

Applications of AI in Cancer Diagnosis and Categorization

Ziling Zhou

Highland Park High School, 4220 Emerson Ave, Dallas, TX 75205, USA

zhouziling07@gmail.com

Abstract. Cancer is a highly individualized disease, as tumors both take on traits from their host and develop their own mutations. Thus, it is critical to better understand the characteristics and behavior of tumors to treat patients more effectively. There are various artificial intelligences (AI) that can and are being used to improve the development of oncological technologies. AI is designed to replicate human intelligence, but not all models are capable of the same tasks. In diagnostics, AI models can be used to analyze multimodal data, to improve the accuracy of cancer diagnostics. AI can also identify cancer stage and grade. A tumor's stage depends on tumor size, while its grade depends on how much the tumor has spread. By subtyping tumors, researchers are able to develop more targeted drugs, as similar tumors most likely arose through similar pathways. However, data from studies is unverified and different, which makes it hard to find formatted data to train AI models. Although the application of AI clinically is promising, AI still needs to go through more trials before it's applied clinically.

Keywords: Artificial Intelligence (AI), imaging, data, classification.

1. Introduction

Each individual has variations within their cells, which results in a unique network of cellular interactions [1]. The development of a tumor is affected by this cellular microenvironment, which includes immune cells, extracellular matrix, and various proteins. The microenvironment surrounding a possible tumor cell can determine if that cell remains localized or spreads aggressively [2]. Tumors also gain a unique combination of genomic and epigenomic features from both their host and random mutations. Thus, each tumor arises through an irreplacable pathway, even though these varying pathways can result in the same type of cancer [1]. Additionally, not every cell within a tumor has the same mutations, causing molecular variability among the tumor cells. Thus, it is critical to develop precise medicine by analyzing the details of the tumor [1]. The information helps decide on the effective treatment for each patient [3].

Early-stage cancer is more likely to be treated successfully [4]. Therefore, early diagnosis of cancer could improve patient prognosis [5]. Additionally, successful treatment often results in less treatment time. This would alleviate the current burden on healthcare systems [4].

With the growth of online databases, artificial intelligence (AI), and digital technology through the years, there is now a plethora of computational tools that can analyze, integrate, and use the data generated by diagnostics [6]. AI and machine learning (ML) approaches, especially various neural network models, have been applied towards discovering cancer biomarkers, detecting cancer-specific

patterns in datasets, and finding prospective statistical classifiers to effectively evaluate diagnostic performance in independent cohorts [2].

This review summarizes the categories of AI used in the biomedical field and the applications of AI in cancer diagnosis, including staging, grading, and subtyping. Moreover, this review discusses the problems of using AI in cancer diagnosis and aims to provide insight into the clinical applications of AI.

2. What is AI?

AI is defined as the intelligence simulated by technology and there are multiple ways for models to be taught (Figure 1). The definitions of AI-related terms are summarized in Table 1.

For supervised learning, through feeding the AI labelled data, the AI will learn to associate certain features with the label. Eventually, the AI will be able to correctly label unlabelled data.

For unsupervised learning, AI finds patterns within unlabelled data. Since there's largely no human input, the AI must discover and understand the data by itself, resulting in the discovery of unusual or undiscovered patterns that could aid in cancer diagnosis and treatment.

For reinforcement learning, AI is rewarded or punished based on its actions and adjusts based on the feedback. Through this loop it eventually learns what to do and what not to do.

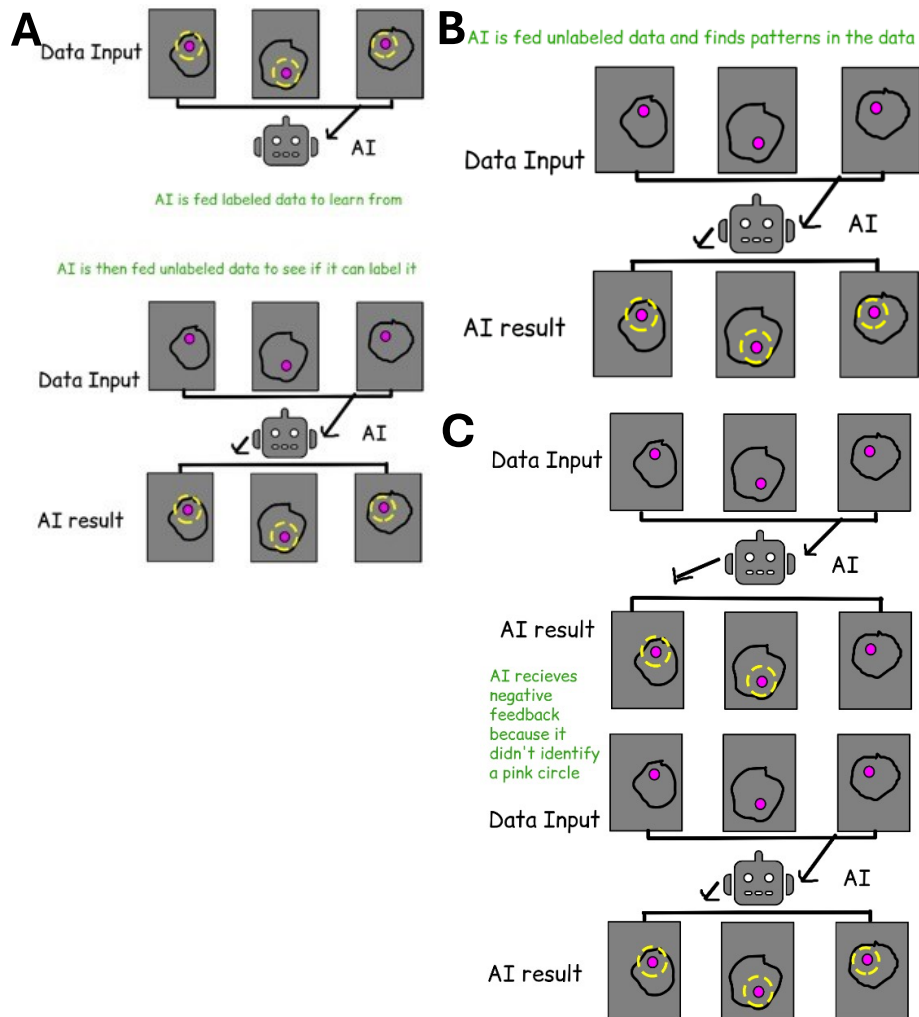


Figure 1. Three types of Artificial Intelligence (AI). A) Supervised learning: Labeled data is inputted into the AI and the AI learns how to label unlabelled data. B) Unsupervised learning: AI recognizes patterns within unlabelled data. C) Reinforcement learning: AI receives feedback based on its actions. It learns from the feedback. Figure credit: original.

Table 1. Definitions of AI-related terms

Term	Definition	Ref
Artificial Intelligence (AI)	A machine simulation of human intelligence	[7]
Machine Learning (ML)	Machines fed data in order to teach them how to emulate human predictions	[7]
Deep learning (DL)	A type of ML using an artificial representation of a human brain	[7]
Neural Network (NN)	The underlying technology used in DL. It is an artificial representation of a human brain, with several layers of “neurons” and separate locations for input and output.	[7]
Convolutional Neural Network (CNN)	A DL model characterized by the use of several partially-connected convolutional layers.	[7]
Fully Convolutional Network (FCN)	A DL model characterized by hierarchical communication among different layers.	[7]
Recurrent Neural Network (RNN)	A DL model that can receive inputs at different times. It uses connected computation units to perform recursive computation on inputs.	[7]

3. Diagnosis

The analysis of histologic sections is the basis of diagnosis and prognostication for cancer [8]. However, unimodal data such as histologic sections cannot explain all clinical behaviors. Analyzing heterogeneous data can improve prediction of disease invasiveness, stratification, and patient outcomes. Therefore, many researchers have attempted to combine genomic and pathological image features to create more accurate diagnostic companion tools [9]. Predictive models can be made by combining several features, making it worthwhile to pursue the merging and integration of various relevant domains [10]. Since AI has been applied to tasks including image processing and classification, AI could also be applied towards relevant novel technologies like multimodal fusion [7]. Additionally, AI could be a tool to help doctors with diagnosis. An AI-aided endoscopy has been shown to have high performance and clinical diagnostic value [11]. In another study, scientists created and tested a promising AI cervical cancer screening system (AICCS) that diagnosed cervical cancer. The system achieved a sensitivity of 0.946, a specificity of 0.890, and an accuracy of 0.892 [5]. However, these systems have been shown to perform better when diagnosing specific tumor types such as breast and prostate cancer, proving that AI might not be an effective tool for all cancers yet [9].

4. Staging and Grading

Risk classification can be accomplished using many different systems such as the metastases (TNM) staging system, molecular features, clinical variables, etc [9]. Although the stage and grade of cancer can sometimes be confused, they’re completely different things. A cancer’s stage is based on how large the tumor is and if it has spread, while the grade of cancer is based on the appearance of the tumor cells [12,13]. To determine the grade of a tumor, pathologists analyze a set of morphological features and images obtained through radiographic and pathological imaging techniques [8,13]. The grade affects a physician’s assessment of a tumor's aggressiveness and a patient's prognosis [8]. The more normal the cells look the less aggressive and harmful the tumor will be and vice versa [14]. A cancer’s stage and grade has been linked to patient outcome, progression, and survival for a number of cancers [13].

DL systems have leverage over other systems since they can be rapidly developed and adapted based on different criteria, including different cancer types and biomarkers. As long as DL systems have been trained enough, these approaches can be highly accurate. Furthermore, with the development of DL-

specific chips and the improvement of computing hardware, DL systems can be extremely efficient and rapidly analyze data [13].

The AICCS system mentioned earlier was also shown to accurately grade cervical tumors as well. The system demonstrated high sensitivity towards abnormal cells, allowing it to identify them. This system could lead to early detection and intervention for cervical cancer [5].

5. Subtyping

The goal of molecular classification is to identify shared characteristics between carcinogenesis pathways. Tumors that are similar may have arisen through similar pathways, and similar tumors may behave similarly. Therefore, these classifications may predict disease course and treatment response [1]. The tumor classification system is both hierarchical and multidimensional with different sets of classifiers, such as genomic, epigenomic, molecular, and interactomic classifiers [1].

Human understanding of cancer has vastly increased by subtyping tumors. Researchers can develop more specific drugs for each subtype, which may improve patient outcomes. Other novel therapeutics may be developed based on the improved understanding. However, these therapeutics are still far from the hospital. They still have to be refined through various studies so that the end product will be reliable, cost-effective, and accessible to local labs

Although the genetic profiles of different cancer types have been rarely shown to overlap, signaling pathways related with specific tumor types are more uniformly represented within distinct molecular subtypes. All currently known driver genes can be sorted into at least 1 of the 12 core signaling pathways that enhance tumor growth [15]. This allows different cancer types to express the same deregulated pathways, even though they have vastly different mutations. In breast, lung, and colorectal tumors, the PI3K and RAS signaling pathways are commonly altered. However, other pathways, such as the TGF- β /mesenchymal and angiogenic pathways, are associated with other cancer types or subtypes [15].

Assessment of molecular subtypes via high - throughput sequencing is not widely available or cost - effective, but AI is. This makes AI a promising novel technology for identifying various tumor subtypes [16].

6. Problems and obstacles

DL algorithms require massive amounts of data, sophisticated imaging technology, top-tier statisticians, and a lot of money. Data collection itself is challenging due to the differences between data collected in studies. Studies vary in sample size, research design, data source, and imaging collecting criteria, undermining attempts to cohesively organize datasets. Furthermore, the clinical applications of AI are limited because prospective datasets lack validation [4]. Validation is important, as researchers might be biased without external validity [10].

Worryingly, AI models disregard social and cultural risk variables, making them inaccurate [10]. Ethnic minorities have worse outcomes in some cancer types when compared to others, which may be caused by differences during the diagnostic phase [17]. To improve models, data about critical risk factors such as demographics, smoking status, and tumor histology should be included [10]. Since all cancers are so different, systems have limited ability to generalize all the types of tumors [10].

7. Conclusion

Cancer is affected by the unique cellular network of each host, so each tumor is different. The individualization of tumors makes cancer treatment a time and resource-consuming task. If developments in AI can be used to improve cancer understanding and diagnosis, many people and resources will be saved. AI is easily accessible, easy to train, and can process high amounts of data, making it the perfect tool for the task. In diagnostics, AI is able to condense massive amounts of multimodal data to increase diagnostic accuracy. Many diagnostic technologies can also be applied towards identifying cancer stage and grade since they all depend on imaging. Stage and grade are often confused, but cancer stage is determined by tumor size while grade is determined by tumor cell appearance. Additionally, AI can be used to subtype tumors. This allows researchers to better understand

the mechanisms behind cancer, develop more targeted drugs, and identify traits that can negatively affect patient prognosis. Subtyping relies on the logic that similar tumors came from similar pathways.

However, AI models aren't perfect. AI cannot take societal factors into account and datasets are hard to verify and organize. As AI tools are gradually developed and approved, various trials are required to ensure that the resulting tools are accessible and accurate. AI technology has the potential to save lives and reduce the healthcare burden.

References

- [1] Ogino S, Fuchs CS and Giovannucci E 2012 *Expert Rev. Mol. Diagn.* 12 621-8
- [2] Crosby D, *et al.* 2022 *Science* 375 eaay9040
- [3] NIH National Cancer Institute, How Cancer Is Diagnosed (2024), Available online at: <https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis>
- [4] Cancer Research UK, Why is early cancer diagnosis important (2023), Available online at: <https://master-7rqtwti-hreqyzlibi4ac.uk-1.platformsh.site/about-cancer/cancer-symptoms/why-is-early-diagnosis-important/1000>
- [5] Wang J, *et al.* 2024 *Nat. Commun.* 15 4369
- [6] Delpierre C and Lefèvre T 2023 *Front. Sociol.* 8 1112159
- [7] Bera K, *et al.* 2019 *Nat. Rev. Clin. Oncol.* 16 703-15
- [8] Beck AH, *et al.* 2011 *Sci. Transl. Med.* 3 108ra113
- [9] Feng X, *et al.* 2024 *J. Transl. Med.* 22 131
- [10] Silva HECD, *et al.* 2023 *PLoS One* 12 e0292063
- [11] Li N, *et al.* 2024 *PLoS One* 19 e0303421
- [12] NIH National Cancer Institute, Cancer Staging (2022), Available online at: <https://www.cancer.gov/about-cancer/diagnosis-staging/staging>
- [13] Bera K, Katz I and Madabhushi A 2020 *KCO Clin. Cancer. Inform.* 4 1039-1050
- [14] NIH National Cancer Institute, Tumor Grade (2022), Available online at: <https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-grade>
- [15] Santos C, *et al.* 2015 *Cell. Oncol.* 38 3-16
- [16] Angeloni M, *et al.* 2024 *J Pathol. Clin. Res.* 10 e12369
- [17] Martins T, Hamilton W and Ukoumunne OC 2013 *BMC Fam. Pract.* 14 167