REVIEW

Pakistan Journal of Medical and Biological Sciences
https://pjmbs.gcuf.edu.pk/index.php/pjmbs

THE ROLE OF ARTIFICIAL INTELLIGENCE IN CYTOPATHOLOGY AND HISTOPATHOLOGY: APPLICATIONS AND LIMITATIONS IN ONCOLOGY DIAGNOSIS AND PROGNOSTICATION

Nimrah Mehmood[1], Naheed Akhtar[1, 2]*, Sidra Huma[3], Muhammad Saleem Qureshi[4]

ABSTRACT: The application of artificial intelligence (AI) in oncologic pathology is transforming cancer diagnostics by utilizing histopathological, cytopathological and radiomic data to improve accuracy, efficiency and clinical decision-making. AI-driven models play a significant role in classifying neoplasms, grading tumors and interpreting biopsies, which in turn aids in early diagnosis and the development of specialized pharmaceutical care plans. However, several challenges persist, including biases in training datasets, inconsistencies in pathological image acquisition and limitations regarding real-world applicability, all of which produce barriers to the widespread integration of AI in digital pathology and early detection of tumor. This review examines the advantages and drawbacks of predictive AI in the assessment of histopathological cancer, analyses of tumor progression and the identification of prognostic biomarkers. It also analyzes the contributions of generative AI models, such as AI integrated chatbots, in enhancing communication between pathologists and clinicians, as well as improving patient comprehension of cytopathological findings. Furthermore, the discussion addresses the need of regulatory considerations, data standardization and the ongoing initiatives aimed at enhancing the reliability and interpretability of AI-driven pathology tools within clinical oncology.

Authors Affiliation:
[1]Department of Pharmaceutics, The Islamia University of Bahawalpur, Bahawalpur, Punjab, Bahawalpur 63100, Pakistan.
[2]Department of Pharmacy, University of Poonch Rawalakot Azad Kashmir
[3]Bahawalpur medical and dental College, Pharmacy department, Pakistan.
[4]Akbar Institute of Medical Sciences Bahawalpur.
Correspondence Author Email:
Naheedakhtar96@gmail.com

KEYWORDS: Artificial Intelligence, Cytopathology, Histopathology, Oncology, Mamagraphy

INTRODUCTION:

Advancements in the field of oncology have led to a significant 33% reduction in cancer mortality rates over the past 32 years (1). The adoption of precise medicine, particularly utilizing artificial intelligence (AI), help oncologists to recognize the previously unidentified patterns in radiological images, disease prognosis and personalized treatment strategies to individual patients. The influence of AI-driven diagnostic tools is particularly notable in breast cancer, where it has achieved a 31% impact as well as in lung and prostate cancer, which provide an 8.5% improvement (2). The advancements and integration of AI technologies into oncology workflows has transformed the analysis of electronic health records (EHRs), thereby enhancing clinical decision-making, reduce diagnostic delays and improving patient management overall. However, the reliability of AI applications in clinical environments leads towards numerous challenges due to variable performance. Biases introduced during the integration of algorithms can lead to diagnostic inaccuracies, which may strain healthcare systems and adversely affect clinical outcomes. This review examined the three significant applications of AI in oncology: diagnostic and histopathological classification, the detection of neoplastic progression and the integration of AI-powered chatbots aimed at improving patient engagement and optimizing clinical workflows.

While the first two applications heavily depend on predictive analytics, the latter utilizes generative AI. Key studies in each area will be discussed to provide AI's current contributions and its future potential for cancer care. Furthermore, this review will address inherent biases, existing limitations and strategies for enhancing AI-driven solutions in oncology. The discussion will conclude with an examination of regulatory considerations for the integration of AI within oncological practice. Self-operating algorithms powered by data that are specifically made to handle complex problems are included in artificial intelligence (AI). Machine learning (ML) and deep learning (DL) are two core AI approaches. As a subset of artificial intelligence, machine learning (ML) enables algorithms to recognize patterns in data on their own without explicit programming (3). By using multilayered mathematical structures that replicate neural networks, DL, on the other hand, is a more advanced extension of ML that analyzes and interprets medical data, such as tumor markers and histological images. Predictive AI and generative AI are the two basic categories into which AI algorithms fall. As demonstrated by the use of mammography images to detect early-stage cancer, predictive AI models are proficient at identifying patterns in training datasets to forecast clinical outcomes. Conversely, generative AI concentrates on creating new data outputs; conversational agents powered by AI that help cancer patients manage their symptoms are a relevant example. Although AI has significantly advancement in oncology, there are still limitations on how these technologies may be used in clinical settings. Widespread adoption is hindered by a number of issues, such as the Food and Drug Administration's (FDA) strict regulations, the high upfront costs of integrating AI into clinical workflows, the complexity of complex algorithms, and inadequate post-deployment monitoring (4). Over 80% of the 71 AI-enabled medical gadgets that the FDA authorized in 2021 were specifically designed to diagnose cancer. Radiation oncology (8.5%), pathology (19.7%), and oncological radiology (54.9%) make up the majority of this category. The use of these AI-powered systems has primarily addressed solid tumors, highlighting AI's revolutionary potential in the fields of therapeutic development and oncological pathology. When integrated into oncology clinical workflows, artificial intelligence (AI) technologies have proven significant potential, especially in their ability to evaluate electronic health records and assist oncologists in reaching well-informed clinical decisions. In addition to improving diagnostic accuracy, this integration speeds up decision-making and strengthens patient care practices. However, using AI-driven technologies raises significant issues, particularly with regard to their reliability after deployment. Machine learning models that contain algorithmic biases may produce erroneous prognostic or diagnostic predictions, which may have a detrimental impact on patient outcomes, multidisciplinary oncology teams, and healthcare systems. The diagnosis of neoplastic diseases and histopathological classification, tumor prognosis and treatment outcome prediction, and the use of AI-driven chatbots to enhance patient engagement and clinical workflow efficiency are the three key applications of AI in oncological care that will be covered in this review. While the third application uses generative AI to improve patient communication and expedite various oncological care processes, the first two applications use predictive analytics to detect cancer early and evaluate its progression. To assess the revolutionary potential of AI in oncology practice, we will look at important primary papers from each field. This review also aims to critically examine systemic barriers, diagnostic limitations, and algorithmic biases in order to suggest practical approaches that will increase AI's effectiveness in this area. Finally, we will discuss regulatory issues related to the use of AI in cancer treatment, making sure that this integration is therapeutically effective, reliable and ethical.

Diagnosis

During radiographic and histological examinations, early-stage cancers and recurring neoplasms after therapy may cause serious diagnostic problems, especially in patients presenting clinical stability. The fact that some clinical abnormalities are subtle and frequently go unnoticed by human observers adds to this intricacy. In areas where traditional methods may not be as accurate, the use of machine learning models, especially those trained on extensive datasets that include radiographic scans, histopathological slides, and dermoscopic images, has demonstrated promise in differentiating between benign and malignant lesions. One of the most widely used AI techniques for picture classification is convolutional neural networks (CNNs), a type of deep learning models that may extract distinctive clinical characteristics that are useful for differentiating between different disease states. By producing probabilistic scores for every diagnostic category, these models make it possible to classify photos according to the category that has the highest probability. Algorithmic predictions are compared against clinician-established "ground truth" diagnoses in order to evaluate the efficacy of AI-assisted diagnosis. In oncology, the use of AI-driven diagnostic tools enhances operational efficiency, facilitates scalable solutions across various healthcare infrastructures, and advances the early diagnosis of diseases with exceptional precision. Notably, the identification and treatment of breast and skin cancer have greatly benefited from developments in AI-assisted diagnostics.

AI in Dermatopathology

The early detection of cutaneous cancers has advanced significantly as a result of the application of artificial intelligence in dermatopathology. Despite being widespread, some types of skin lesions, including melanoma, present serious oncological risks. Even though melanoma accounts for only 5% of skin cancer cases, when it is discovered at a metastatic stage, the five-year survival rate is significantly lower at 32%, and when it is discovered locally, the survival rate is 99% (5). Because of their visual similarities, it can be difficult to distinguish between benign skin lesions and early-stage melanoma, which frequently leads to misclassification or delays in diagnosis. By recognizing detailed pathological differences, AI systems strengthen this distinction and empower clinicians to launch therapeutic actions on time. Researchers looked at a deep learning system designed especially to categorize benign nevi, non-neoplastic skin diseases, and malignant melanoma in a crucial 2017 proof-of-concept study. A collection of 129,450 biopsy-confirmed dermatological pictures was used to train the convolutional neural network (CNN), which was then assessed using the opinions of 21 board-certified dermatologists (6). Surprisingly, the algorithm's classification accuracy was 72.1%, which closely matched the results of two dermatologists who received scores of 66.0% and 65.6%, respectively. These findings demonstrate artificial intelligence's (AI) enormous potential as an adjunct to traditional diagnostic procedures. A histopathology biopsy is used to confirm the diagnosis of melanoma after an in-person dermoscopic examination. However, AI driven systems that can achieve dermatologist-level accuracy without requiring invasive procedures present an unparalleled approach, especially for the underprivileged population with little access to specialized medical care. Additionally, recent multicenter prospective trials have shown the accuracy of AI-driven skin lesion classification from mobile photos, showing diagnostic performance comparable with both experienced and inexperienced doctors (7).

AI in Mammography

AI applications have significantly altered early detection techniques in the field of breast cancer screening. Since 1989, the death rate from breast cancer has decreased by 43% as a result of the development of mammography screening, which has been fueled by strong public health campaigns and technological advancements (8, 9). Currently, methods usually use a two-reader structure in which separate radiologists examine mammograms and consult a third expert in the event that discrepancies are found. However, the limited number of radiologists and screening tools continues to be a problem, made worse by differences in interobserver interpretations and possible human error in scan assessments. According to a thorough analysis of more than 1.6 million mammograms from 359 radiologists, 41% of practitioners showed unpredictable compliance with defined recall standards. Deep learning architectures are used by AI-based computer-aided detection (CAD) systems to identify alarming regions in mammograms, which helps radiologists reduce effort and improve diagnostic accuracy (10). However, differences in performance between various AI models are still evident. The effectiveness of three commercially available AI CAD algorithms in screening 8,805 women for breast cancer was evaluated in an external validation study. Only one algorithm was able to meet the sensitivity and specificity criteria established by the U.S. Breast Cancer Surveillance Consortium, and this top model had a diagnostic accuracy of 95.6%, higher than the average accuracy of 92.1% for the other two models. Additionally, combining AI-based analyses with preliminary radiological evaluations led to an 8% increase in cancer diagnosis rates.

Together, these developments highlight how important AI is to supporting precision oncology by reducing diagnostic inconsistencies, speeding up clinical decision-making, and expanding access to cancer screening programs. In this study, we examined the combination of artificial intelligence (AI) and human clinical knowledge, and found that their combination produced higher effectiveness than either using AI alone or the conventional dual-reader method for mammography evaluations, which requires interpretations by primary and secondary radiologists. The 2020 publication of the DM DREAM study further supported the idea that combining AI with clinical knowledge greatly enhances decision-making in breast cancer screening allocations, particularly when compared to assessments conducted by a single radiologist (11). In this study, the combined interpretation of mammograms by AI algorithms and radiologists was compared to traditional single-radiologist evaluations. The results showed increases in diagnostic precision and a significant 1.5% absolute decrease in recall rates. In addition to better mammography interpretation, AI helps physicians by reducing their burden, which boosts productivity and frees up more time for face-to-face patient interaction. When used as secondary readers, AI-based diagnostic tools performed diagnostically on par with human radiologists in a retrospective analysis of 275,000 cases of breast cancer (12). The validation of AI-based diagnostic tools through prospective clinical trials is essential to verifying their safety and effectiveness in practical situations, in addition to advancements in algorithms and radiological assessments. The first-ever randomized controlled trial (RCT) evaluating the safe incorporation of AI-assisted computer-aided detection (CAD) techniques into clinical workflows was the Mammography Screening with Artificial Intelligence (MASAI) study performed in Sweden (13). An AI CAD system was used in this prospective, noninferiority, single-blinded study to classify mammography interpretations into single-reader and dual-reader frameworks. An evidence-based basis for integrating AI techniques into breast cancer screening methods was established by prioritizing high-risk scans indicated through CAD classification, which led to a 20% improvement in cancer detection rates and a nearly halving of the overall workload.

Bias and Representation Challenging Computer-aided Diagnosis (CAD)

Problems like underrepresentation and inconsistent image acquisition can lead to biases in AI training datasets (14). Because of this, AI models could find it difficult to generalize well to patient groups that are underrepresented in training data. In dermatological oncology, for instance, AI-powered diagnostic algorithms are less successful in detecting cutaneous cancers in people with darker skin tones (15). The fact that many AI models are trained on publicly accessible datasets, which frequently contain biases, complicates this issue. An analysis of 21 publicly available skin lesion datasets with over 100,000 photos, for example, revealed a significant underrepresentation of lesions on darker skin tones (16). In particular, only 11 of the 2,436 photos with skin tone information showed dark or black skin. Furthermore, none of the 1,585 photos that included ethnic background information featured people of South Asian, African, or Afro-Caribbean ancestry. It is essential to make adjustments at different phases of AI model development in order to address these discrepancies. Diversifying training datasets is necessary to guarantee that they contain pictures that accurately represent every demographic group, encompassing a range of skin tones, ages, and body shapes. To increase the reliability of algorithms, pictures should also be taken from a variety of angles, in a range of lighting situations, and with a range of imaging devices. To stay in touch with advancements in image technology, AI models must also be periodically retrained.

Challenges and Future Directions in AI-based Diagnosis

The high-resolution nature of medical imaging, which requires the processing of large amounts of pixelated data, makes the current state of AI-driven picture classification a significant "big data" problem. AI models are able to identify complex relationships in data that traditional regression methods could miss. However, there is a disadvantage to the beneficial performance of sophisticated "black box" AI algorithms: their natural inability to disclose the process by which predictions are made. Some models might produce biased, inaccurate, or non-generalizable interpretations by discovering erroneous connections between image features and diagnostic results. For instance, deep learning algorithms trained on SARS-CoV-2 patients' chest radiographs showed low generalizability upon external validation. Applying explainability techniques to these models showed that, although the AI considered important disease markers such as pulmonary opacities, it also included irrelevant components such image laterality markers and annotations, that are unique to the imaging acquisition protocols of the dataset (17). Therefore, external validation utilizing separate datasets and the application of statistical techniques for model interpretability are essential to guaranteeing the dependability of AI-driven oncologic diagnostics. Finally, to assess the practical dependability of AI-driven diagnostic models in clinical workflows, prospective clinical trials and validation studies are essential. Studying AI algorithms as possible secondary or tertiary readers for diagnostic imaging will shed light on their effectiveness and help guide their broad use in radiological and oncologic diagnoses.

Prognostication in Oncology

In oncology, precise patient outcome prognosis is essential for developing successful treatment plans and directing resources as efficiently as possible. Oncologists, however, have a difficult prognoses; an estimated 63% of practitioners overestimate survival rates, while 17% underestimate them (18). The use of publicly published demographic statistics, like the five-year median survival rates, and clinical precedents, which tend to encourage overgeneralization and result in inaccurate risk assessments, is one reason contributing to this difference. Such inaccurate prognosis can have serious consequences, including increased emotional strain for patients and their caregivers, resource misallocation, deterioration of the patient-physician relationship, and postponements of important therapeutic or end-of-life interventions (19) . Artificial intelligence (AI)-based risk prediction models have surfaced in response to these issues, providing personalized prognostic estimations that improve physicians' risk assessments and assist with individualized oncology care decisions.

Prognostic models are concerned with determining whether a patient would suffer from the progression of their illness or unfavorable outcomes, such hospitalization or death, whereas diagnostic models seek to identify the existence of a disease. Both unstructured data, including radiology reports, clinical notes from electronic health records (EHRs), and pathology findings, as well as structured data, such as patient demographics, test results, and surveys that capture patient-reported outcomes (PROs), are used by these prognostic AI algorithms. There are evident advantages to the amount and scope of information obtained from EHR data, including the possibility of regular, long-term data collecting and continuous patient outcome monitoring. The information obtained from these prediction models is often applied in oncology to group patients along a risk spectrum, allowing for the identification of "high-risk" individuals who are eligible for more intensive treatments. Setting end-of-life care priorities for patients with advanced malignancies is a relevant clinical use of prognostic AI models. For example, in a context where the outcome has a 4% prevalence, EHR-based machine learning algorithms that calculate the 180-day probability of mortality have shown great accuracy (range from 0.95 to 0.96). Compared to traditional prognostic models and decision-making frameworks that base their recommendations on previous randomized controlled trials, this method provides a data-driven alternative (20). Furthermore, a prospective trial that successfully integrated the algorithm with a behavioral health intervention to support serious illness conversations (SICs) for high-risk patients with advanced cancer demonstrated the algorithm's strong applicability to real-world situations. By using the mortality prediction algorithm, the number of patient events with documented critical illness conversations increased by 11%, and end-of-life expenses decreased by an average of $75.33 per day (21, 22). Beyond trial settings, the algorithm's beneficial effects on end-of-life care and serious disease conversation rates have remained consistent, indicating how well it integrates into a big healthcare system.

This algorithm serves as a case study demonstrating how prognostic AI models can improve risk assessments and guide clinical judgment in the oncology. Interestingly, the system was only used and trained at one institution, making it difficult to extrapolate its performance measures to other healthcare environments. Important discussions about algorithmic fairness, reliability, and the long-term safety of using prognostic AI models in oncology practice are caused by the variability in model efficacy across various testing environments, including differences in patient demographics, geographic locations, and temporal factors.

These algorithms must promote equitable allocation of healthcare interventions and reflect algorithmic fairness in order to guarantee equitable performance across different patient demographics, including aspects like gender and ethnicity. Risk prediction models, however, may serve to strengthen preexisting social biases, which would reduce their predictive power for underserved patient groups. The lack of training data representative of medically underserved populations and the improper use of proxy variables that do not fully describe the underlying mechanisms of risk are the main causes of these biases. A notable 2019 study found that a widely used risk prediction algorithm designed to predict healthcare costs, which are influenced by both medical conditions and social determinants of health, performed less well when applied to Black patients with comparable health profiles to White patients. On the other hand, the percentage of Black patients indicated for further healthcare treatments increased by an astounding 28.8% when the incidence of chronic illnesses was predicted (23).

Careful consideration of model inputs is necessary to address the problem of data misrepresentation. Although incorporating socially defined characteristics like race and ethnicity into risk prediction models can have clinical value, it is still a controversial topic in society (24). Incorporating race and ethnicity factors improved algorithmic fairness across various performance criteria by lowering racial prejudice, according to a pertinent study examining four risk models for predicting postoperative cancer recurrence in patients with colorectal cancer (25).

It is important to understand that, despite the potential benefits of include some socially defined variables, relying too much on them may mask underlying and complex drivers that are not sufficiently addressed, such as socioeconomic status and disease-related biomarkers. Disparities in access to high-quality cancer care could be exacerbated if race is included as a variable because it runs the danger of solidifying discriminatory assumptions about protected patient categories. After deployment, performance drift, the gradual deterioration of model performance. presents a serious threat to the accuracy of risk projections. Since many of the models used in oncology are usually deterministic, any changes made to the data generation procedure without a corresponding algorithm update may produce predictions that are inconsistent. Drift in clinical risk models is often caused by two main factors: changes in healthcare practice patterns and modifications to electronic health record (EHR) software and documentation methods.

During the COVID-19 pandemic, the true positive rate dropped by a significant 7%, according to a new study on drift in the six-month mortality prediction model (26). This decline was linked to lower laboratory use that was noted during quarantine. Implementing continuous monitoring and performing sporadic model updates are crucial to addressing the effects of drift on clinical decision-making and resource allocation.

Additionally, the efficacy of prognostic AI tools in real-world scenarios is strongly influenced by the caliber of data collected during the training and pre-deployment stages of AI models. Issues like incompleteness, lack of structure, human recording errors, and inconsistency across various healthcare systems are common problems with several types of EHR data. These worries are mirrored in other patient data sources that are being used more and more for risk prediction, like mobile data and patient-reported outcomes (PROs), which present the further difficulty of imprecise and noisy longitudinal measures. It is essential to use thorough data pretreatment procedures, error correction strategies, and standardization processes in order to enhance the performance of AI models. Furthermore, creating data-sharing guidelines would improve the precision and dependability of AI systems in medical settings.

AI-Driven Chatbots

In this review, we note the significant improvements in the healthcare industry being made by contemporary conversational chatbots. These chatbots are complex computer programs that use cutting-edge algorithms to generate discourse that sounds human. With the development of chatbot technology, generative large language models (LLMs) have replaced predictive natural language processing and speech recognition technologies. These models can effectively generate, forecast, and reproduce content by evaluating large text-based datasets. LLM chatbots give patients a useful tool for improving patient education, promoting patient-clinician communication, and providing mental health assistance. On the other side, doctors can use LLMs to help recruit patients for clinical trials, simplify medical documentation, including informed consent forms, and enhance telemedicine encounters (27, 28). Recent studies on LLMs' use in cancer care management shows that the technology still has issues with the accuracy and quality of the data it provides. A retrospective cross-sectional study evaluated whether the National Comprehensive Cancer Network's (NCCN) guidelines for treating lung, breast, and prostate cancers were followed by OpenAI's commercial LLM, ChatGPT. The results showed that almost one-third of the treatment recommendations made by the chatbot did not fully comply with NCCN criteria (29). Additionally, the suggestions differed according to the questions asked, and discrepancies between the chatbot's answers and accepted criteria were often ascribed to ambiguous or imprecise responses, suggesting that care should be used when using LLM chatbots to provide medical advice. In a parallel examination, researchers compared the quality of information generated by ChatGPT v3.5 with responses from Perplexity, Chatsonic, and AI in response to the most popular Google search questions pertaining to five common malignancies (lung, skin, colorectal, breast, and prostate) (30). With a median DISCERN score of 5, the responses were deemed to be of acceptable quality overall, but they lacked direct applicability due to their complexity and readability level, which was appropriate for college students.

This critical assessment highlights the present restrictions and issues with the application of LLM chatbots in the medical field, especially with regard to the precision of data that is essential for patient care and cancer therapy. Medical chatbots and generative AI technologies have a number of drawbacks, most notably in terms of accuracy, readability, and reliability. Additionally, commercial chatbots like ChatGPT are taught using a variety of text material from the internet, with insufficient quality checks to ensure the accuracy of the data. Large language models (LLMs) also function as "black boxes," which presents significant difficulties because of their low explainability, which makes it hard to understand how these algorithms make their predictions. LLMs are unable to identify the precise training data or sources that were used to generate their answers. This restriction may lead to the dissemination of false information, creating misunderstandings and undermining user confidence, especially that of patients and healthcare professionals. The problem of AI hallucinations, which further erodes confidence in these technologies by producing false information based on faulty premises, is another worry (31). The capacity of chatbots to effectively and concisely convey complicated medical concepts is crucial to their future acceptance in medical settings. This is an area where current algorithms frequently fall short, as their readability scores vary based on how the user phrased requests. These algorithms are not regularly updated to reflect the most recent findings as medical knowledge grows. As a result, chatbots that aren't trained on current data may eventually lose their reliability. Given the speed at which generative AI technology is developing, oncology may see a rise in its use. To put patient safety and privacy first, medical chatbots require stronger regulatory frameworks. Regulating the clinical use of chatbots is more practical, but monitoring the private use of these tools by patients looking for medical information on their own is more difficult. Clinicians in oncology who are aware of these problems can be quite helpful in helping their patients navigate the complex web of medical information produced by chatbots (32).

DISCUSSION

Alan Turing's important question in1950, "Can machines think?" laid the groundwork for the field of artificial intelligence (AI), which led to the creation of early robotic systems that attempted to mimic human decision-making abilities, such as the Tentacle Arm and other industrial assembly robots (33). Over time, the use of AI spread into crucial fields like oncology, which revolutionized cancer care by improving clinical decision-making processes. Today, AI is an essential tool in early diagnosis, risk assessment, and treatment planning, allowing medical professionals to spend more time interacting with their patients. However, there are a number of difficulties in using AI into oncology. Data bias, algorithm accessibility, and effective clinical adoption continue to be major concerns. Training datasets are vital for assessing the efficacy of AI models, and biases included in these datasets may result in incorrect generalizations. Furthermore, reliability of AI systems may be compromised by dataset shift, which occurs when real-world data differs from the training data. The adoption of AI in healthcare is also heavily influenced by human factors; many doctors are hesitant to accept AI results because they are worried about their explainability, possible liability, and the financial consequences of any mistakes. Evaluating interactive AI frameworks, such as human-in-the-loop models, is essential to promoting increased confidence and enhancing accuracy in AI applications (34, 35, 36, 37). Predictions are more reliable thanks to these algorithms' capacity to include real-time physician feedback (38). Furthermore, broadening AI's scope to encompass patient-reported outcomes (PROs), which record symptoms and functional status outside of clinical settings, presents significant possibilities to improve model accuracy. Combining PRO data with electronic health records (EHRs) has been shown to significantly improve prediction performance by 4%. Addressing the ethical and laws that come with the growing use of AI technology is necessary (39). Responsible AI deployment will need the adoption of strong protections, including as data privacy procedures, ongoing algorithm monitoring, and the standardization of clinical operations. Finding a balance between innovation and accountability is essential to maximizing AI's promise to transform cancer treatment while putting patient safety first and preserving public confidence in these novel

REFERENCES

1. Clancy E. ACS Report Shows Prostate Cancer on the Rise, Cervical Cancer on the Decline. Renal & Urology News. 2023:NA-NA.
2. Luchini C, Pea A, Scarpa A. Artificial intelligence in oncology: current applications and future perspectives. British Journal of Cancer. 2022;126(1):4-9.
3. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electronic markets. 2021;31(3):685-95.
4. Elemento O, Leslie C, Lundin J, Tourassi G. Artificial intelligence in cancer research, diagnosis and therapy. Nature Reviews Cancer. 2021;21(12):747-52
5. Fateeva A, Eddy K, Chen S. Current state of melanoma therapy and next steps: battling therapeutic resistance. Cancers. 2024;16(8):1571.
6. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. nature. 2017;542(7639):115-8.
7. Menzies SW, Sinz C, Menzies M, Lo SN, Yolland W, Lingohr J, et al. Comparison of humans versus mobile phone-powered

artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial. The Lancet Digital Health. 2023;5(10):e679-e91.
8. Salim M, Wåhlin E, Dembrower K, Azavedo E, Foukakis T, Liu Y, et al. External evaluation of 3 commercial artificial intelligence algorithms for independent assessment of screening mammograms. JAMA oncology. 2020;6(10):1581-8.
9. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA: a cancer journal for clinicians. 2022;72(1):7-33.
10. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DS, Kerlikowske K, et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. Radiology. 2017;283(1):49-58.
11. Schaffter T, Buist DS, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. JAMA network open. 2020;3(3):e200265-e.
12. Sharma N, Ng AY, James JJ, Khara G, Ambrózay É, Austin CC, et al. Multi-vendor evaluation of artificial intelligence as an independent reader for double reading in breast cancer screening on 275,900 mammograms. BMC cancer. 2023;23(1):460.
13. Lång K, Josefsson V, Larsson A-M, Larsson S, Högberg C, Sartor H, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. The Lancet Oncology. 2023;24(8):936-44.
14. Guo LN, Lee MS, Kassamali B, Mita C, Nambudiri VE. Bias in, bias out: underreporting and underrepresentation of diverse skin types in machine learning research for skin cancer detection—a scoping review. Journal of the American Academy of Dermatology. 2022;87(1):157-9.
15. Adamson AS, Smith A. Machine learning and health care disparities in dermatology. JAMA dermatology. 2018;154(11):1247-8.
16. Wen D, Khan SM, Xu AJ, Ibrahim H, Smith L, Caballero J, et al. Characteristics of publicly available skin cancer image datasets: a systematic review. The Lancet Digital Health. 2022;4(1):e64-e74.
17. DeGrave AJ, Janizek JD, Lee S-I. AI for radiographic COVID-19 detection selects shortcuts over signal. Nature Machine Intelligence. 2021;3(7):610-9.
18. Christakis NA, Smith JL, Parkes CM, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort studyCommentary: Why do doctors overestimate? Commentary: Prognoses should be based on proved indices not intuition. Bmj. 2000;320(7233):469-73.
19. Glare P, Virik K, Jones M, Hudson M, Eychmuller S, Simes J, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. Bmj. 2003;327(7408):195.
20. Parikh RB, Manz C, Chivers C, Regli SH, Braun J, Draugelis ME, et al. Machine learning approaches to predict 6-month mortality among patients with cancer. JAMA network open. 2019;2(10):e1915997-e.
21. Manz CR, Parikh RB, Small DS, Evans CN, Chivers C, Regli SH, et al. Effect of integrating machine learning mortality estimates with behavioral nudges to clinicians on serious illness conversations among patients with cancer: a stepped-wedge cluster randomized clinical trial. JAMA oncology. 2020;6(12):e204759-e.
22. Parikh RB, Chen J, Heintz J, LaPergola M, Bilker W, Patel MS, et al. End-of-life spending analysis of randomized trial of machine learning nudges to prompt serious

illness communication among patients with cancer. American Society of Clinical Oncology; 2023.
23. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53.
24. Manski CF, Mullahy J, Venkataramani AS. Using measures of race to make clinical predictions: Decision making, patient health, and fairness. Proceedings of the National Academy of Sciences. 2023;120(35):e2303370120.
25. Griesler PC, Hu M-C, Wall MM, Kandel DB. Assessment of prescription opioid medical use and misuse among parents and their adolescent offspring in the US. JAMA network open. 2021;4(1):e2031073-e.
26. Parikh RB, Zhang Y, Kolla L, Chivers C, Courtright KR, Zhu J, et al. Performance drift in a mortality prediction algorithm among patients with cancer during the SARS-CoV-2 pandemic. Journal of the American Medical Informatics Association. 2023;30(2):348-54.
27. Decker H, Trang K, Ramirez J, Colley A, Pierce L, Coleman M, et al. Large language model− based chatbot vs surgeon-generated informed consent documentation for common procedures. JAMA network open. 2023;6(10):e2336997-e.
28. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA internal medicine. 2023;183(6):589-96.
29. Chen S, Kann BH, Foote MB, Aerts HJ, Savova GK, Mak RH, et al. Use of artificial intelligence chatbots for cancer treatment information. JAMA oncology. 2023;9(10):1459-62.
30. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence chatbot responses to top searched queries about cancer. JAMA oncology. 2023;9(10):1437-40.
31. Azamfirei R, Kudchadkar SR, Fackler J. Large language models and the perils of their hallucinations. Critical Care. 2023;27(1):120.
32. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ digital medicine. 2023;6(1):120.
33. Turing AM. Computing machinery and intelligence: Springer; 2009.
34. Dlamini Z, Skepu A, Kim N, Mkhabele M, Khanyile R, Molefi T, et al. AI and precision oncology in clinical cancer genomics: From prevention to targeted cancer therapies-an outcomes based patient care. Informatics in Medicine Unlocked. 2022;31:100965.
35. Sebastian AM, Peter D. Artificial intelligence in cancer research: trends, challenges and future directions. Life. 2022;12(12):1991.
36. Istasy P, Lee WS, Iansavichene A, Upshur R, Gyawali B, Burkell J, et al. The impact of artificial intelligence on health equity in oncology: scoping review. Journal of medical Internet research. 2022;24(11):e39748.
37. Chow R, Midroni J, Kaur J, Boldt G, Liu G, Eng L, et al. Use of artificial intelligence for cancer clinical trial enrollment: a systematic review and meta-analysis. JNCI: Journal of the National Cancer Institute. 2023;115(4):365-74.
38. Evans KK, Birdwell RL, Wolfe JM. If you don’t find it often, you often don’t find it: why some cancers are missed in breast cancer screening. PloS one. 2013;8(5):e64366.
39. Parikh RB, Hasler JS, Zhang Y, Liu M, Chivers C, Ferrell W, et al. Development of machine learning algorithms incorporating electronic health record data, patient-reported outcomes, or both to predict mortality for outpatients with cancer. JCO Clinical Cancer Informatics. 2022;6:e2200073.