British Journal of Hospital Medicine

Br. J. Hosp. Med. (Lond) **2026**; 87(7): 51300
https://doi.org/10.31083/BJHM51300

Review

The Use of Artificial Intelligence in Cancer Diagnosis

Arka Banerjee[1,†], Daphne Theresa Chia[2,†], Kai Yuen Wong[2,*]

[1]Department of Plastic Surgery, Guy's and St Thomas' NHS Foundation Trust, SE1 7EH London, UK
[2]Department of Plastic Surgery, Cambridge University Hospitals NHS Foundation Trust, CB2 0QQ Cambridge, UK
*Correspondence: kaiyuen.wong@cantab.net (Kai Yuen Wong)
[†]These authors contributed equally.

Academic Editor: Ganesh Radhakrishna
Submitted: 27 February 2026 Revised: 1 June 2026 Accepted: 30 June 2026 Published: 20 July 2026

Abstract

Artificial intelligence (AI) is increasingly influencing cancer diagnostics, with ongoing advancements into more refined and advanced applications. This review examines current applications of AI in cancer diagnosis. AI has demonstrated considerable success in screening and diagnosis of a range of cancer types, including breast, colorectal, and skin cancer, but challenges with specificity and false-positive rates remain. Within radiology and histopathology, AI shows promise in identifying cancers and augmenting clinician performance. Natural language processing and AI-driven clinical documentation tools can expedite diagnosis by reducing administrative burden and extracting clinically meaningful data from unstructured records. Additionally, AI models using unimodal and multimodal data can improve patient prognostication and risk stratification. Despite these advances, challenges relating to deployment, bias, interpretability, accountability and real-world efficacy persist. Prospective validation and careful implementation are required to ensure safe, equitable, and effective adoption of AI in cancer diagnostics.

Keywords: artificial intelligence; cancer diagnosis; machine learning; deep learning; large language models; natural language processing; prognostication

1. Introduction

Artificial intelligence (AI) describes a broad array of algorithms and systems that are designed to mimic human cognitive processes in performing tasks [1]. Within AI, machine learning (ML)—and its subset deep learning (DL)—are the approaches most relevant to clinical medicine (Fig. 1). ML refers to computational methods which enable systems to learn patterns from data to make predictions [2]. DL is a subfield of machine learning that uses artificial neural networks to model complex patterns from large databases [2]. Within clinical practice, DL is finding applications in drug development [3], health monitoring [4], disease diagnostics [5], medical treatment [6], and surgical treatment [7]. Large language models (LLMs) are a specialised category of DL models that are able to process and generate human language [8]. Examples include OpenAI's ChatGPT and Google's Gemini. LLMs have applications in expediting diagnosis through supporting automated clinical documentation, for example through AI-enabled ambient scribe technology, and by analysing large volumes of written clinical data and integrating multi-modal data [9], for example by analysing electronic health records to identify patients at risk of cancer. The purpose of this article is to explore the current state-of-the-art applications of AI in cancer diagnosis and provide readers with a concise overview of its emerging clinical applications. We start with a focus on AI's application in screening for cancer in asymptomatic individuals, before discussing its applications in diagnosis in radiological and histopathological investigations, and finally moving onto applications of LLMs in expediting diagnosis, and predicting patient outcomes.

2. Screening

There is increasing evidence and recognition for the role of artificial intelligence (AI) in screening asymptomatic individuals for cancers such as in the breast, colorectal, and skin, with many AI deep learning (DL) models achieving high sensitivity and specificity, even when compared to expert clinicians [10,11,12].

A recent Korean study showed how AI models can accurately categorise screening mammogram images into high, intermediate and minimal risk groups [10]. These systems, trained on large, labelled mammography datasets using DL techniques, detect suspicious image patterns and assign risk scores to each area of abnormality. In the study, a combination of two AI models was tested on a dataset of 3012 mammograms and compared with radiologists assessing the same images. The combined AI models achieved significantly higher sensitivity in detecting high-risk mammograms than radiologists ($p = 0.03$), but lower specificity ($p < 0.001$) (Table 1, Ref. [10,12,13,14]). These findings suggest the potential role of AI models in breast cancer screening; however, further research is needed to improve specificity.

AI models have also found applications in endoscopy for colon cancer screening—the Medtronic Gastrointestinal (GI) Genius is a US Food and Drug Administration (FDA)-

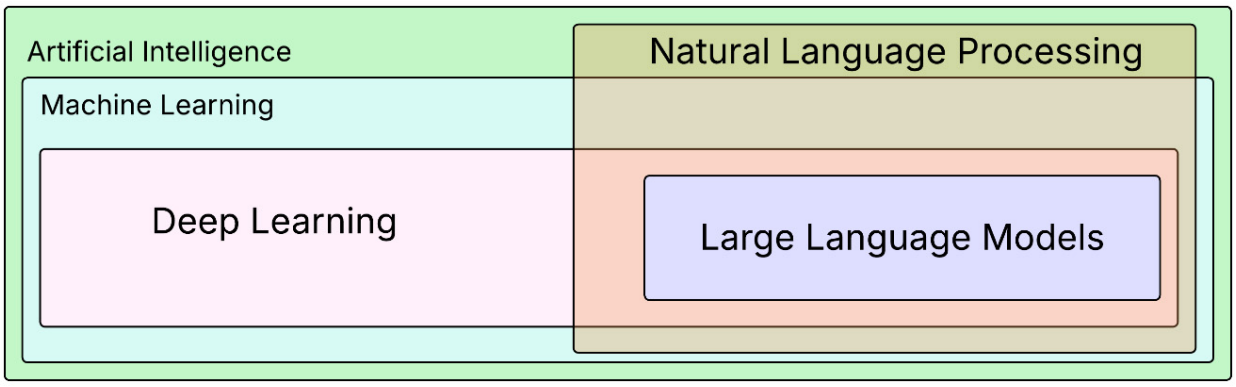

Fig. 1. Overview of artificial intelligence (AI) (relevant to this paper). The figure was created using Microsoft 365 (Microsoft Corporation, Redmond, WA, USA).

Table 1. Sensitivity and specificity figures for AI algorithms in screening for cancer.

Paper	Cancer type	Setting	Investigation	Sensitivity of AI algorithm	Specificity of AI algorithm	Sensitivity of expert control	Specificity of expert control
Kim et al, 2026 [10]	Ductal carcinoma Lobular carcinoma	Secondary Healthcare	Mammography	84.0%	81.3%	81.2%	87.4%
Ameri 2020 [12]	Basal cell carcinoma Actinic keratoses Squamous cell carcinoma *in situ* Melanoma	Secondary Healthcare	Dermoscopy	81.0%	88.0%	NA	NA
Jairath et al, 2024 [13]	Squamous cell carcinoma Melanoma	Secondary Healthcare	Dermoscopy & macroscopic clinical images	87.0%	91.0%	NA	NA
Salinas et al, 2024 [14]	Basal cell carcinoma Actinic keratoses Squamous cell carcinoma Melanoma	Mixed Healthcare (Primary & Secondary)	Dermoscopy & macroscopic clinical images	87.0%	77.1%	79.8%	73.6%

approved, real-time, computer-assisted detection (CADe) system which analyses live colonoscopy videos, highlighting malignant and pre-malignant colon polyps to the endoscopist [11]. A retrospective study by Kim et al. [11] compared adenoma detection rates at a large tertiary hospital during colonoscopy with and without CADe assistance. A total of 798 colonoscopies were reviewed (620 without CADe, 178 with CADe). The use of CADe was associated with a significant increase in adenoma detection rate, rising from approximately 40% to 52.5% (p = 0.05). Although total polyp detection rate increased from 73.8% to 84.3%, this was not statistically significant (p = 0.23). The improvement in adenoma detection rate was primarily driven by increased detection of small adenomas (1–5 mm) from 33.2% to 45.5% (p = 0.03), while there were no additional increases in detection rates of larger adenoma [11]. CADe systems may therefore play a valuable role in improving adenoma detection in colon cancer screening. Further research is necessary to better characterise and refine the models to improve performance in clinical settings. Several clinical trials are currently underway evaluating the efficacy of AI models in identifying colorectal polyps [15,16,17].

For skin cancer, a number of AI models have demonstrated high sensitivity and specificity in differentiating malignant from benign skin lesions (Table 1, [12,13]). A systematic review compared the performance of AI models to clinicians in diagnosing skin cancer from dermoscopic images alone. The review found that AI models achieved statistically significantly higher sensitivity and specificity compared to clinicians overall (Table 1) and generalists. The models achieved comparable performance to expert dermatologists [14]. Compared with colorectal can-

Table 2. Sensitivity and specificity figures for AI algorithms in diagnosing cancer from radiological and histopathological investigations.

Paper	Cancer type	Setting	Investigation	Sensitivity of AI algorithm	Specificity of AI algorithm	Sensitivity of expert control	Specificity of expert control
Kim et al, 2025 [20]	Lung	Secondary Healthcare	Chest radiographs	32.6%	91.0%	41.2%	81.3%
Arvidsson et al, 2024 [21]	Prostate	Secondary Healthcare	Prostate biopsy histopathology	96.0%	73.0%	NA	NA

NA, not applicable.

cer screening, AI models appear to show stronger performance in skin cancer image screening; however, concerns regarding false-negative results persist. Therefore, careful consideration is required when integrating these models into clinical practice, particularly with respect to the false-positive and false-negative rates.

In Asian populations, where there is a high prevalence of pulmonary subsolid nodules, the increased availability of low-dose computerised tomography (CT) screening has led to increased detection of pulmonary ground-glass nodules, raising concerns for overdiagnosis and overtreatment [18]. Whilst early surgical intervention may result in unnecessary treatment of otherwise indolent lesions, delayed intervention would adversely affect prognosis. AI can be used to address this challenge through predictive models which support clinician decision-making regarding optimal timing of intervention, helping clinicians to determine if a nodule warrants surgical intervention or can continue to be managed through active surveillance [18].

3. Radiological & Histopathological Investigations

The ability of artificial intelligence (AI) models to process and analyse large volumes of data makes them ideal for the analysis of radiological and histopathological images for cancer diagnosis, particularly in lung cancer and prostate cancer. Similar to the use case in screening of asymptomatic individuals, AI models have shown high specificity and sensitivity, rivalling that of specialised clinicians, in radiological and histopathological diagnosis of cancers.

Lung cancer is one of the most extensively studied malignancies in the context of the use of AI in diagnosis. This is attributable to its high global incidence and the reliance on radiological imaging in its diagnosis [19]. A recent large-scale study on a large population of 24,370 individuals showed that an AI model achieved higher specificity than experienced radiologists in detecting lung cancer from chest radiographs. However, the model demonstrated lower sensitivity, raising concern for an increased risk of missed diagnoses (Table 2, Ref. [20,21]). Risk difference analysis demonstrated that overall, AI models showed an improvement in sensitivity and specificity over radiologists.

AI models have demonstrated potential in prostate cancer diagnosis through analysis of prostate tissue biopsies. In a retrospective study of 4744 histopathological slides from 180 patients, an AI model achieved high sensitivity and specificity when compared with the original pathology reports as the reference standard. (Table 2) [21]. These results were supported by a study with a different AI model tested on 24,859 prostate biopsy histopathology slides. This AI model achieved an area under the curve (AUC) of 0.991. Interestingly, this particular model was also flexible enough to identify achieve AUCs of 0.989 and 0.965 on novel datasets such as basal cell carcinomas and breast metastases respectively [22]. AUC in this context refers to the probability that the AI model accurately ranks positive and negative instances from the test dataset.

These validation studies have supported the translation of select AI systems into regulated clinical use, primarily designed to augment, rather than replace, clinician interpretation. In 2021, the US Food and Drug Administration (FDA) approved the deployment of PaigeProstate (version: 1.3.0, Paige AI, New York, NY, USA), an AI-based software geared towards prostate cancer diagnosis, to assist pathologists by marking, measuring and grading areas of suspicion according to the Gleason score [23]. The addition of PaigeProstate was found to significantly increase the sensitivity of pathologists from 74% to 90%, with no significant change in specificity [24].

4. Expediting Diagnosis

4.1 Natural Language Processing: Improving Clinical Efficiency and Extracting Clinical Insights

Natural language processing (NLP) is a field of artificial intelligence focused on enabling computers to understand, interpret and generate human language. LLMs are DL models trained on very large collections of text, allowing the model to understand context, summarise texts and even generate clinical notes [25]. Within oncology, clinical documentation is extensive and multidisciplinary. Hence, NLP can enable:

1. Improvement in clinician efficiency through automation of documentation and scribing.

2. Extraction of clinical insight from unstructured data within electronic health records (EHRs), including clinical notes, pathology reports or radiology reports.

4.2 Improving Clinical Efficiency and Expediting Diagnosis With Automated Clinical Documentation

AI models use NLP and LLM technology to "listen" to patient-clinician consultations, transcribe conversations and draft clinical notes or letters. Such AI scribing tools aim to reduce documentation burden, improve accuracy, have the potential to free up clinicians for direct clinical care, and reduce administrative burden. Improvement in clinical efficiency may reduce referral wait-times and expedite time to cancer diagnosis.

A leading system being trialled in the United Kingdom is TORTUS, a National Health Service (NHS) England study across nine London sites, including hospitals and general practice (GP) practices. In the simulated setting, TORTUS has been shown to reduce consultation time by 26.3% through time savings on documentation, without reducing patient-clinician interaction time. Independent senior clinician ratings described 100% of TORTUS AI-produced notes as "good" or "very good" in quality, compared to 43% of standard, manually transcribed notes [26]. To estimate the wider economic and operational impact, the study team conducted modelling via the York Health Economics Consortium on the use of TORTUS in the accident and emergency (A&E) sector. Their analysis suggested scaling TORTUS across A&E departments in England could enable 9259 extra A&E consultations per day, USD 235 million saved in documentation time, and USD 880 million in additional capacity annually [27]. This demonstrates the potential cost and time savings in other medical specialties including oncology.

Beyond TORTUS, other AI scribes are gaining globally. For instance, Heidi Health's AI scribe is similarly used by clinicians for transcription, summarisation, documentation, and letter generation, and is, to date, used in 190 countries across the world [28]. The Dragon Ambient eXperience (DAX) Copilot has been shown via cohort studies [29] and non-randomised trials [30] to not only improve efficiency, but also improve clinician experience and potentially reduce burnout. Whilst these studies are not specific to oncology, it demonstrates the potential for AI-scribes to improve efficiency, reduce administrative workload, and deliver cost savings in oncology care.

4.3 Extracting Clinical Insights From Electronic Health Records to Expedite Cancer Diagnosis

An individual patient's EHR contains a vast amount of clinical information stored in unstructured free-text format. Manually extracting key clinical information and outcomes from unstructured text for direct patient care or research is time-consuming, expensive and error-prone. NLP systems can be used to automate and accelerate extraction and interpretation of information, including clinician notes, blood tests and pathology results [31] and imaging reports [32]. Converting unstructured language into structured clinical variables can help to expedite diagnosis, support earlier triage of patients with suspicious results, and facilitate large-scale research [32]. This is vital for big data research in cancer, such as identifying new biomarkers for diagnosis, improving outcome prediction and supporting precision oncology.

NLP systems have evolved significantly in the past decades and can be classified into three broad categories: (i) Rule-based systems, (ii) Classical machine learning, (iii) Large language models [32]. Table 3 (Ref. [33,34,35,36,37]) summarises how each category system works, providing examples.

LLMs are deep learning models building on previous NLP approaches through neural networks. This enables LLMs to learn contextual meaning within clinical notes or pathology reports, rather than rely on patterns or keywords. LLMs also offer improved scalability and generalisability across different institutions or countries, with diverse language patterns. For example, CancerLLM is a domain-specific LLM within oncology which was trained on 2.7 million clinical notes and 515,000 pathology reports across 17 cancer types. In research evaluation, the model was able to use cancer clinical notes and pathology reports to achieve high performance in phenotype extraction (F1 score 91.78%) and automated diagnosis generation (F1 score 86.81%), identifying features such as hormone receptor type, tumour size and site, cancer grade and histological type [37].

Deep learning models can also be trained for specific tasks within oncology. For example, a Bidirectional Encoder Representations from Transformers (BERT)-based algorithm was trained to identify incidental pulmonary nodules in CT reports, achieving a 98% sensitivity and 99% specificity [38], allowing radiology departments to automatically flag high-risk findings and prompt timely clinical review. Within breast cancer, a PubMedBERT-trained model was shown to achieve an accuracy of 97.4%, in extracting 32 important features of breast cancer phenotypes from pathology reports, outperforming previous rule-based systems [39]. These systems can enable clinicians to prioritise and analyse large volumes of reports more efficiently, with potential to shorten diagnostic timelines and enhance data-driven cancer care.

5. Predicting Patient Outcomes

Prognostication in cancer involves predicting patient outcomes such as survival, recurrence, and treatment response. Conventionally, clinicians relied on clinical, imaging and histological variables (e.g., Tumour-Node-Metastasis [TNM] staging) for prognostication. However, these approaches do not capture the full biological heterogeneity of cancer, in particular the complexity of genetic biomarkers. Advances in DL and genomic profiling now allow for more comprehensive models that integrate diverse data sources to identify subtle patterns and interactions, providing individualised prognostication to patients.

Table 3. Overview and comparison of categories of natural language processing used in cancer diagnostics.

Category	How they work	Example	Pros & Cons
Rule-based systems	Uses predefined rules, keywords or patterns to extract information.	Analysis of Transurethral Resection of a Bladder Tumour (TURBT) pathology reports using keywords such as "carcinoma *in situ*", "muscularis propria", "invading" to determine muscle invasion, a key differentiator when staging and risk stratifying bladder cancer [33].	Transparent and interpretable, but requires manual rule creation, and performs poorly when vocabulary or report formats change [34].
Classical machine learning	Learns statistical associations between words/phrases, and labels from large sets of annotated reports [35,36].	Systems trained to identify diagnoses in breast pathology reports (e.g., "invasive ductal carcinoma", "invasive lobular carcinoma") and to detect negation patterns like "no evidence of" [35].	Works well with abundant labelled data, but struggles with rare findings and ambiguous phrasing [34].
Large language models (LLMs)	Uses deep learning to learn contextual meaning (not just surface keywords) [34].	CancerLLM: oncology-specific LLM trained on millions of clinical notes and pathology reports; high F1 for phenotype extraction and automated diagnosis generation [37].	Scales with data, can generalise across phrasing and tasks. Can be fine-tuned for specific clinical tasks [34].

DL models can be trained on a single data modality (unimodal models) to predict clinical outcomes such as treatment response and prognosis. These approaches have demonstrated utility across multiple cancer types, including breast and ovarian cancer, and across diverse data sources such as genomic profiles and digitised histopathology (Table 4, Ref. [40,41,42]). Unimodal models offer several practical advantages, such as lower data requirements, reduced computational complexity, and easier integration into existing clinical pathways. Therefore, unimodal models may be more feasible in settings where genomic sequencing or multimodal data is unavailable. These models illustrate the potential of modality-specific DL systems to support risk stratification and therapeutic decision-making.

Although unimodal systems may perform well within specific clinical contexts, they may not comprehensively capture the full biological and clinical complexity of cancer. Increasingly, multimodal approaches that integrate different types of data from a single patient have demonstrated improved prognostic performance [43,44]. These methods combine various data types such as histopathology, imaging, molecular profiles, and clinical parameters. This may further enhance prognostication in cancer, hence aiding in patient counselling, individualized treatment planning, and more precise prognostication following diagnosis [45].

Several studies highlight the promise of multimodal deep learning for different individual cancer types such as in glioma [45], oral squamous cell carcinoma [46], and prostate cancer [47]. For example, in prostate cancer, the ArteraAI model combines biopsy pathology images with key clinical variables to generate a prognostic score that informs decisions regarding long-term androgen deprivation therapy. This model has been externally validated using retrospective analysis of datasets derived from multiple phase 3 randomized trials, demonstrating prognostic value beyond conventional predictors of metastatic risk in prostate cancer [47]. However, the ARTERA model has not yet undergone prospective real-world validation, therefore, its impact on clinical workflow and patient outcomes remains uncertain. Nevertheless, given the high quality of validation data and promising results, the model is recommended in the National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology (USA).

Multimodal deep learning has also been explored in models designed for use across multiple cancer types. Chen et al. [43] developed a multimodal fusion (MMF) algorithm that combined histological and molecular data to predict prognosis in 14 cancer types. In 12 of these, the multimodal model achieved higher concordance indices than the corresponding unimodal approaches, with multimodal modelling achieving an overall higher C-index (0.645) than molecular (0.585) or histopathology (0.607) data alone [43]. These findings suggest that integrating diverse data types may improve prognostic accuracy by capturing information that would not be available using a single modality. However, whether the incremental gains in prognostic accuracy justify the greater implementation complexity of multimodal systems remains an important consideration.

Table 4. Overview and comparison of unimodal models used in improving cancer prognostication.

Paper (Author, Year)	Cancer type	Modality	Summary of AI system used	Performance
Kim et al., 2013 [40]	Ovarian cancer	Unimodal (genomics)	The ATHENA model predicted ovarian cancer outcomes using genomic data, which included gene expression, methylation, and miRNA expression [40]	The ATHENA model achieved a balanced accuracy of 72.9% in a retrospective analysis, distinguishing patients with different survival outcomes [40]
Audeh et al, 2019 [41]	Breast cancer	Unimodel (genomics)	The MammaPrint model assesses patient's 10-year risk of distant recurrence, providing guidance on whether adjuvant chemotherapy can be safely omitted [41]	In the prospective randomised Microarray in Node-Negative Disease May Avoid Chemotherapy (MINDACT) trial, 46% of patients initially considered candidates for chemotherapy were categorized as low genomic risk by MammaPrint, with similar 5-year metastasis-free outcomes irrespective of chemotherapy use [41]
Gaury et al., 2025 [42]	Breast cancer	Unimodal (histopathology)	The RlapsRisk model analyses digital histopathology slides to in patients with early-stage, ER-positive, human epidermal growth factor receptor 2 (HER2)-negative, breast cancer [42]	Compared to traditional clinicopathological variables, the RlapsRisk model achieved a higher C-index* (0.81 vs 0.76, $p < 0.05$) in predicting distant recurrence-free interval [42]

*C-index: Measures the model's ability to correctly rank patients according to their risk of experiencing an event (e.g., disease progression or death). C-index of 0.5 represents complete chance, 1.0 indicates perfect risk discrimination.

Compared to unimodal approaches, multimodal systems require greater computational resources, robust infrastructure, and effective integration of clinical, imaging, histopathological and genomic datasets. Consequently, unimodal models may be more practical and accessible, especially in settings where genomic sequencing or multimodal data collection may be unavailable. Hence, the choice between unimodal or multimodal approaches should depend not only on predictive performance, but also considerations such as service availability, infrastructure and the characteristics of the cancer type [43].

6. Limitations and Barriers to Clinical Implementation

Despite substantial advances in AI models for cancer diagnosis, significant limitations remain and should be addressed before widespread clinical adoption. These limitations may be grouped into three broad domains: technological, data- and model- related, and application-related.

6.1 Technological Limitations

The implementation of AI systems in clinical practice requires significant technical infrastructure, such as high-quality microphones, compatible computing hardware, and secure data storage. In clinical settings, background noise, multiple speakers, and network interruptions can reduce transcription accuracy and downstream model performance. In addition, deployment of AI systems often involves direct financial costs, staff training, and ongoing technical support, which may limit feasibility in resource-constrained settings.

6.2 Data- and Model-Related Limitations

Bias and limited generalisability remain significant concerns for AI models trained on data from single institutions, regions, or populations. Such models may perform poorly when applied to other settings with different demographics and diverse patient groups. Performance of AI models may further deteriorate when they are used in a setting with different imaging protocols, electronic health records and clinical workflows from those represented in the training dataset. Systematic biases in training data can reduce predictive accuracy for minority ethnic groups, under-represented tumour types, or patients with atypical disease presentations. This can potentially exacerbate existing health inequities. Generalisability of models is especially important in the diagnosis of cancer, as even small reductions in diagnostic performance may lead to missed cancers or delayed diagnosis. Hence, AI models require robust external validation across a diverse range of real-world healthcare settings prior to clinical implementation.

A further limitation arises from the limited transparency and interpretability of deep learning AI models.

The “black box” nature of these systems makes it difficult for both clinicians and patients to understand how predictions are generated, reducing explainability. A lack of explainability may reduce clinical trust in AI-generated outputs, and hence impede adoption into clinical practice.

Existing evidence is also limited by the heterogeneity between studies in study design, datasets, and validation methods. This variability limits comparison between studies and hinders our ability to draw robust conclusions on the effectiveness and impact of AI models in the clinical setting. Over time, the performance of existing AI models may deteriorate due to ongoing advances in clinical practice and imaging technology, as well as shifts in patient demographics and disease prevalence. These temporal changes necessitate ongoing monitoring and fine-tuning of AI models even after deployment into the clinical setting.

6.3 Application-Related Limitations

Although many AI systems demonstrate high technical performance metrics, even small error rates can have meaningful clinical consequences in cancer diagnosis and prognostication. A key challenge is determining the appropriate confidence level at which an AI system should flag suspected malignancy or high-risk disease. Lower thresholds may reduce false negatives but are likely to increase false positives, potentially resulting in over-investigation and overtreatment, a phenomenon previously observed following the expansion of teledermatology [48]. These trade-offs highlight the importance of model calibration and real-world validation, as incorrect predictions may influence clinical decision-making and adversely affect patient outcomes.

Another major application-related concern is accountability. It remains unclear who bears responsibility when an AI system contributes to an incorrect diagnosis or management decision. While AI tools may inform clinical judgement, responsibility currently rests with clinicians, necessitating ongoing human oversight. This requirement may limit automation and complicate clinical workflows. Regulatory and ethical considerations surrounding data privacy, governance, and consent also need to be addressed.

Most current evidence supporting the use of AI in cancer diagnosis is derived from retrospective studies. Although such studies demonstrate technical feasibility, retrospective validation does not adequately reflect real-world clinical workflows, decision-making processes, or patient outcomes. Further clinical validation of AI models, including well-designed prospective and randomised trials, is therefore required to determine how AI systems influence diagnosis, clinical management and ultimately patient outcomes in cancer care.

7. Conclusion

Artificial intelligence has demonstrated utility across the cancer diagnosis pathway, from screening and image interpretation to prognostication. Advances in NLP and LLM have also expanded the role of AI, enabling automated documentation and extraction of clinical insight from electronic health records. Current evidence shows that AI models have the potential to achieve a high level of diagnostic performance whilst offering opportunities to improve efficiency and streamline clinical workflows.

Future research should therefore prioritise the development and validation of AI models in real-world clinical settings. This includes the use of diverse, multicentre datasets with external validation across different populations and healthcare settings. Prospective studies should be designed to evaluate the impact of AI models in clinical decision-making, healthcare efficiency, and patient outcomes. Future research should also address technical challenges associated with multimodal AI models, such as the integration of heterogeneous data sources, management of missing data, and interoperability between different electronic platforms. Practical challenges relating to infrastructure, accountability and regulatory oversight must also be addressed to foster clinician and patient trust whilst ensuring safe implementation. Only through addressing these challenges can the full potential of AI in cancer diagnosis be realised. The clinical value of AI in cancer diagnosis will ultimately be determined not only by its technical performance, but also by its ability to improve time to diagnosis, healthcare efficiency, and most importantly, patient outcomes.

Key Points

• Artificial Intelligence, including machine learning, deep learning and large language models, has shown promising advancements in the cancer diagnostic pathway by improving screening, diagnostic accuracy, efficiency and prognostication.

• Artificial intelligence-assisted screening systems for various types of cancers such as breast, colorectal and skin cancer have demonstrated high sensitivity and specificity comparable to expert clinicians.

• In radiology imaging and histopathology investigations, artificial intelligence models have shown strong performance in detecting and characterising cancers.

• Natural language processing and large language models can expedite diagnosis and improve clinical efficiency by automating documentation and extracting clinically relevant information from unstructured records.

• Deep learning models are increasingly being used to predict outcomes by integrating genomic, histopathology, imaging and clinical data, hence supporting personalised risk stratification and treatment decisions.

• Implementation of artificial intelligence in cancer diagnostics remains constrained by several important challenges, including data bias, limited generalisability, lack of model transparency, regulatory and ethical concerns, and the need for prospective clinical validation studies.

Availability of Data and Materials

Not applicable.

Author Contributions

AB, DTC and KYW designed the work and drafted the manuscript. All authors contributed to the important editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflicts of Interest

Given his role as an Editorial Board member, Kai Yuen Wong had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to Ganesh Radhakrishna. Other authors declare no conflicts of interest.

References

[1] Tiwari A, Mishra S, Kuo TR. Current AI technologies in cancer diagnostics and treatment. Molecular Cancer. 2025; 24: 159. https://doi.org/10.1186/s12943-025-02369-9

[2] Taulli T. Artificial Intelligence Basics: A Non-Technical Introduction. Apress: Berkeley, CA. 2019. https://doi.org/10.1007/978-1-4842-5028-0

[3] Dasí-Delgado P, Andreu C, Del Olmo ML. Strategies Used for the Discovery of New Microbial Metabolites with Antibiotic Activity. Molecules (Basel, Switzerland). 2025; 30: 2868. https://doi.org/10.3390/molecules30132868

[4] Sands DZ, Finn NB. From Internet to Artificial Intelligence (Al) Bots: Symbiotic Evolutions of Digital Technologies and e-Patients. Journal of Participatory Medicine. 2025; 17: e68911. https://doi.org/10.2196/68911

[5] Jia R, Liu B, Ali M. Establishing an AI-based diagnostic framework for pulmonary nodules in computed tomography. BMC Pulmonary Medicine. 2025; 25: 339. https://doi.org/10.1186/s12890-025-03806-7

[6] Bongurala AR, Save D, Virmani A. Progressive role of artificial intelligence in treatment decision-making in the field of medical oncology. Frontiers in Medicine. 2025; 12: 1533910. https://doi.org/10.3389/fmed.2025.1533910

[7] Kok Wah JN. AI-driven robotic surgery in oncology: advancing precision, personalization, and patient outcomes. Journal of Robotic Surgery. 2025; 19: 382. https://doi.org/10.1007/s11701-025-02555-3

[8] Stryker C. What are large language models (LLMs)? 2025. Available at: https://www.ibm.com/think/topics/large-language-models (Accessed: 14 July 2025).

[9] Ghorbian M, Ghobaei-Arani M, Ghorbian S. Transforming breast cancer diagnosis and treatment with large language Models: A comprehensive survey. Methods (San Diego, Calif.). 2025; 239: 85–110. https://doi.org/10.1016/j.ymeth.2025.04.001

[10] Kim HJ, Kim HH, Eom HJ, Choi WJ, Chae EY, Shin HJ, et al. Combined use of two artificial intelligence-based algorithms for mammography triaging: a retrospective simulation study. European Radiology. 2026; 36: 808–820. https://doi.org/10.1007/s00330-025-11849-x

[11] Kim JH, Wang J, Pence C, Magahis P, Dawod E, Schnoll-Sussman F, et al. GI Genius increases small and right-sided adenoma and sessile serrated lesion detection rate when used with EndoCuff in a real-world setting: a retrospective United States study. Clinical Endoscopy. 2025; 58: 438–447. https://doi.org/10.5946/ce.2024.271

[12] Ameri A. A Deep Learning Approach to Skin Cancer Detection in Dermoscopy Images. Journal of Biomedical Physics & Engineering. 2020; 10: 801–806. https://doi.org/10.31661/jbpe.v0i0.2004-1107

[13] Jairath N, Pahalyants V, Shah R, Weed J, Carucci JA, Criscito MC. Artificial Intelligence in Dermatology: A Systematic Review of Its Applications in Melanoma and Keratinocyte Carcinoma Diagnosis. Dermatologic Surgery. 2024; 50: 791–798. https://doi.org/10.1097/DSS.0000000000004223

[14] Salinas MP, Sepúlveda J, Hidalgo L, Peirano D, Morel M, Uribe P, et al. A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. NPJ Digital Medicine. 2024; 7: 125. https://doi.org/10.1038/s41746-024-01103-x

[15] Petz Aladar County Teaching Hospital. Polyp Artificial Intelligence Study. 2020. Available at: https://clinicaltrials.gov/study/NCT04425941. (Accessed: 19 February 2026).

[16] Neumann H. AI for Colorectal Polyp Detection in Endoscopy. 2020. Available at: https://clinicaltrials.gov/study/NCT04339855. (Accessed: 19 February 2026).

[17] Haji A. EndoBRAIN International Version 11. 2026. Available at: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/endobrain-international-version-11/ (Accessed: 19 February 2026).

[18] Wu FZ, Chou MC, Wu YJ, Tang EK. The Evolving Paradigm in Managing Pulmonary Ground-glass Nodules: From Aggressive Intervention to Precision Surveillance. Journal of Radiological Science. 2026; 51: 1–6. https://doi.org/10.4103/jradiolsci.JRADIOLSCI-D-25-00033

[19] Gillies RJ, Schabath MB. Radiomics Improves Cancer Screening and Early Detection. Cancer Epidemiology, Biomarkers & Prevention. 2020; 29: 2556–2567. https://doi.org/10.1158/1055-9965.EPI-20-0075

[20] Kim T, Shin H, Song YS, Lee JH, Kim H, Shin D. Artificial intelligence software for detecting unsuspected lung cancer on chest radiographs in an asymptomatic population. Radiology Advances. 2025; 2: umae032. https://doi.org/10.1093/radadv/umae032

[21] Arvidsson I, Svanemur E, Marginean F, Simoulis A, Overgaard NC, Åström K, et al. Artificial intelligence for detection of prostate cancer in biopsies during active surveillance. BJU International. 2024; 134: 1001–1009. https://doi.org/10.1111/bju.16456

[22] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine. 2019; 25: 1301–1309. https://doi.org/10.1038/s41591-019-0508-1

[23] Rabilloud N, Allaume P, Acosta O, De Crevoisier R, Bourgade R, Loussouarn D, et al. Deep Learning Methodologies Applied to Digital Pathology in Prostate Cancer: A Systematic Review.

Diagnostics (Basel, Switzerland). 2023; 13: 2676. https://doi.org/10.3390/diagnostics13162676

[24] Raciti P, Sue J, Ceballos R, Godrich R, Kunz JD, Kapur S, et al. Novel artificial intelligence system increases the detection of prostate cancer in whole slide images of core needle biopsies. Modern Pathology. 2020; 33: 2058–2066. https://doi.org/10.1038/s41379-020-0551-y

[25] Chang Y, Wang X, Wang J, Wu Y, Yang L, Zhu K, et al. A Survey on Evaluation of Large Language Models. ACM Transactions on Intelligent Systems and Technology. 2024; 15: 1–45. https://doi.org/10.1145/3641289

[26] Balloch J, Sridharan S, Oldham G, Wray J, Gough P, Robinson R, et al. Use of an ambient artificial intelligence tool to improve quality of clinical documentation. Future Healthcare Journal. 2024; 11: 100157. https://doi.org/10.1016/j.fhj.2024.100157

[27] Great Ormond Street Hospital for Children. GOSH-led trial of AI-scribe technology shows 'transformative' benefits for patients and clinicians across London. 2025. Available at: https://www.gosh.nhs.uk/news/researchgosh-led-trial-of-ai-scribe-technology-shows-transformative-benefits-for-patients-and-clinicians-across-london/ (Accessed: 26 February 2026).

[28] Heidi. Heidi Media Room. 2026. Available at: https://www.heidihealth.com/en-us/about-us/media (Accessed: 26 February 2026).

[29] Haberle T, Cleveland C, Snow GL, Barber C, Stookey N, Thornock C, et al. The impact of nuance DAX ambient listening AI documentation: a cohort study. Journal of the American Medical Informatics Association: JAMIA. 2024; 31: 975–979. https://doi.org/10.1093/jamia/ocae022

[30] Liu TL, Hetherington TC, Stephens C, McWilliams A, Dharod A, Carroll T, et al. AI-Powered Clinical Documentation and Clinicians' Electronic Health Record Experience: A Nonrandomized Clinical Trial. JAMA Network Open. 2024; 7: e2432460. https://doi.org/10.1001/jamanetworkopen.2024.32460

[31] Iqbal MJ, Javed Z, Sadia H, Qureshi IA, Irshad A, Ahmed R, et al. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. Cancer Cell International. 2021; 21: 270. https://doi.org/10.1186/s12935-021-01981-1

[32] Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. Journal of Biomedical Informatics. 2017; 73: 14–29. https://doi.org/10.1016/j.jbi.2017.07.012

[33] Glaser AP, Jordan BJ, Cohen J, Desai A, Silberman P, Meeks JJ. Automated Extraction of Grade, Stage, and Quality Information From Transurethral Resection of Bladder Tumor Pathology Reports Using Natural Language Processing. JCO Clinical Cancer Informatics. 2018; 2: 1–8. https://doi.org/10.1200/CCI.17.00128

[34] Santos T, Tariq A, Gichoya JW, Trivedi H, Banerjee I. Automatic Classification of Cancer Pathology Reports: A Systematic Review. Journal of Pathology Informatics. 2022; 13: 100003. https://doi.org/10.1016/j.jpi.2022.100003

[35] Buckley JM, Coopey SB, Sharko J, Polubriaginof F, Drohan B, Belli AK, et al. The feasibility of using natural language processing to extract clinical information from breast pathology reports. Journal of Pathology Informatics. 2012; 3: 23. https://doi.org/10.4103/2153-3539.97788

[36] Wieneke AE, Bowles EJA, Cronkite D, Wernli KJ, Gao H, Carrell D, et al. Validation of natural language processing to extract breast cancer pathology procedures and results. Journal of Pathology Informatics. 2015; 6: 38. https://doi.org/10.4103/2153-3539.159215

[37] Li M, Zhan Z, Huang J, Yeung J, Ding K, Blaes A, et al. CancerLLM: a large language model in cancer domain. NPJ Digital Medicine. 2026; 9: 266. https://doi.org/10.1038/s41746-026-02441-8

[38] Grolleau E, Couraud S, Jupin Delevaux E, Piegay C, Mansuy A, de Bermont J, et al. Incidental pulmonary nodules: Natural language processing analysis of radiology reports. Respiratory Medicine and Research. 2024; 86: 101136. https://doi.org/10.1016/j.resmer.2024.101136

[39] Kwok COT, Arbour G, Zhang A, Hiemstra J, Nichol AM, Ng R, et al. Development of a machine learning model for automatic data extraction from breast cancer pathology reports. Scientific Reports. 2025; 15: 39855. https://doi.org/10.1038/s41598-025-23461-6

[40] Kim D, Li R, Dudek SM, Ritchie MD. ATHENA: Identifying interactions between different levels of genomic data associated with cancer clinical outcomes using grammatical evolution neural network. BioData Mining. 2013; 6: 23. https://doi.org/10.1186/1756-0381-6-23

[41] Audeh W, Blumencranz L, Kling H, Trivedi H, Srkalovic G. Prospective Validation of a Genomic Assay in Breast Cancer: The 70-gene MammaPrint Assay and the MINDACT Trial. Acta Medica Academica. 2019; 48: 18–34. https://doi.org/10.5644/ama2006-124.239

[42] Gaury V, Vaquette Z, Aubert V, Barcenas C, Jones LJ, Jacobs D, et al. Clinical Validation of RlapsRisk BC in an international multi-cohorts setting. medRxiv. 2025. https://doi.org/10.1101/2025.07.18.25331788 (preprint)

[43] Chen RJ, Lu MY, Williamson DFK, Chen TY, Lipkova J, Noor Z, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell. 2022; 40: 865–878.e6. https://doi.org/10.1016/j.ccell.2022.07.004

[44] Unger M, Kather JN. Deep learning in cancer genomics and histopathology. Genome Medicine. 2024; 16: 44. https://doi.org/10.1186/s13073-024-01315-6

[45] Alleman K, Knecht E, Huang J, Zhang L, Lam S, DeCuypere M. Multimodal Deep Learning-Based Prognostication in Glioma Patients: A Systematic Review. Cancers. 2023; 15: 545. https://doi.org/10.3390/cancers15020545

[46] Alabi RO, Bello IO, Youssef O, Elmusrati M, Mäkitie AA, Almangush A. Utilizing Deep Machine Learning for Prognostication of Oral Squamous Cell Carcinoma-A Systematic Review. Frontiers in Oral Health. 2021; 2: 686863. https://doi.org/10.3389/froh.2021.686863

[47] Parker CTA, Mendes L, Liu VYT, Grist E, Joun S, Yamashita R, et al. External validation of a digital pathology-based multimodal artificial intelligence-derived prognostic model in patients with advanced prostate cancer starting long-term androgen deprivation therapy: a post-hoc ancillary biomarker study of four phase 3 randomised controlled trials of the STAMPEDE platform protocol. The Lancet. Digital Health. 2025; 7: 100885. https://doi.org/10.1016/j.landig.2025.100885

[48] Chuchu N, Dinnes J, Takwoingi Y, Matin RN, Bayliss SE, Davenport C, et al. Teledermatology for diagnosing skin cancer in adults. The Cochrane Database of Systematic Reviews. 2018; 12: CD013193. https://doi.org/10.1002/14651858.CD013193