-Associated Clusters Predictor provide insights into CRC pathophysiology by identifying metabolic clusters related to cancer progression [4].

Attention-based regression models in gene expression analysis offer accurate estimates from histology images, aiding patient stratification and personalized treatment planning [41]. In prognostic modeling, deep learning techniques integrate multiple T cell-related genes to enhance predictions, improving understanding of immune responses in CRC [24]. The DTI-MD method in gene sequence analysis effectively identifies discriminative motifs vital for understanding CRC at the molecular level [15].

Deep learning models have shown promise in automatically detecting colon cancer from confocal laser microscopy images, achieving high accuracy in identifying cancerous tissues [26]. Performance metrics such as accuracy, precision, recall, and F1-score are crucial in assessing these models' effectiveness, ensuring reliability in clinical settings [14].

These advancements underscore deep learning's critical role in CRC research, paving the way for personalized and effective therapeutic interventions. Innovative algorithms like CRCNet utilize whole-slide images to predict survival outcomes and chemotherapy effectiveness, enhancing diagnostic precision and identifying high-risk subgroups likely to benefit from chemotherapy. CNNs extract clinically relevant biomarkers from routine histological samples, providing a cost-effective alternative to traditional genetic testing methods, improving patient stratification and outcomes in CRC management [42, 43].

3.3 Advancements in Predictive Modeling

Advancements in predictive modeling using deep learning have significantly transformed cancer diagnostics and treatment, particularly for colorectal cancer (CRC). Recent developments focus on ensemble learning methods to enhance prediction accuracy by combining multiple models' strengths, mitigating overfitting risks, especially in small datasets [44]. Models have achieved up to 82.5% accuracy in predicting survival rates when consensus is reached [45].

Integrating deep learning with multi-omics data shows promise in identifying novel biomarkers and therapeutic targets, uncovering insights into cancer progression [36]. Techniques like the ConvNet-CRC model apply stain normalization to enhance predictive modeling for CRC tissue classification, improving histopathological analysis accuracy [46]. This aligns with findings emphasizing robustness against staining variations and interpretability in model evaluation [47].

Innovative methods such as the Activity Score Generation Method (ASGM) analyze metabolic behaviors across CRC patient cohorts by computing metabolic reaction scores based on transcript levels [48]. Continual learning approaches in tumor classification incorporate diverse tumor types without exhaustive model retraining, offering scalable solutions for digital pathology [49].

Advanced predictive models like the Deep Learning System (DLS) have achieved a 5-year disease-specific survival area under the curve (AUC) of 0.70 and 0.69 in validation datasets, marking significant progress in CRC diagnostics [50]. The Divide-and-Rule (DnR) self-supervised learning method enhances prognostic predictions by capturing interactions in histopathological patterns [12]. Similarly, the DeepDisMISL model innovates by utilizing a distribution-based approach in multiple-instance learning, contrasting with methods focusing solely on extreme scoring patches [19].

The CAPRESE model demonstrates robustness in reconstructing cancer progression models, outperforming existing methods in noisy environments with limited samples [51]. Scalable entropic approaches like eSPA have shown a 30-fold improvement in classification performance over traditional methods, offering efficient solutions for predictive modeling in cancer research [33].

These advancements highlight deep learning's transformative potential in predictive modeling, paving the way for personalized and effective cancer diagnostics and treatment strategies. By integrating clinical and omics covariates, models like FusedTree enhance prognostic accuracy, effectively managing high-dimensional data [25]. The scRank method exemplifies this by leveraging relative expression orderings (REOs) to create stable prognostic signatures, robust against batch effects and normalization issues [11]. These developments continue to drive the field forward, opening new avenues for precision medicine in oncology.

4 Single-cell Omics and Colorectal Cancer

The integration of single-cell omics into cancer research significantly enhances our understanding of colorectal cancer (CRC), offering detailed insights into tumor biology and aiding in the development of novel therapeutic strategies. Figure 3 illustrates this integration, highlighting key technological advancements that contribute to our understanding of cancer heterogeneity, the identification of novel biomarkers, and the discovery of therapeutic targets. This figure demonstrates how single-cell omics technologies and computational methods provide a comprehensive framework for enhancing

precision medicine in colorectal cancer, thereby emphasizing the critical role these advancements play in shaping future research directions and therapeutic approaches.

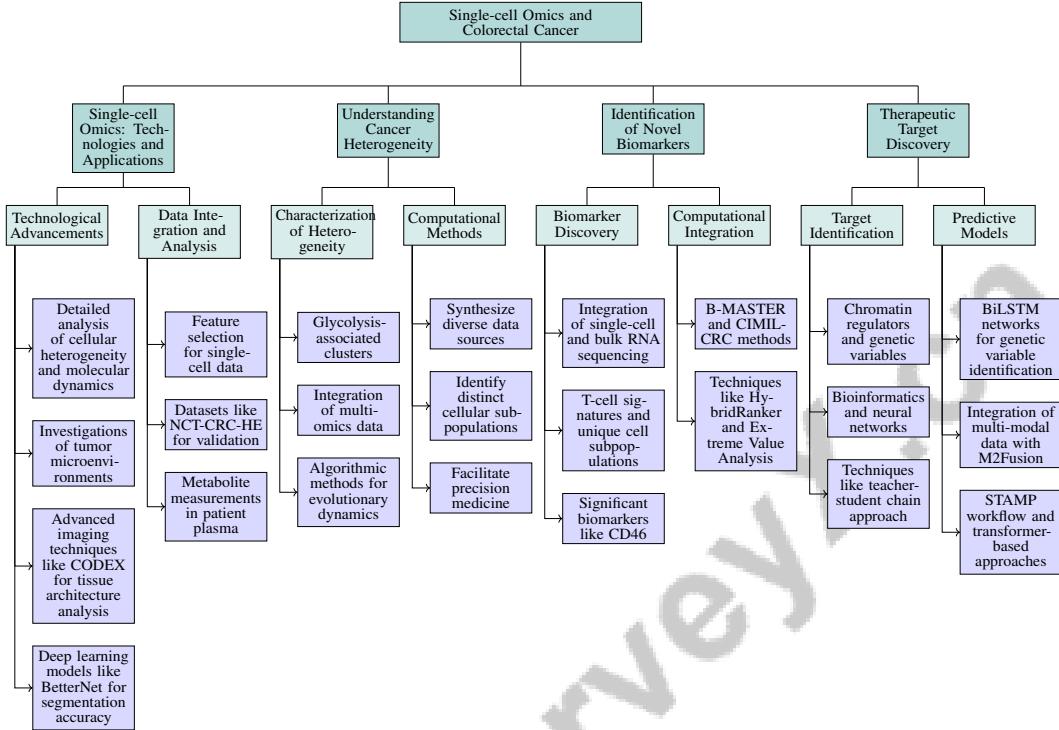


Figure 3: This figure illustrates the integration of single-cell omics in colorectal cancer research, highlighting key technological advancements, understanding of cancer heterogeneity, identification of novel biomarkers, and therapeutic target discovery. It demonstrates how single-cell omics technologies and computational methods provide a comprehensive framework for enhancing precision medicine in colorectal cancer.

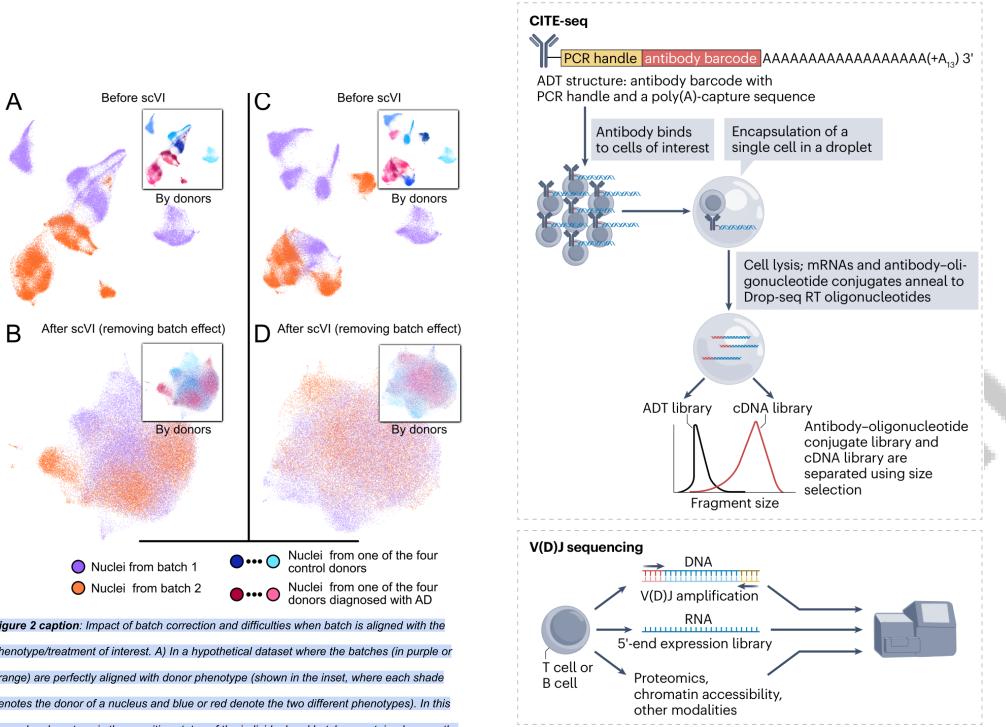
4.1 Single-cell Omics: Technologies and Applications

Single-cell omics technologies revolutionize cancer research by providing detailed analyses of cellular heterogeneity and molecular dynamics within tumors like CRC. These advancements enable comprehensive investigations of tumor microenvironments, identifying diverse cellular subpopulations and their roles in tumor progression, drug resistance, and immune response [20, 52]. Integration with computational techniques enhances CRC biology understanding and supports precision medicine.

Advanced imaging techniques, such as CODEX, capture high-dimensional data, facilitating detailed analyses of tissue architectures in CRC [10]. Deep learning models, like BetterNet, improve segmentation accuracy through attention mechanisms [13]. Feature selection methods tailored for single-cell data extract meaningful insights from complex datasets, aiding in tissue-specific gene expression analysis [27, 15].

Datasets like NCT-CRC-HE and multimodal datasets from CRC100K and PatchCamelyon provide frameworks for validating single-cell omics methodologies [23, 53]. Single-cell omics also extends to metabolite measurements in patient plasma, revealing insights into disease mechanisms and potential therapeutic targets [54].

As shown in Figure 4, single-cell omics technologies provide insights into cellular heterogeneity and disease progression in CRC. Batch correction challenges and CITE-seq methodologies illustrate the power of these technologies in unraveling CRC's molecular underpinnings, informing the development of targeted therapeutic strategies [55, 56].



(b) Cellular Transcriptome Analysis Using CITE-seq and V(D)J Sequencing[56]

Figure 4: Examples of Single-cell Omics: Technologies and Applications

4.2 Understanding Cancer Heterogeneity

Single-cell omics technologies enhance our understanding of CRC heterogeneity by characterizing cellular diversity and interactions within the tumor microenvironment. These technologies reveal glycolysis-associated clusters contributing to cancer cell metabolic reprogramming [4].

Integrating multi-omics data improves the understanding of CRC heterogeneity by elucidating molecular pathways and cellular interactions. Advanced computational methods synthesize diverse data sources, enhancing tumor biology understanding [57, 58]. Algorithmic methods infer evolutionary dynamics, providing insights into CRC's clonal architecture and evolutionary trajectories [59, 37, 51, 60, 61].

Single-cell omics technologies, combined with computational methods, provide a framework for unraveling CRC's intratumoral heterogeneity, identifying distinct cellular subpopulations and their genomic, epigenomic, and transcriptomic profiles. This approach facilitates precision medicine by uncovering prognostic biomarkers and individualized treatment strategies [20, 2, 52, 11].

4.3 Identification of Novel Biomarkers

Single-cell omics technologies are crucial in identifying novel biomarkers for CRC, offering insights into the tumor microenvironment and cellular heterogeneity. By integrating single-cell and bulk RNA sequencing, researchers can elucidate CRC's immune microenvironment, revealing T-cell signatures as prognostic factors and identifying unique cell subpopulations linked to therapeutic responses [2, 24, 62, 11, 36].

Significant biomarkers like CD46 illustrate single-cell omics' capacity to clarify CRC heterogeneity [63]. Computational methods like B-MASTER and CIMIL-CRC integrate microbiome and clinical data, enriching insights into CRC's metabolic landscape and molecular subtype identification [64, 65].

Techniques like HybridRanker and Extreme Value Analysis identify candidate genes and microbial signatures associated with CRC [66, 28].

Single-cell omics, combined with computational methods, establishes a framework for identifying novel CRC biomarkers by analyzing intratumoral heterogeneity. These technologies enhance therapeutic resistance understanding and facilitate precise diagnostic and treatment strategies [20, 2].

4.4 Therapeutic Target Discovery

Single-cell omics technologies advance therapeutic target discovery in CRC by examining cellular and molecular heterogeneity within tumors. These technologies identify chromatin regulators and genetic variables crucial for understanding disease mechanisms and developing targeted therapies [21].

Bioinformatics and neural networks enhance target discovery by identifying cancer types based on mutations [67]. Techniques like the teacher-student chain approach facilitate CRC tissue sample classification, aiding in therapeutic target identification [68]. Predictive models, including BiLSTM networks, enhance genetic variable identification, offering efficient solutions for cancer research [69].

Integrating multi-modal data from pathology and radiology, as demonstrated by M2Fusion, improves prediction accuracy and therapeutic target identification [70]. The STAMP workflow and transformer-based approaches enhance MSI status prediction, aiding biomarker discovery and improving patient outcomes [71, 17].

Single-cell omics, combined with computational methods, provides a framework for discovering therapeutic targets in CRC. These technologies offer insights into tumor biology and genetic diversity, facilitating targeted therapeutic strategies and enhancing precision medicine effectiveness [60, 36, 72, 2].

5 Integrating Deep Learning with Single-cell Omics

5.1 Interdisciplinary Approaches: Integrating Deep Learning and Single-cell Omics

The convergence of deep learning and single-cell omics in colorectal cancer (CRC) research marks a significant leap in understanding tumor biology and advancing precision medicine. This interdisciplinary synergy leverages deep learning's computational prowess to decode complex single-cell omics datasets, enhancing insights into CRC heterogeneity. For instance, BetterNet employs residual connections and attention mechanisms to accurately segment polyps, capturing both local and global contextual cues essential for identifying subtle tumor microenvironment variations that signal disease progression and therapeutic responses [13].

Context-aware convolutional neural networks (CNNs) further improve spatial context capture and local feature extraction, boosting grading and classification accuracy in histopathological images. Standardized benchmarking platforms ensure the reliability of polyp detection models in clinical settings, validating their applicability [1]. Attention-based models facilitate gene expression analysis integration with histology images, identifying gene expression hotspots crucial for understanding CRC molecular subtypes and patient stratification, thereby enhancing clinical decision-making [65, 73, 74, 23, 41]. Multimodal fusion techniques exemplify the integration of imaging and clinical data, enriching predictive capabilities and providing comprehensive CRC insights.

Addressing challenges like batch effects and data quality variability requires robust computational tools. Future research should focus on validating findings through rigorous studies, emphasizing AI-augmented systems like SmartPath, which optimize DNA yield and tumor purity in CRC samples, thereby enhancing histopathologic review reliability and reducing tissue waste and costs. Integrating single-cell multiomics data can propel individualized medicine by elucidating disease mechanisms and identifying optimal biomarkers, necessitating a thorough evaluation of tissue environment and gene expression interplay [37, 36, 75]. By harnessing deep learning and single-cell omics, researchers can gain profound insights into CRC heterogeneity, identify novel therapeutic targets, and advance precision medicine in oncology.

5.2 Enhancing Biomarker Prediction and Classification

The fusion of deep learning with single-cell omics technologies significantly enhances biomarker prediction and classification in CRC research. This integration facilitates the extraction of meaningful insights from complex datasets, improving the accuracy and interpretability of biomarker prediction models. Regression-based deep learning techniques predict continuous biomarker scores from histopathological images, offering insights into CRC's molecular underpinnings [76].

Probabilistic assessments of cell presence and intercellular distances enrich analyses by quantitatively reflecting cell interaction dynamics, crucial for capturing complex tumor cellular landscapes [77]. Kernel Herding in distribution-based sketching methods effectively preserves cellular landscapes, enhancing classification task performance [78].

Benchmark development for cellular segmentation and composition analysis is vital for advancing algorithms in automating histology image analysis, significantly impacting cancer diagnosis and treatment by providing a standardized framework for evaluating and improving algorithm performance [73]. The two-stage resampling approach, involving oversampling in image space and undersampling in feature space, enhances model training and fine-tuning, leading to more robust biomarker prediction models [79].

Integrating immune dynamics into CRC models, as demonstrated by the CRC-ICM benchmark, supports personalized medicine and immunotherapy research by enhancing understanding of immune interactions within the tumor microenvironment [62]. BiLSTM mechanisms in models like DPSeq capture long-range dependencies in image data, crucial for accurate biomarker prediction and classification [69].

The integration of deep learning and single-cell omics provides a powerful framework for enhancing biomarker prediction and classification in CRC. Advanced sequencing technologies and bioinformatics methods enable comprehensive tumor biology understanding, uncover novel biomarkers, and formulate targeted therapeutic strategies. These methodologies, including single-cell multi-omics and large-scale knowledge graphs, facilitate the identification of genetic alterations and exploration of complex biological mechanisms, significantly advancing precision medicine in oncology, enhancing drug discovery, and improving individualized treatment plans [60, 58, 36].

5.3 Integration with Radiomic and Histopathology Data

Integrating omics data with radiomic and histopathology information offers a multifaceted approach to CRC analysis, enhancing diagnostic accuracy and informing therapeutic strategies. This methodology leverages advanced imaging techniques and machine learning to analyze tumor heterogeneity and biological characteristics, incorporating predictive modeling from extensive medical records. By overcoming limitations of traditional diagnostic methods, such as variability in histopathological assessments and polyp characterization challenges, this integrated approach improves CRC risk assessments and treatment decisions, contributing to personalized patient management [80, 22, 18]. Radiomics, involving the extraction of quantitative features from medical imaging, complements molecular insights from omics data, facilitating a holistic understanding of tumor biology and personalized treatment strategies.

Advanced computational techniques, particularly deep learning models, bridge the gap between radiomic features and histopathological data, enabling the identification of novel biomarkers and therapeutic targets. Models like M2Fusion use a Bayesian-based multimodal framework to integrate radiomic and histopathological features, significantly enhancing prediction accuracy and providing a robust platform for biomarker discovery [70]. This underscores the potential of combining diverse data modalities to enhance predictive model interpretability and clinical utility.

Moreover, deep learning applications in analyzing whole-slide images (WSIs) and CT scans demonstrate the ability to capture intricate patterns within the tumor microenvironment, indicative of disease progression and therapeutic response. Techniques employing fully transformer-based models for biomarker prediction exemplify the integration of imaging data with molecular profiles, achieving clinical-grade performance in predicting key biomarkers like microsatellite instability (MSI) [17].

Innovative frameworks incorporating immune dynamics into predictive models significantly advance the integration of radiomics and histopathology. These frameworks leverage spatial omics technologies and artificial intelligence to provide detailed insights into tumor-immune interactions,

highlighting the tumor microenvironment's spatial organization and its impact on immune cell localization. By utilizing deep learning algorithms to analyze histopathological images and radiomic features, researchers can enhance predictions of T-cell infiltration and other immune responses, improving therapeutic strategies for various cancers, including immunologically cold solid tumors [80, 81, 82, 72]. Such integrative approaches are crucial for advancing precision medicine, offering a detailed understanding of the tumor landscape, enabling the identification of potential therapeutic targets, and improving patient stratification.

The integration of omics data with radiomics and histopathology marks a transformative advancement in CRC research, combining non-invasive imaging techniques with comprehensive biological insights to enhance diagnostic accuracy, prognostic predictions, and treatment responses. This multifaceted approach leverages radiomics' strengths in assessing tumor heterogeneity through advanced imaging modalities alongside detailed histopathological characteristics and patient medical records, ultimately facilitating personalized management strategies for CRC patients and improving risk assessment methodologies [80, 37, 18]. By leveraging the strengths of each data modality, researchers can gain deeper insights into tumor biology, identify novel biomarkers, and develop targeted therapeutic strategies, ultimately improving patient outcomes in oncology.

6 Challenges and Future Directions

The complexities in colorectal cancer (CRC) research necessitate addressing multifaceted challenges in single-cell data analysis. This section delves into specific obstacles in single-cell omics, highlighting limitations that impede the clinical translation of these technologies. Understanding these challenges is crucial for enhancing model performance and developing innovative solutions to improve the reliability and applicability of single-cell analyses in CRC.

6.1 Challenges in Single-cell Data Analysis

Single-cell omics data analysis in CRC research faces significant challenges that hinder effective clinical application. A primary issue is the under-sampling of specific clinical conditions in training datasets, which affects model performance across diverse patient populations [14]. This is compounded by the limited availability of public data and the lack of standardized laboratory protocols, essential for reproducibility and reliability [2]. Benchmarks often lack diversity and representation of polyp types, leading to poor model generalization [1]. The high dimensionality and complexity of single-cell data complicate CRC categorization based on glycolysis-related molecular subtypes [4]. Robust feature selection and multi-omics integration methods are needed to maintain model generalizability across datasets.

AI model performance variability, due to small sample sizes and high dimensionality, complicates CRC diagnostics integration. Advanced algorithms are needed to accurately capture gene expression spatial heterogeneity, overcoming current predictive model limitations. Understanding intratumoral genomic and transcriptomic heterogeneity is critical for advancing single-cell omics research, leading to effective precision medicine strategies. Single-cell multi-omics technologies, integrating various omics modalities at the individual cell level, promise to identify novel prognostic biomarkers and therapeutic targets, improving patient outcomes [2, 20, 56, 11, 36].

6.2 Integration with Other Omics and Clinical Data

Integrating single-cell omics with other omics and clinical data in CRC research presents both challenges and opportunities. A significant hurdle is the lack of structured documentation in electronic health records, complicating reliable colorectal cancer status classification [38]. This underscores the need for improved data management and standardization to facilitate seamless integration of diverse datasets. The high dimensionality of single-cell omics data poses overfitting risks when combined with other omics data, as seen in models like FCS-Net, necessitating advanced computational techniques for managing complexity and ensuring model robustness [83]. Datasets may inadequately represent rare tumor types, affecting model generalization [84]. Comprehensive dataset representation is essential for improving model applicability across clinical scenarios.

Innovative approaches, such as the PiCnIC pipeline, offer promising solutions for analyzing tumor heterogeneity [59]. These integrative methods enhance tumor biology understanding by combining

genetic, transcriptomic, and proteomic data, facilitating novel biomarker and therapeutic target identification. The NCT-CRC-HE dataset reveals critical issues limiting its applicability for biomedical tool design, emphasizing the need for careful interpretation of results from models trained on this dataset [23]. Addressing these challenges requires robust validation frameworks and diverse data type incorporation to enhance integrative model generalizability and clinical utility.

Integrating single-cell omics with other data holds significant potential for advancing precision medicine in CRC. By overcoming current obstacles and employing cutting-edge methodologies, researchers can deepen tumor biology understanding, refine diagnostic precision, and develop targeted therapeutic strategies. Enhancements in natural language processing (NLP) have improved precancerous lesion characterization, enhancing early cancer detection diagnostic tests, particularly in colorectal cancer. Integrating sequencing technologies and bioinformatics facilitates genetic alteration identification driving tumorigenesis, while AI-augmented pathology systems optimize tissue extraction for next-generation sequencing, reducing waste and costs. Single-cell multi-omics data application supports personalized medicine by enabling tailored treatment plans based on individual genetic profiles, improving patient outcomes [60, 37, 38, 36].

6.3 Computational Complexity and Resource Requirements

Integrating deep learning with single-cell omics in CRC research involves substantial computational complexity and resource requirements, posing challenges for advancing precision medicine. A primary concern is hyperparameter tuning complexity and potential overfitting in high-dimensional settings, critical for developing robust predictive models [25]. This complexity is compounded by the need to process high-resolution images, such as digitized HE-stained slides, requiring significant computational resources and expertise for tasks like automatic segmentation [85].

Reliance on 'off-the-shelf' features from models like ImageNet, not optimized for histopathological data, results in suboptimal performance in tissue classification and molecular feature prediction [74]. Tailored computational strategies are needed to manage medical data's unique characteristics. Implementing deep learning methods like 'ikarus' in single-cell omics necessitates advanced computational infrastructures [52]. Integrating diverse data modalities, such as transcriptomics and epigenomics, complicates seamless integration due to inherent dataset differences [86]. Deep learning model effectiveness is often limited by the availability of large, annotated datasets, essential for training robust predictive models. The lack of detailed annotated datasets hampers existing computational pathology methods' performance and limits models' generalizability trained on smaller or less diverse datasets [87].

Innovative approaches, like eSPA, offer solutions by providing reliable predictions even in small data scenarios without extensive data preprocessing, breaching the overfitting barrier [33]. These methods highlight deep learning's potential to enhance complex biological data analysis but also underscore the substantial computational resources required for implementation. Additionally, analyzing large datasets from digitized slides for predicting optimal DNA yields exemplifies the resource-intensive nature of these tasks [37].

BetterNet's current implementation focuses solely on binary segmentation, limiting applicability in multi-class scenarios and necessitating further development to broaden its utility [13]. Future research should refine analytical methods to address computational challenges, explore cellular diversity's functional implications, and develop scalable solutions for efficiently integrating multi-modal data. By addressing these challenges, researchers can enhance deep learning's utility in single-cell omics, advancing precision medicine in CRC.

6.4 Model Interpretability and Robustness

Ensuring model interpretability and robustness in deep learning applications for CRC research is a multifaceted challenge, particularly when integrating diverse data types. Current studies often lack robustness and may not scale well with increasing data dimensionality, posing challenges in reproducibility and generalization [27]. Extreme Value Analysis (EVA) suggests that extreme values in microbial abundance distributions are often the most informative, providing insights into the role of specific taxa in CRC progression [28]. However, integrating these insights into deep learning models highlights the need for improved interpretability.

The NCT-CRC-HE dataset may not fully account for all variations in image quality and processing, affecting model predictions [23]. This limitation emphasizes the importance of developing models that can handle data quality variability and ensure reliable predictions across conditions. The DTI-MD method exemplifies the challenge of model interpretability in deep learning applications, relying on directed information to capture the causal influence of motifs on gene expression [15].

Future research should focus on enhancing model interpretability, addressing data quality and interoperability issues, and exploring deep learning integration within broader healthcare systems [35]. Improving transparency in model decision-making processes fosters trust and reliability in single-cell omics applications, advancing precision medicine in oncology. Addressing these challenges requires developing models that provide accurate predictions and insights into underlying biological processes, ensuring successful clinical implementation and enhancing their utility in CRC research.

6.5 Limited Data and Generalizability

Limited data availability and generalizability significantly hinder the application of deep learning models in CRC research. The scarcity of labeled data, crucial for training supervised learning models, remains a primary concern, as these models require extensive annotated datasets to achieve high predictive accuracy [82]. This issue is compounded by the high costs and time-consuming nature of data labeling, restricting the application of deep learning methods in digital pathology [82]. Moreover, reliance on accurate classification steps affects overall segmentation performance in adaptive segmentation pipelines, underscoring the need for robust data handling [32].

The generalizability of deep learning models is further challenged by limited diversity in existing datasets, which often fail to capture the full spectrum of clinical scenarios [34]. This limitation is evident in studies relying heavily on single datasets, such as TCGA, which may not adequately represent CRC heterogeneity [2]. Variability in patient responses and the high dimensionality of single-cell RNA sequencing (scRNA-seq) data present additional hurdles, as traditional analysis methods often inadequately manage these complexities [11].

While data augmentation techniques generally improve performance, they may not always yield significant benefits, particularly for well-separated tasks [88]. The performance of in-context learning varies based on the quality and representativeness of examples, with certain classes still posing challenges [53]. These issues highlight the need for better representation in training data to enhance model robustness and accuracy [89].

Future research should focus on expanding datasets to include more diverse patient populations, enhancing model generalizability [2]. Integrating additional omics data for improved prognostic modeling and refining frameworks like scRank are also recommended [11]. By overcoming these limitations, researchers can enhance the applicability and reliability of deep learning models in CRC research, ultimately advancing precision medicine in oncology.

6.6 Integration of Multi-modal and Multi-omics Data

Integrating multi-modal and multi-omics data in CRC research is pivotal for enhancing our understanding of tumor biology and advancing precision medicine strategies. Future research should focus on developing robust frameworks for effectively integrating diverse data types, including genomic, transcriptomic, proteomic, and imaging data, to provide a comprehensive view of tumor biology. This integration is crucial for uncovering complex interactions and molecular mechanisms driving CRC progression, enabling the identification of novel biomarkers and therapeutic targets [54].

Advancements in deep learning methodologies, such as exploring transformer models in protein function classification, could enhance multi-omics data integration [90]. Leveraging multi-modal data to improve predictive accuracy and developing effective non-parametric and hybrid approaches can address unique challenges posed by single-cell RNA-Seq data [57]. Enhancing computational methods for data analysis and exploring mechanistic studies of cellular interactions are essential for translating findings into clinical applications and improving patient outcomes [10].

Developing robust deep learning models that utilize single-cell multi-omics data is another promising avenue for future research [91]. Refining methods for application to other types of omics data and investigating the biological implications of identified tissue environmental effects can provide deeper insights into CRC heterogeneity [75]. Acquiring more data and improving detection methods for

malignant tissue in the colon are critical for enhancing model accuracy and reliability in clinical settings [26].

Future research should also explore integrating additional genomic features and demographic factors, such as age and gender, to enhance model accuracy [16]. By fostering interdisciplinary collaboration and leveraging the strengths of multi-modal and multi-omics data, researchers can gain a deeper understanding of cancer mechanisms and improve precision medicine interventions in CRC.

The integration of multi-modal and multi-omics data holds significant potential for advancing CRC research. By addressing contemporary challenges and utilizing cutting-edge technologies such as natural language processing (NLP), sequencing methods, and artificial intelligence, researchers can deepen their understanding of tumor biology. These innovative approaches enhance the characterization of precancerous lesions, improve diagnostic accuracy through better classification of cancer stages, and facilitate the identification of optimal therapeutic targets. Consequently, these advancements have the potential to lead to more effective treatment strategies, ultimately improving patient outcomes in oncology [37, 38, 72, 60, 36].

7 Conclusion

The fusion of deep learning and single-cell omics has marked a significant leap forward in colorectal cancer (CRC) research, refining both diagnostic and therapeutic methodologies. Lightweight deep learning models, exemplified by architectures like SqueezeNet, have shown remarkable promise in advancing CRC diagnosis and treatment. Concurrently, hyperdimensional computing (HDC) has emerged as a robust tool for managing extensive biological datasets, providing clarity and accommodating varied data types within a unified framework.

Artificial intelligence (AI) continues to be instrumental in CRC screening, enhancing early detection and improving patient prognoses. Notable models such as DeepDisMISL have leveraged comprehensive patch information to substantially enhance CRC survival predictions, surpassing the performance of existing algorithms. Furthermore, elucidating glycolysis-related molecular subtypes is pivotal for progressing personalized medicine in CRC, enabling more precise therapeutic interventions.

The role of single-cell technologies is crucial in advancing early diagnosis, predicting therapeutic responses, and identifying novel drug targets, thus fostering more effective treatment plans. The identification of TRGS as an independent prognostic biomarker highlights its potential in forecasting patient responses to immunotherapy and chemotherapy, underscoring the importance of integrating multi-omics data in CRC research.

Moving forward, research efforts should prioritize the development of multi-class segmentation models and the exploration of uncertainty estimation techniques to bolster model reliability. By harnessing these sophisticated methodologies, researchers can gain deeper insights into tumor biology, enhance diagnostic precision, and formulate targeted therapeutic strategies, ultimately propelling precision medicine in CRC research to new heights.

References

- [1] Kaidong Li, Mohammad I. Fathan, Krushi Patel, Tianxiao Zhang, Cuncong Zhong, Ajay Bansal, Amit Rastogi, Jean S. Wang, and Guanghui Wang. Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations, 2021.
- [2] Andrea Angius, Antonio Mario Scanu, Caterina Arru, Maria Rosaria Muroni, Ciriaco Carru, Alberto Porcu, Paolo Cossu-Rocca, and Maria Rosaria De Miglio. A portrait of intratumoral genomic and transcriptomic heterogeneity at single-cell level in colorectal cancer. *Medicina*, 57(11):1257, 2021.
- [3] Panagiotis Katrakazas, Aristotelis Ballas, Marco Anisetti, and Ilias Spais. An artificial intelligence outlook for colorectal cancer screening, 2022.
- [4] Zhenling Wang, Yu Shao, Hongqiang Zhang, Yunfei Lu, Yang Chen, Hengyang Shen, Changzhi Huang, Jingyu Wu, and Zan Fu. Machine learning-based glycolysis-associated molecular classification reveals differences in prognosis, tme, and immunotherapy for colorectal cancer patients. *Frontiers in Immunology*, 14:1181985, 2023.
- [5] Jason W. Wei, Arief A. Suriawinata, Louis J. Vaickus, Bing Ren, Xiaoying Liu, Mikhail Lisovsky, Naofumi Tomita, Behnaz Abdollahi, Adam S. Kim, Dale C. Snover, John A. Baron, Elizabeth L. Barry, and Saeed Hassanpour. Deep neural networks for automated classification of colorectal polyps on histopathology slides: A multi-institutional evaluation, 2019.
- [6] Yoav Kan-Tor, Michael Morris Danziger, Eden Zohar, Matan Ninio, and Yishai Shimon. Does your model understand genes? a benchmark of gene properties for biological and text models, 2024.
- [7] Doyeon Ha, JungHo Kong, Donghyo Kim, Kwanghwan Lee, Juhun Lee, Minhyuk Park, Hyunsoo Ahn, Youngchul Oh, and Sanguk Kim. Development of bioinformatics and multi-omics analyses in organoids. *BMB reports*, 56(1):43, 2022.
- [8] Srinath Jayachandran and Ashlin Ghosh. Deep transfer learning for texture classification in colorectal cancer histology, 2020.
- [9] Ricardo Romero. Integration of genetic algorithms and deep learning for the generation and bioactivity prediction of novel tyrosine kinase inhibitors, 2024.
- [10] Wilson Kuswanto, Garry Nolan, and Guolan Lu. Highly multiplexed spatial profiling with codex: bioinformatic analysis and application in human disease. In *Seminars in immunopathology*, volume 45, pages 145–157. Springer, 2023.
- [11] Mengsha Tong, Yuxiang Lin, Wenxian Yang, Jinsheng Song, Zheyang Zhang, Jiajing Xie, Jingyi Tian, Shijie Luo, Chenyu Liang, Jiali Huang, et al. Prioritizing prognostic-associated subpopulations and individualized recurrence risk signatures from single-cell transcriptomes of colorectal cancer. *Briefings in Bioinformatics*, 24(3):bbad078, 2023.
- [12] Christian Abbet, Inti Zlobec, Behzad Bozorgtabar, and Jean-Philippe Thiran. Divide-and-rule: Self-supervised learning for survival analysis in colorectal cancer, 2020.
- [13] Owen Singh and Sandeep Singh Sengar. Betternet: An efficient cnn architecture with residual learning and attention for precision polyp segmentation, 2024.
- [14] Shashank Reddy Vadyala and Eric A. Sherer. Natural language processing accurately categorizes indications, findings and pathology reports from multicenter colonoscopy, 2021.
- [15] Arvind Rao, Alfred O. Hero III au2, David J. States, and James Douglas Engel. Finding sequence features in tissue-specific sequences, 2007.
- [16] Hadar Hezi, Daniel Shats, Daniel Gurevich, Yosef E. Maruvka, and Moti Freiman. Exploring the interplay between colorectal cancer subtypes genomic variants and cellular morphology: A deep-learning approach, 2024.

-
- [17] Sophia J. Wagner, Daniel Reisenbüchler, Nicholas P. West, Jan Moritz Niehues, Gregory Patrick Veldhuizen, Philip Quirke, Heike I. Grabsch, Piet A. van den Brandt, Gordon G. A. Hutchins, Susan D. Richman, Tanwei Yuan, Rupert Langer, Josien Christina Anna Jenniskens, Kelly Offermans, Wolfram Mueller, Richard Gray, Stephen B. Gruber, Joel K. Greenson, Gad Rennert, Joseph D. Bonner, Daniel Schmolze, Jacqueline A. James, Maurice B. Loughrey, Manuel Salto-Tellez, Hermann Brenner, Michael Hoffmeister, Daniel Truhn, Julia A. Schnabel, Melanie Boxberg, Tingying Peng, and Jakob Nikolas Kather. Fully transformer-based biomarker prediction from colorectal cancer histology: a large-scale multicentric study, 2023.
 - [18] Shuai Jiang, Christina Robinson, Joseph Anderson, William Hisey, Lynn Butterly, Arief Suriawinata, and Saeed Hassanpour. Improving colorectal cancer screening and risk assessment through predictive modeling on medical images and records, 2024.
 - [19] Xingyu Li, Jitendra Jonnagaddala, Min Cen, Hong Zhang, and Xu Steven Xu. Colorectal cancer survival prediction using deep distribution based multiple-instance learning, 2022.
 - [20] Anjun Ma, Gang Xin, and Qin Ma. The use of single-cell multi-omics in immuno-oncology. *Nature Communications*, 13(1):2728, 2022.
 - [21] Yichao Ma, Fang Fang, Kai Liao, Jingqiu Zhang, Chen Wei, Yiqun Liao, Bin Zhao, Yongkun Fang, Yuji Chen, Xinyue Zhang, et al. Identification and validation of the clinical prediction model and biomarkers based on chromatin regulators in colon cancer by integrated analysis of bulk-and single-cell rna sequencing data. *Translational Cancer Research*, 13(3):1290, 2024.
 - [22] Mary Adewunmi and Reem Abdel-Salam. Pubtrend: General overview of artificial intelligence for colorectal cancer diagnosis from 2010-2022, 2024.
 - [23] Andrey Ignatov and Grigory Malivenko. Nct-crc-he: Not all histopathological datasets are equally useful, 2024.
 - [24] Peng Cui, Haibo Wang, and Zhigang Bai. Integrated single-cell and bulk rna-seq analysis identifies a prognostic t-cell signature in colorectal cancer. *Scientific Reports*, 14(1):20177, 2024.
 - [25] Jeroen M. Goedhart, Mark A. van de Wiel, Wessel N. van Wieringen, and Thomas Klausch. Fusion of tree-induced regressions for clinico-genomic data, 2024.
 - [26] Nils Gessert, Lukas Wittig, Daniel Drömann, Tobias Keck, Alexander Schlaefer, and David B. Ellebrecht. Feasibility of colon cancer detection in confocal laser microscopy images using convolution neural networks, 2018.
 - [27] Pengyi Yang, Hao Huang, and Chunlei Liu. Feature selection revisited in the single-cell era, 2021.
 - [28] Stephanie Danni Song, Patricio Jeraldo, Jun Chen, and Nicholas Chia. Extreme value analysis of gut microbial alterations in colorectal cancer, 2019.
 - [29] M D'Orazio, M Murdocca, A Mencattini, P Casti, J Filippi, G Antonelli, D Di Giuseppe, MC Comes, C Di Natale, F Sangiuolo, et al. Machine learning phenomics (mlp) combining deep learning with time-lapse-microscopy for monitoring colorectal adenocarcinoma cells gene expression and drug-response. *Scientific Reports*, 12(1):8545, 2022.
 - [30] Manoj M Wagle, Siqu Long, Carissa Chen, Chunlei Liu, and Pengyi Yang. Interpretable deep learning in single-cell omics, 2024.
 - [31] Min Cen, Xingyu Li, Bangwei Guo, Jitendra Jonnagaddala, Hong Zhang, and Xu Steven Xu. Time to embrace natural language processing (nlp)-based digital pathology: Benchmarking nlp- and convolutional neural network-based deep learning pipelines, 2023.
 - [32] Peilong Wang, Timothy L. Kline, Andy D. Missett, Cole J. Cook, Matthew R. Callstrom, Alex Chan, Robert P. Hartman, Zachary S. Kelm, and Panagiotis Korfiatis. A classification-based adaptive segmentation pipeline: Feasibility study using polycystic liver disease and metastases from colorectal cancer ct images, 2024.

-
- [33] Illia Horenko. On a scalable entropic breaching of the overfitting barrier in machine learning, 2020.
 - [34] Seonwoo Min, Byunghan Lee, and Sungroh Yoon. Deep learning in bioinformatics, 2016.
 - [35] Farzan Shenavarmasouleh, Farid Ghareh Mohammadi, Khaled M. Rasheed, and Hamid R. Arabnia. Deep learning in healthcare: An in-depth analysis, 2023.
 - [36] Xinying Zhang, Panwen Wang, and Jing Qin. Enhanced biological mechanism study, drug discovery and individualized medicine with single-cell multiomics data and integrative analysis. *Frontiers in Genetics*, 14:1236126, 2023.
 - [37] Bolesław L. Osinski, Aïcha BenTaieb, Irvin Ho, Ryan D. Jones, Rohan P. Joshi, Andrew Westley, Michael Carlson, Caleb Willis, Luke Schleicher, Brett M. Mahon, and Martin C. Stumpe. Ai-augmented histopathologic review using image analysis to optimize dna yield and tumor purity from ffpe slides, 2022.
 - [38] Yizhen Zhong, Jiajie Xiao, Thomas Vetterli, Mahan Matin, Ellen Loo, Jimmy Lin, Richard Bourgon, and Ofer Shapira. Improving precancerous case characterization via transformer-based ensemble learning, 2022.
 - [39] Shams Forruque Ahmed, Md. Sakib Bin Alam, Maliha Kabir, Shaila Afrin, Sabiha Jannat Rafa, Aanushka Mehjabin, and Amir H. Gandomi. Unveiling the frontiers of deep learning: innovations shaping diverse domains, 2023.
 - [40] Yu Li, Chao Huang, Lizhong Ding, Zhongxiao Li, Yijie Pan, and Xin Gao. Deep learning in bioinformatics: introduction, application, and perspective in big data era, 2019.
 - [41] Mara Graziani, Niccolò Marini, Nicolas Deutschmann, Nikita Janakarajan, Henning Müller, and María Rodríguez Martínez. Attention-based interpretable regression of gene expression in histology, 2022.
 - [42] Fabi Prezja, Leevi Annala, Sampsa Kiiskinen, Suvi Lahtinen, Timo Ojala, Pekka Ruusuvuori, and Teijo Kuopio. Improving performance in colorectal cancer histology decomposition using deep and ensemble machine learning, 2024.
 - [43] Xingyu Li, Jitendra Jonnagaddala, Shuhua Yang, Hong Zhang, and Xu Steven Xu. A retrospective analysis using deep-learning models for prediction of survival outcome and benefit of adjuvant chemotherapy in stage ii/iii colorectal cancer, 2021.
 - [44] Ralph Saber, David Henault, Rolando Rebolledo, Simon Turcotte, and Samuel Kadoury. Prediction of a t-cell/mhc-i-based immune profile for colorectal liver metastases from ct images using ensemble learning, 2023.
 - [45] Chris Roadknight, Uwe Aickelin, John Scholefield, and Lindy Durrant. Ensemble learning of colorectal cancer survival rates, 2014.
 - [46] Francesco Ciompi, Oscar Geessink, Babak Ehteshami Bejnordi, Gabriel Silva de Souza, Alexi Baidoshvili, Geert Litjens, Bram van Ginneken, Iris Nagtegaal, and Jeroen van der Laak. The importance of stain normalization in colorectal tissue classification with convolutional networks, 2017.
 - [47] Maximilian Springenberg, Annika Frommholz, Markus Wenzel, Eva Weicken, Jackie Ma, and Nils Strodthoff. From modern cnns to vision transformers: Assessing the performance, robustness, and classification strategies of deep learning models in histopathology, 2023.
 - [48] Davide Maspero, Claudio Isella, Marzia Di Filippo, Alex Graudenzi, Sara Erika Bellomo, Marco Antoniotti, Giancarlo Mauri, Enzo Medico, and Chiara Damiani. Metabolic enrichment through functional gene rules, 2017.
 - [49] Veena Kaustaban, Qinle Ba, Ipsita Bhattacharya, Nahil Sobh, Satarupa Mukherjee, Jim Martin, Mohammad Saleh Miri, Christoph Guetter, and Amal Chaturvedi. Continual learning for tumor classification in histopathology images, 2022.

-
- [50] Ellery Wulczyn, David F. Steiner, Melissa Moran, Markus Plass, Robert Reihs, Fraser Tan, Isabelle Flament-Auvigne, Trissia Brown, Peter Regitnig, Po-Hsuan Cameron Chen, Narayan Hegde, Apaar Sadhwani, Robert MacDonald, Benny Ayalew, Greg S. Corrado, Lily H. Peng, Daniel Tse, Heimo Müller, Zhaoyang Xu, Yun Liu, Martin C. Stumpe, Kurt Zatloukal, and Craig H. Mermel. Interpretable survival prediction for colorectal cancer using deep learning, 2020.
- [51] Daniele Ramazzotti. A model of selective advantage for the efficient inference of cancer clonal evolution, 2016.
- [52] Jan Dohmen, Artem Baranovskii, Jonathan Ronen, Bora Uyar, Vedran Franke, and Altuna Akalin. Identifying tumor cells at the single-cell level using machine learning. *Genome biology*, 23(1):123, 2022.
- [53] Dyke Ferber, Georg Wölflein, Isabella C. Wiest, Marta Ligero, Srividhya Sainath, Narmin Ghafari Laleh, Omar S. M. El Nahhas, Gustav Müller-Franzes, Dirk Jäger, Daniel Truhn, and Jakob Nikolas Kather. In-context learning enables multimodal large language models to classify cancer pathology images, 2024.
- [54] Ali Cakmak and M. Hasan Celik. Personalized metabolic analysis of diseases, 2020.
- [55] Katherine E. Prater and Kevin Z. Lin. All the single cells: single-cell transcriptomics/epigenomics experimental design and analysis considerations for glial biologists, 2024.
- [56] Alev Baysoy, Zhiliang Bai, Rahul Satija, and Rong Fan. The technological landscape and applications of single-cell multi-omics. *Nature Reviews Molecular Cell Biology*, 24(10):695–713, 2023.
- [57] Juliana Costa-Silva, Douglas S. Domingues, David Menotti, Mariangela Hungria, and Fabricio M Lopes. Computational methods for differentially expressed gene analysis from rna-seq: an overview, 2021.
- [58] Md. Rezaul Karim, Lina Molinas Comet, Md Shajalal, Oya Deniz Beyan, Dietrich Rebholz-Schuhmann, and Stefan Decker. From large language models to knowledge graphs for biomarker discovery in cancer, 2023.
- [59] Giulio Caravagna, Alex Graudenzi, Daniele Ramazzotti, Rebeca Sanz-Pamplona, Luca De Sano, Giancarlo Mauri, Victor Moreno, Marco Antoniotti, and Bud Mishra. Algorithmic methods to infer the evolutionary trajectories in cancer progression, 2017.
- [60] Geng Chen, Lu Xie, Fangqing Zhao, and David P Kreil. The application of sequencing technologies and bioinformatics methods in cancer biology, 2022.
- [61] Mona Nourbakhsh, Kristine Degn, Astrid Saksager, Matteo Tiberti, and Elena Papaleo. Prediction of cancer driver genes and mutations: the potential of integrative computational frameworks, 2023.
- [62] Zahra Mokhtari, Elham Amjadi, Hamidreza Bolhasani, Zahra Faghah, AmirReza Dehghanian, and Marzieh Rezaei. Crc-icm: Colorectal cancer immune cell markers pattern dataset, 2023.
- [63] Yihui Liu, Uwe Aickelin, Jan Feyereisl, and Lindy G. Durrant. Wavelet feature extraction and genetic algorithm for biomarker detection in colorectal cancer data, 2013.
- [64] Priyam Das, Tanujit Dey, Christine Peterson, and Sounak Chakraborty. B-master: Scalable bayesian multivariate regression analysis for selecting targeted essential regressors to identify the key genera in microbiome-metabolite relation dynamics, 2024.
- [65] Hadar Hezi, Matan Gelber, Alexander Balabanov, Yosef E. Maruvka, and Moti Freiman. Cimil-crc: a clinically-informed multiple instance learning framework for patient-level colorectal cancer molecular subtypes classification from h&e stained images, 2024.
- [66] Zahra Razaghi-Moghadama, Razieh Abdollahia, Sama Goliaeib, and Morteza Ebrahimia. Hybirdranker: Integrating network structure and disease knowledge to prioritize cancer candidate genes, 2016.

-
- [67] Ayad Ghany Ismaeel and Dina Yousif Mikhail. Effective data mining technique for classification cancers via mutations in gene using neural network, 2016.
 - [68] Shayne Shaw, Maciej Pajak, Aneta Lisowska, Sotirios A Tsaftaris, and Alison Q O’Neil. Teacher-student chain for efficient semi-supervised histology image classification, 2020.
 - [69] Min Cen, Xingyu Li, Bangwei Guo, Jitendra Jonnagaddala, Hong Zhang, and Xu Steven Xu. Dpseq: A novel and efficient digital pathology classifier for predicting cancer biomarkers using sequencer architecture, 2023.
 - [70] Quan Liu, Jiawen Yao, Lisha Yao, Xin Chen, Jingren Zhou, Le Lu, Ling Zhang, Zaiyi Liu, and Yuankai Huo. m^2 fusion: Bayesian-based multimodal multi-level fusion on colorectal cancer microsatellite instability prediction, 2024.
 - [71] Omar S. M. El Nahhas, Marko van Treeck, Georg Wöllein, Michaela Unger, Marta Ligero, Tim Lenz, Sophia J. Wagner, Katherine J. Hewitt, Firas Khader, Sebastian Foersch, Daniel Truhn, and Jakob Nikolas Kather. From whole-slide image to biomarker prediction: A protocol for end-to-end deep learning in computational pathology, 2023.
 - [72] Ying Xu, Guan-Hua Su, Ding Ma, Yi Xiao, Zhi-Ming Shao, and Yi-Zhou Jiang. Technological advances in cancer immunity: from immunogenomics to single-cell analysis and artificial intelligence. *Signal Transduction and Targeted Therapy*, 6(1):312, 2021.
 - [73] Muhammad Dawood, Raja Muhammad Saad Bashir, Srijay Deshpande, Manahil Raza, and Adam Shephard. Cellular segmentation and composition in routine histology images using deep learning, 2022.
 - [74] Xingyu Li, Min Cen, Jinfeng Xu, Hong Zhang, and Xu Steven Xu. Improving feature extraction from histopathological images through a fine-tuning imagenet model, 2022.
 - [75] Daigo Okada, Jianshen Zhu, Kan Shota, Yuuki Nishimura, and Kazuya Haraguchi. Data mining method of single-cell omics data to evaluate a pure tissue environmental effect on gene expression level, 2024.
 - [76] Omar S. M. El Nahhas, Chiara M. L. Loeffler, Zunamys I. Carrero, Marko van Treeck, Fiona R. Kolbinger, Katherine J. Hewitt, Hannah S. Muti, Mara Graziani, Qinghe Zeng, Julien Calderaro, Nadina Ortiz-Brüchle, Tanwei Yuan, Michael Hoffmeister, Hermann Brenner, Alexander Brobeil, Jorge S. Reis-Filho, and Jakob Nikolas Kather. Regression-based deep-learning predicts molecular biomarkers from pathology slides, 2023.
 - [77] Toru Nagasaka, Kimihiro Yamashita, and Mitsugu Fujita. Novel methods for analyzing cellular interactions in deep learning-based image cytometry: Spatial interaction potential and co-localization index, 2024.
 - [78] Vishal Athreya Baskaran, Jolene Ranek, Siyuan Shan, Natalie Stanley, and Junier B. Oliva. Distribution-based sketching of single-cell samples, 2022.
 - [79] Michał Koziarski. Two-stage resampling for convolutional neural network training in the imbalanced colorectal cancer image classification, 2021.
 - [80] Parsa Karami and Reza Elahi. Radiomics-based artificial intelligence (ai) models in colorectal cancer (crc) diagnosis, metastasis detection, prognosis, and treatment response, 2024.
 - [81] Zitong Jerry Wang, Alexander M. Xu, Aman Bhargava, and Matt W. Thomson. Generating counterfactual explanations of tumor spatial proteomes to discover effective strategies for enhancing immune infiltration, 2023.
 - [82] Adalberto Claudio Quiros, Roderick Murray-Smith, and Ke Yuan. Pathologygan: Learning deep representations of cancer tissue, 2021.
 - [83] Zhendong Sha, Yuanzhu Chen, and Ting Hu. Genetic heterogeneity analysis using genetic algorithm and network science, 2023.

-
- [84] Liyu Shi, Xiaoyan Li, Weiming Hu, Haoyuan Chen, Jing Chen, Zizhen Fan, Minghe Gao, Yujie Jing, Guotao Lu, Deguo Ma, Zhiyu Ma, Qingtao Meng, Dechao Tang, Hongzan Sun, Marcin Grzegorzek, Shouliang Qi, Yueyang Teng, and Chen Li. Ebhi-seg: A novel enteroscope biopsy histopathological haematoxylin and eosin image dataset for image segmentation tasks, 2022.
 - [85] Nisarg A. Shah, Divij Gupta, Romil Lodha, Ujjwal Baid, and Sanjay Talbar. Colorectal cancer segmentation using atrous convolution and residual enhanced unet, 2021.
 - [86] Tianwei Yu. Autoencoder-based integrative multi-omics data embedding that allows for confounder adjustments, 2021.
 - [87] Trinh Thi Le Vuong, Quoc Dang Vu, Mostafa Jahanifar, Simon Graham, Jin Tae Kwak, and Nasir Rajpoot. Impash: A novel domain-shift resistant representation for colorectal cancer tissue classification, 2022.
 - [88] Elliott Gordon-Rodriguez, Thomas P. Quinn, and John P. Cunningham. Data augmentation for compositional data: Advancing predictive models of the microbiome, 2022.
 - [89] Lydia A. Schoenpflug, Maxime W. Lafarge, Anja L. Frei, and Viktor H. Koelzer. Multi-task learning for tissue segmentation and tumor detection in colorectal cancer histology slides, 2023.
 - [90] Divyanshu Aggarwal and Yasha Hasija. A review of deep learning techniques for protein function prediction, 2022.
 - [91] Zhenyu Wu, Patrick J Lawrence, Anjun Ma, Jian Zhu, Dong Xu, and Qin Ma. Single-cell techniques and deep learning in predicting drug response. *Trends in Pharmacological Sciences*, 41(12):1050–1065, 2020.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

www.SurveyX.Cn