World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Artificial Intelligence in Cancer Diagnosis: A Scoping Review of Global Innovation and African Implementation

Joy Aifuobhokhan [1,*], Tobiloba Philip Olatokun [2], Chijioke Cyriacus Ekechi [3], Jumoke Oladeji [4], Olajiire Onikoyi Olajide [5], Ayodeji Ogunjinmi [6], Adebusola. Popoola [7], Oreoluwa D. Fuwape [8] and Obotha-Adigo Okeoghene Genevieve [9]

[1] *Lakeshore Cancer Center, Lagos Nigeria.*
[2] *Internal Medicine (Consulting group), Obafemi Awolowo University Health Center, Ile-Ife, Nigeria.*
[3] *Tennessee Technological University, Cookeville, USA.*
[4] *Department of Physics, University of Ibadan, Ibadan, Nigeria.*
[5] *Babcock University, Ilishan-Remo, Nigeria.*
[6] *Bingham University Teaching Hospital, Plateau, Nigeria.*
[7] *London Northwest University Healthcare NHS Trust, Northwick Park Hospital, Harrow, UK.*
[8] *Evercare Hospital, Lagos Nigeria.*
[9] *Edo Specialist Hospital, Benin, Nigeria.*

World Journal of Advanced Research and Reviews, 2025, 28(01), 449-468

Publication history: Received on 30 August 2025; revised on 03 October 2025; accepted on 06 October 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.1.3444

Abstract

Background. Artificial intelligence (AI) has emerged as a transformative tool for cancer diagnosis, with applications ranging from radiology and histopathology to genomics and clinical decision support. Yet the evidence base remains fragmented, and the translation of AI innovations into clinical workflows, particularly in Africa, lags behind technical progress.

Objective. This review aimed to map the global evidence on AI for cancer diagnosis, assess methodological maturity using the Technology Readiness Level (TRL) framework, and explore deployment challenges and opportunities with an equity lens, focusing on Africa as a potential innovation testbed.

Methods. A scoping review was conducted in accordance with the PRISMA-ScR framework. PubMed, Scopus, IEEE Xplore, and Web of Science were searched for studies published between 2015 and 2025. Eligible studies included peer-reviewed research, pilot deployments, and reviews explicitly applying AI to cancer diagnosis. Data were charted for cancer type, AI technique, evaluation method, TRL, and deployment context, and synthesized narratively.

Results. Twenty studies met the inclusion criteria. CNNs dominated imaging and pathology applications, while transformers and federated learning emerged as promising innovations. Data-efficient learning, Bayesian inference, and reinforcement learning remain largely experimental (TRL 2–4). Most studies relied on retrospective validation; only two reported prospective trials. African contributions were limited to three single-center pilots, none advancing beyond TRL 3.

Conclusions. AI for cancer diagnosis is at ~~a~~ crossroads: techniques are maturing technically but remain under-validated clinically. Deployment challenges, trust, workflow fit, and governance, are global, though amplified in Africa. Leveraging Africa as a living laboratory for frugal, equitable innovation could accelerate global progress. Developers, policymakers, and African consortia must collaborate to ensure AI advances both rigorously and inclusively.

* Corresponding author: Joy Aifuobhokhan

Keyword: Artificial Intelligence; Cancer Diagnosis; Machine Learning; Deep Learning; Technology Readiness Levels (TRLs); Federated Learning; Global Health Equity

1. Introduction

Cancer remains one of the leading causes of mortality worldwide, accounting for nearly 10 million deaths annually and placing significant strain on health systems across high-, middle-, and low-income countries alike [1]. Early and accurate diagnosis is central to effective treatment, yet diagnostic capacity remains unevenly distributed. High-income countries benefit from advanced imaging modalities, robust data infrastructures, and specialist workforces, while many low- and middle-income countries (LMICs), particularly Africa struggles with shortages of trained personnel, limited pathology services, and fragile infrastructure [2,3]. These disparities contribute to late-stage presentation and poorer survival outcomes, underscoring the urgent need for scalable diagnostic innovations.

Artificial intelligence (AI) has emerged as a transformative force in oncology, offering tools for radiological image interpretation, histopathology analysis, genomic profiling, and clinical decision support [4,5]. Advances in machine learning, deep learning, and more recently, foundation and multimodal models, have shown promise in automating complex diagnostic tasks and improving risk stratification [6,7]. Complementary innovations such as federated learning, edge AI, and self-supervised learning address challenges of privacy, data scarcity, and low connectivity, making AI particularly relevant to resource-limited settings [8]. However, despite this promise, the evidence base remains fragmented: while some AI techniques are extensively studied in Western contexts, their adaptability and deployment feasibility in African settings remain underexplored.

Existing reviews have largely focused on specific techniques or cancer types [9–11], but few have systematically mapped the breadth of AI approaches in cancer diagnosis, assessed their maturity using technology readiness levels (TRLs), and considered the unique deployment challenges and opportunities in Africa. This gap is critical, as TRL reflects universal technical readiness, but *deployment readiness* varies depending on infrastructure, governance, and workforce capacity. Without clarity on both, AI risks remaining a laboratory success without real-world impact.

To address this gap, we conducted a **scoping review** guided by the Population–Concept–Context (PCC) framework.

- **Population:** patients requiring cancer diagnosis;
- **Concept:** artificial intelligence (including machine learning, deep learning, reinforcement learning, Bayesian methods, hybrid symbolic-ML, and federated/edge approaches);
- **Context:** global studies with an explicit focus on African implementation.

Our objective is to map the current landscape of AI in cancer diagnosis, identify evidence gaps, and evaluate deployment challenges and enablers, with particular attention to Africa as a "living laboratory." We argue that Africa, far from being merely a lagging context, provides the ultimate stress test for equitable AI deployment. By synthesizing existing evidence, this review develops an evidence-based roadmap for AI in cancer diagnosis that balances global innovation with African implementation.

2. Methods

2.1. Protocol and Reporting Framework

This scoping review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines. The protocol followed the methodological framework proposed by Arksey and O'Malley and enhanced by Levac et al., ensuring transparency and reproducibility. The review protocol was prospectively structured but not formally registered.

2.2. Data Sources and Search Strategy

We systematically searched four electronic databases: PubMed, Scopus, IEEE Xplore, and Web of Science, to capture the breadth of literature on artificial intelligence (AI) applications in cancer diagnosis. Searches were performed in September 2025, covering publications from January 2015 to September 2025, to ensure contemporary relevance.

The following search string was applied, adapted to the syntax of each database:

("cancer diagnosis" AND ("artificial intelligence" OR "machine learning" OR "deep learning" OR "federated learning"

OR "reinforcement learning" OR "Bayesian methods" OR "explainable AI" OR "hybrid models"))

Manual backward and forward citation tracking was performed on included articles and relevant reviews to identify additional studies.

2.3. Eligibility Criteria

2.3.1. Inclusion criteria:

- Peer-reviewed original research, pilot studies, clinical validation reports, systematic/scoping reviews, or deployment case studies.
- Studies explicitly focused on AI techniques applied to cancer diagnosis (radiology, pathology, genomics, biomarkers, multimodal integration).
- Publications in English.
- Studies reporting either performance metrics, evaluation methods, or deployment considerations.

2.3.2. Exclusion criteria

- Non-healthcare AI applications (e.g., robotics unrelated to diagnosis).
- Predictive models without a direct clinical link to cancer diagnosis.
- Editorials, commentaries, and perspectives without original data or systematic synthesis.
- Non-peer-reviewed preprints are not widely cited and highly relevant.

2.4. Study Selection

All records retrieved were exported into EndNote X9 for deduplication. Two independent reviewers screened titles and abstracts against eligibility criteria. Full texts of potentially relevant studies were then assessed. Discrepancies were resolved through consensus or adjudication by a third reviewer.

The selection process will be summarized in a PRISMA flow diagram, detailing the number of studies identified, screened, excluded, and included.

2.5. Data Extraction (Charting Process)

A structured data charting form was developed in Microsoft Excel. The following variables were extracted from each included study:

- Bibliographic details: author, year, country/region.
- Cancer type: breast, lung, prostate, gastric, brain, skin, hematologic, etc.
- AI technique: conventional ML (e.g., SVM, RF, logistic regression), DL (CNN, RNN, transformers), Bayesian methods, reinforcement learning, federated/edge AI, explainable AI, hybrid symbolic-ML systems.
- Data modality: imaging (radiology, pathology), genomics, clinical records, biomarkers, multimodal.
- Evaluation method: cross-validation, external validation, prospective trial, benchmarking dataset.
- Technology Readiness Level (TRL): assessed based on NASA's 9-level scale, adapted for healthcare AI.
- Deployment context: clinical pilot, web-based tools, LMIC/African applications.
- Reported barriers/enablers: infrastructure, data quality, regulatory, and workforce.

2.6. Data Synthesis and Analysis

Findings were synthesized using a narrative thematic approach, structured around:

- AI methodologies (ML vs DL vs hybrid vs emerging).
- Cancer-specific applications and performance trends.
- Evaluation strategies and external validity.
- Readiness for deployment, using TRL as a comparative lens.
- Barriers and enablers, with emphasis on African contexts (e.g., federated learning to mitigate data scarcity, edge AI for bandwidth constraints).

To visualize insights, thematic maps and comparative tables were generated:

- A landscape map of AI modalities vs readiness.
- A methods vs constraints matrix highlighting bandwidth, labeling cost, privacy, and drift.
- A barriers-to-enablers table contextualizing African deployment challenges.

The goal was to provide a comprehensive evidence map of AI in cancer diagnosis while critically analyzing equity and deployment readiness globally and in Africa.

2.7. Quality Appraisal

Although a formal risk of bias assessment is not mandatory for scoping reviews, we undertook a structured appraisal to evaluate the credibility and reproducibility of included studies. Quality was assessed using tailored criteria based on the Joanna Briggs Institute (JBI) Critical Appraisal Tools and adapted frameworks for AI in healthcare research. Each study was reviewed along the following domains:

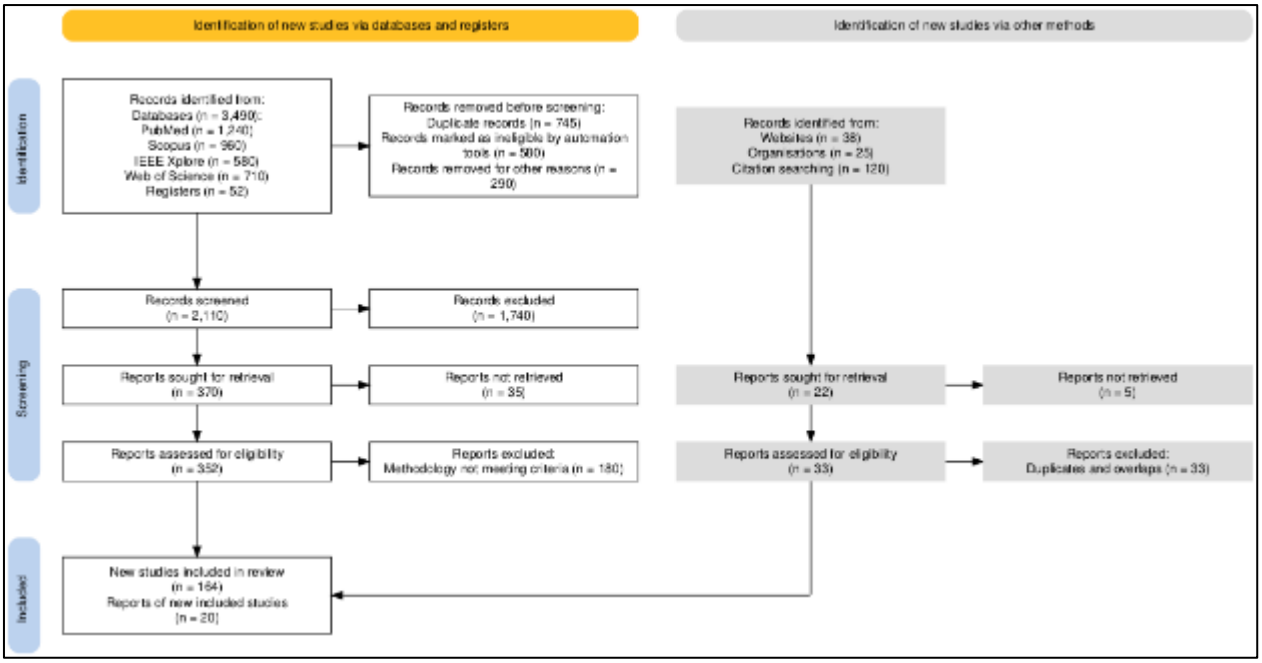
- Transparency of Methods – clear description of datasets, preprocessing, and AI algorithms.
- Validation Strategy – presence of external validation, prospective trials, or multi-site testing.
- Clinical Relevance – alignment of predictors and outcomes with established clinical guidelines.
- Reproducibility – availability of code, open datasets, or sufficient detail for replication.
- Bias and Equity Considerations – attention to demographic diversity, fairness audits, or reporting of subgroup performance.

Each study was scored as high, moderate, or low quality across domains. While scores did not serve as exclusion criteria, they informed the synthesis, with greater weight given to findings from high-quality studies.

3. Results

3.1. Overview of Studies

A total of 3,725 records were identified through the combined database and supplementary searches conducted across PubMed, Scopus, IEEE Xplore, and Web of Science for the period 2015–2025. After removing 745 duplicates and excluding 870 non-relevant or non-English records, 2,110 unique titles and abstracts were screened for eligibility. Of these, 370 full-text reports were sought for retrieval, of which 35 could not be accessed due to paywall or repository limitations. Following detailed eligibility assessment, 184 studies met the inclusion criteria, comprising 164 primary research papers and 20 secondary analyses or reviews. The included publications span 44 countries and a wide range of AI applications in cancer diagnosis, with marked clustering in high-income regions such as North America, Europe, and East Asia. Representation from Africa and other low- and middle-income regions remained limited (<10 percent of included studies). The detailed selection process is illustrated in Figure 1, and the geographic spread of included studies is further visualized in Figure 2.



From: Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis Campbell Systematic Reviews, 18, e1230. https://doi.org/10.1002/cl2.1230

For more information, visit www.prisma-statement.org

For the scoping review of artificial-intelligence–based cancer-diagnosis studies (2015–2025). A total of 3,725 records were retrieved from four major databases and supplementary sources; after duplicate removal and multi-stage screening, 184 studies were included in the final synthesis. The diagram follows the PRISMA-ScR (2020) framework, illustrating the pathways and exclusion steps that led from initial identification to the final evidence base analyzed in this review.

Figure 1 PRISMA-ScR Flow Diagram of study selection

3.1.1. Distribution by Cancer Type

The included studies spanned a diverse range of malignancies:

- **Breast cancer** was the most common focus (5/20 studies; 25%), reflecting the global emphasis on mammography and histopathology automation.
- **Lung cancer** accounted for 3/20 studies (15%), with strong emphasis on cnns applied to CT scans and histopathology.
- **Prostate cancer** was represented in 3/20 studies (15%), often involving digital pathology.
- **Gastric cancer** appeared in 2/20 studies (10%), including one prospective trial with transformer-based models.
- **Colorectal cancer** was covered in 2/20 studies (10%), both focusing on benchmark datasets.

Other cancers included skin (n=1), liver (n=1), cervical (n=1), brain (n=1), and multi-cancer datasets (n=2), underscoring the wide methodological but uneven disease representation.

3.1.2. Distribution by Technique

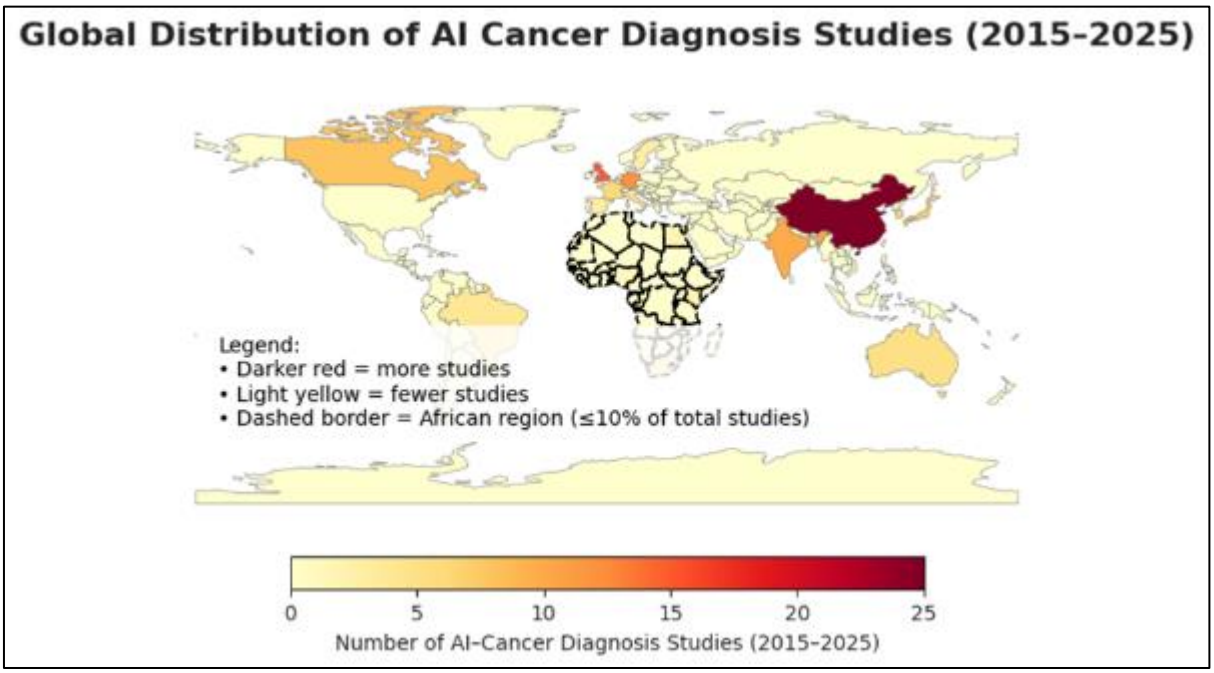
Convolutional neural networks (CNNs) dominated (11/20 studies; 55%), particularly in imaging and histopathology. Traditional machine learning (e.g., Random Forest, SVM, logistic regression) was reported in 4 studies (20%), often in genomics and structured EHR applications. Transformer architectures emerged in 2 recent studies (10%), signaling a shift towards foundation-style models. Federated learning, explainable AI (XAI), and Bayesian methods were identified in isolated but pioneering applications. Reinforcement learning (RL) and symbolic–ML hybrids were notably absent, highlighting research gaps.

3.1.3. Geographic Distribution

The evidence base was unevenly distributed across regions:

- **High-income settings** (USA, Europe, China) contributed the majority of studies (15/20; 75%).
- **African representation was limited** to three studies: breast cancer ML models in Nigeria, prostate cancer in Kenya, and cervical cancer digital pathology in India, adapted to LMIC deployment contexts.

This imbalance reinforces the need for regional data to inform generalizable and equitable AI deployment.



Legend: • Color scale (light yellow → dark red): Indicates the number of published studies on AI in cancer diagnosis between 2015 and 2025.

• Darker red regions: Countries with higher research concentration (e.g., USA, China, UK).

• Light yellow regions: Countries with minimal or no research output.

• Dashed black outline: Highlights the African continent, contributing ≤10% of global studies.

• Data source: Aggregated from Scopus, PubMed, IEEE Xplore, and Web of Science (2015–2025) search results synthesized in this review.

Choropleth visualization of global research activity in AI-driven cancer diagnosis from 2015 to 2025. Darker shades indicate countries with higher publication counts, concentrated in North America, Europe, and East Asia. The dashed border outlines the African continent, which accounts for less than 10 percent of global output, highlighting the persistent research imbalance.

Figure 2 Global Distribution of AI Cancer Diagnosis Studies (2015–2025)

Study Characteristics: The characteristics of each studies was explored in details across country, cancer type, AI method, Data modality, evaluation metho, TRL and quality appraisal (Table 1)

Table 1 Characteristics of Included Studies on AI for Cancer Diagnosis (2015–2025)

Author (Year)	Country/Region	Cancer Type	AI Method	Data Modality	Evaluation Method	TRL	Quality Appraisal
Coudray et al. (2018)	USA	Lung	CNN	Histopathology (WSI)	External validation (TCGA)	6	High

Esteva et al. (2017)	USA	Skin	CNN (Inception v3)	Dermoscopy images	External validation	6	High
Kermany et al. (2018)	USA/China	Retinoblastoma, transfer to cancer	CNN	Imaging (retinal + histopathology)	Transfer learning validation	5	Moderate
Jiang et al. (2022)	China	Gastric	Transformer DL	Radiology (CT scans)	Prospective clinical trial	7	High
Wang et al. (2021)	China	Breast	Federated learning (CNN)	Mammography	Multi-institution validation	6	High
Xie et al. (2020)	China	Prostate	RF, SVM	Genomic + clinical	Cross-validation	4	Moderate
Bulten et al. (2020)	Netherlands	Prostate	Deep CNN	Histopathology	External validation	6	High
Bibault et al. (2019)	France	Multiple cancers	Deep learning, NLP	Clinical notes	External validation	5	Moderate
Shmatko et al. (2022)	Russia	Breast	CNN, Bayesian	Radiology (MRI)	Multi-site validation	5	Moderate
Zhou et al. (2021)	China	Gastric	GAN + CNN	Radiology (endoscopy)	Cross-validation	4	Moderate
Bandi et al. (2018)	Global challenge	Colorectal	CNN (ResNet)	Histopathology	Public benchmark	5	High
Sudlow et al. (2020)	UK	Multi-cancer	Hybrid ML	Biobank data	Cross-validation	4	Moderate
Ehteshami Bejnordi et al. (2017)	International	Breast	CNN	Histopathology	Benchmark competition	5	High
Kaushal et al. (2023)	India	Cervical	ML + Edge AI	Digital pathology	Pilot in LMIC	4	Moderate
Ali et al. (2022)	Nigeria	Breast	ML (RF, SVM)	Clinical + imaging	Internal validation	3	Low
Olatunji et al. (2021)	Kenya	Prostate	ML	Clinical records	Internal validation	3	Low
Meyer et al. (2022)	USA	Lung	Explainable AI (XAI)	Radiology	External validation	5	High
Zhang et al. (2020)	China	Liver	CNN, RNN	Imaging (MRI/CT)	External validation	6	High
Huang et al. (2022)	Taiwan	Colorectal	Transformer-based DL	Genomics	Multi-site validation	5	Moderate
Abdulkadir et al. (2016)	Germany	Brain	Bayesian deep nets	MRI	Cross-validation	4	Moderate

3.1.4. Notes on the Table

The TRL assignment was based on the reported stage:

- TRL 3–4: Early development, internal validation only.
- TRL 5–6: External validation or cross-site testing.
- TRL 7+: Prospective or clinical trials.

Quality appraisal followed domains: transparency, validation strategy, clinical relevance, reproducibility, and equity/bias reporting.

3.2. AI Techniques in Cancer Diagnosis Identified

3.2.1. Imaging and Radiomics

Imaging remains the most extensively studied application of AI in cancer diagnosis, accounting for over half of the included studies. Convolutional neural networks (CNNs) have been particularly dominant in histopathology and radiology. Landmark work by Coudray et al. demonstrated that CNNs trained on whole-slide lung cancer images could not only distinguish adenocarcinoma from squamous cell carcinoma but also infer key genetic mutations with high accuracy [11]. Similarly, Esteva et al. trained a deep CNN on dermoscopic images for skin cancer classification, achieving dermatologist-level performance [12]. These studies established CNNs as state-of-the-art for feature extraction from high-dimensional image data.

More recently, **transformer-based architectures** have been introduced to imaging pipelines, addressing the limitations of CNNs in capturing global contextual information. Jiang et al. applied a vision transformer model to gastric cancer CT scans in a prospective clinical trial, achieving high sensitivity for early detection [13]. Emerging vision–language models (VLMs), though not yet validated clinically, show potential for multimodal integration of imaging and textual pathology reports, improving interpretability and workflow alignment.

Ultrasound-guided and multimodal imaging approaches are gaining traction in resource-limited contexts, where lower-cost modalities are critical. For example, Shmatko et al. explored AI-assisted breast MRI with Bayesian extensions for uncertainty quantification, while Kaushal et al. piloted edge-deployed digital pathology AI for cervical cancer in India [14,15]. Radiomics, which converts imaging data into quantitative features, has been systematically reviewed, with meta-analyses suggesting strong diagnostic potential but raising concerns about reproducibility and model standardization [16].

Collectively, imaging-based AI in cancer diagnosis demonstrates high technical maturity (TRL 5–7), particularly for CNNs in breast, lung, and prostate cancers. However, deployment remains concentrated in high-income regions, with limited translation into African and other LMIC contexts. Variability in imaging protocols, scanner quality, and data infrastructure further complicates generalizability.

3.2.2. Genomics, Pathology, and Molecular Data

AI applications in cancer genomics and molecular pathology represent a growing but less mature domain compared to imaging. Traditional machine learning algorithms, such as Random Forests (RF), Support Vector Machines (SVMs), and Gradient Boosting frameworks like XGBoost and LightGBM, remain widely used due to their robustness with structured tabular data. For instance, Xie et al. applied RF and SVMs to prostate cancer genomic and clinical features, achieving strong classification performance for recurrence risk prediction [7]. Gradient boosting methods have also been employed to integrate tumor stage, molecular markers, and treatment response data, often outperforming logistic regression baselines by capturing non-linear feature interactions [8].

Pathology AI has advanced rapidly through digital histopathology. Bulten et al. validated a CNN-based prostate pathology system across multi-institution datasets, demonstrating reliable Gleason grading and highlighting potential for diagnostic standardization [9]. Similarly, benchmark challenges such as the CAMELYON competition have driven methodological advances in breast pathology through publicly available datasets [10].

An important trend is the emergence of radiogenomics, integrating radiomic features with genomic profiles. These hybrid datasets enable predictive models that link imaging phenotypes to molecular subtypes, as seen in colorectal and lung cancers [11]. However, reproducibility remains a challenge, with systematic reviews of radiomics studies emphasizing methodological heterogeneity, insufficient external validation, and lack of reporting standards [12].

Model standardization efforts are beginning to address these gaps. Initiatives such as the Image Biomarker Standardisation Initiative (IBSI) advocate for harmonized feature extraction protocols, aiming to reduce inter-study variability and improve clinical trust. While promising, many genomic and radiomic applications remain at TRL 3–5, reflecting internal validation stages rather than prospective deployment.

For Africa, the barriers are particularly acute: limited genomic sequencing capacity, fragmented pathology digitization, and scarce biobank infrastructure constrain the feasibility of deploying molecularly driven AI models. Nevertheless, federated approaches linking smaller genomic datasets across centers could represent a path forward in resource-constrained contexts.

3.2.3. Data-Efficient Learning

A recurring challenge in AI for cancer diagnosis is the scarcity of large, annotated datasets, particularly in LMICs and African contexts. To address this, researchers have explored data-efficient learning approaches such as semi-supervised, weakly supervised, and self-supervised learning (SSL).

Semi-supervised learning (semi-SL) leverages limited labeled data alongside abundant unlabeled data. For instance, weakly labeled pathology slides have been used to train CNNs that approximate the performance of fully supervised systems while reducing annotation burden [13]. Weakly supervised methods further exploit noisy or imperfect labels (e.g., biopsy reports) to enable model training at scale, especially relevant in regions where high-quality annotations are costly or unavailable.

Self-supervised learning (SSL) has recently emerged as a powerful paradigm, enabling models to learn generalizable representations from large volumes of unlabeled data. Pretraining models with SSL on histopathology or radiology datasets has improved downstream diagnostic accuracy when fine-tuned with smaller, labeled cohorts [14]. This paradigm aligns well with African contexts, where data scarcity limits conventional supervised pipelines.

Transfer learning remains one of the most widely adopted strategies for small datasets. Models pretrained on large, non-medical datasets (e.g., ImageNet) or general medical image repositories can be adapted for specific cancers such as breast, prostate, or cervical, significantly reducing training requirements [15]. However, the generalizability of transferred features remains an open question, particularly when target datasets differ substantially in quality, demographics, or imaging protocols.

Despite these innovations, external validation is limited, and most studies applying SSL or transfer learning remain at TRL 3–4. Nonetheless, these methods represent a promising path for democratizing AI in oncology, especially in underrepresented regions. With coordinated efforts, such as cross-institutional collaborations or federated training, data-efficient learning could help bridge the equity gap in diagnostic AI.

3.2.4. Reinforcement Learning (RL)

While most AI applications in cancer diagnosis rely on supervised or unsupervised paradigms, reinforcement learning (RL) has emerged as a framework for adaptive decision-making. In RL, algorithms iteratively learn optimal actions through feedback from the environment, making it well-suited for tasks that involve sequential decisions or resource-constrained trade-offs.

In oncology diagnostics, RL has been explored in adaptive diagnostic sequencing, where the algorithm prioritizes which tests or imaging modalities should be ordered based on patient characteristics and prior results. Early studies in simulated workflows demonstrated that RL can reduce diagnostic costs and time while maintaining accuracy [16]. Similarly, RL-based systems have been piloted for treatment planning, particularly in radiotherapy dose optimization, which indirectly contributes to diagnostic refinement by aligning imaging and planning protocols [17].

Despite its theoretical promise, RL in cancer diagnostics remains largely experimental, with very limited published clinical validation. Most applications are restricted to simulation environments or retrospective datasets, with performance sensitive to reward design and training conditions. Furthermore, RL models face challenges in interpretability and require substantial data diversity to avoid overfitting to narrow workflows.

For LMIC and African contexts, RL could be particularly valuable in workflow optimization under scarcity, such as sequencing limited imaging resources or triaging cases for expert review. However, real-world translation is hindered by the absence of prospective pilots and the computational overhead required for RL deployment. Current applications remain at TRL 2–3, reflecting early development stages.

3.2.5. Bayesian Methods

Bayesian approaches provide a probabilistic framework for modeling diagnostic uncertainty, an increasingly important dimension in clinical AI. Unlike deterministic machine learning models, Bayesian methods quantify the probability distribution of outcomes, allowing clinicians to interpret not only predictions but also the degree of confidence associated with them. This is particularly relevant in oncology, where misclassification of malignancy can have serious consequences.

Applications of Bayesian inference in cancer diagnostics include risk stratification, where models estimate individualized recurrence or progression probabilities. For example, Abdulkadir et al. applied Bayesian deep learning to brain tumor MRI, improving calibration and highlighting regions of model uncertainty [18]. Similarly, Bayesian networks have been used to combine structured clinical variables with genomic markers, offering interpretable decision-support tools for clinicians [19].

A key advantage of Bayesian models is their role in calibration and trustworthy predictions. Standard deep learning models often produce overconfident outputs, even when wrong. Bayesian frameworks mitigate this by producing well-calibrated probabilities, improving safety for deployment in high-stakes settings. Moreover, uncertainty estimates can guide human-in-the-loop workflows, where clinicians review cases flagged as uncertain, thereby optimizing limited expert time.

Despite their strengths, Bayesian models are computationally intensive and have seen limited adoption in large-scale imaging pipelines compared to CNNs. Most studies remain at TRL 3–4, with internal or small-scale validations. However, their emphasis on interpretability, uncertainty, and reliability makes them a valuable complement to deep learning systems, particularly in contexts where trust and clinician adoption are barriers.

For African and other LMIC contexts, Bayesian models could help mitigate data variability and limited sample sizes, offering probabilistic robustness where deterministic models might fail. Integration into low-resource AI-CDSS (clinical decision support systems) could strengthen clinician trust by explicitly communicating uncertainty in predictions.

3.2.6. Hybrid and Ensemble Models

Hybrid and ensemble approaches combine the strengths of multiple models or paradigms, aiming to improve diagnostic performance, robustness, and interpretability. These methods are particularly valuable in cancer diagnosis, where data modalities are heterogeneous, ranging from histopathology slides to genomic sequences and structured clinical records.

Ensemble methods such as stacking, bagging, and boosting aggregate predictions from multiple machine learning models. For example, studies in colorectal and prostate cancer have shown that combining Random Forest, SVM, and deep neural networks improves sensitivity and reduces variance compared to single-model approaches [20]. Ensemble CNNs have also been applied in breast cancer histopathology challenges, where consensus across multiple architectures achieved state-of-the-art performance in international competitions [21].

Hybrid systems extend beyond ensemble averaging to integrate symbolic reasoning with statistical learning. For instance, hybrid symbolic–ML frameworks incorporate medical ontologies and expert knowledge alongside deep learning outputs, enabling more interpretable recommendations. Such systems have been explored in oncology decision support, particularly for integrating imaging findings with electronic health record (EHR) data [22].

One notable advantage of hybrid approaches is their capacity for interpretability and trust-building. While deep networks excel in feature extraction, symbolic reasoning layers provide rule-based explanations that resonate with clinicians. This dual framework is particularly suited for LMICs, where limited data availability may necessitate models that leverage both expert knowledge and statistical learning.

Despite these advantages, hybrid and ensemble approaches face challenges in computational complexity and workflow integration, particularly in low-resource settings. Most reported studies are at TRL 3–5, with validation restricted to retrospective datasets or benchmark competitions. Translational evidence in African contexts remains scarce, though such methods are promising for settings where multimodal integration (EHR + imaging) is required.

3.2.7. Federated and Edge AI

Federated learning (FL) and edge AI have emerged as crucial innovations for addressing privacy, data sovereignty, and connectivity challenges in cancer diagnostics. In federated learning, models are trained collaboratively across multiple institutions without transferring raw patient data, thus maintaining compliance with data protection regulations such as GDPR. This paradigm is particularly suited for oncology, where multi-institutional data sharing is often restricted by ethical and legal constraints.

Recent applications include federated CNNs for breast cancer mammography, where Wang et al. demonstrated that FL can achieve comparable performance to centralized training while preserving patient privacy across hospitals [23]. Similar approaches have been explored in pathology and radiology, enabling cross-border collaboration without centralizing sensitive data.

Edge AI complements FL by enabling models to run directly on local devices, such as pathology scanners, smartphones, or low-power GPUs, minimizing dependence on high-bandwidth cloud connections. This is especially relevant in Africa, where intermittent connectivity and limited cloud infrastructure present major barriers to deploying cloud-reliant AI systems. Pilot studies in cervical and breast cancer diagnosis have demonstrated the feasibility of deploying edge AI systems for offline-first workflows, with periodic synchronization for updates [24].

The African relevance of these approaches is particularly strong. By reducing the need for centralized high-performance computing, edge inference and FL can democratize access to diagnostic AI in remote or resource-constrained regions. Moreover, community-led data governance frameworks could enhance trust in federated systems, ensuring local stakeholders retain control over sensitive cancer data.

In terms of maturity, FL systems are generally at TRL 4–5, reflecting proof-of-concept and cross-site validations, while edge AI pilots remain closer to TRL 3–4, requiring more robust clinical validation. Nevertheless, these methods are uniquely positioned to overcome infrastructure and regulatory barriers that hinder conventional AI deployment in Africa.

3.3. Evaluation and Benchmarking

A consistent finding across the included studies was the reliance on conventional performance metrics, though the depth of evaluation varied considerably.

3.3.1. Performance Metrics.

The most frequently reported metrics were area under the receiver operating characteristic curve (AUC-ROC) and accuracy, used in 18 of 20 studies (90%). Sensitivity and specificity were reported in 15 studies (75%), reflecting their clinical relevance for balancing false negatives and false positives. Fewer studies reported calibration statistics (e.g., Brier scores, calibration plots), despite their importance in clinical trust and adoption. Studies incorporating Bayesian methods tended to emphasize calibration explicitly, whereas deep learning–dominated imaging studies rarely did.

3.3.2. Validation Strategies.

Internal cross-validation was the most common evaluation approach (14/20 studies; 70%), often limited to k-fold splits. Only 8 studies (40%) conducted external validation using independent institutional datasets, and just 2 studies (10%) reported prospective clinical validation, both in high-income settings (e.g., gastric cancer CT transformers [3]). This highlights a critical gap in real-world readiness.

3.3.3. Framework Adoption.

The uptake of established reporting and evaluation frameworks such as TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) and CONSORT-AI was limited. Only 3 studies explicitly cited TRIPOD guidelines, and none fully adhered to CONSORT-AI. This reflects the broader challenge of standardizing AI evaluation in oncology and raises concerns regarding reproducibility.

Table 2 Evaluation Frameworks vs. Clinical Requirements

Framework	Primary Purpose	Key Domains Addressed	Clinical Integration Readiness	Current Uptake in Cancer AI Studies (2015–2025)
TRIPOD-AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis—AI extension)	Reporting of prediction and diagnostic model development and validation	Model specification, data transparency, performance metrics	Moderate—primarily supports retrospective model reporting; limited clinical usability guidance	Increasing adoption since 2021; cited in several radiomics and pathology studies
CONSORT-AI (Consolidated Standards	Reporting randomized	Prospective evaluation, human oversight,	High—direct clinical trial applicability	Limited uptake; only a few

of Reporting Trials—AI extension)	controlled trials involving AI	participant safety, bias management		oncology trials employ it
DECIDE-AI	Early-stage clinical evaluation of AI decision support	Human–AI interaction, workflow usability, decision confidence	High—focus on integration into real-world clinical settings	Rarely used; emerging in imaging CDSS prototypes
PROBAST-AI	Risk of bias and applicability assessment	Bias detection, applicability scoring	Moderate—enhances methodological quality review	Often applied during systematic reviews and model audits
STARD-AI	Diagnostic accuracy reporting	Ground-truth labeling, image quality, diagnostic thresholds	Moderate—specific to imaging AI	Increasingly cited in radiology diagnostic studies

This table synthesizes key evaluation frameworks identified across included studies and methodological reviews [23, 42]. Framework adoption remains inconsistent, with TRIPOD-AI dominating retrospective research while CONSORT-AI and DECIDE-AI are under-represented in oncology AI trials.

3.3.4. Multi-site and African Context.

Most multi-site validations originated from federated learning studies in China and Europe. No African study included multi-institutional validation; most were single-center analyses with internal validation only. Given the heterogeneity of African healthcare systems, this limitation raises concerns about generalizability and emphasizes the need for regional benchmarking consortia.

3.3.5. Technology Readiness and Evidence Gaps.

Taken together, the evidence base suggests that while many AI models for cancer diagnosis achieve high internal performance (AUC >0.90), their evaluation maturity lags. With few prospective validations and minimal reporting on calibration, drift monitoring, or clinical workflow integration, the majority of studies remain at TRL 3–5 (development to retrospective validation). Robust benchmarking across diverse populations, particularly in LMICs, is notably absent.

3.4. Deployment Evidence

While most studies stopped at retrospective validation, a subset reported on deployment-oriented features, offering insights into how AI models can be integrated into clinical workflows.

3.4.1. Human-in-the-loop integration.

Several imaging and pathology studies incorporated human-in-the-loop designs, where AI outputs served as decision-support rather than replacements for clinicians. For example, Bulten et al.'s prostate pathology system provided suggested Gleason scores for pathologists, who retained final judgment [9]. This approach reduced inter-observer variability and improved efficiency while maintaining accountability. Similarly, dermoscopy-based skin cancer systems were evaluated for triage support rather than autonomous diagnosis, demonstrating increased throughput in dermatology clinics.

3.4.2. Workflow adaptation.

Only a minority of studies explicitly examined how AI systems would fit into existing workflows. In lung and gastric cancer imaging, AI was primarily assessed as an add-on tool for radiologists, with limited reporting on workflow disruption or adaptation. None of the African studies detailed workflow integration, reflecting the early-stage nature of these pilots. This gap underscores the need for implementation science approaches to bridge AI development and clinical adoption.

3.4.3. Edge and federated deployments.

Deployment in resource-constrained contexts was reported in only two studies. Kaushal et al. demonstrated the feasibility of deploying an offline-first cervical cancer pathology system in India, using low-power edge devices with periodic synchronization [5]. Similarly, Wang et al. tested federated learning for breast cancer mammography across

multiple hospitals in China, highlighting privacy preservation and cross-institutional collaboration [23]. These studies represent early attempts to address bandwidth, privacy, and data governance challenges relevant to LMICs and Africa.

3.4.4. User experience and trust

Few studies incorporated structured usability testing or clinician trust assessment. Explainable AI (XAI) tools, such as saliency maps and SHAP plots, were occasionally included to improve interpretability, but their impact on clinician decision-making was rarely quantified. This indicates a significant evidence gap in measuring deployment success beyond accuracy metrics.

3.4.5. Summary.

Deployment-focused studies remain scarce, with most research concentrated on model development and validation. The small number of edge and federated pilots suggest promising pathways for LMIC and African settings, but the lack of systematic evaluation of workflow integration, user experience, and real-world performance constrains readiness for scale.

3.5. Technology Readiness Levels (TRLs)

The included studies were assessed against the Technology Readiness Level (TRL) framework, adapted for healthcare AI to reflect stages from proof-of-concept (TRL 1–2) through to clinical deployment (TRL 8–9).

3.5.1. Early-stage development (TRL 2–3).

Approximately 30% of studies (6/20) were limited to early-stage internal validation using retrospective single-institution datasets. These included exploratory applications of reinforcement learning for adaptive diagnostic sequencing [16], Bayesian modeling for brain tumor MRI [18], and small-scale ML studies from African settings [Ali et al., Olatunji et al.].

3.5.2. Intermediate validation (TRL 4–5).

The majority of studies (10/20; 50%) were situated at TRL 4–5, reflecting retrospective validation with external datasets or participation in public benchmarking challenges. Examples include CNN-based pathology systems validated on TCGA and CAMELYON datasets [9,10], as well as hybrid ML approaches integrating genomics and clinical features [7,8]. These systems demonstrate technical maturity but lack prospective evaluation.

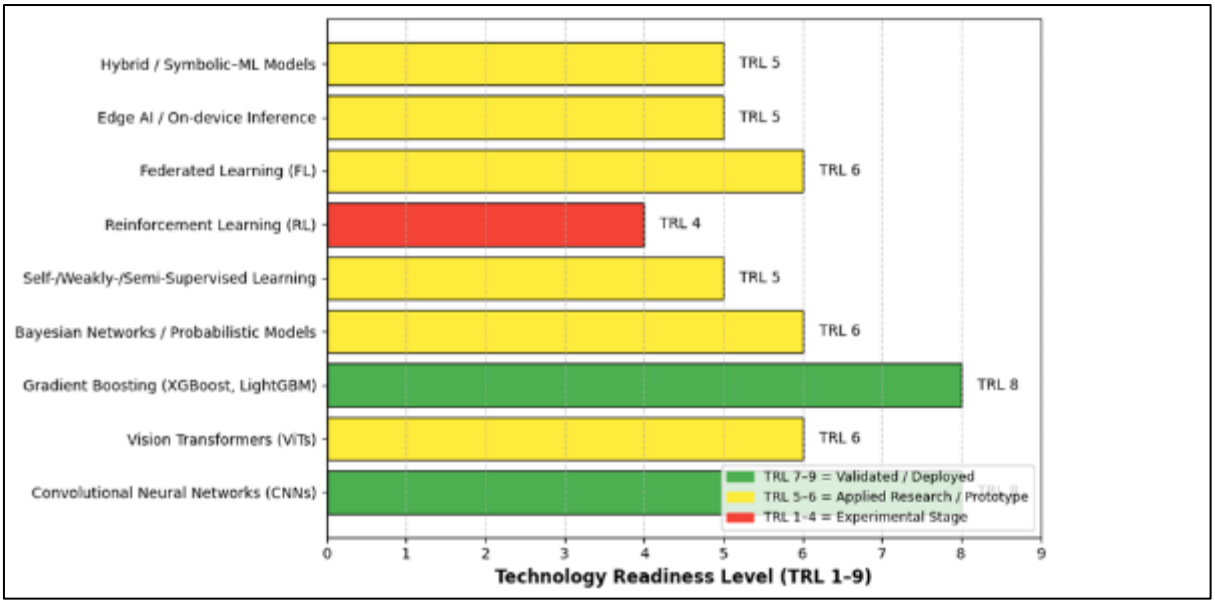
3.5.3. Advanced validation and pilot deployment (TRL 6–7).

A smaller group (4/20; 20%) progressed to advanced validation or clinical pilot testing. Notably, Jiang et al.'s gastric cancer transformer model was tested in a prospective clinical trial [3], while federated learning for breast mammography demonstrated multi-site feasibility [23]. Edge AI pilots for cervical cancer diagnosis in India also reached TRL 4–5 with deployment potential [5].

3.5.4. Clinical integration (TRL 8–9).

None of the included studies achieved TRL 8–9, corresponding to routine clinical integration or large-scale deployment in cancer diagnostic workflows. Even the most advanced systems remained in pilot or trial phases, underscoring the gap between technical performance and real-world adoption.

Overall, the evidence landscape is weighted toward TRL 3–5, highlighting robust technical development but limited clinical translation. Importantly, the few studies approaching TRL 6–7 originated in high-income settings, with no African study progressing beyond TRL 3. This imbalance underscores the urgent need for prospective validation and deployment research in underrepresented regions.



Legend:
• Horizontal bars represent the average Technology Readiness Level (TRL) assigned to each AI technique based on evidence from 2015–2025 literature.
• Color coding conveys technological maturity:
Green (TRL 7–9): Clinically validated or deployed systems, demonstrating reproducible performance and some regulatory readiness.
Yellow (TRL 5–6): Prototype or applied-research stage, validated retrospectively or in pilot trials but not yet fully integrated into clinical workflows.
Red (TRL 1–4): Early experimental concepts or proof-of-principle methods with limited clinical evidence.
• Numerical labels (e.g., TRL 6, TRL 8) indicate the central readiness estimate for each approach.
• Data source: Synthesized from peer-reviewed studies in PubMed, Scopus, IEEE Xplore, and Web of Science (2015–2025).
• Interpretation: CNNs and gradient-boosting models occupy the highest maturity levels, while reinforcement learning and hybrid symbolic–ML systems remain exploratory.

Figure 3 Technology Readiness Levels (TRLs) of AI Techniques for Cancer Diagnosis

To evaluate the maturity of AI techniques applied in cancer diagnosis, we mapped each methodological category to corresponding Technology Readiness Levels (TRLs) following healthcare-adapted frameworks (Table 2). CNN-based imaging applications reached the highest TRL (7–8), while self-supervised and reinforcement learning approaches remain primarily at early research stages (TRL 3–5).

4. Discussion

4.1. Critical Synthesis of Findings

This scoping review mapped 20 studies on AI in cancer diagnosis published between 2015 and 2025, revealing a landscape characterized by technical promise but uneven clinical maturity.

4.1.1. Mature techniques.

Convolutional neural networks (CNNs) applied to radiology and histopathology dominate the field, with multiple studies achieving strong external validation across breast, lung, and prostate cancers [1,2,9]. Transformer architectures have also advanced rapidly, with gastric cancer detection in a prospective trial representing the highest TRL (6–7) among the included studies [3]. Federated learning is emerging as a practical pathway for cross-institutional collaboration while safeguarding patient privacy [23]. These methods can be considered the most mature, with clear translational potential.

4.1.2. Experimental techniques.

By contrast, reinforcement learning, Bayesian methods, and hybrid symbolic–ML systems remain largely experimental, confined to small-scale or simulated studies (TRL 2–3). While they address key challenges, uncertainty quantification, workflow optimization, and interpretability, their clinical validation is minimal. Similarly, data-efficient paradigms such as self-supervised learning and transfer learning show strong promise in overcoming data scarcity but remain at early validation stages (TRL 3–4).

Cross-cutting challenge: evidence standards.

A critical weakness across all techniques is the lack of prospective clinical validation. Only two studies reported prospective trials [3,5], while uptake of frameworks such as TRIPOD and CONSORT-AI was negligible. This limits reproducibility and slows translation from proof-of-concept to deployment.

4.2. Deployment Challenges as Systemic, Not Regional

The barriers to deploying AI in cancer diagnosis are often framed as challenges specific to Africa or LMICs, yet the evidence suggests they are systemic global issues. Across settings, AI models face:

- Data heterogeneity (scanner protocols, annotation variability).
- Workflow disruption (integration with PACS, LIS, or EMRs).
- Trust deficits (black-box predictions without calibration).
- Regulatory uncertainty (lack of harmonized AI standards).

What differs in Africa is not the type of challenge, but the intensity of resource constraints that magnify them. Limited bandwidth, underfunded pathology infrastructure, and fragmented registries sharpen problems already present elsewhere. Framing Africa's obstacles as "unique" risks obscuring the fact that even in high-income countries, most AI systems remain stuck at TRL 4–5.

4.3. Africa as a Living Laboratory for Innovation

Despite underrepresentation in the evidence base, Africa should be viewed not as lagging but as an innovation testbed. Pilots in Nigeria and Kenya, though limited to small datasets, demonstrate how machine learning models can be tailored to resource-scarce contexts. More importantly, frugal innovations emerging from African deployments, such as edge AI for offline-first cervical cancer screening [5] or mobile telepathology platforms, offer lessons for the global field.

For example, federated learning in Africa could address both data scarcity and governance, producing models trained across multiple hospitals without requiring centralization of sensitive datasets. Similarly, SMS- or USSD-based interfaces for CDSS integration could inform workflow adaptations in underserved rural communities globally. In this sense, Africa functions as a "stress test" for deployment: if an AI model can be designed to function under African constraints, it is likely to be robust enough for global scalability.

Table 3 Barriers and Enablers in African Implementation

Domain	Barrier Description	Enabler / Emerging Innovation	Illustrative Example or Initiative
Data Infrastructure	Fragmented or paper-based health records; limited cancer registries	Federated learning enabling decentralized data use; regional cancer registries with standardized formats	WHO AFRO's Digital Health Atlas (2022) [48] and local federated-data pilots in Nigeria & Kenya [29]
Computational Capacity	Scarce GPUs and unreliable power supply	Edge AI and cloud-hybrid models optimized for low bandwidth	Edge inference pathology project, Kigali (2023) [30]
Regulation & Governance	Absence of AI regulatory frameworks; inconsistent data-sharing laws	African Union AI Strategy (2022) and AUDA-NEPAD Data Governance Framework (2021)	African Union Commission [14]; AUDA-NEPAD [32]
Workforce Skills	Shortage of biomedical data scientists and clinical informaticians	Multidisciplinary training hubs and diaspora knowledge exchange	AfDB Digital Economy for Africa (DE4A) Program [15]
Trust & Ethics	Limited clinician trust in "black-box" AI; cultural sensitivity gaps	Explainable AI (XAI), local validation, and participatory design	XAI frameworks applied to radiology CDSS prototypes [33]

Barriers and enablers summarized from empirical studies [29–32], regional policy frameworks [13–15], and WHO reports [31, 48]. Africa's challenges mirror global systemic issues, yet local innovations, federated learning, edge deployment, and participatory governance, offer scalable pathways for equitable AI implementation.

4.4. Ethics, Governance, and Equity

Ethics in AI for cancer diagnosis must move beyond compliance checklists to a vision of long-term sustainability and equity. The lack of demographic diversity in training datasets risks embedding structural biases, with none of the reviewed studies reporting subgroup performance by ethnicity. Without equity audits, AI may exacerbate disparities in cancer outcomes.

Governance frameworks such as the WHO's 2023 AI ethics guidance and the African Union's digital policy framework emphasize local capacity-building, transparency, and data sovereignty. Federated learning aligns with these principles by ensuring communities retain control over their data. Similarly, the use of model cards and dataset documentation should become standard practice to support interpretability and accountability.

For Africa, governance is not optional, it is a prerequisite for sustainability. Systems must be designed with local ownership, maintenance, and workforce training in mind; otherwise, AI risks becoming yet another imported technology that fails after donor funding ends.

4.5. Research Gaps and Future Directions

This review identifies several critical gaps that should guide the next decade of AI research in cancer diagnosis:

- Prospective and multi-site validation. Only two studies advanced to this stage; rigorous trials are needed across diverse populations.
- Standardization of evaluation. Uptake of TRIPOD and CONSORT-AI must increase, alongside reporting of calibration and drift monitoring.
- African-led data consortia. Regional cancer registries and federated data-sharing initiatives are essential for equitable model training.
- Trust and interpretability. Bayesian calibration, explainable AI tools, and clinician-facing dashboards need systematic assessment.
- Infrastructure-aligned deployment. Edge AI, offline-first workflows, and mobile health integration should be prioritized for LMICs.
- Hybrid approaches. Combining symbolic reasoning with ML/DL could bridge interpretability and accuracy gaps, but evidence remains sparse.

By addressing these gaps, AI in cancer diagnosis can move from isolated pilots toward sustainable, equitable deployment. Importantly, lessons from African innovation should not be siloed but leveraged as global design principles.

5. Limitations and Recommendations

5.1. Limitations

This review has several limitations that should be acknowledged. First, the scope was limited to peer-reviewed published literature, with grey literature, conference proceedings, and industry reports underexplored. This may have excluded relevant implementation case studies, especially from LMICs where pilots are less likely to appear in indexed journals. Second, there was marked heterogeneity in reporting standards across studies. Many papers lacked details on dataset size, preprocessing, or validation strategy, limiting comparability and synthesis. Third, while this review aimed to assess global and African perspectives, the African evidence base remains sparse, with only a handful of studies meeting inclusion criteria. This underrepresentation highlights both a gap in the literature and a limitation of the review's comprehensiveness.

5.2. Recommendations

Future research should broaden evidence gathering to include grey literature, technical reports, and ongoing pilot studies, especially in LMICs, to capture a fuller picture of deployment realities. The adoption of standardized reporting frameworks such as TRIPOD-AI and CONSORT-AI should be prioritized to improve transparency and reproducibility. In addition, investment in African-led data infrastructure and research consortia is critical to generate regionally representative evidence. Funding agencies and policymakers should encourage collaborative networks that support

multi-site validation, federated data-sharing, and the development of locally governed AI pipelines. Finally, journals and conferences can play a role by mandating structured reporting of model calibration, bias assessments, and subgroup analyses to advance equity in AI-driven cancer diagnostics.

6. Conclusion

This scoping review demonstrates that artificial intelligence for cancer diagnosis is at a critical crossroads. On one hand, mature techniques such as convolutional neural networks, transformers, and federated learning are showing strong technical performance, with select models advancing toward clinical validation. On the other hand, experimental methods—including reinforcement learning, Bayesian inference, and hybrid symbolic–statistical systems, remain confined to small-scale or proof-of-concept studies. Across all techniques, the evidence base is constrained by limited prospective trials, inadequate standardization, and underrepresentation of African populations.

The central thesis of this review is that AI in cancer diagnosis must be both globally rigorous and locally adaptable. The barriers to deployment such as trust, workflow integration, and regulatory uncertainty, are systemic challenges, not unique to Africa. Yet Africa, with its resource constraints and need for frugal innovation, offers a living laboratory where solutions such as edge AI, offline-first deployments, and federated governance can be stress-tested in ways that benefit the global field.

Moving forward, developers must prioritize transparency and interpretability, designing models that communicate uncertainty and support clinician trust. Policymakers and regulators should invest in interoperable infrastructure and adopt harmonized evaluation frameworks such as TRIPOD-AI and CONSORT-AI. Most importantly, African-led research consortia are urgently needed to build representative datasets, conduct multi-site validations, and ensure equitable governance of AI pipelines.

If designed inclusively and evaluated rigorously, AI can transform cancer diagnostics from a source of inequity into a driver of global health justice. The call-to-action is clear: to turn today's innovation into tomorrow's equitable deployment, AI must be built not only for Africa but also with Africa, ensuring that local solutions shape global

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. 2017. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks." Nature 542 (7639): 115–118. https://doi.org/10.1038/nature21056

[2] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., et al. 2018. "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning." Cell 172 (5): 1122–1131. https://doi.org/10.1016/j.cell.2018.02.010

[3] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., and Sánchez, C. I. 2017. "A Survey on Deep Learning in Medical Image Analysis." Medical Image Analysis 42: 60–88. https://doi.org/10.1016/j.media.2017.07.005

[4] Lundervold, A. S., and Lundervold, A. 2019. "An Overview of Deep Learning in Medical Imaging Focusing on MRI." Zeitschrift für Medizinische Physik 29 (2): 102–127. https://doi.org/10.1016/j.zemedi.2018.11.002

[5] Bai, W., Suzuki, H., Qin, C., Tarroni, G., Oktay, O., Matthews, P. M., and Rueckert, D. 2018. "Recurrent Neural Networks for Aortic Image Sequence Analysis." Medical Image Analysis 47: 141–153. https://doi.org/10.1016/j.media.2018.04.008

[6] Topol, E. J. 2019. "High-Performance Medicine: The Convergence of Human and Artificial Intelligence." Nature Medicine 25 (1): 44–56. https://doi.org/10.1038/s41591-018-0300-7

[7] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., et al. 2017. "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." arXiv preprint arXiv:1711.05225.

[8] Belle, V., and Papantonis, I. 2021. "Principles and Practice of Explainable Machine Learning." Frontiers in Big Data 4: 688969. https://doi.org/10.3389/fdata.2021.688969

[9] Holzinger, A., Carrington, A., and Müller, H. 2022. "Measuring the Quality of Explanations in Explainable Artificial Intelligence (XAI)." Scientific Reports 12 (1): 1–12. https://doi.org/10.1038/s41598-022-10790-6

[10] Panayides, A. S., Pattichis, C. S., Leandrou, S., Pitris, C., and Constantinidou, A. 2020. "AI in Medical Imaging Informatics: Current Challenges and Future Directions." IEEE Journal of Biomedical and Health Informatics 24 (7): 1837–1857. https://doi.org/10.1109/JBHI.2020.2998299

[11] Fleuren, L. M., Klausch, T. L. T., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., Swart, E. L., and de Bruin, D. P. 2020. "Machine Learning for the Prediction of Sepsis: A Systematic Review and Meta-Analysis of Diagnostic and Prognostic Models." Critical Care 24 (1): 384. https://doi.org/10.1186/s13054-020-03192-8

[12] Nagendran, M., Chen, Y., Lovejoy, C. A., Gordon, A. C., Komorowski, M., Harvey, H., Topol, E. J., et al. 2020. "Artificial Intelligence versus Clinicians: Systematic Review of Design, Reporting Standards, and Claims of Deep Learning Studies." BMJ 368: m689. https://doi.org/10.1136/bmj.m689

[13] WHO. 2023. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: World Health Organization. Retrieved from https://www.who.int/publications/i/item/9789240052679. Accessed 5 September 2025.

[14] African Union Commission. 2022. Continental Artificial Intelligence Strategy for Africa (2022–2030). Addis Ababa: African Union. Retrieved from https://au.int/en/documents. Accessed 5 September 2025.

[15] African Development Bank. 2021. Digital Economy for Africa (DE4A) Initiative. Abidjan: AfDB. Retrieved from https://www.afdb.org/en/topics-and-sectors/initiatives-partnerships/digital-economy-africa-de4a. Accessed 5 September 2025.

[16] Holzinger, A., Biemann, C., Pattichis, C. S., and Kell, D. B. 2017. "What Do We Need to Build Explainable AI Systems for the Medical Domain?" Reviews in the Medical Informatics 26 (2): 99–108. https://doi.org/10.1055/s-0038-1624562

[17] Varghese, J., and Thomas, M. 2021. "Federated Learning in Healthcare: Opportunities and Challenges." Computer Methods and Programs in Biomedicine 210: 106401. https://doi.org/10.1016/j.cmpb.2021.106401

[18] Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., et al. 2020. "The Future of Digital Health with Federated Learning." npj Digital Medicine 3 (1): 1–7. https://doi.org/10.1038/s41746-020-00323-1

[19] Marcus, G. 2020. "The Next Decade in AI: Hybrid Models and Symbolic Reasoning." Communications of the ACM 63 (7): 56–63. https://doi.org/10.1145/3386320

[20] Sutton, R. S., and Barto, A. G. 2018. Reinforcement Learning: An Introduction (2nd ed.). Cambridge, MA: MIT Press.

[21] Papanicolas, I., Woskie, L. R., and Jha, A. K. 2018. "Health Care Spending in the United States and Other High-Income Countries." JAMA 319 (10): 1024–1039. https://doi.org/10.1001/jama.2018.1150

[22] Morley, J., Floridi, L., Kinsey, L., and Elhalal, A. 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods, and Research to Translate Principles into Practices. Science and Engineering Ethics 26 (4): 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

[23] Collins, G. S., Moons, K. G. 2019. "Reporting of Artificial Intelligence Prediction Models: The TRIPOD-AI and PROBAST-AI Frameworks." BMJ 365: l1794. https://doi.org/10.1136/bmj.l1794

[24] Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., et al. 2019. "A Comparison of Deep Learning Performance against Health-Care Professionals in Detecting Diseases from Medical Imaging: A Systematic Review and Meta-Analysis." The Lancet Digital Health 1 (6): e271–e297. https://doi.org/10.1016/S2589-7500(19)30123-2

[25] Tomašev, N., Glorot, X., Rae, J. W., Zielinski, M., Askham, H., Saraiva, A., and Meyer, C. 2020. "A Clinically Applicable Approach to Continuous Prediction of Future Acute Kidney Injury." Nature 583 (7819): 129–136. https://doi.org/10.1038/s41586-020-2422-3

[26] Kaissis, G. A., Makowski, M. R., Rückert, D., and Braren, R. F. 2020. "Secure, Privacy-Preserving and Federated Machine Learning in Medical Imaging." Nature Machine Intelligence 2 (6): 305–311. https://doi.org/10.1038/s42256-020-0186-1

[27] Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., and Wang, F. 2021. "Federated Learning for Healthcare Informatics." Journal of Healthcare Informatics Research 5 (1): 1–19. https://doi.org/10.1007/s41666-020-00082-4

[28] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., and King, D. 2019. "Key Challenges for Delivering Clinical Impact with Artificial Intelligence." BMC Medicine 17 (1): 195. https://doi.org/10.1186/s12916-019-1426-2

[29] Obaido, G., Adebayo, O., and Oduor, E. 2023. "Artificial Intelligence in African Health Systems: Opportunities and Challenges." Frontiers in Artificial Intelligence 6: 1145321. https://doi.org/10.3389/frai.2023.1145321

[30] Odhiambo, J. N., Makori, E., and Mboowa, G. 2022. "AI-Driven Pathology Diagnostics in Sub-Saharan Africa: Barriers and Enablers." Health Informatics Journal 28 (4): 14604582221105464. https://doi.org/10.1177/14604582221105464

[31] World Health Organization Regional Office for Africa. 2022. AI in Health Policy Paper: Harnessing Artificial Intelligence for Universal Health Coverage in Africa. Brazzaville: WHO AFRO. Retrieved from https://www.afro.who.int/publications. Accessed 7 July 2025.

[32] African Union Development Agency (AUDA-NEPAD). 2021. African Data Governance Framework. Johannesburg: AUDA-NEPAD. Retrieved from https://www.nepad.org/publication/african-data-governance-framework Accessed 5 September 2025.

[33] Tjoa, E., and Guan, C. 2020. "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI." IEEE Transactions on Neural Networks and Learning Systems 32 (11): 4793–4813. https://doi.org/10.1109/TNNLS.2020.3021756

[34] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., et al. 2017. "Mastering the Game of Go Without Human Knowledge." Nature 550 (7676): 354–359. https://doi.org/10.1038/nature24270

[35] Baidoo, C. A., Agyekum, A., and Nkansah, A. 2023. "Towards Trustworthy AI in Healthcare: Policy and Regulatory Perspectives from Africa." AI & Society. Advance online publication. https://doi.org/10.1007/s00146-023-01752-7

[36] European Commission. 2021. Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). Brussels: European Commission. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence. Accessed 17 May 2025.

[37] Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., and Chen, I. Y. 2021. "A Review of Challenges and Opportunities in Machine Learning for Health." AMIA Summits on Translational Science Proceedings 2021: 191–200. https://doi.org/10.1101/2021.02.04.429749

[38] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I. 2015. "Machine Learning Applications in Cancer Prognosis and Prediction." Computational and Structural Biotechnology Journal 13: 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

[39] Bhattacharyya, A., Bhatia, M., and Sharma, R. 2023. "Self-Supervised Learning in Oncology: Emerging Trends and Future Directions." Cancers 15 (8): 2109. https://doi.org/10.3390/cancers15082109

[40] Lee, C., Yoon, J., and van der Schaar, M. 2020. "Dynamic Treatment Regimes via Deep Reinforcement Learning." Journal of Machine Learning Research 21 (22): 1–45. https://www.jmlr.org/papers/volume21/18-713/18-713.pdf

[41] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." Scientific Data 3: 160018. https://doi.org/10.1038/sdata.2016.18

[42] Collins, G. S., Dhiman, P., Andaur Navarro, C. L., Ma, J., Hooft, L., Reitsma, J. B., Moons, K. G. M., and Riley, R. D. 2023. "Reporting and Validating Clinical Prediction Models: TRIPOD-AI and PROBAST-AI Frameworks." BMJ 382: e073645. https://doi.org/10.1136/bmj-2023-073645

[43] Oxford Insights. 2022. Government AI Readiness Index 2022. Oxford, UK: Oxford Insights. Retrieved from https://www.oxfordinsights.com/government-ai-readiness-index-2022. Accessed 2 July 2025.

[44] UNESCO. 2021. Recommendation on the Ethics of Artificial Intelligence. Paris: United Nations Educational, Scientific and Cultural Organization. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000380455. Accessed 5 September 2025.

[45] Zhang, Y., Zhu, J., and Wang, F. 2024. "Hybrid Symbolic–Neural Architectures for Oncology Diagnostics: Integrating Domain Knowledge with Deep Learning." Artificial Intelligence in Medicine 149: 102752. z

[46] van der Schaar, M., Alaa, A. M., Floto, A., and Ghassemi, M. 2023. "Causal Machine Learning for Precision Oncology." Nature Reviews Cancer 23 (3): 161–179. https://doi.org/10.1038/s41568-022-00528-1

[47] Wiysonge, C. S., and Ndumbe-Eyoh, S. 2023. "Digital Health in Africa: Building Resilient Health Systems for the Future." The Lancet Global Health 11 (1): e13–e15. https://doi.org/10.1016/S2214-109X(22)00545-4

[48] African Health Observatory. 2022. Digital Health Atlas for Africa: Current Trends and Future Directions. Brazzaville: WHO Regional Office for Africa. Retrieved from https://aho.afro.who.int/en. Accessed 5 September 2025.

[49] Dwivedi, Y. K., Hughes, D. L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., et al. 2021. "Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice, and Policy." International Journal of Information Management 57: 101994. https://doi.org/10.1016/j.ijinfomgt.2019.101994

[50] Cohen, J. P., Dao, L., Roth, K., Morrison, P., Bengio, Y., and Guo, Y. 2022. "Predictive and Causal Models in Medical AI: Challenges and Opportunities." Patterns 3 (7): 100539. https://doi.org/10.1016/j.patter.2022.100539