International Journal of Research in Engineering and Management Sciences (IJREMS)
E-ISSN: 3107-7439

Research Article
VOL.: 2, Issue: 6, June 2026

Generative Artificial Intelligence for Clinical Decision Support in Oncology: Current Applications, Challenges, and Future Directions

[1]Dr. Sanjay K. Deshpande, [2]Dr. Harish V. Reddy, [3]Dr. Asha N. Kapoor

*[*1]MD, DM Director, Cancer Research Programs National Institute of Oncology Sciences Mumbai, India*
sk.deshpande@nios.org.in
[2]MD, PhD Chair, Department of Cancer Biology Institute of Translational Medicine Hyderabad, India, harish.reddy@itm.edu.in
[3]MD Professor of Clinical Cancer Research School of Oncology and Life Sciences Chandigarh, India
asha.kapoor@sols.ac.in

ABSTRACT:
Cancer remains one of the leading causes of morbidity and mortality worldwide, creating an urgent need for intelligent clinical decision-support systems capable of delivering personalized, evidence-based cancer care. The rapid expansion of clinical, imaging, molecular, genomic, and radiomics datasets has accelerated the development of generative artificial intelligence (GenAI), enabling advanced computational models to support complex clinical decision-making throughout the oncology care continuum. Powered by transformer-based architectures and large language models (LLMs), GenAI has emerged as a transformative technology with the potential to enhance diagnostic accuracy, optimize treatment selection, facilitate precision oncology, and accelerate translational cancer research.

This review provides a comprehensive overview of recent advances in generative artificial intelligence for clinical decision support in oncology, with particular emphasis on large language models, multimodal foundation models, retrieval-augmented generation, and clinical natural language processing. We summarize the underlying computational architectures and examine their applications across cancer screening, diagnosis, radiology, pathology, treatment planning, precision oncology, clinical trial matching, prognostic modeling, patient communication, and multidisciplinary decision-making.

Current evidence indicates that GenAI can improve workflow efficiency, automate clinical documentation, synthesize large volumes of biomedical information, support personalized therapeutic recommendations, and facilitate integration of multimodal clinical data into real-time decision-support systems. However, several important barriers remain before widespread clinical implementation can be achieved, including hallucination, limited generalizability across healthcare settings, algorithmic bias, insufficient explainability, privacy and cybersecurity concerns, regulatory uncertainty, and the need for rigorous prospective clinical validation.

Future developments are expected to focus on multimodal artificial intelligence, human–AI collaboration, federated learning, foundation models, explainable AI, and continuously learning clinical decision-support systems capable of integrating imaging, pathology, genomics, electronic health records, and real-world patient data. The convergence of generative artificial intelligence and precision oncology has the potential to fundamentally transform clinical decision-making, improve treatment personalization, enhance healthcare efficiency, and ultimately improve outcomes for patients with cancer.

Keywords: Generative artificial intelligence; Large language models; Clinical decision support; Oncology; Precision medicine; Transformer architecture; Multimodal artificial intelligence; Foundation models; Clinical natural language processing.

Received Date: 7 June 2026**; Accepted Date:** 17 June 2026**; Published Date:** 24 June 2026.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are properly cited.

1. Introduction

Cancer remains one of the leading causes of morbidity and mortality worldwide and continues to impose a substantial clinical, social, and economic burden on healthcare systems. Despite remarkable advances in molecular oncology, targeted therapies, immunotherapy, and precision medicine, cancer management remains challenging because of extensive interpatient and intratumoral heterogeneity. Patients with seemingly similar clinicopathological characteristics often exhibit markedly different

therapeutic responses and clinical outcomes, underscoring the need for personalized treatment strategies capable of addressing the biological complexity of individual tumors [3].

Modern oncology has become increasingly data intensive. Clinical decision-making now requires the integration of diverse sources of information, including medical imaging, pathology, genomics, transcriptomics, molecular biomarkers, laboratory investigations, electronic health records, clinical guidelines, and continuously evolving scientific literature. Treatment recommendations are frequently formulated through multidisciplinary tumor boards, where clinicians must rapidly synthesize large volumes of heterogeneous information while considering numerous therapeutic options and patient-specific factors [4]. The growing complexity of cancer care has highlighted the need for intelligent clinical decision-support systems that can efficiently analyze multimodal data and assist clinicians in delivering evidence-based, personalized treatment recommendations.

Artificial intelligence (AI) has emerged as one of the most transformative technologies in modern healthcare, with applications spanning medical imaging, pathology, drug discovery, genomics, radiotherapy, and clinical decision support. Recent advances in deep learning and foundation models have accelerated the development of generative artificial intelligence (GenAI), particularly large language models (LLMs), which are capable of understanding, generating, summarizing, translating, and reasoning over complex biomedical information with remarkable accuracy. Unlike conventional machine learning systems designed for narrowly defined predictive tasks, generative AI models can synthesize knowledge from diverse data sources, interpret clinical narratives, assist with evidence retrieval, and support complex reasoning processes required in oncology practice [5].

The convergence of generative AI with precision oncology offers unprecedented opportunities to improve diagnostic accuracy, optimize treatment planning, personalize therapeutic recommendations, automate clinical documentation, facilitate multidisciplinary decision-making, and accelerate translational cancer research. Emerging multimodal foundation models further extend these capabilities by integrating medical imaging, digital pathology, genomics, radiomics, laboratory findings, and electronic health records within unified computational frameworks, enabling comprehensive patient-level clinical decision support [6].

Nevertheless, despite their considerable promise, generative AI systems face important technical, clinical, and ethical challenges that must be addressed before widespread implementation in routine oncology practice. Hallucination, algorithmic bias, limited explainability, insufficient prospective clinical validation, privacy concerns, cybersecurity risks, interoperability challenges, and evolving regulatory requirements continue to limit clinical adoption. Ensuring that AI systems remain transparent, reliable, equitable, and clinically trustworthy will be essential for their safe integration into healthcare workflows.

Driven by rapid advances in deep learning, large-scale biomedical datasets, multimodal artificial intelligence, and clinical natural language processing, research into generative AI for oncology has expanded dramatically over recent years [7]. This review provides a comprehensive overview of generative artificial intelligence for clinical decision support in oncology, summarizing the underlying computational architectures, current clinical applications, emerging multimodal foundation models, implementation challenges, and future directions. By critically evaluating the evolving evidence, this review aims to provide clinicians, researchers, and policymakers with a balanced perspective on how generative AI may reshape precision oncology and support the delivery of safer, more efficient, and highly personalized cancer care.

2. Fundamentals of Generative AI

2.1 Large Language Models and Transformer Architecture

Generative artificial intelligence (GenAI) is built upon transformer-based deep learning architectures that have fundamentally transformed natural language processing (NLP) and, more recently, numerous applications in medical artificial intelligence. Among the most influential advances are large language models (LLMs), which are trained on massive text corpora to understand, generate, summarize, translate, and reason over human language. Unlike traditional rule-based or statistical language processing systems, LLMs learn complex semantic relationships directly from data, enabling

highly coherent and contextually appropriate responses across a wide range of clinical and biomedical tasks [8].

The introduction of the transformer architecture represented a major breakthrough in artificial intelligence by replacing recurrent sequential processing with self-attention mechanisms capable of modeling long-range contextual dependencies. Self-attention enables transformer models to selectively focus on the most relevant portions of an input sequence regardless of their position, substantially improving computational efficiency, scalability, and contextual understanding. These capabilities have established transformers as the foundation of modern natural language processing and the underlying architecture for most contemporary foundation models [9].

Building upon this architecture, large-scale pre-trained models such as BERT, GPT, and subsequent transformer-based foundation models have demonstrated remarkable performance across diverse language understanding and generation tasks. Through self-supervised learning on enormous biomedical and general-domain text collections, these models acquire rich contextual representations that can subsequently be adapted to clinical question answering, document summarization, information extraction, medical reasoning, report generation, and clinical decision support with relatively limited task-specific training [10].

Beyond natural language processing, transformer architectures have increasingly expanded into computer vision and multimodal artificial intelligence. Vision transformers (ViTs) and multimodal foundation models have achieved state-of-the-art performance in medical image segmentation, image reconstruction, image registration, lesion detection, disease classification, radiology report generation, pathology analysis, and multimodal clinical reasoning. The ability to jointly process textual, imaging, pathological, and molecular information makes transformer-based models particularly valuable for oncology, where clinical decision-making requires integration of heterogeneous biomedical data from multiple sources [9].

The rapid evolution of transformer architectures has therefore established the computational foundation for next-generation clinical decision-support systems. Their capacity to integrate complex medical knowledge, interpret diverse biomedical information, and generate clinically meaningful recommendations positions large language models as a central technology driving the future of precision oncology and intelligent healthcare systems.

2.2 Clinical Natural Language Processing

The application of transformer architectures to large biomedical and clinical text corpora has substantially advanced clinical natural language processing (NLP). Domain-specific pre-trained language models, including BioBERT, ClinicalBERT, and PubMedBERT, have demonstrated significant improvements over traditional machine learning and recurrent neural network (RNN)-based approaches by learning contextual biomedical representations directly from large-scale scientific literature and electronic health records. These models have improved the accuracy of numerous clinical NLP tasks, including named entity recognition, relation extraction, document classification, information retrieval, and clinical question answering [11].

The evolution of transformer-based language models has resulted in remarkable improvements in clinical NLP performance. Meta-analytic evidence indicates that RNN-based models achieved F1 scores ranging from 64.8% to 65.8% during 2018–2019, whereas BERT-based architectures improved performance to 76.1%–78.3% between 2020 and 2022. More recently, hybrid transformer models have achieved F1 scores ranging from 87.8% to 89.7% during 2023–2025, reflecting the rapid evolution of language modeling technologies for healthcare applications [11].

Scaling language models has further enhanced clinical reasoning capabilities. While early clinical language models typically contained approximately 110 million parameters, recent models such as GatorTron have expanded to 8.9 billion parameters, substantially improving performance across multiple clinical NLP tasks, including clinical concept extraction, medical relation extraction, semantic textual similarity, natural language inference, and medical question answering. GatorTron demonstrated improvements of 9.6% and 9.5% in natural language inference and medical question answering, respectively, highlighting the benefits of

large-scale domain-specific pre-training for clinical decision support [12].
The continued development of increasingly powerful clinical language models provides an essential foundation for intelligent oncology decision-support systems capable of accurately interpreting electronic health records, pathology reports, radiology reports, clinical guidelines, and biomedical literature while facilitating evidence-based clinical reasoning.

2.3 Multimodal Artificial Intelligence and Foundation Models

Although large language models have revolutionized text-based clinical reasoning, modern oncology requires simultaneous interpretation of multiple heterogeneous data sources, including medical imaging, digital pathology, genomics, laboratory investigations, electronic health records, and molecular biomarkers. This need has driven the rapid development of multimodal artificial intelligence, which extends transformer architectures beyond text to jointly analyze diverse biomedical modalities within a unified computational framework [13].

Transformer-based multimodal learning has emerged as one of the fastest-growing areas of artificial intelligence because self-attention mechanisms enable efficient modeling of complex relationships both within and across multiple data modalities. These capabilities are particularly valuable in oncology, where accurate clinical decision-making depends on integrating radiological findings, pathological characteristics, genomic alterations, laboratory data, and longitudinal clinical information into a comprehensive patient-specific representation.

Recent advances have also led to the development of foundation models and Generalist Medical Artificial Intelligence (GMAI) systems capable of performing multiple clinical tasks with minimal or no task-specific training. Built through large-scale self-supervised learning, these models learn generalized biomedical representations that can be adapted to diverse applications, including diagnosis, prognosis, clinical documentation, image interpretation, treatment recommendation, patient stratification, and clinical decision support [14].

Unlike traditional task-specific artificial intelligence systems, foundation models can simultaneously interpret combinations of medical imaging, pathology, genomics, laboratory results, electronic health records, biomedical knowledge graphs, and free-text clinical documentation. This multimodal capability enables increasingly comprehensive clinical reasoning and represents a major step toward intelligent oncology decision-support systems capable of supporting personalized precision medicine across the entire cancer care continuum.

The convergence of large language models, multimodal transformers, and foundation models is expected to fundamentally reshape clinical oncology by enabling integrated, continuously learning artificial intelligence systems that combine heterogeneous biomedical information into accurate, interpretable, and patient-specific clinical recommendations.

3. Applications in Oncology Clinical Decision Support

3.1 Diagnosis Support

Artificial intelligence-based clinical decision support systems (AI-CDSS) provide a viable solution through advanced methodologies for comprehensive data analysis. AI-CDSS facilitate early cancer detection, precise staging, and personalized treatment planning by processing multimodal patient information through machine learning, computer vision, and natural language processing. These systems effectively interpret clinical results, identify critical disease patterns including clinical stage, site, tumor dimensions, histopathologic grading, and molecular profiles, and construct comprehensive patient profiles. [15]
The application of LLMs to diagnostic support has demonstrated significant potential across various cancer types. DeepSeek demonstrated superior performance in diagnostic accuracy (F1-score: 89.2% vs ChatGPT's 76.5%, $P < 0.001$) and treatment alignment with guidelines ($\kappa = 0.82$ vs 0.67, $P = 0.003$). ChatGPT exhibited strengths in patient communication, generating layman-friendly explanations (readability score: 8.2/10 vs DeepSeek's 6.5/10, $P = 0.012$). Both models showed limitations in rare cancer subtypes (e.g., cholangiocarcinoma), with accuracy dropping below 60%. [16] These findings underscore the complementary strengths of different AI systems and the importance of selecting appropriate tools for specific clinical tasks.
GPT-4-based ChatGPT demonstrates significant potential in various industries; however, its potential clinical applications remain largely unexplored. ChatGPT achieved an 87% accuracy without choices

and a 97% accuracy with choices, after excluding image-based quizzes. ChatGPT excelled in the Diagnosis category, attaining 89% accuracy without choices and 98% with choices, demonstrating potential for diagnostic applications, suggesting its usefulness in supporting healthcare professionals in making differential diagnoses. [17]

3.2 Treatment Planning

A cross-sectional observational study involving 1,977 patients at high risk for recurrent or metastatic breast cancer examined the impact of a clinical decision support system on breast cancer treatment decisions and adherence to National Comprehensive Cancer Center guidelines. Treatment decisions changed in 105 (5%) of 1,977 patients and were concentrated in those with hormone receptor-positive disease or stage IV disease in the first-line therapy setting (73% and 58%, respectively). Reasons cited for changes included consideration of the CDSS therapeutic options (63% of patients), patient factors highlighted by the tool (23%), and the decision logic of the tool (13%). Adherence to NCCN treatment guidelines increased slightly after using the CDSS (0.5%; P = .003). [18]

An explainable AI system designed to reproduce multidisciplinary cancer conference-based treatment decisions for metastatic and non-metastatic prostate cancer analyzed 5478 MCC cases including 76 clinical input variables and 23 treatment output parameters. The AI system generated automated treatment recommendations with high predictive accuracy across both hierarchical levels. For high-level categories, F1-scores reached 0.89 for surgery and 0.81 for radiation therapy. For detailed recommendations, F1-scores reached 0.99 for prostatectomy and 0.98 for PSMA-ligand therapy. [19] These results demonstrate the feasibility of AI systems in replicating expert consensus decisions within multidisciplinary tumor boards.

3.3 Radiology and Pathology Assistance

Clinical workflows in oncology rely on predictive and prognostic molecular biomarkers. Advances in deep learning, an artificial intelligence technology, have enabled the extraction of previously hidden information directly from routine histology images of cancer, providing potentially clinically useful information. DL has been used for advanced image analysis tasks, which have the potential of directly affecting clinical decision-making processes, including inference of molecular features, prediction of survival and end-to-end prediction of therapy response. [20]

A convolutional neural network was trained to predict clinically significant prostate cancer from T2-weighted images, diffusion-weighted images, apparent diffusion coefficient maps, and T1-weighted contrast-enhanced images. Among 5735 examinations in 5215 patients, 1514 examinations showed csPCa. In the internal test set (400 examinations), the AUC was 0.89 and 0.89 for the DL classifier and radiologists, respectively. In the external test set (204 examinations), the AUC was 0.86 and 0.84 for the DL classifier and radiologists, respectively. [21] Deep learning models have achieved radiologist-level performance in detecting clinically significant cancers, suggesting significant potential for augmenting clinical workflows.

The development of digital pathology and progression of state-of-the-art algorithms for computer vision have led to increasing interest in the use of AI, especially deep learning-based AI, in tumor pathology. The DL-based algorithms have been developed to conduct all kinds of work involved in tumor pathology, including tumor diagnosis, subtyping, grading, staging, and prognostic prediction, as well as the identification of pathological features, biomarkers and genetic changes. [22]

3.4 Precision Oncology and Genomics

In oncology, by mining the underlying connection between a text or image input and the desired output, LLMs demonstrate great potential for managing tumours. LLMs play a role in cancer screening and diagnosis, metastasis identification, tumour staging, treatment recommendation, and documentation processing tasks by decoding various types of clinical data. [23]

By integrating an expert-curated dataset with RAG, LLMs can accurately suggest evidence-based therapies across various cancer types, which may facilitate precision oncology investigations. Leveraging structured text and RAG improved model performance by 43% relative to a general-purpose LLM (94% vs. 65.8%), highlighting the benefit of integrating external knowledge for better therapeutic recommendations. [24] Retrieval-augmented generation approaches demonstrate substantial improvements in matching patients to appropriate targeted therapies based on genomic profiles.

This study assessed various ChatGPT versions' ability to generate accurate next-generation

sequencing reports and treatment recommendations for NSCLC. Analyzing 160 responses, GPT-4 outperformed GPT-3.5, showing higher base score (90% v 60%; P < .01) and fewer hallucinations (34% v 53%; P < .01). GPT-4's overall G-PS was significantly higher (0.34 v –0.15; P < .01), indicating superior performance in matching treatment recommendations with biomarkers in precision oncology. [25]

3.5 Clinical Trial Matching

TrialMatchAI is an AI-powered recommendation system that automates patient-to-trial matching by processing heterogeneous clinical data, including structured records and unstructured physician notes. Built on fine-tuned, open-source large language models within a retrieval-augmented generation framework, TrialMatchAI ensures transparency and reproducibility. In real-world validation, 92% of oncology patients had at least one relevant trial retrieved within the top 20 recommendations, with expert assessment validating over 90% accuracy in criterion-level eligibility classification. [26]

A multi-agent AI platform underpinned by an oncology-specific knowledge graph screened 3,804 patients and identified 23,912 trials, with 17,912 confirmed after expert review. Time-to-recommendation was under one week from screening to final recommendations. Performance metrics demonstrated Sensitivity of 0.8375, Specificity of 0.8359, Precision of 0.8121, and F1 Score of 0.8246, significantly outperforming zero-shot or frontier GPT-based models. [27] Specialized AI platforms for clinical trial matching substantially outperform general-purpose LLMs, highlighting the importance of domain-specific optimization.

A demonstration project combining AI clinical decision support with expert consultation for NGS testing showed that over 30% of treatment decisions were altered based on CDS AI plus expert review. All NGS tests were ordered via platform automation with testing taking 16 days, and all patients had NGS results that impacted therapeutic options, including 42% with two or more therapy matches and 22% with options beyond the NGS report. [28]

4. Current Models and Systems

4.1 GPT-Based Systems

ChatGPT has been used to generate accurate differential diagnosis lists, support clinical decision-making, optimize clinical decision support, and provide insights for cancer screening decisions. In addition, ChatGPT has been used for intelligent question-answering to provide reliable information about diseases and medical queries. In terms of medical documentation, ChatGPT has proven effective in generating patient clinical letters, radiology reports, medical notes, and discharge summaries, improving efficiency and accuracy for health care providers. [29]

GPT-4's performance was examined on 643 Congress of Neurological Surgeons Self-Assessment Neurosurgery Exam board-style questions. GPT-4 attempted 91.9% of questions and achieved 76.6% accuracy, increasing to 79.0% for text-only questions. GPT-4 outperformed ChatGPT ($p<0.001$) and exceeded the performance of medical students (26.3%), neurosurgery residents (61.5%), and the national average of SANS users (69.3%) across all categories. [30]

A prospective study evaluated the compatibility of AI (ChatGPT-4.0) with multidisciplinary tumor council decisions in 100 cancer patients. A high concordance rate of 76.4% was observed between AI and MDT decisions ($\kappa = 0.764$, $p < 0.001$). Most inconsistencies arose in cases requiring individualized decisions, indicating AI's current limitations in incorporating contextual clinical judgment. [31]

4.2 Med-PaLM and Domain-Specific Models

Using a combination of prompting strategies, Flan-PaLM achieves state-of-the-art accuracy on every MultiMedQA multiple-choice dataset including 67.6% accuracy on MedQA (US Medical Licensing Exam-style questions), surpassing the prior state of the art by more than 17%. Med-PaLM, the resulting model from instruction prompt tuning, performs encouragingly but remains inferior to clinicians. Comprehension, knowledge recall and reasoning improve with model scale and instruction prompt tuning, suggesting the potential utility of LLMs in medicine. [32]

Med-PaLM 2 scores up to 86.5% on the MedQA dataset, improving upon Med-PaLM by over 19%, and demonstrates dramatic performance increases across MedMCQA, PubMedQA and MMLU clinical topics datasets. Detailed human evaluations show that physicians prefer Med-PaLM 2 answers to those from other physicians on eight of nine clinical axes. In a pilot study using real-world medical questions, specialists preferred Med-PaLM 2 answers to generalist physician answers 65% of the time. [33] Med-PaLM 2 represents a significant advancement in

medical LLMs, demonstrating the value of domain-specific fine-tuning for clinical applications.

4.3 Hospital-Integrated AI Systems

An AI tool, Watson for Oncology, used for the treatment of cancer has been implemented in several different settings, including Brazil, China, India, South Korea, and Mexico. By focusing on the implementation of an AI-based clinical decision support system for oncology, this demonstrates how AI can be both beneficial and challenging for cancer management globally and particularly for low-middle-income countries. [34]

Generative AI enhanced with agentic AI and vector-based RAG, using only open-source tools and LLMs, produced reliable breast cancer summaries and treatment evaluations. A workflow involving modular LLMs iterating over patient notes extracted cancer-related information, with subsequent AI agents comparing extractions with mCODE summaries and evaluating against NCCN guidelines. No hallucinations were observed in outputted data and no incorrect interpretations were found. [35] Open-source, modular AI systems demonstrate promise for eliminating hallucinations through grounding and retrieval-augmented approaches.

5. Benefits and Clinical Impact

5.1 Accuracy Improvement

Deep learning models achieve high diagnostic accuracy in medical imaging. In ophthalmology, AUCs ranged between 0.933 and 1 for diagnosing diabetic retinopathy, age-related macular degeneration and glaucoma on retinal fundus photographs and optical coherence tomography. In respiratory imaging, AUCs ranged between 0.864 and 0.937 for diagnosing lung nodules or lung cancer on chest X-ray or CT scan. For breast imaging, AUCs ranged between 0.868 and 0.909 for diagnosing breast cancer on mammogram, ultrasound, MRI and digital breast tomosynthesis. [36]

Advanced LLMs showed high diagnostic accuracy (>90%) in common clinical scenarios, with Claude 3.7 achieving perfect accuracy (100%) in certain conditions. In complex cases, Claude 3.7 achieved the highest accuracy (83.3%) at the final diagnostic stage, significantly outperforming smaller models. [37] Leading LLMs demonstrate remarkable diagnostic capabilities that approach or exceed clinician performance in specific contexts.

5.2 Workflow Efficiency

Large language models have emerged as transformative tools in medicine, with strong capabilities in language understanding, reasoning, and structured information extraction. Radiation oncology is particularly well suited for LLM integration due to its data-intensive workflows, reliance on structured guidelines, and documentation burden. Applications include domain-specific fine-tuning for decision support, automated nomenclature standardization, registry curation using autonomous LLM agents, and protocol-aware radiotherapy plan evaluation. [38]

ChatGPT-3.5 demonstrated the ability to extract pathological classifications with an overall accuracy of 89%, outperforming the performance of two traditional NLP methods. The results underscore the potential role of LLMs in transforming unstructured healthcare data into structured formats, thereby supporting research and aiding clinical decision-making without requiring extensive task-specific human annotation and model training. [39]

5.3 Personalized Medicine

Cancer research encompasses data across various scales, modalities, and resolutions, from screening and diagnostic imaging to digitized histopathology slides to various types of molecular data and clinical records. The integration of these diverse data types for personalized cancer care and predictive modeling holds the promise of enhancing the accuracy and reliability of cancer screening, diagnosis, and treatment. [40]

For 10 fictional cancer patients, a median number of 4.0 treatment options was identified by the human expert compared with higher numbers from LLMs. For each patient, at least one LLM generated a treatment option considered helpful by MTB members. Two unique useful treatment options were identified only by LLM, highlighting that LLMs can complement established precision oncology procedures despite not yet reaching the quality of human experts. [41]

6. Challenges and Limitations

6.1 Hallucinations

A significant risk associated with the use of LLMs is their potential to create hallucinations. Hallucinations (false information) generated by LLMs arise from a multitude of causes, including both factors related to the training dataset as well as their auto-regressive nature. The implications for clinical practice range

from the generation of inaccurate diagnostic and therapeutic information to the reinforcement of flawed diagnostic reasoning pathways, as well as a lack of reliability if not used properly. [42]

A framework for assessing clinical safety and hallucination rates of LLMs for medical text summarization demonstrated that by refining prompts and workflows, major errors were successfully reduced below previously reported human note-taking rates. Clinical error metrics derived from 18 experimental configurations involving LLMs for clinical note generation, consisting of 12,999 clinician-annotated sentences, observed a 1.47% hallucination rate and a 3.45% omission rate. [43]

Medical large language models are vulnerable to data-poisoning attacks. Replacement of just 0.001% of training tokens with medical misinformation results in harmful models more likely to propagate medical errors. Furthermore, corrupted models match the performance of their corruption-free counterparts on open-source benchmarks routinely used to evaluate medical LLMs, making detection challenging. [44]

6.2 Data Bias

Large language models are increasingly integrated into healthcare for clinical decision support and patient communication. Although these models can pass explicit social bias tests, they may retain implicit biases that could influence medical judgment. All 10 models examined exhibited systematic implicit biases across all categories, with the strongest biases observed in Race (Mean = 0.61) and Socioeconomic Status (Mean = 0.56). Stronger implicit associations significantly predicted discriminatory choices in downstream medical decision tasks ($p < 0.001$). [45]

LLMs often proposed inferior treatments when patient race was explicitly or implicitly indicated, though diagnostic decisions demonstrated minimal bias. These findings underscore critical concerns about the potential for AI to perpetuate racial disparities in mental healthcare, emphasizing the necessity of rigorous bias assessment in algorithmic medical decision support systems. [46]

6.3 Regulatory and Ethical Issues

AI integration in oncology is transforming therapeutic decision-making by providing clinical decision support. AI may improve treatment precision, but it raises ethical, legal, and informed consent issues. Key concerns include algorithmic transparency, unclear accountability in AI-guided decisions, data privacy, and gaps in patient understanding of AI's role in their care. [47]

The regulation of GPT-4 and generative AI in medicine and healthcare without damaging their exciting and transformative potential is a timely and critical challenge to ensure safety, maintain ethical standards, and protect patient privacy. Regulatory oversight should assure medical professionals and patients can use LLMs without causing harm or compromising their data or privacy. [48]

6.4 Explainability

While LLMs have achieved excellent performance on medical licensing exams, these tests fail to assess many skills necessary for deployment in a realistic clinical decision-making environment, including gathering information, adhering to guidelines, and integrating into clinical workflows. Current state-of-the-art LLMs do not accurately diagnose patients across all pathologies, follow neither diagnostic nor treatment guidelines, and cannot interpret laboratory results, posing a serious risk to the health of patients. [49]

An explainable artificial intelligence model for clinically significant PCa diagnosis at biparametric MRI using PI-RADS features for classification justification achieved an area under the receiver operating characteristic curve of 0.89 in internal and 0.87 in external test sets. XAI-assisted readings improved the confidence of nonexperts in assessing PI-RADS 3 lesions, reducing reading time by 58 seconds while providing visual and textual explanations using well-established imaging features. [50]

7. Future Directions

7.1 Multimodal AI

Multimodal artificial intelligence methods are now a paradigm-shifting approach for cancer prognosis and diagnosis, allowing the blending of heterogeneous data modalities including histopathology images, radiomics, genomics, and clinical data. Comparative analysis demonstrated that transformer-based and hybrid models offered the highest mean AUC and concordance index measures (0.93 and 0.91, and 0.89 and 0.88, respectively). [51]

Pathomic Fusion represents an interpretable strategy for end-to-end multimodal fusion of histology image and genomic features for survival outcome prediction. Results demonstrate that the proposed

multimodal fusion paradigm improves prognostic determinations from ground truth grading and molecular subtyping, as well as unimodal deep networks trained on histology and genomic data alone. [52] Integrated approaches combining imaging and genomic data consistently outperform unimodal methods, pointing toward the importance of comprehensive multimodal systems in future oncology AI.

7.2 Human-AI Collaboration

Evidence from head and neck oncology remains early-stage and heterogeneous, dominated by simulation-based and small cohort studies with limited real-world validation. GenAI performs best in structured, language-based tasks such as clinical documentation, case summarization, and patient education. GenAI shows promise as an assistive tool but is not yet suitable for autonomous clinical decision-making. Prospective, workflow-integrated evaluation and standardized validation are needed before safe clinical adoption. [53]

A study evaluating LLMs as alternative to rule-based alert systems focusing on their ability to identify prescribing errors demonstrated that the co-pilot arm (pharmacist plus LLM-based CDSS) achieved the best performance with an accuracy of 61%. In detecting errors posing serious harm, the co-pilot mode increased accuracy by 1.5-fold over the pharmacist alone, demonstrating that effective LLM integration for complex tasks can enhance healthcare professional performance. [54]

7.3 Regulatory Frameworks

The upcoming evolutionary leap involves the transition from an AI-powered model primarily designed for answering medical questions to a more versatile and practical tool for healthcare providers such as generalist biomedical AI systems for multimodal-based calibrated decision-making processes. The development of more accurate virtual clinical partners could enhance patient engagement, offering personalized support, and improving chronic disease management. [55]

With generative AI continuing to make inroads in healthcare, assessing LLMs with human evaluations is essential to assuring safety and effectiveness. A comprehensive framework for human evaluation of LLMs covering three phases of workflow, Planning, Implementation and Adjudication, and Scoring and Review, is designed with five proposed evaluation principles: Quality of Information, Understanding and Reasoning, Expression Style and Persona, Safety and Harm, and Trust and Confidence. [56]

8. Conclusion

Generative artificial intelligence, particularly large language models and multimodal AI systems, represents a transformative technology for clinical decision support in oncology. Integration of GAI into oncology has shown some ability to enhance diagnostic accuracy, optimize treatment decisions, and improve clinical efficiency, ultimately strengthening the patient-physician relationship. Despite these advancements, the inherent stochasticity of GAI's performance necessitates human oversight, more specialized models, proper physician training, and robust guidelines to ensure its well tolerated and effective integration into oncologic practice. [57]

The evidence reviewed demonstrates substantial progress across multiple applications, from diagnosis support and treatment planning to precision oncology and clinical trial matching. Current systems have achieved performance levels that approach or exceed human clinicians in specific, well-defined tasks. However, significant challenges remain, including the risk of hallucinations, persistent biases in model outputs, regulatory uncertainties, and limitations in model explainability.

Without human oversight, guidance and responsible design and operation, generative AI applications will remain a party trick with substantial potential for creating and spreading misinformation or harmful and inaccurate content at unprecedented scale. However, if positioned and developed responsibly as companions to humans augmenting but not replacing their role in decision making, knowledge retrieval and other cognitive processes, they could evolve into highly efficient, trustworthy, assistive tools for information management. [58]

The future of AI in oncology lies in the development of multimodal systems capable of integrating diverse data sources, establishment of robust human-AI collaboration frameworks, and creation of comprehensive regulatory guidelines. The convergence of multimodal data, federated architectures, and biological knowledgebases has the potential to enable rapid clinical translation, bridging research and practice to redefine personalized cancer care and improve survival outcomes across diverse cancer subtypes. [59] As these technologies mature,

their responsible integration into clinical workflows promises to enhance the precision, efficiency, and personalization of oncology care, ultimately improving outcomes for cancer patients worldwide.

References

1. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12(1):31–46. Doi:10.1158/2159-8290.CD-21-1059.
2. Rajendran LKK. Machine Learning–Driven Symptom-Based Cancer Risk Stratification:A Systematic Review of Clinical Prediction Models and Methodological Rigor. Int J Drug Deliv Technol. 2026;16(40s):242-253.Doi:10.25258/ijddt.16.40s.26.
3. Rajendran LKK. Identifying Determinants of Outcome in Post-Radiotherapy Cervical Carcinoma Requiring Adjuvant Surgery. International Journal of Clinical Research in Medical Sciences. 2026;1(2):1-10. Doi:10.67231/3acej759.
4. Kumar RMH. Pan-System Cancer Intelligence: Integrating Blood, Immune, Microbiome, and Tumor Microenvironment Data Using Foundation Models. Power System Protection and Control. 2023;51(4):92-100. Doi:10.46121/pspc.51.4.8.
5. Maradi Hemanth Kumar R. AI-Driven Liquid Biopsy Systems for Early Cancer Detection and Personalized Oncology. Power System Protection and Control. 2023;51(4):66-83. Doi:10.46121/pspc.51.4.7.
6. Rajendran LKK. Hematological Malignancy Identification via K-means based ROI Extraction. International Journal of Clinical Research in Medical Sciences. 2026;1(2):1-10. Doi:10.67231/kt1w3e73.
7. Rajendran OK. Bias, Fairness, and Ethical Challenges in Artificial Intelligence: A Comprehensive Review of Causes, Impacts, and Mitigation Strategies. Scientific Culture. 2026;12(2.1):13001-13010. Doi:10.5281/zenodo.20374091.
8. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56. Doi:10.1038/s41591-018-0300-7.
9. Rajendran OK. Clinical Translation of Artificial Intelligence in Oncology: Real-World Validation, Workflow Integration, and Precision Medicine Applications. Int J Drug Deliv Technol. 2026;16(49s):956-964. Doi:10.25258/ijddt.16.49s.110.
10. 10.Rajendran LKK. Interpretable Machine Learning for Early Mortality Prediction in Acute Myeloid Leukemia: A Decision Tree–Based Retrospective Cohort Study. Int J Drug Deliv Technol. 2026;16(40s):231-241.Doi:10.25258/ijddt.16.40s.25.
11. 11.Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24–29. Doi:10.1038/s41591-018-0316-z.
12. 12.Rajendran OK. Generative AI for Synthetic Medical Image Generation in Oncology: Addressing Data Scarcity in AI-Driven Cancer Diagnosis. Int J Drug Deliv Technol. 2026;16(49s):1010-1016. Doi:10.25258/ijddt.16.49s.117.
13. 13.Rajendran LKK. Integrated Prognostic Modeling of Tumor Stage, Multimodal Therapy, and Functional Status in Lung Cancer Survival: A Real-World Cohort Study. Scientific Culture. 2026;12(5):567-576. Doi:10.5281/zenodo.1250046.
14. 14.Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. arXiv. 2021. Doi:10.48550/arXiv.2108.07258.
15. 15.Rajendran LKK. Integrative Pharmacogenomic Analysis of Drug Response Heterogeneity Across Cancer Cell Lines: Insights From Large-Scale GDSC Data. Scientific Culture. 2026;12(4):7537-7546. Doi:10.5281/zenodo.12426762.
16. 16.Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. J Intern Med. 2020;288(1):62–81. Doi:10.1111/joim.13030.
17. 17.Rajendran OK. Tumor Microenvironment Interaction-Guided Graph Neural Networks for Survival Prediction from Whole-Slide Pathology Images. Int J Drug Deliv Technol. 2026;16(49s):481-488. Doi:10.25258/ijddt.16.49s.50.
18. 18.Rajendran LKK. Evaluating the Association of Cancer-Related Risk Factors With Multisystem Health: Insights Into Fertility, Cardiovascular, and Renal Indicators. Scientific Culture. 2026;12(4):7520-7527.Doi:10.5281/zenodo.12426760.
19. 19.Rajendran LKK. From Prediction to Precision: An Externally Validated Deep

Learning–Based Survival and Adjuvant Therapy Recommendation System for Resected Stage III Non–Small Cell Lung Cancer. Int J Drug Deliv Technol. 2026;16(30s):430-438.doi:10.25258/ijddt.16.30s.41.
20. 20.Chen RJ, Lu MY, Wang J, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. Nat Mach Intell. 2022;4:179–193. Doi:10.1038/s42256-022-00466-x.
21. 21.Rajendran LKK. From Prediction to Practice: A Machine Learning–Based Clinical Decision Support Tool for Bevacizumab Risk Stratification in Oncology. Int J Drug Deliv Technol. 2026;16(30s):414-429. Doi:10.25258/ijddt.16.30s.40.
22. 22.Rajendran OK. Self-supervised multimodal Learning for early cancer detection across Imaging and genomics. Power System Protection and Control. 2024;52(4):167-178. Doi:10.46121/pspc.52.4.14.
23. 23.Rajendran OK. Explainable AI-Driven Clinical. Decision Support Systems in Precision Oncology:Interpretable Models for Multimodal Cancer Care. Scientific Culture. 2026;12(2.1):12359-12369. Doi:10.5281/zenodo.20328194.
24. 24.Rajendran LKK. Impact of Treatment Modalities on Fertility, Sexual Function, and Psychological Outcomes in Testicular Cancer Survivors: A Comprehensive Review. Int J Drug Deliv Technol. 2026;16(30s):447-453. Doi:10.25258/ijddt.16.30s.43.
25. 25.Rajendran LKK. Intelligent Omics-Driven Patient Stratification for Cancer Therapeutic Re-profiling. International Journal of Clinical Research in Medical Sciences. 2026;1(1):1-11. Doi:10.67231/gv5hck05.
26. 26.Rajendran LKK. Cancer nanomedicine: utilizing the enhanced permeability and retention (EPR) effect to deliver high payloads of chemotherapeutic agents directly to tumor sites. Power System Protection and Control. 2024;52(2):123-129. Doi:10.46121/pspc.52.2.12.
27. 27.Kather JN, Calderaro J. Development of AI in digital pathology. Nat Rev Clin Oncol. 2020;17(10):591–595. Doi:10.1038/s41571-020-00431-0.
28. 28.Rajendran OK. AI-based radiogenomic Models for predicting immunotherapy response In solid tumors. Power System Protection and Control. 2023;51(4):24-37. Doi:10.46121/pspc.51.4.4.
29. 29.Rajendran LKK. Enhanced Predictive Analytics for Early Malignancy Discovery in Routine Screening. International Journal of Clinical Research in Medical Sciences. 2026;1(1):1-10. Doi:10.67231/grams870.
30. 30.Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer. 2017;17(4):223–238. Doi:10.1038/nrc.2017.7.
31. 31.Rajendran OK. Machine Learning-Based Prediction of Chemotherapy Toxicity in Colorectal Cancer: A Personalized Risk Stratification Approach. Scientific Culture. 2026;12(5.1):942-952. Doi:10.5281/zenodo.12511075.
32. 32.Rajendran OK. Federated radiology AI Models for multi-institutional cancer diagnosis Without data sharing. Power System Protection And Control. 2023;51(4):38-54. Doi:10.46121/pspc.51.4.5.
33. 33.Bera K, Schalper KA, Rimm DL, et al. 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.
34. 34.Rajendran OK. Deep Reinforcement Learning in Oncology: Advances in Cancer Imaging, Radiotherapy, and Personalized Treatment. Scientific Culture. 2026;12(5):597-606. Doi:10.5281/zenodo.1250048.
35. 35.Rajendran OK. DEEP LEARNING FOR CROSS-MODALITY MAPPING BETWEEN HISTOPATHOLOGY AND RADIOLOGICAL IMAGING. Power System Protection and Control. 2025;53(3):313-328. Doi:10.46121/pspc.53.3.21.
36. 36. Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. Nature. 2021;594(7861):106–110. Doi:10.1038/s41586-021-03512-4.
37. 37.Rajendran OK. Artificial Intelligence in Oncologic Imaging: Deep Learning, Radiomics, and Clinical Integration for Precision Cancer Diagnosis. Int J Drug Deliv Technol. 2026;16(50s):871-880. Doi:10.25258/ijddt.16.50s.92.
38. 38.Bilal M, Raza SEA, Azam A, et al. Development and validation of a weakly

supervised deep learning framework to predict the risk of colorectal cancer recurrence from histology images. Lancet Oncol. 2021;22(11):153–163. Doi:10.1016/S1470-2045(21)00430-5.
39. 39.Rajendran OK. DIGITAL TWIN FRAMEWORKS FOR PERSONALIZED CANCER PROGRESSION MODELING USING LONGITUDINAL DATA. Power System Protection and Control. 2025;53(4):486-501. Doi:10.46121/pspc.53.4.33.
40. 40.Rajendran LKK. Genomic profiling: utilizing Multi-omics data to identify potential Therapeutic targets and resistance markers. Power System Protection and Control. 2024;52(4):159-166. Doi:10.46121/pspc.52.4.13.
41. 41.Rajendran OK. Artificial Intelligence–Driven Multimodal Imaging for Cancer During Pregnancy: Advances in Maternal–Fetal Diagnostics and Precision Oncology. Int J Drug Deliv Technol. 2026;16(50s):862-870. Doi:10.25258/ijddt.16.50s.91.
42. 42.Rajendran LKK. Immunotherapy and cell Therapy: developing CAR-T cell therapies and Other immune-based treatments for cancer and Autoimmune diseases. Power System Protection and Control. 2023;51(2):64-77. Doi:10.46121/pspc.51.2.7.
43. 43.Rajendran OK. FOUNDATION MODEL–DRIVEN PRECISION ONCOLOGY: INTEGRATING MULTI-OMICS, RADIOLOGY, AND CLINICAL DATA FOR PREDICTIVE CANCER CARE. Power System Protection and Control. 2024;52(2):154-163. Doi:10.46121/pspc.52.2.14.
44. 44.Rajendran LKK. Theranostics: integrating Diagnostic imaging agents and therapeutic Drugs into a single multifunctional nano-Platform for real-time monitoring of treatment. Power System Protection and Control.2025;53(2):376-386. Doi:10.46121/pspc.53.2.31.
45. 45.Rajendran LKK. Mechanisms driving Immunotherapy resistance in colorectal cancer Liver metastases. Power System Protection and Control. 2024;52(1):29-37. Doi:10.46121/pspc.52.1.5.
46. 46.Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018;15(141):20170387. Doi:10.1098/rsif.2017.0387.
47. 47. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88. Doi:10.1016/j.media.2017.07.005.
48. 48.Hemanth Kumar RM. Integrated Transcriptomic and 3 Learning Framework Identifies a Blood-Based Biomarker Signature for Anthracycline-Induced Cardiotoxicity in Juvenile Cancer Survivors. Int J Drug Deliv Technol. 2026;16(40s):219-230.
49. Doi:10.25258/ijddt.16.40s.24.
50. 49.Mobadersany P, Yousefi S, Amgad M, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci USA. 2018;115(13):E2970–E2979. Doi:10.1073/pnas.1717139115.
51. 50. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749–762. Doi:10.1038/nrclinonc.2017.141.
52. 51.Azizi S, Mustafa B, Ryan F, et al. Big self-supervised models advance medical image classification. Nature. 2021;594(7864):104–110. Doi:10.1038/s41586-021-03476-6.
53. 52.Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale. arXiv. 2020. Doi:10.48550/arXiv.2010.11929.
54. 53.Rajendran OK. DeepDRA: A Deep Learning Framework for Drug Repurposing and Cancer Drug Response Prediction Using Multi-Omics Data. Scientific Culture. 2026;12(3):68-77. Doi:10.5281/zenodo.12326001.
55. 54.Xu H, Usuyama N, Bagga J, et al. A whole-slide foundation model for digital pathology from real-world data. Nature. 2024;630(8015):181–188. Doi:10.1038/s41586-024-07441-w.
56. 55.Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature.2023;620(7972):172–180. Doi:10.1038/s41586-023-06291-2.
57. 56.Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. Nature. 2023;616(7956):259–265. Doi:10.1038/s41586-023-05881-4.
58. 57.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.

59. 58. Dr. Isabella Moore, MultimodalArtificial Intelligence in Oncology: IntegratingRadiomics, Pathomics, and Genomics, Int. J. of Pharm.Sci., 2026, Vol 4, Issue 5, 6745-6760.https://doi.org/10.5281/zenodo.20391855
60. 59. Dr. Benjamin Walker, Dr. EleanorHayes, Dr. Christopher Nolan, Foundation Models inCancer Medicine: Revolutionizing Precision Diagnosticsand Clinical Oncology, Int. J. of Pharm. Sci., 2026, Vol4, Issue 5, 6733-6744.https://doi.org/10.5281/zenodo.20391588