Received: 15 July 2024 | Accepted: 1 August 2024

DOI: 10.1111/cts.70001

POSITION PAPER

From organs to algorithms: Redefining cancer classification in the age of artificial intelligence

Sean Khozin[1,2,3]

[1]CEO Roundtable on Cancer, Morrisville, North Carolina, USA

[2]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[3]Phyusion Bio, LLC, New York, New York, USA

Correspondence
Sean Khozin, CEO Roundtable on Cancer, Morrisville, NC, USA.
Email: khozin@mit.edu

Abstract

Traditional cancer classification based on organ of origin and histology is increasingly at odds with precision oncology. Tumors in different organs can share molecular features, while those in the same organ can be heterogeneous. This disconnect impacts clinical trials, drug development, and patient care. Recent advances in artificial intelligence (AI), particularly machine learning and deep learning, offer promising avenues for reclassifying cancers through comprehensive integration of molecular, histopathological, imaging, and clinical characteristics. AI-driven approaches have the potential to reveal novel cancer subtypes, identify new prognostic variables, and guide more precise treatment strategies for improving patient outcomes.

INTRODUCTION

The classification of cancer has traditionally relied on organ of origin and histological characteristics observable through microscopy. This approach has been foundational to our understanding of cancer biology and treatment for over a century, shaping how we diagnose, research, and treat various forms of cancer. However, as our knowledge of cancer biology has advanced, it has become increasingly clear that the traditional classification system is at odds with the principles of modern precision oncology.

Recent developments in molecular profiling and advanced imaging techniques have revealed a more complex picture of cancer biology. We now understand that tumors arising in different organs can share key molecular and morphological features, while tumors in the same organ can be strikingly heterogeneous at multiple levels of analysis. This heterogeneity extends beyond simple histological differences to encompass a wide range of genetic, epigenetic, and functional characteristics.

The disconnect between traditional organ-based classification and emerging multimodal approaches has significant implications for various aspects of cancer research and treatment. It affects the design and interpretation of clinical trials, potentially obscuring important subgroup effects or leading to the development of therapies that are effective only in a subset of patients within a broadly defined cancer type. In drug development, this discrepancy can result in missed opportunities or the pursuit of therapies that may not be optimally targeted. Ultimately, this impacts patient care, as treatment decisions based on broad classifications may not always align with the specific biological characteristics of an individual's tumor.

In recent years, artificial intelligence (AI), particularly machine learning and deep learning, has emerged as an important tool for addressing these challenges. AI offers new avenues for reclassifying cancers based on a comprehensive integration of molecular, imaging, histopathological, and clinical characteristics. Advanced computational methods can analyze vast amounts of multimodal data, identifying patterns and relationships that may not be

apparent through traditional disease classification methods. By embracing the full complexity of tumor biology through AI-enabled multimodal analyses, we can move beyond the historical constraints of organ-based and histological classifications and towards a new era that truly delivers on the promise of personalized medicine for patients with cancer.

THE EVOLUTION OF CANCER CLASSIFICATION

The evolution of cancer diagnosis reflects the progression of medical science, from ancient empirical observations to contemporary molecular techniques. The earliest documented description of cancer dates to ancient Egyptian papyri, with Hippocrates subsequently introducing the terms "carcinos" and "carcinoma" to describe solid tumors.[1] Advances in surgical exploration, microscopy, and immunohistochemistry have incrementally enhanced our capacity to diagnose and classify neoplasms.

The surgical treatment of cancer patients in the 18th and 19th centuries led to the codification of anatomical tumor classification, as surgeons categorized tumors based on their location and gross appearance. This approach was further refined with the advent of microscopy in the late 19th century, which enabled the development of histological classification systems. Recent years have witnessed the emergence of sophisticated molecular profiling techniques, including next-generation sequencing, transcriptomics, and proteomics at single-cell resolution, providing unprecedented insights into cellular heterogeneity and molecular pathways.

Despite the recent advances, current disease classification systems remain predominantly anchored in anatomical location and histological features. This framework, rooted in historical medical practices rather than reflective of our evolving understanding of underlying biological processes, constrains the full exploitation of cutting-edge diagnostic modalities, impeding the development of more precise therapeutic strategies.

LIMITATIONS OF CURRENT CLASSIFICATION SYSTEMS

The current cancer classification system, based primarily on anatomical origin, is a historical artifact with significant limitations. For example, terms such as "lung cancer" stem from early surgical observations of anatomic tumor locations and past diagnostic constraints. This approach can be misleading in modern oncology, as cancers from the same organ often exhibit diverse molecular drivers and biological behaviors. Non-small cell lung cancer (NSCLC) exemplifies this problem, paradoxically defining what the cancer is not, rather than what it is. This classification, based on human visual inspection of histopathology, groups together a heterogeneous array of malignancies with distinct molecular profiles and treatment responses, obscuring crucial differences at the mechanistic and molecular levels of analysis.

The essential tenet of precision medicine postulates that administering the right therapy to the right patient at the right time and dose should theoretically result in 100% response to treatment. However, in practice, we often fall short of this goal, not because the principle is flawed, but because our current disease classification methods frequently fail in matching the right therapy to the right patient. This is demonstrated by the inherent inter- and intra-observer variability in human sensory-based assessments across multiple diagnostic modalities, including histopathology and radiology.[2]

The distribution of tumor responses observed across cancer clinical trials and the point of care consistently reveals a heterogeneous patient population with variable degrees of clinical benefit, with a substantial proportion of patients with metastatic disease deriving minimal to no therapeutic advantage due to reliance on traditional disease classification schemas for treatment selection.

THE PROMISE OF AI-DRIVEN MULTIMODAL DISEASE CLASSIFICATION

AI is emerging as a critical tool in oncology, offering unprecedented opportunities for reimagining current cancer classification standards. The integration of AI methodologies with diverse data modalities can take us beyond the limitations of traditional classification systems and toward a more nuanced, mechanistic understanding of cancer biology.

Digital pathology and AI-assisted histopathology

Convolutional neural networks (CNNs) have demonstrated remarkable capabilities in analyzing whole slide images (WSIs). These deep learning models, inspired by the human visual cortex, employ multiple layers of interconnected nodes to automatically learn hierarchical features from images, enabling them to recognize complex patterns with high accuracy.[3] For instance, a deep learning system for Gleason grading of prostate biopsies achieved a level of agreement with expert

uropathologists, and a higher performance compared with general pathologists.[3] Such level of performance not only enhances diagnostic accuracy but also has the potential to standardize grading across institutions, addressing a critical need in clinical practice. Moreover, AI-assisted digital histopathology can dramatically improve workflow efficiency, having been shown, for example, to reduce diagnosis time by up to 62% for plasma cell identification and 33% for lymph node metastasis detection, while also decreasing diagnostic uncertainty and additional testing requests by over 60%.[4]

Perhaps the most transformative potential of AI in digital histopathology for disease reclassification lies in its ability to identify de novo features that speak to new disease phenotypes, transcending the limitations of human visual assessment and traditional staining methods. Deep learning models can analyze WSIs at a level of detail and dimensionality far beyond human capability. These models can identify subtle patterns and features in tissue architecture, cellular morphology, and spatial relationships that may be imperceptible or overlooked by human observers.[5] For instance, deep learning methods have identified a novel prognostic marker for patients with colorectal cancer by classifying variations in the stromal microenvironment of tumors.[6] A recent pan-cancer analysis applying deep learning to histopathology images across multiple cancer types discovered novel morphological features predictive of specific gene mutations.[7] These *mutation-associated features* represented previously unrecognized patterns in tissue architecture and cellular morphology. In endometrial cancer, an AI-powered analysis of WSIs has identified a distinct subset of patients with significantly inferior progression-free and disease-specific survival.[8]

Bridging the gap between morphology and molecular features

The ability of AI to integrate information across multiple scales, from subcellular features to tissue-wide patterns, offers a unique opportunity to bridge the gap between morphological and molecular classifications. Deep learning model have been able to predict a range of molecular features, including gene mutations and microsatellite instability, directly from H&E-stained WSIs in colorectal cancer.[9] A recent study has shown that a deep learning framework can predict genome-wide tumor mRNA expression from WSI while being capable of predicting responses to targeted and immune therapies based on the inferred expression values.[10]

The convergence of morphological and molecular data facilitated by AI-driven analysis has significant implications for reclassifying oncological diseases. By correlating visual patterns observed in histological samples with specific genomic alterations and expression profiles, these methods can yield classifications that are both biologically meaningful and clinically relevant. This integrative approach enable the identification of new cancer phenotypes characterized by unique combinations of morphological and molecular features.

Radiomics and AI-enhanced imaging analysis

AI-enhanced radiomics has demonstrated remarkable efficacy in improving the accuracy and consistency of cancer detection, characterization, and staging across various imaging modalities. Deep learning models, particularly CNNs, have shown exceptional performance in tasks such as lung nodule detection and classification on CT scans. For example, a lung cancer screening study utilizing CNNs achieved a performance surpassing that of experienced radiologists in tumor detection.[11] Moreover, AI-driven radiomics has shown promise in identifying new patients groups with superior response to radiotherapy, offering valuable information to guide clinical decision-making.[12]

AI-enhanced radiomics can also identify de novo imaging features and patterns that may reveal new disease phenotypes or provide insights into underlying biological processes. A recent study identified four novel radiomics features common to three malignancies, with each feature having unique molecular characteristics and prognostic implications.[13] Such studies demonstrate the value of quantitative imaging features in identification of novel phenotypes in traditionally-defined tumor types, paving the way for the development of more mechanistically-aligned predictive and prognostic biomarkers.

Multimodal and multi-omics integration

Multimodal deep learning approaches, such as attention-based networks and graph neural networks (GNNs), can combine imaging, histopathology, and molecular data to offer a more comprehensive view of tumor biology. In multi-omics analysis, techniques such as autoencoders and variational autoencoders (VAEs) have demonstrated remarkable capabilities in integrating genomic, transcriptomic, and proteomic data. These methods can compress high-dimensional data into lower-dimensional latent spaces, revealing novel cancer subtypes.[14,15] Advanced dimensionality reduction techniques, including self-organizing maps (SOMs), t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP) have been successfully applied

TABLE 1 Comparison of current and artificial intelligence (AI)-driven cancer classification systems.

Feature	Current classification system	AI-driven classification
Primary basis	Anatomical location and histological features	Integration of molecular, imaging, histopathological, and clinical characteristics
Origin	Historical medical practices	Reflective of evolving understanding of biological processes
Precision	Limited by inter- and intra-observer variability	Potentially more objective and reproducible
Molecular insight	Limited integration of molecular data	Comprehensive integration of multi-omics data
Novel features	Limited ability to identify new subtypes	Can identify de novo features and patterns
Data integration	Primarily relies on single modality	Integrates multiple data modalities (imaging, pathology, genomics, etc.)
Treatment matching	Often leads to suboptimal therapy matching	Potential for improved precision in treatment selection
Biological relevance	May group heterogeneous cancers together	Aims to reflect underlying disease mechanisms more accurately
Standardization	Varies across institutions	Potential for improved standardization of diagnostic criteria
Prognostic value	Limited by broad groupings of heterogenous patient populations	Can reveal new prognostic indicators

to visualize high-dimensional omics data. These methods project complex, multi-dimensional datasets onto two-dimensional spaces, facilitating the identification of patterns and relationships within the data.[16] Such visualizations can reveal intricate molecular landscapes and novel subgroups, challenging existing classification paradigms and providing insights into biological processes and disease mechanisms.

DISCUSSION

AI is a powerful tool for reimagining existing cancer classification standards, offering the potential to bridge the gap between the promise of precision oncology and its full clinical realization. AI-powered cancer classification inverts the current diagnostic paradigm by leveraging high-resolution, multi-omic profiling strategies and de novo features extracted using computational methods that aim to closely reflect underlying disease mechanisms, forming the foundation of a new disease classification system. This bottom-up methodology reduces reliance on subjective human interpretation of diagnostic results, prioritizing molecular and functional characteristics over traditional anatomical and histological features. By anchoring classification in comprehensive multi-omics profiles and objectively reproducible image-based assessments, we can achieve a more nuanced and biologically relevant taxonomy of cancer (Table 1).

The implementation of a computationally-powered cancer classification system necessitates addressing several challenges, including data standardization and integration, algorithm interpretability, appropriate methods for clinical validation, ethical considerations, and computational infrastructure requirements. The path forward requires close collaboration among clinicians, researchers, and AI specialists to ensure that the applied methodologies are rigorously validated and appropriately integrated into clinical practice. This interdisciplinary effort is essential to ensure the biological relevance and clinical utility of the derived classifications. By moving beyond the historical constraints of organ-based and histological classification systems, we can move closer to a new era of precision oncology that truly delivers on the promise of personalized medicine for patients with cancer.

FUNDING INFORMATION

No funding was received for this work.

CONFLICT OF INTEREST STATEMENT

Sean Khozin is the cofounder of Phyusion LLC, a firm specializing in advancing innovations at the intersection of biology, technology, and AI.

ORCID

Sean Khozin https://orcid.org/0000-0003-2206-7620

REFERENCES

1. Wagener DJT. *The History of Oncology*. Bohn Stafleu van Loghum; 2009.
2. Benchoufi M, Matzner-Lober E, Molinari N, Jannot AS, Soyer P. Interobserver agreement issues in radiology. *Diagn Interv Imaging*. 2020;101(10):639-641. doi:10.1016/j.diii.2020.09.001
3. Bulten W, Kartasalo K, Chen PHC, et al. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat Med*. 2022;28(1):154-163. doi:10.1038/s41591-021-01620-2
4. Huang Z, Yang E, Shen J, et al. A pathologist–AI collaboration framework for enhancing diagnostic accuracies and efficiencies. *Nat Biomed Eng*. 2024;1-16. doi:10.1038/s41551-024-01223-5
5. Talaei Khoei T, Ould Slimane H, Kaabouch N. Deep learning: systematic review, models, challenges, and research directions. *Neural Comput Applic*. 2023;35(31):23103-23124. doi:10.1007/s00521-023-08957-4
6. Kather JN, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med*. 2019;16(1):e1002730. doi:10.1371/journal.pmed.1002730
7. Fu Y, Jung AW, Torne RV, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat Can*. 2020;1(8):800-810. doi:10.1038/s43018-020-0085-8
8. Darbandsari A, Farahani H, Asadi M, et al. AI-based histopathology image analysis reveals a distinct subset of endometrial cancers. *Nat Commun*. 2024;15(1):4973. doi:10.1038/s41467-024-49017-2
9. Gustav M, Reitsam NG, Carrero ZI, et al. Deep learning for dual detection of microsatellite instability and POLE mutations in colorectal cancer histopathology. *Npj Precis Oncol*. 2024;8(1):1-11. doi:10.1038/s41698-024-00592-z
10. Hoang DT, Dinstag G, Shulman ED, et al. A deep-learning framework to predict cancer treatment response from histopathology images through imputed transcriptomics. *Nat Can*. 2024;1-13. doi:10.1038/s43018-024-00793-2
11. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019;25(6):954-961. doi:10.1038/s41591-019-0447-x
12. Lou B, Doken S, Zhuang T, et al. An image-based deep learning framework for individualizing radiotherapy dose. *Lancet Digit Health*. 2019;1(3):e136-e147. doi:10.1016/S2589-7500(19)30058-5
13. Wu J, Li C, Gensheimer M, et al. Radiological tumour classification across imaging modality and histology. *Nat Mach Intell*. 2021;3(9):787-798. doi:10.1038/s42256-021-00377-0
14. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res*. 2018;24(6):1248-1259. doi:10.1158/1078-0432.CCR-17-0853
15. Vale-Silva LA, Rohr K. Long-term cancer survival prediction using multimodal deep learning. *Sci Rep*. 2021;11(1):13505. doi:10.1038/s41598-021-92799-4
16. Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2019;37(1):38-44. doi:10.1038/nbt.4314

How to cite this article: Khozin S. From organs to algorithms: Redefining cancer classification in the age of artificial intelligence. *Clin Transl Sci*. 2024;17:e70001. doi:10.1111/cts.70001