IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Review On Artificial Intelligence (AI) Use In Cancer Diagnosis

1ARJUN YADAV, 2SAFIYA BEE, 3DR.VIKASH JAIN

1STUDENT , 2ASSITANT PROFESSOR, 3PRINCIPAL

1MAHAKAL INSTITUDE OF PHARMACETICAL STUDIES, UJJAIN,MADHYA PRADESH,

2MAHAKAL INSTITUDE OF PHARMACETICAL STUDIES, UJJAIN,MADHYA PRADESH,

3MAHAKAL INSTITUDE OF PHARMACETICAL STUDIES, UJJAIN,MADHYA PRADESH

Abstract

Artificial intelligence (AI) is increasingly recognized for its potential to enhance cancer care, particularly amid challenges such as an ageing population, workforce shortages, and rapid technological advancements. While AI has shown promise in cancer diagnostics, its broader impact on the quality, efficiency, and equity of care throughout the cancer care continuum remains uncertain. This systematic review aims to evaluate the clinical readiness and practical deployability of AI applications in post-diagnostic cancer care through an analysis of prospective studies. A comprehensive search of PubMed and Web of Science (1 January 2013 to 1 May 2023) identified 15 studies that prospectively assessed AI interventions following cancer diagnosis. Each study was critically appraised using the Risk of Bias tools for randomized and non-randomized studies, along with implementation analyses focusing on time, cost, and resource requirements. Our findings indicate that most AI research in oncology remains in the experimental phase, lacking prospective clinical validation or real-world deployment. Many studies failed to demonstrate clinical validity or translate AI-generated efficacy into meaningful patient outcomes. Key limitations include insufficient standardization, poor interoperability across health systems, and a lack of implementation and equity-focused evaluations. To address the existing gaps—namely, low clinical evidence, limited outcome translation, and research ecosystem fragmentation—future efforts should prioritize multicenter, collaborative implementation studies that align with current methodological standards and are tailored to specific healthcare settings.

Key word - Artificial Intelligence (AI),Cancer Care,Post-Diagnostic Pathway,Clinical Readiness Implementation,Science Health System Interoperability,Prospective Studies,Oncology, Quality of Care, AI Deployment Equity in Healthcare,Systematic Review

INTRODUCTION

Cancer care is becoming increasingly complex due to aging populations, growing socioeconomic disparities, and the rapid evolution of novel treatment technologies.[1] [2] These challenges, coupled with persistent healthcare workforce shortages and infrastructure limitations, present significant opportunities for artificial intelligence (AI) to transform cancer care across multiple stages of the patient journey.[3]

To date, the most substantial impact of AI in oncology has been observed in diagnostics—particularly in radiology.[4] For example, AI-powered breast cancer screening systems have demonstrated performance comparable to expert radiologists in real-world mammogram interpretation.[5] Similarly, *Paige Prostate*, an AI software that enhances the accuracy and efficiency of prostate biopsy interpretation, has received FDA approval in the United States.[6] AI has also made significant strides in treatment planning. Notably, OSAIRIS, an open-source AI tool for medical image analysis, was piloted in a UK hospital and shown to markedly reduce radiotherapy planning time.[7]

Beyond diagnostics and planning, AI is being explored in diverse areas such as patient monitoring, precision oncology, behavioral interventions, and treatment response prediction. For instance, machine learning models analyzing cell-free DNA have improved cancer detection rates and enabled more effective disease monitoring, contributing to advances in liquid biopsy.[8]

Despite these advances, integrating AI into routine oncology care remains challenging. Uncertainty persists around its real-world impact on care quality, efficiency, and equity.[9] A major concern is the lack of large-scale, prospective clinical evaluations of AI tools across diverse healthcare settings, which are necessary to establish clinical validity, cost-effectiveness, and system-level benefits. Many AI algorithms suffer from bias, particularly when trained on limited, retrospective, or non-representative datasets.[10–12] IBM Watson Health's AI system for cancer treatment exemplifies these limitations; although it has been deployed in countries like Brazil, China, India, South Korea, and Mexico, a 2019 concordance study revealed inconsistent alignment between its recommendations and clinical standards.[13] [14]

Additionally, the implementation of AI in healthcare is hindered by issues such as data privacy, inconsistent regulatory frameworks, administrative burden, and a lack of standardization in AI research.[15] [16] The World Health Organization (WHO) has also expressed concern that poorly designed AI systems may exacerbate existing healthcare disparities, reinforcing biases in care delivery.[17]

To our knowledge, no comprehensive effort has been made to systematically assess the clinical readiness and deployability of AI in oncology beyond the diagnostic phase. This systematic review seeks to fill that gap by examining prospective evaluations of AI tools used in the post-diagnostic cancer care pathway. Our aim is to map the current research landscape, assess the quality and breadth of available evidence, and identify key barriers and priorities for future implementation-focused research.

METHODS

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A comprehensive literature search was performed to identify relevant studies published between 1 January 2013 and 1 May 2023 across two major electronic databases: PubMed (including Medline) and Embase.

The search strategy incorporated a combination of keywords and Boolean operators to capture studies involving artificial intelligence in oncology. The following terms were used: "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Network" AND "Cancer" OR "Tumour" OR "Malignant".

A detailed account of the full search strategy is provided in Online Supplemental Appendix A.

Inclusion Criteria

Studies were eligible for inclusion if they evaluated validated AI algorithms within the post-diagnostic cancer care pathway, specifically assessing the efficacy, quality, or efficiency of AI in improving patient workflows. Eligible studies focused on adult patients with solid organ malignancies or multiple tumour types, including haematological cancers, and were written in English. Only studies involving human participants were considered. Furthermore, studies had to be prospective in design, encompassing phase I–IV clinical trials, case-control studies, or observational studies that investigated real-world implementation or outcomes of AI technologies.

Exclusion Criteria

Publications were excluded if they employed a retrospective study design or focused solely on haematological malignancies or paediatric cancer populations. The review also excluded non-primary research formats such as review articles, editorials, letters, conference abstracts or proceedings, trial protocols, and preclinical studies. Additionally, studies published in languages other than English were not considered. Studies limited to the development or initial validation of AI models, as well as those that trained or evaluated AI tools using retrospective data from a single institution, were excluded due to limited generalizability and lack of clinical deployment relevance.

Data selection

Titles and abstracts of all retrieved studies were screened for relevance. Full texts were obtained for articles that met the initial inclusion criteria or where eligibility could not be determined from the abstract alone. Initial study selection was performed by SM, with AA independently reviewing and confirming excluded studies to ensure consistency. SM and PYN conducted data extraction from each eligible study, and any discrepancies or uncertainties were resolved through discussion with AA and RS.

Data Extraction

Data were systematically extracted using a predefined template. The following variables were collected from each included study:

- Study location
- Study characteristics (including funding source, clinical setting, study design, and sample size)
- Tumour type or site
- Purpose of the AI intervention and the stage of the cancer care pathway in which it was applied
- Evaluation outcomes
- Implementation analysis, including considerations of time, cost, and resource utilization

Data Analysis

Risk of bias assessments were conducted for all included studies. PYN performed quality appraisals using the appropriate tools: the Risk of Bias 2 (ROB-2) tool for randomized controlled trials and the Risk of Bias in Non-randomised Studies of Interventions (ROBINS-I) for non-randomised interventional studies.

RESULTS

Search strategy

A total of 12,457 publications were initially identified in the PubMed (inclusive of Medline) and Embase database. 32 studies were selected after the abstracts and titles were screened. Of these, 17 were excluded because of

(1) being in the development and validation phase,

(2) not directly investigating an AI tool,

(3) preclinical phase studies or

(4) clinical trial protocols.

Ultimately, 15 studies met all inclusion criteria and were included in the final analysis. These studies are summarized in Supplemental Table S1 (Online Supplemental Appendix B). The full search and selection process is detailed in Figure 1.

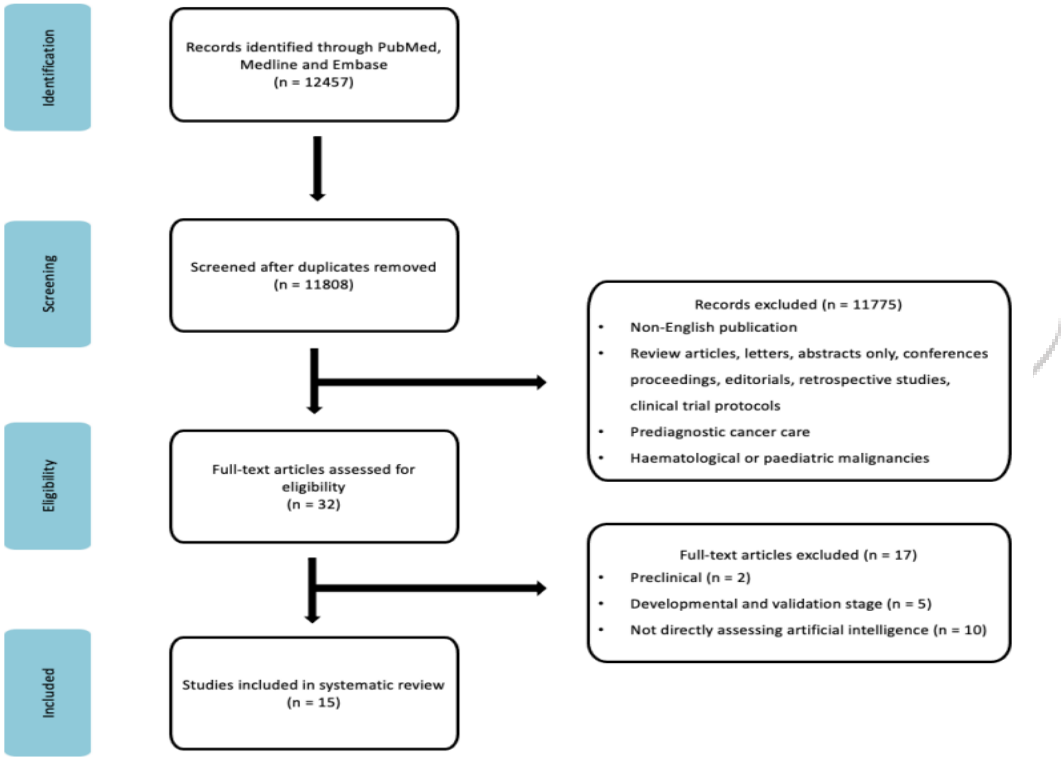

Figure 1 The Preferred Reporting Items for Systematic reviews and Meta-analyses flow chart of identification for articles for inclusion.

TUMOUR TYPE, LOCATION, TYPE OF STUDIES

The selected studies demonstrated a diverse application of artificial intelligence across multiple cancer types. A significant proportion of the research focused on multiple tumour sites, types, breast cancer and gastrointestinal cancers (including colorectal, pancreatic, and hepatobiliary malignancies) were the most commonly investigated. These areas likely reflect both the global disease burden and the availability of structured imaging and pathology data that facilitate AI integration.

Other cancer types studied included prostate cancer, which has seen growing interest due to the complexity of diagnostic interpretation and biopsy planning; gynaecological cancers, where AI tools were applied in areas such as treatment decision-making and monitoring; and thyroid and head and neck cancers, where AI supported imaging analysis and surgical planning.

In terms of geographic distribution, the United States led with the highest number of published studies, reflecting both its advanced digital infrastructure and investment in AI-driven research. Other countries represented included Canada, where national health data integration has enabled multi-institutional studies, and China and South Korea, where rapid AI innovation and government support for digital health have facilitated clinical trials and real-world evaluations.

All included studies were prospective in design, reflecting a shift toward real-world assessment of AI in clinical settings rather than retrospective model development. Approximately half of the studies were designed as randomised controlled trials, underscoring a commitment to rigorous evidence generation. These trials typically compared AI-assisted care with standard clinical practices to determine added value in terms of accuracy, efficiency, or patient outcomes. The remaining studies were observational, involving prospective cohorts or case-control methodologies, and often focused on workflow integration, feasibility, and preliminary outcome assessments.

This distribution of study types highlights the growing momentum to evaluate AI tools not just in development environments, but within active healthcare systems. However, the relatively small number of prospective studies also suggests that widespread clinical validation and implementation of AI in oncology remains in its early stages.

TYPES OF AI INVESTIGATED AND CLINICAL PATHWAYS

Influence on Clinician Behaviour A subset of studies explored the application of AI tools in guiding or altering clinician behaviour in oncology care. Several investigations assessed the impact of machine learning models that predict short-term mortality (e.g., 180-day risk) in cancer patients and examined whether these predictions influence clinical decisions, such as initiating serious illness conversations or advanced care planning. These studies observed a meaningful increase in clinician-initiated end-of-life discussions and a decrease in the use of aggressive systemic therapies near the end of life. Despite promising findings, these studies were limited by the use of a single healthcare electronic record system, reducing generalisability.

One study evaluated an AI-powered Clinical Decision Support System (CDSS) that offered treatment recommendations for breast cancer patients based on real-world clinical data. The results indicated that the CDSS influenced treatment decisions in a minority of cases, particularly for hormone receptor-positive and advanced-stage cancers.

Impact on Patient Behaviour AI technologies have also been examined for their influence on patient behaviour, especially concerning physical activity and health literacy. A short-term pilot study demonstrated that an AI-based voice-coaching system significantly increased daily step counts among overweight and sedentary cancer survivors. However, its findings were constrained by a limited participant pool and brief follow-up period.

Another study explored AI's capacity to estimate physical activity accurately, addressing discrepancies between self-reported activity and objective accelerometer data. Results showed similar baseline measurements between the two methods, validating the machine learning model's accuracy in tracking behavioural change.

Additionally, AI-enhanced educational tools, such as deep neural network-generated 3D thyroid models, were employed to improve patient understanding of thyroid surgery. Patients exposed to these visual aids demonstrated better comprehension of their condition and associated surgical risks, although the study's lack of blinding may have introduced bias.

Survival Prediction Studies investigating survival prediction models used AI to estimate prognosis in patients with advanced cancers. The NEAT (Number of active tumors, ECOG performance status, Albumin level, and Tumour site) model demonstrated superior prognostic accuracy compared to

experienced oncology professionals. However, its external validity remains limited due to being tested in a single institution. Conversely, another study validated a machine learning model across multiple centres, highlighting its potential for scalable implementation in predicting short-term survival.

AI in Treatment Planning AI applications in treatment, particularly radiotherapy and brachytherapy, were extensively explored. Multiple studies demonstrated that AI-assisted systems could achieve non-inferior dosimetric outcomes compared to manual planning while significantly reducing planning time. For instance, AI-supported brachytherapy for prostate cancer and auto-segmentation tools in radiotherapy showed efficient workflow integration and high clinician satisfaction. Notable limitations included a lack of comprehensive imaging data (e.g., PET scans) and modest generalisability.

One prospective Canadian study identified a discrepancy between retrospective evaluations of AI and its real-world clinical adoption, with reduced clinician acceptance during live deployment phases. Another study reinforced the feasibility of integrating deep learning-based auto-segmentation into clinical workflows, showing minimal editing requirements and strong alignment with expert standards.

Surgical Applications and Acute Care Planning Within surgical oncology, a machine learning model improved the accuracy of surgical case time predictions in gynaecological and colorectal surgeries, aiding in resource allocation. However, this model's functionality was limited to short-term forecasts and required precise data input.

Finally, an AI algorithm designed to identify high-risk cancer patients undergoing radiotherapy or chemoradiotherapy demonstrated that more frequent clinical reviews reduced unplanned acute care visits. Despite encouraging outcomes, the findings were restricted to a single-centre implementation, underscoring the need for broader validation.

QUALITY ASSESSMENT OF STUDIES

Randomised Controlled Trials

The methodological quality of the randomised controlled trials (RCTs) included in this review was evaluated using the ROB-2 (Risk of Bias 2) tool. Of the RCTs assessed, four were rated as having a low overall risk of bias across all five domains. These trials demonstrated rigorous randomisation procedures, minimal loss to follow-up, and clear outcome measurement and reporting.

However, several studies exhibited notable limitations. One trial experienced post-randomisation exclusions affecting nearly 10% of the study population, raising concerns about potential attrition bias and imbalance between groups. Another study was assessed as having a high risk of bias due to substantial missing outcome data—over one-third of participants lacked hospice enrolment data, and outcome ascertainment was dependent on incomplete cancer registry sources. A third trial was rated as high risk because it conducted per-protocol analysis instead of intention-to-treat, with a dropout rate exceeding 13%, potentially skewing the results. Furthermore, in some trials, blinding was not implemented, which could have influenced both participant and clinician behaviours and subjective outcome assessment.

These findings highlight the variability in study rigour and the importance of robust trial design when evaluating AI interventions in clinical settings. Consistent application of intention-to-treat principles, full data reporting, and blinding where feasible would strengthen future RCT evidence in this field.

Table 1 Risk of Bias Assessment of Randomised Controlled Trials (ROB-2)

	Domain 1	Domain 2		Domain 3	Domain 4	Domain 5	
Study	Risk of bias arising from the randomisation process	Risk of bias due to deviations from the intended interventions (effect of assignment to intervention)	Risk of bias due to deviations from the intended interventions (effect of adhering to intervention)	Missing outcome data	Risk of bias in measurement of the outcome	Risk of bias in selection of the reported result	Overall risk of bias
Manz *et al*[18]	Low	Low (unblinded participants)	Low	Low	Low	Low	Low
Manz *et al*[19]	Low	Low (unblinded participants)	Low	High risk (missing data for secondary outcome)	Low	Low	High risk (unpredictable)
Hassoon *et al*[20]	Low	Low (unblinded participants)	Low	Low	Low	Low	Low
Strömblad *et al*[21]	Low	Some concerns (Hawthorne's effect and unequal arm size after exclusion postrandomisation)	Low	Low	Low	Low	Some concerns (unpredictable)
Seok *et al*[31]	Low	Low (unblinded participants)	Low	Low	Low	Low	Low
Nicolae *et al*[29]	Low	Low	Low	Low	Low	Low	Low
Nelson *et al*[26]	Low	High risk (per-protocol analysis with 13% dropout rate)	Low	High risk (13% missing outcome data)	Low	Low	High risk (unpredictable)

The risk of bias for each domain of the study is colour-coded. High risk of bias: red; moderate risk of bias: orange and low risk of bias: green.

Observational Studies

The non-randomised interventional and observational studies were assessed using the ROBINS-I (Risk of Bias In Non-randomised Studies of Interventions) tool. Most observational studies were prospective cohort designs, and while they offer valuable real-world insights, they also presented varying levels of risk across different domains.

Several studies were rated as having moderate risk of bias due to confounding variables not being adequately controlled, particularly in studies where AI tools were implemented across diverse clinical settings without adjusting for site-level or clinician-level differences. In addition, participant selection processes in some cases lacked clarity, leading to potential selection bias. Measurement of outcomes was generally sound, but a few studies relied on self-reported data (e.g., physical activity), which may have introduced reporting bias.

Some studies also demonstrated limitations in follow-up completeness and transparency regarding missing data, which could influence the observed effects of the AI intervention. Additionally, in certain cases, the study protocol or registration was not publicly available, making it difficult to assess the risk of selective outcome reporting.

Despite these limitations, several observational studies employed robust data extraction and analysis methods, with efforts to triangulate AI performance with clinical outcomes. Still, improvements in study transparency, control of confounding, and use of standardised outcome metrics are essential for enhancing the validity of future non-randomised evaluations of AI in oncology.

DISCUSSION

The integration of artificial intelligence (AI) into oncology practice has generated considerable enthusiasm, yet its tangible clinical impact remains limited. This systematic review highlights a critical gap between AI development and its real-world application in postdiagnostic cancer care. Despite the rapid advancements in AI algorithms and their potential to enhance clinical decision-making, treatment planning, and patient engagement, only 15 studies met the inclusion criteria for this review, reflecting a sparse evidence base.

The majority of included studies were conducted in single-centre settings with relatively small patient cohorts, often involving ≤50 participants. This limited scale raises concerns regarding statistical power, generalisability, and the ability to detect meaningful clinical differences. Moreover, no studies were

conducted in low- or middle-income countries, underscoring a global disparity in AI research and access. This lack of geographic diversity restricts the applicability of AI tools across varied healthcare systems and populations, particularly where resource constraints might alter implementation feasibility or outcomes.

Most studies focused on early-stage clinical validation or pilot testing of AI tools rather than large-scale deployment or health economic evaluation. Randomised controlled trials and real-world implementation studies were notably scarce, and those that existed frequently suffered from methodological limitations such as high risk of bias, incomplete outcome data, or lack of blinding. Furthermore, the cost-effectiveness of AI tools—an essential consideration for health system adoption—was not addressed in any of the reviewed studies.

Another key challenge identified is the complexity of AI integration into existing clinical workflows. Even when AI tools demonstrated technical efficacy, real-world adoption was often hindered by limited clinician trust, workflow disruption, or lack of interoperability with electronic health record systems. One study revealed that clinicians were less likely to use AI-generated radiotherapy plans during deployment than they were in retrospective evaluations, illustrating the gap between theoretical potential and practical usability.

The scarcity of robust, prospective validation trials and implementation studies indicates that AI in oncology remains at an experimental stage. Until issues of clinical utility, user acceptability, equity, and cost are comprehensively addressed, AI's role in the postdiagnostic cancer pathway will likely remain fragmented and underutilised. Overcoming these barriers requires interdisciplinary collaboration, inclusive research design, transparent reporting, and policy-level support to ensure that AI innovations translate into meaningful improvements in cancer care globally.

Low Level of Evidence

A major limitation across the reviewed studies was the generally low level of evidence, with most investigations situated in the pilot or early clinical deployment phase. These studies were typically conducted under tightly controlled trial conditions, with limited follow-up durations, thereby offering little insight into the long-term clinical utility and system-wide integration of AI tools. The variability in study design and execution further compromised the strength and consistency of the evidence generated.

Half of the included studies were randomised controlled trials (RCTs), while the remaining were observational in nature. Quality assessments using established tools—ROB-2 for RCTs and ROBINS-I for observational studies—revealed significant concerns. Nearly half of the studies (n = 6; 42.9%) exhibited either some concerns or a serious risk of bias. Common issues included unaddressed confounding variables, missing outcome data, selective outcome reporting, and inappropriate analytical methods.

Furthermore, the majority of studies (n = 9; 64.3%) were conducted at single clinical sites, limiting the external validity and generalisability of findings. This narrow scope likely reflects underlying technical and organisational barriers to multicentre research, particularly the lack of standardised, interoperable healthcare data systems. The adoption of unified data exchange formats, such as Fast Healthcare Interoperability Resources (FHIR), along with standardised clinical coding frameworks like the Minimal Common Oncology Data Elements (mCODE), is essential for scaling AI integration across institutions. These tools were exemplified in the multicentre studies led by Manz et al., which were conducted within a unified health system and demonstrated more scalable AI applications.

Sample sizes across studies varied widely; however, several investigations enrolled fewer than 50 participants, which is insufficient to support statistically robust conclusions or clinical deployment decisions. Building AI-ready data infrastructures is crucial to enable more rigorous model training and

validation. Ontology-based frameworks, such as the Operational Ontology for Oncology, can assist in harmonising disparate real-world datasets to support the development and evaluation of advanced AI algorithms.

Encouragingly, there are emerging examples of high-quality, large-scale prospective AI evaluations in oncology. Notably, studies by Dembrower et al. and Lång et al. validated AI-assisted mammography screening in Sweden, representing a gold standard for real-world AI research. These studies illustrate that with appropriate data governance, standardisation, and design, AI tools can be rigorously assessed and successfully deployed in routine clinical workflows.

Gap Between AI Efficacy and Clinical Outcomes

Despite advances in algorithmic accuracy, a significant gap remains between demonstrated AI efficacy in controlled environments and meaningful improvements in patient care outcomes. This disconnect, often referred to as the "AI chasm," underscores the challenge of translating technical performance into tangible clinical benefits.

A key example is the use of serious illness conversations (SICs) as surrogate markers for improved goal-concordant care and reduced end-of-life treatment intensity. In initial studies, AI-generated predictions increased SICs, which were assumed to align care with patient preferences and reduce aggressive interventions near death. However, when alternative endpoints—such as hospice enrolment, hospital length of stay, inpatient death, and ICU admission—were evaluated using the same AI tool, no measurable improvements were observed. This suggests that proxy indicators may not reliably reflect patient-centered or system-level benefits.

To ensure clinical relevance, outcome measures used in AI evaluation must be both intuitive to clinicians and meaningful to patients. Indicators should reflect critical aspects of care quality, with changes in these metrics corresponding to improvements in healthcare delivery. Delphi consensus methods can help define clinically relevant endpoints across stakeholders. A strong example of this approach is the ongoing international ARCHERY (ARtificial intelligenCe-based radiotHERapY) trial, which incorporates pre-specified clinical metrics—including contouring accuracy and dosimetric performance—alongside workflow efficiency and economic impact assessments.

Secondly, many AI models were trained and validated on retrospective data, which often suffer from missing, inconsistent, or poor-quality entries. These limitations can impair model performance during real-world deployment. The study by Strömblad et al. exemplifies this issue: their algorithm for predicting surgery duration performed poorly when intraoperative events deviated from preoperative plans, highlighting the need for adaptive models that can account for dynamic clinical variables.

Third, few studies investigated the experiences of end-users—both healthcare providers and patients. The studies by Hosny et al. and Wong et al. represent rare exceptions in evaluating user interaction with AI tools. Most deployments involved limited numbers of users over short durations, offering little insight into long-term acceptability, feasibility, and sustainability. For instance, McIntosh et al. reported a 21% decline in clinician adoption of AI-generated radiotherapy plans from simulation to actual clinical use, despite objective superiority in expert assessments. This discrepancy underscores the influence of user trust and perception on adoption.

To address these challenges, implementation science frameworks—such as the Consolidated Framework for Implementation Research (CFIR)—should be embedded in early-stage study designs. These frameworks allow researchers to systematically assess the contextual, cultural, and operational barriers that may influence AI adoption and to develop tailored strategies to address them, including clinician bias and workflow integration.

Finally, economic considerations were largely absent. Only three studies—by Nicolae et al., Hosny et al., and Hong et al.—extended their analyses to include time savings or potential resource efficiencies. However, comprehensive economic evaluations remain rare, despite being critical for informed decision-making, particularly in resource-constrained healthcare settings. Future research must incorporate implementation-focused validation to examine the broader system-level impacts of AI adoption, including changes in healthcare delivery structures, processes, and workforce requirements. Additionally, equity considerations were neglected across all studies. Tools such as the WHO-endorsed RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) framework can guide structured implementation research, ensuring scalability and sustainability of AI interventions across diverse settings.

Table 2 The Risk of Bias In Non-randomised Studies-of Interventions (ROBIN-I) assessment

Study	Risk of bias due to confounding	Risk of bias in selection of participants into the study	Risk of bias in classification of interventions	Risk of bias due to deviations from the intended interventions	Risk of bias due to missing data	Risk of bias in measurement of the outcome	Risk of bias in selection of the reported result	Overall risk of bias
Kao *et al*[22]	Low	Low	Low	Low	Low	Low	Low	Low
Manz *et al*[23]	Low	Low	Low	Low	Serious (missing ECOG status, excluded from analysis)	Low	Moderate (multiple subgroup analysis)	Serious
Hong *et al*[24]	Serious (unadjusted for confounding variables)	Low	Low	Low	Low	Low	Low	Serious
Feng *et al*[28]	Low	Low	Low	Low	Low	Low	Low	Low
Xu *et al*[27]	Serious (small sample size unadjusted for user acceptability of AI and clinical experience)	Low	Low	Low	Low	Low	Moderate (multiple subgroup analysis)	Serious
Hosny *et al*[32]	Low	Low	Low	Low	Low	Low	Low	Low
McIntosh *et al*[30]	Low	Low	Low	Low	Low	Low	Low	Low
Wong *et al*[25]	Low	Low	Low	Low	Low	Serious (unblinded participants with two subjective outcome measures)	Low	Serious

The risk of bias for each domain of the study is colour-coded. High risk of bias: red; moderate risk of bias: orange and low risk of bias: green.
AI, artificial intelligence; ECOG, Eastern Cooperative Oncology Group.

Research Ecosystem for AI

Our review identified a wide variation in methodologies used to evaluate the same types of AI interventions for similar clinical objectives. This inconsistency reflects a broader lack of standardisation in how AI efficacy is assessed across studies. While metrics like the area under the curve (AUC) of the receiver operating characteristic are commonly reported to describe algorithmic performance, such statistical measures alone are insufficient to establish clinical utility.

To ensure robust and generalisable findings, it is critical to consider the full lifecycle of AI deployment. This includes the development environment (the clinical setting and data source used to train the model), the operational environment (how the AI integrates with existing health record systems and institutional infrastructure), and the nature of human-AI interactions. Research protocols should incorporate these factors to provide transparent, context-rich evaluations of AI interventions. The SPIRIT-AI extension to the Consolidated Standards of Reporting Trials (CONSORT) offers valuable guidance for structuring such protocols.

In addition, reporting standards should be tailored to the specific type and purpose of AI. For example, AI-powered image analysis tools should include visual aids such as class activation maps to show which image features most influenced predictions. For predictive models, established guidelines like the TRIPOD-ML (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis – Machine Learning) and STARD-AI (Standards for Reporting Diagnostic Accuracy Studies –

Artificial Intelligence) should be followed. Despite the availability of these resources, only two studies in our review adhered to up-to-date reporting guidelines.

Ethical and regulatory oversight remains an underexplored area. Health data used for AI training and deployment are often sensitive, large-scale, and private, necessitating robust safeguards for collection, storage, and use. For instance, one AI intervention involving a voice coaching application monitored participants' daily physical activity to generate feedback, raising concerns about data confidentiality and user privacy. Securing such data against breaches and misuse requires sustained investment, vigilant oversight, and comprehensive legal frameworks.

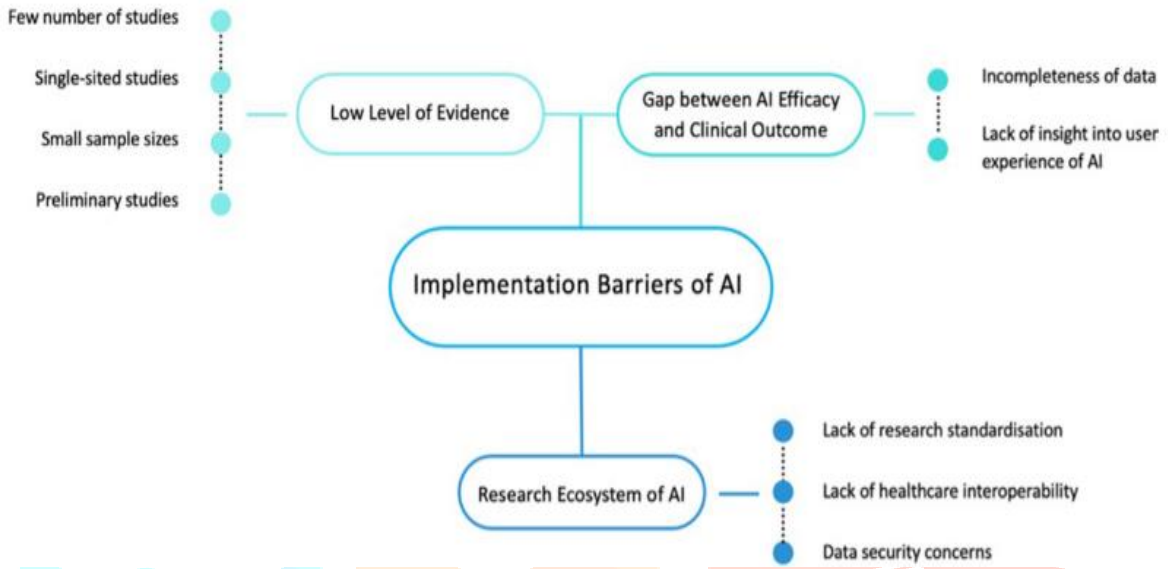

Figure 2 Implementation barriers of artificial intelligence (AI).

Recommendations

To address the triad of challenges—limited clinical evidence, the gap between AI efficacy and patient outcomes, and the fragmented research ecosystem—key actions must be taken by stakeholders across the healthcare and research continuum.

First, there is a pressing need for increased investment in AI research, with a particular emphasis on implementation science. While the overall volume of funding for AI has grown, relatively little is allocated to studies that explore how AI tools can be effectively deployed in real-world clinical settings. Implementation research is essential to understand contextual factors such as workflow integration, clinician adoption, and patient engagement, all of which influence the success of AI in routine oncology care.

Second, coordinated, multidisciplinary efforts are required to drive progress. In the UK, the NHS AI Lab exemplifies a promising approach. It brings together clinicians, data scientists, healthcare providers, and regulators in a collaborative space. Through initiatives such as the AI Health and Care Award, the lab supports AI projects at various stages of development and deployment. This model of stakeholder engagement, underpinned by targeted funding, provides a valuable blueprint for other health systems seeking to scale AI innovation responsibly and effectively.

Such initiatives should be expanded and adapted globally, particularly in resource-limited settings where the potential for AI to improve care is substantial but underutilised. Strengthening collaboration, standardising evaluation frameworks, and prioritising meaningful, patient-centered outcomes will be critical to bridging the gap between AI research and its clinical application in cancer care.

Strengths and Limitations of the Review

The strengths of this review lie in its comprehensive yet focused inclusion criteria, which targeted prospective studies in the postdiagnostic phase of cancer care. This approach allowed for a detailed examination of AI applications beyond diagnosis, an area often overshadowed by the more advanced diagnostic domain. Additionally, the methodology employed was rigorous, combining systematic review techniques with established quality assessment tools. This ensured a critical appraisal of evidence and enhanced the validity of the conclusions drawn. Furthermore, the review goes beyond summarising findings by offering stakeholder-specific recommendations, grounded in practical examples and broader implementation insights.

However, certain limitations must be acknowledged. The review intentionally excluded studies focusing on AI in cancer diagnostics—currently the most mature and widely researched area of oncology AI. As a result, high-quality prospective studies that have been deployed across diverse healthcare settings were not included in the analysis, which may limit generalisability. Furthermore, earlier computational or predictive models that were not explicitly labeled as AI, machine learning, deep learning, or neural networks were also excluded. These models, although different in classification, may still offer valuable insights or parallels relevant to AI implementation and should be considered in future comparative analyses.

Table 3 Action plans tailored to stakeholders for addressing specific issues

Current issues	Action plans	Stakeholders
Lack of interoperability between hospitals	► Engage with medical informatics system vendors to facilitate integration of AI and secure data storage	Healthcare providers
Lack of validation of AI quality/efficacy	► Conduct tests using independent external data to validate, optimise and audit AI efficacy	
Lack of standardisation in evaluation and validation of AI	► Develop and mandate the use of standard oncology terminologies and ontologies ► Set the standards required to evaluate the performance of AI-based tools systematically ► Establish an up-to-date regulatory and legal frameworks for different AI based on implementation risks	Commissioners and regulators
Lack of integration of implementation science framework	► Establish consensus regarding trial protocol involving AI to standardise reporting ► Conduct AI studies that validate patient-centred outcomes and cost/time/resource effectiveness ► Promote implementation science research to learn optimal methods to AI deployment in cancer care	Academics and healthcare providers
Lack of workforce training	► Level up on knowledge of AI and basics of medical informatics ► Prepare for disruption and adapt to changes in nature of work with the integration of AI	Healthcare professionals

AI, artificial intelligence.

CONCLUSION

Artificial intelligence is a rapidly evolving technology with the potential to significantly transform cancer care beyond diagnostics. However, despite the surge in research exploring AI applications in the postdiagnostic phase, only a limited number of tools have undergone prospective clinical evaluation. Concerns remain regarding study size, participant diversity, and methodological robustness. Additionally,

key aspects such as cost-effectiveness, time efficiency, and qualitative assessments of user acceptability, feasibility, and long-term sustainability are notably underexplored.

To bridge the gap between technological promise and clinical reality, future research must prioritize implementation-focused studies. These efforts should be multidisciplinary and co-developed by academic institutions, healthcare providers, policymakers, and patients. Adherence to current research standards and alignment with local health system needs will be essential to comprehensively evaluate AI tools and determine their readiness for safe, effective, and sustainable integration into routine cancer care.

REFERENCES

1. Chua IS, Gaziel-Yablowitz M, Korach ZT, et al. Artificial intelligence in oncology: path to implementation. Cancer Med 2021;10:4138–49.
2. Somashekhar SP, Sepúlveda M-J, Puglielli S, et al. Watson for oncology and breast cancer treatment recommendations: agreement with an expert Multidisciplinary tumor board. Ann Oncol 2018;29:418–23.
3. Mak RH, Endres MG, Paik JH, et al. Use of crowd innovation to develop an artificial intelligence–based solution for radiation therapy targeting. JAMA Oncol 2019;5:654–61.
4. Luchini C, Pea A, Scarpa A. Artificial intelligence in oncology: Current applications and future perspectives. Br J Cancer 2022;126:4–9.
5. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. Nature 2020;577:89–94.
6. Eloy C, Marques A, Pinto J, et al. Artificial intelligence–assisted cancer diagnosis improves the efficiency of Pathologists in Prostatic biopsies. Virchows Arch 2023;482:595–604.
7. Senior K. NHS embraces AI-assisted radiotherapy technology. Lancet Oncol 2023;24.
8. Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature 2019;570:385–9.
9. Cabral BP, Braga LAM, Syed-Abdul S, et al. Future of artificial intelligence applications in cancer care: A global cross-sectional survey of researchers. Curr Oncol 2023;30:3432–46.
10. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17.
11. Crawford K, Calo R. There is a blind spot in AI research. Nature 2016;538:311–3.
12. Chen I, Frederik J, Sontag D. Why is my Classifier discriminatory? Adv Neural Inf Process Syst 2018.
13. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56.
14. Zhou N, Zhang C-T, Lv H-Y, et al. Concordance study between IBM Watson for oncology and clinical practice for patients with cancer in China. Oncologist 2019;24:812–9.
15. Hutson M. Artificial intelligence faces reproducibility crisis. Science 2018;359:725–6.
16. Aristidou A, Jena R, Topol EJ. Bridging the chasm between AI and clinical implementation. The Lancet 2022;399:620.
17. Jandoo T. WHO guidance for Digital health: what it means for researchers. Digit Health 2020;6.
18. Manz CR, Parikh RB, Small DS, 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 Oncol 2020;6:e204759.
19. Manz CR, Zhang Y, Chen K, et al. Long-term effect of machine learning–triggered behavioral Nudges on serious illness conversations and end-of-life outcomes among patients with cancer: A randomized clinical trial. JAMA Oncol 2023;9:414–8.

20. Hassoon A, Baig Y, Naiman DQ, et al. Randomized trial of two artificial intelligence coaching interventions to increase physical activity in cancer survivors. NPJ Digit Med 2021;4.
21. Kao J, Zucker A, Urso M, et al. Improving survival prognostication in patients with metastatic cancer through clinical judgment. Anticancer Res 2022;42:1397–401.
22. Manz CR, Chen J, Liu M, et al. Validation of a machine learning algorithm to predict 180-day mortality for outpatients with cancer. JAMA Oncol 2020;6:1723–30.
23. Hong JC, Eclov NCW, Dalal NH, et al. System for high-intensity evaluation during radiation therapy (SHIELD-RT): A prospective randomized study of machine learning–directed clinical evaluations during radiation and Chemoradiation. J Clin Oncol 2020;38:3652–61.
24. Wong J, Huang V, Wells D, et al. Implementation of deep learningbased auto Segmentation for radiotherapy planning structures: a Workflow study at two cancer centers. Radiat Oncol 2021;16.
25. 26 Nelson SH, Natarajan L, Patterson RE, et al. Physical activity change in an RCT: comparison of measurement methods. Am J Health Behav 2019;43:543–55.
26. Xu F, Sepúlveda M-J, Jiang Z, et al. Effect of an artificial intelligence clinical decision support system on treatment decisions for complex breast cancer. JCO Clin Cancer Inform 2020;4:824–38.
27. Feng L, Liu Z, Li C, et al. Development and validation of a Radiopathomics model to predict pathological complete response to Neoadjuvant Chemoradiotherapy in locally advanced Rectal cancer: a Multicentre observational study. Lancet Digit Health