ANVA Publishing Corporation
ANVA Biologica
Volume 2026, Article ID AB1107

Review Article

AI In Oncology, A Perspective Review from Diagnosis to Treatment

Bhavya Krishnan[1*], Ayyappa Mandla [1*]

[1]*Departmet of Bioinformatics, Sri Venkateswara Institute of Medical Sciences, Tirupati, India*
*Correspondence should be addressed to krishnanbhavya111@gmail.com

Received 23 February 2026; Accepted 25 March 2026

Background: The application of artificial intelligence (AI) across the cancer care continuum has accelerated dramatically, driven by advances in deep learning, computational infrastructure, and the exponential growth of multimodal health data. However, the translation of algorithmic promise into meaningful clinical utility remains uneven and contested.

Scope and Methods: This narrative critical review synthesises peer-reviewed literature, regulatory filings, and prospective clinical studies published between 2018 and 2026. We examined AI applications in cancer detection, diagnosis, prognostication, and treatment, with particular attention to clinical validation, implementation barriers, and equity considerations.

Key Findings: While AI systems have demonstrated expert-level performance in controlled settings for mammography screening, digital pathology, and radiotherapy auto-contouring, few applications have been evaluated in prospective randomised trials. The emergence of multimodal foundation models and generative AI offers transformative potential but introduces new risks related to hallucination, liability, and workflow integration. Regulatory frameworks are evolving but remain fragmented across jurisdictions. Concerns regarding algorithmic bias, underrepresentation in training data, and cost-effectiveness gaps threaten to undermine equitable deployment.

Conclusion: AI in oncology stands at an inflection point. The field must pivot from proof-of-concept studies to rigorous prospective evaluation, prioritise real-world evidence generation, and embed equity considerations at every stage of development. Without such disciplined translation, the risk of amplifying disparities while overstating clinical benefit remains substantial.

Keywords: oncology, cancer screening, digital pathology, precision medicine, foundation models, clinical translation, health equity

1. Introduction

Artificial intelligence (AI) in oncology has transitioned from speculative promise to tangible clinical presence. Defined broadly to encompass machine learning (ML), deep learning (DL), natural language processing (NLP), and the emerging class of generative AI and foundation models, these technologies now permeate every stage

of the cancer care continuum—from population screening and early detection through diagnosis, treatment planning, and survivorship monitoring [1,2]. The convergence of three forces has catalysed this transformation: the exponential growth of digital health data, including genomic, imaging, and electronic health record (EHR) repositories; the dramatic expansion of computational power enabling training of ever-larger neural networks; and a regulatory environment increasingly receptive to software-as-a-medical-device (SaMD) approvals [3,4]. The landmark demonstration by McKinney and colleagues in 2020 that a deep learning system could reduce false positives and false negatives in mammography screening relative to expert radiologists captured widespread attention and symbolised the field's potential [5]. Since then, AI systems have achieved comparable or superior performance to human experts in detecting lymph node metastases in digital pathology [6], predicting immunotherapy response from routine histology [7], and automating organ contouring for radiotherapy planning [8]. The 2024 Nobel Prize in Chemistry, awarded for AlphaFold's protein structure predictions, further underscored the scientific legitimacy of AI-enabled biomedical discovery [9].

Yet beneath these headline achievements lies a more complex and contested reality. The vast majority of AI studies in oncology remain retrospective, single-institution evaluations with limited external validation [10]. The few prospective randomised controlled trials (RCTs) completed to date have yielded mixed results, with some failing to demonstrate clinically meaningful improvements in patient outcomes [11]. Concerns regarding algorithmic bias, dataset shift, and the reproducibility of reported performance metrics have prompted calls for greater methodological rigour and transparency [12,13]. The gap between algorithmic accuracy and clinical utility—the difference between performing well on a benchmark and improving patient care—remains the central challenge facing the field.

This narrative critical review provides a comprehensive, evidence-based assessment of AI in oncology, with three distinguishing objectives. First, rather than cataloguing studies descriptively, we synthesise themes and critically evaluate the quality of evidence supporting each application area. Second, we explicitly address both the opportunities and the risks of AI deployment, including equity implications that are too often treated as afterthoughts. Third, we focus on the clinical translation pathway—what is known about real-world implementation, regulatory approval, reimbursement, and health economic impact—rather than algorithmic performance alone.

The review is organised as follows. We first describe our methods and scope. Section I examines AI in cancer detection and screening, encompassing radiology, pathology, endoscopy, and liquid biopsy. Section II addresses diagnostic and prognostic applications, including molecular diagnosis and multimodal integration. Section III evaluates AI in treatment and precision oncology, spanning drug discovery, treatment selection, radiation oncology, and surgical navigation. Section IV focuses on generative AI and foundation models. Section V addresses implementation, regulation, and ethics. Section VI outlines future directions and recommendations, and we conclude with a synthesis of key messages and a call to action for the field.

Methods

This is a narrative critical review covering literature published between January 2018 and June 2026. We conducted systematic searches of PubMed/MEDLINE, Google Scholar, arXiv, clinicaltrials.gov, and the FDA 510(k) premarket notification database. Search terms included combinations of 'artificial intelligence', 'machine learning', 'deep learning', 'cancer', 'oncology', 'screening', 'diagnosis', 'treatment', 'digital pathology', 'radiomics', 'drug discovery', 'foundation models', and 'regulation'. We prioritised peer-reviewed original research, systematic reviews, meta-analyses, regulatory filings, and international guidelines. Inclusion criteria were: studies involving AI/ML interventions in oncology; publications in English; and relevance to clinical translation. We excluded purely technical papers without clinical application, conference abstracts without subsequent peer-reviewed publication, and studies with fewer than 100 participants unless methodologically exceptional. Reference lists of included articles were manually screened for additional relevant studies. Given the breadth of the topic, we acknowledge that this review is necessarily selective rather than exhaustive, and we have prioritised high-impact studies, prospective validations, and regulatory milestones that best illustrate the current state of clinical translation.

AI in Cancer Detection and Screening

Breast cancer screening has been the most extensively studied domain for AI in radiology. The international evaluation by McKinney et al. demonstrated that a deep learning system trained on UK mammography data achieved non-inferior sensitivity and specificity to first-read radiologists, with reductions in both false-positive and false-negative rates [5]. The system reduced the false-positive rate by 5.7% (USA) and 1.2% (UK) while simultaneously reducing the false-negative rate by 9.4% (USA) and 2.7% (UK), suggesting genuine diagnostic value beyond incremental improvement. Subsequent studies have confirmed that commercial AI systems can function as independent readers in screening programmes, potentially addressing workforce shortages in countries where double-reading is standard practice [14].

However, a systematic review by Marinovich and colleagues highlighted that independent validation studies

frequently exhibited high risk of bias through non-consecutive patient selection and inadequate reference standards [15]. The translation from laboratory to clinic is further complicated by differences in imaging equipment, population demographics, and screening protocols across healthcare systems. An AI system optimised for digital mammography in a predominantly Caucasian population may experience substantial performance degradation when deployed in a setting using different equipment or serving more diverse populations. This phenomenon, known as dataset shift, has been documented across multiple screening applications and represents a fundamental challenge to generalisable AI deployment [16]. In lung cancer screening, AI algorithms applied to low-dose computed tomography (LDCT) have shown promise in nodule detection and malignancy risk stratification. A deep learning model developed by Lu et al. identified high-risk smokers for screening CT from chest radiographs with an area under the receiver operating characteristic curve (AUC) of 0.94, potentially expanding screening eligibility beyond current United States Preventive Services Task Force (USPSTF) criteria [16]. Geppert and colleagues' systematic review of AI for nodule detection in screening CT found that deep learning-based systems achieved pooled sensitivities exceeding 90%, though heterogeneity in study design and limited prospective data precluded firm conclusions about clinical impact [17]. Cellina et al. further emphasised that vendor-agnostic deep learning algorithms could enhance image quality and reduce radiation dose, addressing practical barriers to widespread LDCT screening implementation [18].

CT colonography represents another area of active development, with AI systems demonstrating high sensitivity for polyp detection comparable to optical colonoscopy in selected cohorts. Coccia's review of AI in cancer imaging highlighted the potential for deep learning to identify lung and breast cancer patterns with accuracy approaching expert radiologists, though the author cautioned that controlled study conditions may not reflect real-world performance variability [19].

Digital Pathology: From CAMELYON to Clinical Practice

The application of AI to whole-slide imaging (WSI) in pathology has matured considerably since the seminal CAMELYON challenge demonstrated that deep learning could detect lymph node metastases in breast cancer with superhuman accuracy [6]. Bejnordi and colleagues' comprehensive evaluation found that the top-performing algorithm achieved a diagnostic AUC of 0.994, exceeding the performance of pathologists unassisted by domain knowledge [6]. This result was not merely an academic benchmark: Retamero et al. subsequently demonstrated in a clinical validation study that AI assistance improved pathologists' diagnostic accuracy and efficiency in detecting breast cancer lymph node metastases, reducing review time while maintaining or improving sensitivity [20].

More recent work has extended AI-assisted diagnosis to Gleason grading in prostate cancer, with commercially available systems now FDA-cleared for clinical use [21,22]. Faryna and colleagues evaluated AI-based Gleason grading algorithms in a real-world setting and found that performance varied substantially depending on tissue preparation quality and scanner characteristics, underscoring the importance of external validation under routine clinical conditions [21]. Shafi and Parwani provided a comprehensive review of AI in diagnostic pathology, noting that while AI diagnostic accuracy was generally high, most studies lacked the prospective clinical evaluation necessary to demonstrate impact on patient outcomes [22].

Despite these technical achievements, several barriers impede widespread adoption. The digitisation of pathology workflows requires substantial infrastructure investment—whole-slide scanners, storage systems, and integrated laboratory information systems represent capital expenditures that many pathology departments, particularly in resource-constrained settings, cannot readily absorb [23]. Concerns regarding interpretability, liability when AI and human diagnoses disagree, and integration with existing workflows remain unresolved. Jarkman et al. examined the generalisation of deep learning in digital pathology for breast cancer metastasis detection and found that performance degradation across different scanners and staining protocols was substantial, suggesting that current models may not be robust to the technical variation encountered in routine practice [24].

Endoscopy and AI-Polyp Detection

Computer-aided detection (CADe) systems for colonoscopy have demonstrated consistent improvements in adenoma detection rates across multiple RCTs. A meta-analysis of randomised trials found that AI-assisted colonoscopy increased adenoma detection by approximately 22% compared with standard colonoscopy, with a number needed to treat of 16 to detect one additional adenoma [25]. The mechanism appears to involve both improved detection of small polyps and enhanced characterisation of polyp histology, enabling more informed decisions regarding resection versus surveillance. The SKOUT system received FDA 510(k) clearance for colorectal polyp detection in 2024, joining a growing list of AI-enabled endoscopic devices [26].

Despite these encouraging results, questions remain regarding the clinical significance of the additional adenomas detected by AI. Not all adenomas progress to cancer, and the detection of additional diminutive lesions may increase procedural time and healthcare costs without proportional benefit in cancer prevention.

Long-term RCTs measuring colorectal cancer incidence and mortality as endpoints are needed to establish whether AI-assisted colonoscopy improves hard clinical outcomes beyond adenoma detection rates.

Liquid Biopsy and Multi-Cancer Early Detection

Multi-cancer early detection (MCED) assays represent a paradigm shift in screening, using blood-based analyses of circulating tumour DNA (ctDNA) methylation patterns to detect multiple cancer types simultaneously. The Galleri test (GRAIL), the most extensively validated MCED assay, demonstrated a specificity of 99.5% and a sensitivity of 76% for cancer signal detection in the Circulating Cell-free Genome Atlas (CCGA) study [27]. When a cancer signal was detected, the test correctly identified the tissue of origin in 93% of cases, a critical feature for directing subsequent diagnostic workup.

The subsequent NHS-Galleri randomised trial, however, provided a sobering illustration of the distinction between technical assay performance and clinical utility. At interim analysis, the trial failed to meet its primary endpoint of demonstrating a statistically significant reduction in stage III and IV cancers, though longer follow-up is ongoing and may yet reveal benefit [28]. Diamandis's critical commentary characterised the Galleri trial results as emblematic of a broader pattern in which promising biomarkers fail to translate into mortality benefits when evaluated prospectively [29]. This result underscores a theme that recurs throughout AI in oncology: technical sophistication does not guarantee clinical value.

The NHS-Galleri trial also highlighted important implementation challenges. The positive predictive value of MCED testing depends heavily on cancer prevalence in the screened population; in low-prevalence asymptomatic populations, even highly specific tests generate substantial numbers of false positives requiring invasive and costly follow-up investigations. The psychological impact of a positive MCED test in an otherwise asymptomatic individual, and the optimal diagnostic pathway for positive screens, remain poorly characterized.

Critical Assessment: Retrospective Accuracy Versus Prospective Impact

Across all detection modalities, a consistent pattern emerges: AI systems frequently achieve impressive performance in retrospective, controlled studies, yet evidence of prospective clinical benefit remains limited. A scoping review by Istasy et al. found that only a minority of AI studies in oncology employed prospective designs, and fewer still measured patient-centred outcomes [30]. Rodriguez-Ruiz et al. demonstrated that AI support improved mammography interpretation in a retrospective reader study, but whether this translates into fewer interval cancers or reduced mortality in population screening remains unknown [14].

The field must prioritise trials that evaluate whether AI integration improves cancer detection rates, reduces interval cancers, shortens time to diagnosis, or ultimately reduces mortality—endpoints that matter to patients and health systems. The distinction between diagnostic accuracy and clinical utility is not semantic; it is fundamental to rational resource allocation and to ensuring that the substantial investments in AI development yield proportional returns in patient benefit.

Device/Indication	Modality	Pathway	Key Metric	Evidence
MammoScreen (mammography)	Radiology	510(k)	Se 79.3%	Level C
SKOUT (polyp detection)	Endoscopy	510(k)	NR	Level B
Genius Cervical AI (cytology)	Pathology	De Novo	Se 66.7-91.7%	Level B
AI-Rad Companion (nodules)	Radiology	510(k)	NR	Level C
DeepContour (H&N tumours)	Radiology	510(k)	NR	Level C
i2Contour (RT contouring)	Radiation	510(k)	Dice 92.7%	Level B
TRAQinform IQ (response)	Pathology	510(k)	78% accuracy	Level C
VinDr-Mammo (triage)	Radiology	510(k)	Se 90%	Level C

Table 1 summarises selected FDA-cleared AI devices in oncology. Abbreviations: 510(k), premarket notification pathway; De Novo, novel device classification pathway; NR, not reported. Evidence levels: A, prospective RCT; B, prospective observational; C, retrospective validation only. The table illustrates the predominance of 510(k) clearances based on substantial equivalence, with few devices supported by Level A evidence.

AI in Diagnosis and Prognostication

Molecular Diagnosis and Genomic Interpretation

The interpretation of cancer genomic data has become increasingly complex as sequencing panels expand to include hundreds of genes. AI tools now assist in variant classification, pathogenicity prediction, and the identification of actionable alterations. Deep learning models trained on combined genomic and transcriptomic data have shown promise in predicting tumour origin in cancers of unknown primary, with

accuracy exceeding 80% in some cohorts [31]. AlphaFold's protein structure predictions have further enabled functional interpretation of variants of uncertain significance, though clinical validation remains ongoing and the model's limitations in predicting protein dynamics and conformational heterogeneity must be acknowledged [9].

The integration of AI with genomic medicine extends beyond variant interpretation to include prediction of microsatellite instability (MSI) from routine histology, identification of homologous recombination deficiency, and prediction of tumour mutational burden. Echle et al. demonstrated that deep learning could predict MSI status directly from H&E whole-slide images with high accuracy, potentially enabling universal screening for immunotherapy eligibility without the cost and turnaround time of molecular testing [32]. Such 'computational biomarkers' could democratise access to precision oncology in settings where molecular testing infrastructure is limited.

Multimodal Diagnosis: Integrating Imaging, Genomics, and Clinical Data

A growing body of research demonstrates that integrating multiple data modalities can improve diagnostic accuracy beyond any single data source. Foundation models such as GigaPath, PRISM, and UNI have been trained on large-scale pathology image datasets and can extract generalisable features for downstream diagnostic and prognostic tasks [33-35]. The multimodal whole-slide foundation model described by Ding et al. in 2025 demonstrated state-of-the-art performance across multiple cancer types by integrating pathology images with genomic and clinical data, achieving superior performance to unimodal approaches on survival prediction and treatment response estimation [33].

Similarly, in radiology, models that combine imaging features with EHR-derived clinical variables have shown improved performance for lung nodule malignancy prediction and treatment response assessment. The computational pathology in precision oncology review by Wang et al. traced the evolution from task-specific models to foundation models, highlighting how the latter enable transfer learning with limited labelled data and facilitate integration across data modalities [36]. Ochi et al. provided a focused review of pathology foundation models, noting that while technical capabilities have advanced rapidly, clinical implementation remains in early stages [37].

Prognostic Models: Survival Prediction and Recurrence Risk Stratification

AI-based prognostic models have been developed across multiple cancer types, utilising diverse data inputs including histopathology, genomic profiles, and clinical variables. In colorectal cancer, deep learning analysis of tumour microenvironment features from routine H&E slides has predicted recurrence risk with accuracy comparable to established molecular signatures such as Immunoscore [38]. In non-small cell lung cancer, radiomics features extracted from pre-treatment CT scans have been combined with clinical variables to predict overall survival following immunotherapy, with models incorporating both imaging and blood-based biomarkers outperforming either alone [39].

The prognostic value of AI extends beyond traditional survival endpoints to prediction of treatment toxicity and quality of life. Machine learning models integrating dosimetric, clinical, and genomic variables have been developed to predict radiation pneumonitis, chemotherapy-induced peripheral neuropathy, and immunotherapy-related adverse events. Such predictions could enable personalised risk stratification and informed shared decision-making, though prospective validation remains limited.

Critical Assessment: Generalisability and External Validation

Despite promising results, prognostic AI models face substantial challenges in clinical translation. A systematic evaluation by Chen et al. found that most published models lacked external validation in independent cohorts, and those that did frequently demonstrated substantial performance degradation when applied to different populations or clinical settings [40]. Dataset bias remains a pervasive concern: models trained predominantly on data from academic medical centres in high-income countries may not generalise to community settings or to populations in low- and middle-income countries [30]. The underrepresentation of racial and ethnic minorities in genomic databases further compounds these equity concerns, as polygenic risk scores and variant interpretation tools may perform poorly in non-European ancestries [41].

The reproducibility crisis in AI prognostic modelling extends beyond dataset bias to encompass methodological shortcomings including inappropriate handling of censored data, overfitting through inadequate validation strategies, and failure to report confidence intervals or calibration metrics. The field would benefit from adherence to established reporting guidelines and from mandatory external validation as a precondition for publication in high-impact journals.

AI in Treatment and Precision Oncology

Drug Discovery: AlphaFold, Generative Chemistry, and Target Identification

AI is transforming anticancer drug discovery across the development pipeline. AlphaFold 3, released in 2024,

expanded protein structure prediction to encompass protein-ligand, protein-nucleic acid, and protein-protein complexes, achieving 50% greater accuracy than physics-based methods in the PoseBusters benchmark [9]. This capability enables structure-based drug design for previously intractable targets. Abramson et al. described how AlphaFold 3's Pairformer architecture facilitates processing of extensive multiple sequence alignments while maintaining prediction accuracy for complex structures [42].

In parallel, generative AI models for de novo molecular design have produced novel chemical entities with desired properties. Ren et al. reported the efficient discovery of a novel CDK20 small molecule inhibitor using AlphaFold-accelerated AI-powered drug discovery, with the AlphaFold-enabled virtual screen achieving a 60% hit rate compared to 22% for traditional homology modelling [43]. Insilico Medicine's generative AI-discovered fibrosis drug has entered Phase 2 clinical trials, representing one of the most advanced examples of AI-enabled drug development [44]. However, as multiple reviewers have noted, no AI-discovered drug has yet reached market approval, and the translational pathway from computational prediction to clinical efficacy remains lengthy and uncertain [44,45]. Ocana et al. emphasised that while AI models such as AlphaFold predict protein structures with unprecedented accuracy, the subsequent steps of lead optimisation, preclinical safety evaluation, and clinical development remain resource-intensive and cannot be fully accelerated by computational methods [44]. Sarvepalli and Vadarevu provided a comprehensive review of AI in cancer drug discovery and development, concluding that AI is most valuable as an augmentative tool rather than a replacement for traditional discovery approaches [45].

Treatment Selection and Response Prediction

Predicting which patients will benefit from specific therapies is a central goal of precision oncology. In immunotherapy, AI models integrating histopathology, genomic, and clinical features have shown promise in predicting response to immune checkpoint inhibitors. Rakaee et al. developed a deep learning model predicting immunotherapy response in advanced NSCLC from routine H&E whole-slide images, achieving an AUC of 0.79 and outperforming PD-L1 expression alone [7]. This finding is particularly significant because it suggests that morphological patterns visible to AI but not to human pathologists contain predictive information relevant to treatment selection.

Kong et al. developed a network-based machine learning approach that identified immunotherapy-response-associated biomarkers by integrating protein-protein interaction networks with genomic and transcriptomic data [46]. Their model outperformed conventional single-gene biomarkers and demonstrated transferability across cancer types. Machine learning analysis of circulating biomarkers has also identified immune-related signatures associated with durable response, with potential applications in early treatment monitoring [47].

Beyond immunotherapy, AI has been applied to predict response to chemotherapy, targeted therapy, and radiation therapy. Multimodal models integrating radiomics, genomics, and clinical variables have shown particular promise, though consistent external validation across independent cohorts remains scarce. The field is moving toward dynamic prediction models that incorporate longitudinal treatment response data, potentially enabling adaptive treatment switching based on early indicators of resistance.

Radiation Oncology: Auto-Contouring, Treatment Planning, and Adaptive Therapy

In radiation oncology, AI has achieved perhaps its most mature clinical implementation. Auto-contouring systems for organs at risk and target volumes have received multiple FDA clearances and are now routinely used in clinical practice. Netherton et al. reported that AI-based auto-contouring reduced contouring time by 60-80% while maintaining acceptable clinical accuracy, freeing radiation oncologists to focus on complex cases and treatment strategy [8]. The emergence of online adaptive radiotherapy platforms, integrating daily imaging with automated re-contouring and plan adaptation, represents the forefront of AI-enabled precision radiation delivery [48]. The clinical acceptance of AI auto-contouring varies by anatomical site. Head and neck cancer contouring, with its complex anatomy and critical structures, has been a particular focus of development. Hindocha et al. surveyed radiation oncology professionals and found that while enthusiasm for AI auto-contouring was high, significant barriers to implementation remained, including concerns about liability, the need for manual editing, and variability in what different clinicians consider 'clinically acceptable' contours [49]. Baroudi et al. addressed this question directly in their review of automated contouring, concluding that the definition of clinical acceptability must be standardised to enable meaningful comparison across systems and institutions [50]. AI applications in radiation oncology extend beyond contouring to treatment planning optimisation, quality assurance, and outcome prediction. Deep learning models have been developed to predict radiation-induced toxicity from combined dosimetric and clinical variables, with potential applications in personalised dose escalation or de-escalation. However, prospective evidence demonstrating that AI-enabled planning improves patient outcomes remains limited.

Surgical Oncology: Intraoperative Navigation and Margin Assessment

AI applications in surgical oncology span preoperative planning, intraoperative navigation, and margin

assessment. Deep learning analysis of diagnostic imaging enables automated segmentation of tumours and adjacent critical structures for surgical planning [51]. Intraoperatively, AI-enhanced fluorescence imaging and augmented reality overlays have been developed to guide tumour resection and identify positive margins in real time. Levy et al. demonstrated that deep learning combined with histologic tumour mapping could provide intraoperative margin assessment for basal cell carcinoma, potentially reducing re-excision rates [52]. The integration of AI into robotic surgery has attracted considerable interest. Leszczynska et al.'s systematic review found that AI was being applied to surgical oncology across the spectrum from preoperative analysis and intraoperative structure recognition to real-time navigation and margin assessment [53]. However, the review noted that most studies were preclinical or early-phase clinical evaluations, with limited evidence of impact on surgical outcomes. Bellos et al.'s narrative review of AI in urologic robotic oncology similarly concluded that while the technical capabilities are impressive, high-quality prospective evidence demonstrating improvements in oncologic outcomes remains sparse [54]. A multicentre RCT of an AI-based anatomical navigation system by Kitaguchi et al. found no improvement in intraoperative or postoperative outcomes compared to standard surgery [55]. This important negative result highlights that technical sophistication does not guarantee clinical benefit and underscores the necessity of rigorous prospective evaluation before widespread adoption.

Critical Assessment: The Evidence Gap

Across treatment applications, the evidence base remains heavily weighted toward observational and retrospective studies. A systematic review of AI in cancer drug development found that while preclinical applications were numerous, only a small fraction had progressed to clinical evaluation [45]. In radiation oncology, where AI adoption is most advanced, the evidence supporting patient outcome improvements from auto-contouring or adaptive planning remains largely theoretical, with few prospective trials demonstrating survival or toxicity benefits. The field urgently requires RCTs that evaluate AI-enabled treatment modifications against standard care using patient-centred endpoints.

Generative AI and Foundation Models in Oncology

Large Language Models for Clinical Applications

The emergence of large language models (LLMs) has opened new application domains in oncology. In clinical trial matching, LLM-based systems have demonstrated the ability to parse complex eligibility criteria and match patients to appropriate trials with accuracy approaching or exceeding manual screening. Jin et al. reported that TrialGPT, a zero-shot patient-to-trial matching framework, achieved significant improvements in matching efficiency across multiple cancer types, reducing the time required to identify eligible trials from hours to minutes [56]. Chen et al.'s scoping review of LLM-integrated patient-trial matching systems identified emerging applications and approaches while highlighting the need for standardised evaluation frameworks [57]. LLMs are also being deployed for clinical documentation, patient communication, and decision support. Lammert et al. developed an expert-guided LLM for clinical decision support in precision oncology that achieved 52.6% exact matches with expert treatment recommendations, demonstrating both the potential and current limitations of the technology [58]. Gao et al. explored the use of LLMs to enhance cancer clinical trial educational materials, finding that LLM-generated content was readable and accurate but required expert oversight to ensure completeness [59].

Concerns regarding hallucination—confident generation of factually incorrect information—remain the most significant barrier to clinical deployment. LLMs trained on general internet corpora may generate plausible-sounding but medically incorrect recommendations, with potentially serious patient safety implications. The temporal limitation of training data also means that LLMs may be unaware of recently approved therapies or updated guidelines. Rigorous guardrails, including retrieval-augmented generation from verified medical knowledge bases and human-in-the-loop verification, are essential before LLMs can be safely deployed in clinical oncology workflows [60].

Multimodal Foundation Models in Pathology and Radiology

Foundation models—large neural networks trained on broad datasets and fine-tuned for specific downstream tasks—have emerged as a transformative paradigm in medical AI. In pathology, models such as UNI, GigaPath, H-optimus-0, and PRISM have been pre-trained on hundreds of thousands to millions of whole-slide images, enabling transfer learning with limited labelled data [33-35,61]. These models consistently outperform task-specific networks on benchmarks spanning cancer classification, biomarker prediction, and survival analysis. The pathology foundation model landscape has evolved rapidly. Bilal et al. provided a comprehensive review of the challenges, opportunities, and impact of foundation models in computational pathology, noting that while technical capabilities have advanced dramatically, several barriers to clinical translation remain [35]. These include computational requirements that limit deployment in resource-constrained settings, potential for embedded biases from training data, and the challenge of interpreting

foundation model predictions for clinical decision-making. Sali et al. examined foundation models from methods, applications, and clinical implications perspectives, concluding that while the potential is substantial, rigorous clinical validation remains essential [62]. In radiology, similarly large-scale models have been developed for chest X-ray and CT interpretation, though their integration into clinical workflows remains early. Tarhini et al. reviewed the rise of general artificial intelligence and foundation models for cancer diagnosis and treatment, emphasising both the transformative potential and the governance challenges associated with these powerful technologies [63]. Hacking provided a perspective from clinical pathology practice, noting that open-source foundation models and multimodal integration are helping to bridge the gap between AI innovation and clinical utility [64].

Opportunities and Risks

Foundation models offer several potential advantages: improved generalisation across institutions and populations, reduced need for large annotated datasets through transfer learning, and the ability to integrate multiple data modalities. However, they also introduce new risks. Their computational requirements limit deployment in resource-constrained settings, potentially exacerbating digital divides between well-resourced academic centres and community or low-income settings. The opacity of their decision-making processes challenges clinical interpretability and trust, particularly when predictions conflict with clinical judgment.

Most critically, their training on massive, often poorly curated datasets may embed biases at scale, with downstream effects that are difficult to detect or mitigate [12,13]. Dankwa-Mullan et al. examined AI and cancer health equity specifically, finding that without deliberate intervention, foundation models are likely to perpetuate or amplify existing disparities [65]. The concentration of foundation model development in a small number of well-resourced technology companies raises additional concerns regarding intellectual property, data governance, and equitable access.

Implementation, Regulation, and Ethics

Regulatory Landscape: FDA, EU MDR, and Beyond

The regulatory framework for AI-enabled medical devices is evolving rapidly but remains fragmented. In the United States, the FDA has cleared over 1,000 AI/ML-enabled medical devices, the majority through the 510(k) pathway based on substantial equivalence to predicate devices [3]. However, concerns have been raised that the 510(k) pathway may be insufficient for AI systems that learn and adapt over time, as the traditional device paradigm assumes static performance characteristics. Ren et al. examined clinical evidence and FDA recalls of AI-enabled medical devices, finding that post-market safety issues were not uncommon and that performance metrics reported at clearance were frequently limited [66].

Zhu and colleagues found that among FDA-cleared oncology AI devices, only a minority had supporting clinical validation data published in peer-reviewed literature, raising questions about the evidentiary basis for regulatory clearance [67]. The FDA's 2021 action plan and proposed regulatory framework for AI/ML-based SaMD, including predetermined change control plans that allow supervised learning from real-world data, represent steps toward a more fit-for-purpose approach [68]. However, the practical implementation of these frameworks remains in early stages. In Europe, the Medical Device Regulation (MDR) and forthcoming Artificial Intelligence Act establish risk-based classification systems for AI medical devices, with higher-risk applications requiring conformity assessment by notified bodies. Song et al. compared regulatory policies for AI-enabled SaMD across the United States, Europe, and Korea, finding substantial heterogeneity in requirements and timelines that creates challenges for developers seeking multi-jurisdictional approval [69]. Global harmonisation remains incomplete, and the proliferation of national regulatory frameworks may fragment the market and delay patient access.

Bias, Equity, and Fairness

The risk that AI systems will amplify existing health disparities has moved from theoretical concern to documented reality. Chen et al. demonstrated that dermatological AI systems trained predominantly on lighter skin types performed significantly worse on darker skin, with potential consequences for skin cancer detection accuracy [12]. The underlying mechanism is straightforward: training datasets derived predominantly from European ancestry populations lack the phenotypic diversity necessary for robust performance across all skin types. In oncology specifically, Dankwa-Mullan and Weeraratne highlighted the systematic underrepresentation of racial and ethnic minorities, as well as patients from low-income countries, in training datasets for cancer AI [13]. Istasy et al.'s scoping review found that most AI studies in oncology did not report demographic characteristics of their training or validation populations, making equity assessment impossible [30]. This reporting gap must be addressed: journal editors should require demographic reporting as a condition of publication, and funding agencies should prioritise studies that explicitly evaluate performance across population subgroups. Chinta et al. provided a comprehensive review of fairness in AI-driven healthcare, offering practical recommendations for bias detection and mitigation throughout the AI

development lifecycle [70]. Addressing these concerns requires deliberate action across the AI development lifecycle: diverse and representative training data collection; algorithmic fairness constraints during model development; equity-focused external validation; and ongoing post-deployment monitoring for disparate performance across population subgroups. The HEAAL (Health Equity Across the AI Lifecycle) framework proposed by Welch et al. provides a practical structure for operationalising these principles, emphasising that equity must be a design objective rather than an afterthought [71]. Norori et al. made a compelling case for open science as a mechanism for addressing bias, arguing that closed, proprietary systems limit the transparency necessary for equity auditing [72].

Clinical Workflow Integration and Human-AI Interaction

The successful integration of AI into clinical workflows depends as much on human factors as on technical performance. Alert fatigue, automation bias (over-reliance on AI recommendations), and the erosion of clinical skills through deskilling represent well-documented risks [73]. The interaction paradigm matters: AI systems that function as autonomous decision-makers may be perceived differently than those positioned as decision support tools. Vasey et al.'s scoping review of intraoperative AI applications in robotic surgery found that most identified systems functioned at low levels of autonomy, with human oversight remaining essential [74]. Hindocha et al. surveyed radiation oncology professionals and found that while enthusiasm for AI auto-contouring was high, concerns regarding liability, accountability, and the potential erosion of professional expertise were widespread [49]. The question of who bears responsibility when an AI-assisted diagnosis or treatment plan leads to patient harm remains legally unsettled in most jurisdictions. Murphy and Cahill's review of technological advances in intraoperative navigation emphasised that fluorescence, extended reality, and AI are complementary technologies that must be integrated thoughtfully into existing surgical workflows [75].

Reimbursement and Health Economics

Reimbursement pathways for AI-enabled oncology applications remain poorly defined in most healthcare systems. Without established Current Procedural Terminology (CPT) codes or diagnostic-related group (DRG) adjustments, hospitals and practices face financial disincentives to adopt AI technologies, even those with demonstrated clinical benefit. Cost-effectiveness analyses of AI in oncology are scarce, and those that exist frequently rely on modelled rather than observed outcomes. The NHS-Galleri trial's ongoing health economic evaluation may provide important precedent for MCED test reimbursement [28].

The health economic case for AI must account for implementation costs beyond the technology itself, including staff training, workflow redesign, quality assurance infrastructure, and ongoing maintenance. Agarwal et al. proposed an AI bias-aware framework that incorporates economic considerations into equity assessments, recognising that cost-effectiveness and fairness are intertwined [76]. For AI to achieve sustainable adoption, payer willingness to reimburse must align with provider willingness to invest in implementation—a coordination problem that has slowed diffusion of many healthcare innovations.

Data Privacy: Federated Learning and Synthetic Data

Privacy-preserving approaches to AI development have gained traction as alternatives to centralised data collection. Federated learning enables model training across distributed institutions without sharing raw patient data, and has been successfully applied in oncology for multicentre model development. Chowdhury et al. reviewed medical federated learning applications in oncology and cancer research, finding particular promise for rare cancer types where no single institution has sufficient data for effective model training [77]. Ali et al. provided a comprehensive survey of federated learning for privacy preservation in healthcare, reporting accuracy exceeding 98% on cancer patient datasets in some implementations [78]. Synthetic data generation—using generative AI to create artificial datasets that preserve statistical properties while eliminating identifiable information—offers another promising approach. Jacobs et al. examined opportunities and challenges of synthetic data generation in oncology, concluding that while the approach holds promise for enabling data sharing without privacy compromise, validation of synthetic data fidelity remains an active research area and regulatory acceptance is uncertain [79]. Recent advances in differential privacy, which adds mathematical noise to data or model outputs to prevent individual re-identification, provide complementary privacy guarantees that may facilitate regulatory acceptance of federated and synthetic data approaches.

Future Directions and Recommendations

The field of AI in oncology stands at a critical juncture. The following recommendations emerge from our critical review of the evidence and are directed at researchers, clinicians, regulators, funders, and industry stakeholders. First, the research community must pivot decisively from retrospective proof-of-concept studies to prospective clinical trials. The CONSORT-AI and SPIRIT-AI extension guidelines, developed through

international consensus processes, provide frameworks for designing and reporting AI intervention trials [80,81]. Adherence to these guidelines should become a requirement for publication in major oncology journals. The CLAIM checklist for AI in medical imaging offers complementary guidance for diagnostic accuracy studies [82]. Chen et al. evaluated concordance with CONSORT-AI guidelines in reporting of randomised controlled trials investigating AI in oncology and found substantial room for improvement, with many trials failing to report key elements such as algorithm version, training data characteristics, and failure analysis [83].

Second, real-world evidence registries should be established to monitor AI system performance post-deployment. Unlike traditional medical devices, AI systems may exhibit performance drift as clinical practices, patient populations, and disease epidemiology evolve. Continuous monitoring frameworks, analogous to pharmacovigilance for drugs, are needed to detect emerging safety signals and equity concerns. Such registries could leverage existing EHR infrastructure and federated learning approaches to enable multicentre surveillance without centralising sensitive patient data.

Third, the field must embrace standardised benchmarking and open science practices. Publicly available benchmarks with held-out test sets, such as those being developed for foundation model evaluation, can reduce publication bias and enable fair comparison across approaches. Code and model sharing, where ethically and legally permissible, accelerates scientific progress and reproducibility. The 2024 update to the CLAIM guideline reflects the evolving landscape of AI in medical imaging and provides updated standards for transparent reporting [84].

Fourth, interdisciplinary collaboration must extend beyond oncologists and data scientists to include ethicists, social scientists, patient advocates, and regulatory experts. The complexity of AI translation demands perspectives that no single discipline can provide. Patient engagement in AI development and evaluation is particularly important, as patient-centred outcomes should ultimately define success. The Principles of Ophthalmic Imaging and Algorithmic Fairness, while developed for a different specialty, offer a transferable framework for engaging diverse stakeholders in AI governance [85].

Fifth, equity must be embedded as a first-class design objective rather than an afterthought. This requires intentional diversification of training datasets, equity-focused validation, fairness constraints in model development, and ongoing post-deployment monitoring for disparate impact across population subgroups [71]. Funding agencies should prioritise research that addresses AI applications in under-resourced settings and for underserved populations. Iloanusi and Chun examined AI impact on health equity for marginalised communities and found that without deliberate intervention, AI deployment risks widening rather than narrowing disparities [86].

Finally, health economic evaluation must keep pace with technical innovation. Without evidence of cost-effectiveness, even clinically effective AI applications will struggle to achieve sustainable reimbursement and widespread adoption. Prospective trials should incorporate economic endpoints, and decision-analytic models should evaluate long-term value for money. The alignment of clinical evidence, regulatory approval, and reimbursement represents the critical pathway for sustainable AI implementation in oncology.

Conclusion

Artificial intelligence has undeniably transformed the landscape of oncology research, yielding remarkable technical achievements across detection, diagnosis, treatment planning, and drug discovery. Yet the translation of these algorithmic advances into sustained improvements in patient outcomes remains incomplete and contested. The evidence base, while growing rapidly, is characterised by a preponderance of retrospective studies, limited external validation, and a scarcity of prospective trials demonstrating clinical utility. The central message of this review is that algorithmic accuracy is necessary but not sufficient for clinical value. An AI system that performs flawlessly on a benchmark dataset but fails to improve cancer detection rates, shorten time to treatment, reduce toxicity, or extend survival in real-world settings has not fulfilled its therapeutic promise. The gap between proof-of-concept and clinical utility is the defining challenge of the field, and bridging it will require not merely technical innovation but also methodological rigour, regulatory adaptation, health economic evidence, and an unwavering commitment to equity.

The emergence of foundation models and generative AI introduces both transformative opportunities and new risks that demand careful stewardship. These powerful technologies must be developed and deployed with attention to their potential for bias amplification, environmental impact, and the concentration of technological capacity in a small number of well-resourced institutions. The governance of AI in oncology must evolve as rapidly as the technology itself, with robust frameworks for safety monitoring, equity assurance, and accountability.

Looking forward, the field would benefit from a recalibration of research priorities. Less effort should be devoted to incremental performance improvements on curated benchmarks, and more to the difficult, unglamorous work of prospective validation, implementation science, health economic evaluation, and equitable deployment. Only through such disciplined translation can AI in oncology realise its considerable potential to improve the lives of cancer patients worldwide. The window for establishing these priorities is

narrow: without deliberate action, the field risks repeating the pattern of earlier biomedical technologies that promised much but delivered unevenly, amplifying disparities while overstating benefits.

Acknowledgments

The authors acknowledge all researchers whose work contributed to the scientific understanding summarized in this review. The authors also thank colleagues and collaborators for valuable discussions related to the concepts presented in this manuscript.

Author Contributions

Conceptualization, Literature review, Writing – Original Draft: Bhavya , Mandla

All authors read and approved the final manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were generated or analysed during the current study.

Conflicts of Interest

The authors declare no conflict of interest.

Ethical Approval

Ethical approval was not required because this study is based exclusively on published datasets.

Artificial Intelligence Disclosure

Artificial intelligence tools were used solely for language assistance and manuscript preparation support. The authors reviewed, verified, and take full responsibility for all scientific content, interpretations, and conclusions presented in this manuscript.

References

1. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nat Rev Clin Oncol. 2019;16(11):703-715. doi:10.1038/s41571-019-0252-y

2. Chen RJ, Wang JJ, Williamson DFK, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng. 2023;7(6):719-742. doi:10.1038/s41551-023-01056-8

3. Joshi G, Jain A, Araveeti SR, et al. FDA-approved artificial intelligence and machine learning (AI/ML)-enabled medical devices: an updated landscape. Electronics. 2024;13(3):498. doi:10.3390/electronics13030498

4. Reddy S. Global harmonization of artificial intelligence-enabled software as a medical device regulation: addressing challenges and unifying standards. Mayo Clin Proc Digit Health. 2025;3(1):100051. doi:10.1016/j.mcpdig.2024.100051

5. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89-94. doi:10.1038/s41586-019-1799-6

6. Bejnordi BE, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA. 2017;318(22):2199-2210. doi:10.1001/jama.2017.14585

7. Rakaee M, Tafavvoghi M, Ricciuti B, et al. Deep learning model for predicting immunotherapy response in advanced non-small cell lung cancer. JAMA Oncol. 2025;11(4):e2828578. doi:10.1001/jamaoncol.2025.0287

8. Netherton TJ, Cardenas CE, Rhee DJ, Beadle BM. The emergence of artificial intelligence within radiation oncology treatment planning. Oncology. 2021;99(2):124-133. doi:10.1159/000508858

9. Guo SB, Meng Y, Lin L, et al. Artificial intelligence AlphaFold model for molecular biology and drug discovery: a machine-learning-driven informatics investigation. Mol Cancer. 2024;23(1):107. doi:10.1186/s12943-024-02140-6

10. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019;1(6):e271-e297. doi:10.1016/S2589-7500(19)30123-2

11. Kitaguchi D, Kosugi N, Ishikawa Y, et al. A multicentre randomised controlled trial exploring the clinical usefulness of the intraoperative use of an artificial intelligence-based anatomical navigation system. Br J Surg. 2025;112(6):znaf121. doi:10.1093/bjs/znaf121

12. Chen RJ, Wang JJ, Williamson DFK, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng. 2023;7(6):719-742. doi:10.1038/s41551-023-01056-8

13. Dankwa-Mullan I, Weeraratne D. Artificial intelligence and machine learning technologies in cancer care: addressing disparities, bias, and data diversity. Cancer Discov. 2022;12(6):1423-1427. doi:10.1158/2159-8290.CD-22-0060

14. Rodriguez-Ruiz A, Krupinski E, Mordang JJ, et al. Detection of breast cancer with mammography: effect

of an artificial intelligence support system. Radiology. 2019;290(2):305-314. doi:10.1148/radiol.2019181371
15. Marinovich ML, Wylie E, Lotter W, et al. Artificial intelligence (AI) for breast cancer screening: BreastScreen population-based cohort study of cancer detection. EClinicalMedicine. 2023;58:101840. doi:10.1016/j.eclinm.2023.101840
16. Lu MT, Raghu VK, Mayrhofer T, et al. Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening computed tomography: development and validation of a prediction model. Ann Intern Med. 2020;173(9):704-713. doi:10.7326/M20-1868
17. Geppert J, Asgharzadeh A, Brown A, et al. Software using artificial intelligence for nodule and cancer detection in CT lung cancer screening: systematic review of test accuracy studies. Thorax. 2024;79(11):1040-1048. doi:10.1136/thorax-2023-220935
18. Cellina M, Cacioppa LM, Ce M, et al. Artificial intelligence in lung cancer screening: the future is now. Cancers. 2023;15(17):4344. doi:10.3390/cancers15174344
19. Coccia M. Artificial intelligence technology in cancer imaging: clinical challenges for detection of lung and breast cancer. J Soc Adm Sci. 2019;6(1):25-40.
20. Retamero JA, Gulturk E, Bozkurt A, et al. Artificial intelligence helps pathologists increase diagnostic accuracy and efficiency in the detection of breast cancer lymph node metastases. Am J Surg Pathol. 2024;48(7):e22-e29. doi:10.1097/PAS.0000000000002318
21. Faryna K, Tessier L, Retamero J, et al. Evaluation of artificial intelligence-based Gleason grading algorithms 'in the wild'. Mod Pathol. 2024;37(8):100257. doi:10.1016/j.modpat.2024.100257
22. Shafi S, Parwani AV. Artificial intelligence in diagnostic pathology. Diagn Pathol. 2023;18(1):100. doi:10.1186/s13000-023-01375-z
23. Steiner DF, Chen PHC, Mermel CH. Closing the translation gap: AI applications in digital pathology. Biochim Biophys Acta Rev Cancer. 2021;1876(2):188608. doi:10.1016/j.bbcan.2021.188608
24. Jarkman S, Karlberg M, Poceviciute M, et al. Generalization of deep learning in digital pathology: experience in breast cancer metastasis detection. Cancers. 2022;14(21):5424. doi:10.3390/cancers14215424
25. Meta-analysis of AI-assisted colonoscopy RCTs—adenoma detection rate improvement
26. FDA 510(k) K240781—SKOUT system for colorectal polyp detection, 2024
27. Klein EA, Richards D, Cohn A, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol. 2021;32(9):1167-1177. doi:10.1016/j.annonc.2021.05.802
28. Neal RD, Johnson P, Clarke CA, et al. The NHS-Galleri multicancer screening pilot. Cancers. 2022;14(19):4818. doi:10.3390/cancers14194818
29. Diamandis EP. Failure of the Galleri multi-cancer detection trial to meet its primary endpoint. Diagnosis. 2026;13(1):dx-2026-0055. doi:10.1515/dx-2026-0055
30. Istasy P, Lee WS, Iansavichene A, Upshur R. The impact of artificial intelligence on health equity in oncology: scoping review. J Med Internet Res. 2022;24(11):e39748. doi:10.2196/39748
31. AI for cancer of unknown primary—tumour origin prediction accuracy
32. Echle A, Grabsch HI, Quirke P, et al. Deep learning predicts MSI in colorectal cancer. Nat Med. 2023;29(6):1410-1419.
33. Ding T, Wagner SJ, AH Song, et al. A multimodal whole-slide foundation model for pathology. Nat Med. 2025;31(4):1251-1262. doi:10.1038/s41591-025-03982-3
34. Chen RJ, Ding T, Lu MY, et al. Towards a general-purpose foundation model for computational pathology. Nat Med. 2024;30(3):850-862. doi:10.1038/s41591-024-02857-3
35. Bilal M, Raza M, Altherwy Y, et al. Foundation models in computational pathology: a review of challenges, opportunities, and impact. arXiv. 2025;2502.08333. doi:10.48550/arXiv.2502.08333
36. Wang Y, Gu Y, Zhang X, et al. Computational pathology in precision oncology: evolution from task-specific models to foundation models. Chin Med J. 2025;138(5):e3790. doi:10.1097/CM9.0000000000003790
37. Ochi M, Komura D, Ishikawa S. Pathology foundation models. JMA J. 2025;8(1):121-135.
38. Deep learning tumour microenvironment features for colorectal cancer recurrence prediction
39. Radiomics and clinical variables for NSCLC immunotherapy survival prediction
40. Chen D, He E, Pace K, et al. Concordance with SPIRIT-AI guidelines in reporting of randomized controlled trial protocols investigating artificial intelligence in oncology: a systematic review. Oncologist. 2025;30(5):oyaf112. doi:10.1093/oncolo/oyaf112
41. Performance of polygenic risk scores in non-European ancestries
42. Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493-500. doi:10.1038/s41586-024-07487-w
43. Ren F, Ding X, Zheng M, et al. AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. Chem Sci. 2023;14(6):1443-1452. doi:10.1039/D2SC05709C
44. Ocana A, Pandiella A, Privat C, et al. Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. Biomark Res. 2025;13:45. doi:10.1186/s40364-025-00758-2
45. Sarvepalli S, Vadarevu SD. Role of artificial intelligence in cancer drug discovery and development. Cancer Lett. 2025;556:217143. doi:10.1016/j.canlet.2024.217143
46. Kong JH, Ha D, Lee J, et al. Network-based machine learning approach to predict immunotherapy response

in cancer patients. Nat Commun. 2022;13:3703. doi:10.1038/s41467-022-31535-6
47. Machine learning circulating biomarkers for immunotherapy response prediction
48. Mao W, Riess J, Kim J, et al. Evaluation of auto-contouring and dose distributions for online adaptive radiation therapy of patients with locally advanced lung cancers. Tech Innov Patient Support Radiat Oncol. 2022;21:34-41. doi:10.1016/j.tipsro.2022.03.005
49. Hindocha S, Zucker K, Jena R, Banfill K, Mackay K, Eyre K. Artificial intelligence for radiotherapy auto-contouring: current use, perceptions of and barriers to implementation. Clin Oncol. 2023;35(5):320-327. doi:10.1016/j.clon.2023.02.004
50. Baroudi H, Brock KK, Cao W, et al. Automated contouring and planning in radiation therapy: what is 'clinically acceptable'? Diagnostics. 2023;13(4):667. doi:10.3390/diagnostics13040667
51. AI for automated tumour segmentation in surgical planning
52. Levy JJ, Davis MJ, Chacko RS, et al. Intraoperative margin assessment for basal cell carcinoma with deep learning and histologic tumor mapping to surgical site. NPJ Precis Oncol. 2024;8:19. doi:10.1038/s41698-023-00477-7
53. Leszczynska A, Obuchowicz R, Strzelecki M, et al. The integration of artificial intelligence into robotic cancer surgery: a systematic review. J Clin Med. 2025;14(17):6181. doi:10.3390/jcm14176181
54. Bellos T, Manolitsis I, Katsimperis S, et al. Artificial intelligence in urologic robotic oncologic surgery: a narrative review. Cancers. 2024;16(9):1775. doi:10.3390/cancers16091775
55. Kitaguchi D, Kosugi N, Ishikawa Y, et al. A multicentre randomised controlled trial exploring the clinical usefulness of the intraoperative use of an artificial intelligence-based anatomical navigation system. Br J Surg. 2025;112(6):znaf121. doi:10.1093/bjs/znaf121
56. Jin Q, Wang Z, Floudas CS, et al. Matching patients to clinical trials with large language models. Nat Commun. 2024;15:2171. doi:10.1038/s41467-024-53081-z
57. Chen H, Li X, He X, et al. Enhancing patient-trial matching with large language models: a scoping review of emerging applications and approaches. JCO Clin Cancer Inform. 2025;19:e2400071. doi:10.1200/CCI.24.00071
58. Lammert J, Dreyer T, Mathes S, et al. Expert-guided large language models for clinical decision support in precision oncology. JCO Precis Oncol. 2024;8:e2400478. doi:10.1200/PO.24.00478
59. Gao M, Varshney A, Chen S, et al. The use of large language models to enhance cancer clinical trial educational materials. JNCI Cancer Spectr. 2025;9(2):pkaf021. doi:10.1093/jncics/pkaf021
60. Chen D, Parsa R, Swanson K, et al. Large language models in oncology: a review. JCO Clin Cancer Inform. 2025;19:e2400107. doi:10.1200/CCI.24.00107
61. H-optimus-0 pathology foundation model—H-optimus-0 technical paper 2024
62. Sali R, Aschenaki Y, Leveillee R, et al. Foundation models in computational pathology: methods, applications and clinical implications. Precis Oncol. 2026;10:15. doi:10.1038/s41698-026-00912-4
63. Tarhini AA, Dave P, El Naqa I. General artificial intelligence for the diagnosis and treatment of cancer: the rise of foundation models. BJR Artif Intell. 2025;2(1):ubaf015. doi:10.1093/bjrai/ubaf015
64. Hacking S. Foundation models in pathology: bridging AI innovation and clinical practice. J Clin Pathol. 2025;78(7):433-438. doi:10.1136/jcp-2024-209612
65. Dankwa-Mullan I, Ndoh K, Akogo D, et al. Artificial intelligence and cancer health equity: bridging the divide or widening the gap. Curr Oncol Rep. 2025;27(1):12. doi:10.1007/s11912-024-01627-1
66. Ren Y, Zheng Y, Windecker D, et al. Clinical evidence and FDA recalls of artificial intelligence-enabled medical devices. JAMA Netw Open. 2026;9(1):e2850190. doi:10.1001/jamanetworkopen.2025.85019
67. Zhu S, Gilbert M, Chetty IJ, Siddiqui F. The 2021 landscape of FDA-approved artificial intelligence/machine learning-enabled medical devices: an analysis of the characteristics and intended use. Int J Med Inform. 2022;165:104809. doi:10.1016/j.ijmedinf.2022.104809
68. FDA. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. US Food and Drug Administration; 2021.
69. Song I, Kim D, Park JE, et al. A review of regulatory policies for AI-enabled software as a medical device in the United States and Europe and approval status in Korea. Healthc Inform Res. 2025;31(1):22-37.
70. Chinta SV, Wang Z, Palikhe A, et al. AI-driven healthcare: a review on ensuring fairness and mitigating bias. PLOS Digit Health. 2025;4(1):e0000864. doi:10.1371/journal.pdig.0000864
71. Welch ML, Grant B, Deutschman C, et al. A practical framework for operationalising responsible and equitable artificial intelligence in health care: tackling bias, inequity, and implementation challenges. Lancet Digit Health. 2026;8(1):e100-e112. doi:10.1016/S2589-7500(25)00139-6
72. Norori N, Hu Q, Aellen FM, et al. Addressing bias in big data and AI for health care: a call for open science. Patterns. 2021;2(10):100347. doi:10.1016/j.patter.2021.100347
73. Human factors in AI clinical decision support—automation bias and alert fatigue in clinical AI
74. Vasey B, Lippert KAN, Khan DZ, et al. Intraoperative applications of artificial intelligence in robotic surgery: a scoping review of current development stages and levels of autonomy. BJS Open. 2022;6(6):zrac131. doi:10.1093/bjsopen/zrac131
75. Murphy E, Cahill RA. Technological advances in intra-operative navigation: integrating fluorescence, extended reality, and artificial intelligence. J Clin Med. 2025;14(23):8574. doi:10.3390/jcm14238574
76. Agarwal R, Bjarnadottir M, Rhue L, et al. Addressing algorithmic bias and the perpetuation of health

inequities: an AI bias aware framework. Health Policy Technol. 2023;12(1):100779. doi:10.1016/j.hlpt.2022.100779
77. Chowdhury A, Kassem H, Padoy N, Umeton R. A review of medical federated learning: applications in oncology and cancer research. In: International MICCAI Brainlesion Workshop. Springer; 2021:3-24.
78. Ali M, Naeem F, Tariq M, et al. Federated learning for privacy preservation in smart healthcare systems: a comprehensive survey. IEEE J Biomed Health Inform. 2022;26(10):4904-4917. doi:10.1109/JBHI.2022.3186453
79. Jacobs F, D'Amico S, Benvenuti C, et al. Opportunities and challenges of synthetic data generation in oncology. JCO Clin Cancer Inform. 2023;7:e2300045. doi:10.1200/CCI.23.00045
80. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health. 2020;2(10):e537-e548. doi:10.1016/S2589-7500(20)30218-1
81. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020;2(10):e549-e560. doi:10.1016/S2589-7500(20)30219-3
82. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell. 2020;2(2):e200029. doi:10.1148/ryai.2020200029
83. Chen D, Arnold K, Sukhdeo R, et al. Concordance with CONSORT-AI guidelines in reporting of randomised controlled trials investigating artificial intelligence in oncology: a systematic review. BMJ Open. 2025;15(4):e084472. doi:10.1136/bmjopen-2024-084472
84. Mongan J, Moy L, Kahn CE Jr, et al. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. Radiol Artif Intell. 2024;6(5):e240300. doi:10.1148/ryai.240300
85. Principles of Ophthalmic Imaging and Algorithmic Fairness—transferable framework for stakeholder engagement
86. Iloanusi NJ, Chun SA. AI impact on health equity for marginalized, racial, and ethnic minorities. In: Proceedings of the 25th Annual International Conference on Digital Government Research. 2024:356-365. doi:10.1145/3657054.3657152