Comment

AI-driven biomarker discovery: enhancing precision in cancer diagnosis and prognosis

Esther Ugo Alum[1] 🆔

Abstract

Cancer remains a significant health issue, resulting in around 10 million deaths per year, particularly in developing nations. Demographic changes, socio-economic variables, and lifestyle choices are responsible for the rise in cancer cases. Despite the potential to mitigate the adverse effects of cancer by early detection and the implementation of cancer prevention methods, several nations have limited screening facilities. In oncology, the use of artificial intelligence (AI) represents a transformative advancement in cancer diagnosis, prognosis, and treatment. The use of AI in biomarker discovery improves precision medicine by uncovering biomarker signatures that are essential for early detection and treatment of diseases within vast and diverse datasets. Deep learning and machine learning diagnostics are two examples of AI technologies that are changing the way biomarkers are made by finding patterns in large datasets and making new technologies that make it possible to deliver accurate and effective therapies. Existing gaps include data quality, algorithmic transparency, and ethical concerns around privacy, among others. The advancement of biomarker discovery methodologies with AI seeks to transform cancer by improving patient survival rates through enhanced early diagnosis and targeted therapy. This commentary aims to clarify how AI is improving the identification of novel biomarkers for optimal early diagnosis, focused treatment, and improved clinical outcomes, while also addressing certain obstacles and ethical issues related to the application of artificial intelligence in oncology. Data from reputable scientific databases such as PubMed, Scopus, and ScienceDirect were utilized.

Highlights

- AI, deep learning and machine learning models in particular, is revolutionizing biomarker discovery by identifying complex, non-intuitive patterns from vast multiple technologies that explain biological information at molecular levels, enhancing the precision of cancer screening and prognosis.
- AI biomarkers give information about the patient's reaction to a treatment especially in immunotherapy, help in the cancer therapy and in predicting the progression of a disease and response to treatment.
- Nevertheless, biomarker discovery utilizing AI model has some issues, such as, a need for large and high quality data, understanding of the model, privacy, and ensuring diversity to avoid exacerbating health disparities. By addressing these issues, AI's potential in biomarker discovery and enhanced cancer prognosis could revolutionize oncology, enhancing patient outcomes and survival rates, despite these challenges.

✉ Esther Ugo Alum, esther.alum@kiu.ac.ug; alumesther79@gmail.com | [1]Department of Research and Publications, Kampala International University, P. O. Box 20000, Kampala, Uganda.

1 Introduction

Cancer has turned out to be a health danger of great concern around the world, with about 10 million deaths occurring due to this ailment each year [1]. Increased population growth, rising life expectancy, and risk factors like tobacco consumption, poor nutrition, and pollution fuel cancer [2]. However, low- and middle-income countries shoulder and unequal burden, recording 70% of cancer-related mortalities. Some of the frequently affected portions of the human body by cancer are the lungs, breasts, colon, brain, and stomach. Lifestyle changes such as smoking cessation, vaccinations, and dietary changes can significantly reduce the incidence of cancer disease [3]. While it is possible, and indeed necessary, to detect diseases in an easily recognizable stage, screening programs are not available or implemented effectively in all countries. The global cancer burden has increased, necessitating the development of safe, cost-effective, accessible, and effective methods to prevent the disease [4]. The introduction of artificial intelligence (AI) in oncology has created a new strategy to cancer diagnosis, prognosis, and treatment. This change has come about due to the growing need to practice precise medicine. The search for an appropriate treatment takes into account not only the disease but also the individual, their genetics, environment, lifestyle, and other factors [5]. One of AI's most constructive roles is the search for biomarkers, which is one of the most critical elements in precision oncology. For quite a long time, cancer diagnostics and treatment planning have used biomarkers, which are defined as genes, proteins, or other substances that inform on the presence or evolution of a disease. However, most of the conventional biomarker discovery approaches are hypothesis-driven, which is not practical given the intricacy of cancer biology and the quantity of data that modern 'omics' technologies bring [6]. The rapid integration of AI, principally deep learning and machine learning, is revolutionizing the development of biomarkers by enabling the derivation of ideas from patterns in big datasets and enhancing the development of new technologies that facilitate the provision of precise and effective therapies [7]. Recently, Barioni et al. [8], in a systematic review, reported that the above tools demonstrated potential in enhancing cancer detection, diagnosis, prognosis, and therapeutic strategies. However, challenges persist, particularly in some cancers like ovarian and pancreatic cancer, where limited data availability and quality exist [9]. To deal with these problems, researchers are coming up with new ways to use computers. For example, Rodland et al. [10] described specific vocabularies for AI research that protect patient privacy, and advanced Bayesian algorithms for early cancer detection using longitudinal measures. Despite the potential, we must resolve concerns about AI transparency, explainability, and trustworthiness to ensure its effective deployment in clinical environments [9]. This commentary focuses on the technology of AI-based biomarker discovery, aiming to comprehend the impact of biomarker discovery on oncology, specifically in terms of cancer diagnosis and prognosis. It aims to clarify how AI is enhancing the discovery of new biomarker signatures for effective early diagnosis, targeted therapy, and favorable clinical outcomes while addressing some of the barriers and ethical dilemmas surrounding the use of artificial intelligence in a medical setting.

2 The promise of AI in uncovering novel biomarkers

AI makes it easier to use advanced algorithms to look at huge genomic, transcriptomic, proteomic, and metabolomic datasets. This fundamentally changes the way biomarker identification is usually done. Unlike classical approaches based on hypothesized hypotheses, AI-based models uncover innovative and surprising connections within high-dimensional datasets that common statistical methods could easily miss [11]. How deep learning models demonstrate expertise in decoding complex data patterns is one example. These models decode information from a variety of datasets, such as tumor biopsies samples, blood tests, and medical images, to identify biomarkers associated with patient outcomes, such as survival rates and medication reactions [12]. Bioinformatics techniques are used by AI-driven technologies like PandaOmics to look at multimodal omics data. This helps find the therapeutic targets and biomarkers that are needed in cancer care. These platforms use high-throughput technologies to amass extensive biological datasets over the last two decades [13]. Researchers can uncover intricate correlations within these datasets, typically overlooked by traditional analytical techniques, facilitated by advanced statistical methods and machine learning algorithms. For instance, explainable AI (XAI) frameworks are particularly significant. They enrich the interpretability of AI frameworks, helping clinicians better recognize the connection between particular biomarkers and patient results [14]. A study showcases an XAI-based deep learning framework for biomarker discovery in non-small cell lung cancer (NSCLC), demonstrating how

explainable models can assist in clinical decision-making. This strategy improves diagnosis accuracy and boosts health professionals' confidence in AI-generated results [15]. In biomarker research, accuracy refers to the ability of a biomarker test or tool to correctly identify or measure the intended biological condition or state. It reflects how closely the test results align with the true value or actual condition, encompassing both sensitivity (true positive rate) and specificity (true negative rate). High accuracy ensures reliable and valid results for diagnostic, prognostic, or therapeutic purposes.

The use of artificial intelligence also creates chances to bring in data from various 'omics'—genomics, epigenomics, and proteomics—to understand multi-omics biomarkers that provide more complete insights into tumor biology. Adopting an integrative method, we might attain exceptional information surrounding cancer pathways and mechanisms, which may possibly result in better, personalized therapies. Using AI to coordinate and assess complicated data forms will help researchers distance themselves from single-biomarker strategies and advance their exploration of cancer biology in its entirety.

3 Enhancing diagnostic precision with AI-derived biomarkers

The integration of AI into biomarker research is expected to significantly improve cancer diagnosis precision. Although typically valuable, biomarker-based diagnostics historically face challenges with both sensitivity and specificity. Sensitivity refers to the ability of a biomarker to correctly identify individuals with a particular condition (true positive rate). A highly sensitive test minimizes false negatives. Specificity refers to the ability of a biomarker to correctly identify individuals without the condition (true negative rate). A highly specific test minimizes false positives. Both sensitivity and specificity are critical for evaluating the diagnostic accuracy of biomarkers. For example, screening methods currently used for cancer, such as mammography for breast cancer or prostate specific antigen testing for prostate cancer, face challenges with false readings that lead to either overtreatment or missed diagnoses [16]. By contrast, AI-driven models have the capability to detect highly precise biomarker signatures linked to different cancer subtypes, enhancing diagnostic methods. By interpreting complex histopathological images, AI can further elevate diagnostic accuracy [17]. AI techniques have proven effective in categorising cancer cases and healthy individuals, with ramifications for precision medicine and tailored cancer therapies [18]. Next-generation sequencing (NGS) technologies and AI have changed precision oncology by making it easier to assess risk, diagnose cancer early, and find biomarkers [19]. As shown by Ozaki et al. [20], AI applications can help find cancer earlier and predict how it will progress by analysing multi-omics data, radiomics, pathomics, and clinical records. Machine learning and deep learning models have demonstrated superior efficacy in categorizing cancer types and stages, especially for breast, lung, brain, and skin cancers [21]. Deep learning algorithms, trained on a vast collection of histological images, have consistently demonstrated remarkable accuracy in identifying cancerous tissues, often surpassing the performance of human pathologists. By combining these image-based analyses with biomarker data, AI can provide a more comprehensive diagnostic method that significantly improves early detection and diagnostic accuracy. Lung cancer diagnosis using AI-enhanced diagnostics has shown promising results in analyzing radiological and genomic data, which could significantly improve patient outcomes through earlier identification and focused intervention [22].

4 Prognostic value of AI-discovered biomarkers

The prognostic potential of AI-derived biomarkers is of considerable importance in predicting patient outcomes and informing therapeutic choices. Oncologists can make more informed treatment decisions using models based on biomarkers and AI, which can predict the likely response of patients to specific therapies. It is especially important within the field of cancer immunotherapy, as patient responses are unpredictably variable. AI can pinpoint biomarker signatures, which help to determine certain patients who are more predisposed to react to immunotherapies like checkpoint inhibitors, thus aiding customized and more effective treatment plans [23]. Furthermore, prognostic biomarkers based on AI provide dynamic insights into the evolution of cancer. Over time, AI systems can pick up on minor changes seen in patient data—including the levels of circulating tumor DNA or RNA—allowing detection of disease recurrence or treatment resistance before these conditions become clinically detectable [24]. The power to foresee results in real time gives oncologists the opportunity to alter treatment approaches in a proactive manner, which could enhance both survival statistics and the standard of living for cancer patients. In oncology, distinguishing predictive from prognostic biomarkers is critical. Predictive biomarkers report how effective a treatment is for a patient, whereas prognostic biomarkers indicate

the likely outcome of a case independent of treatment. The most recent advancements in this area have resulted in the establishment of the Predictive Biomarker Modeling Framework (PBMF), which uses contrastive learning to systematically extract predictive biomarkers from rich clinical data [25]. Retrospective studies have demonstrated the potential of this framework, revealing significant improvements in patient survival rates through its predictive capabilities.

AI models can amalgamate several data modalities, including radiography, histology, genomics, and electronic health records, to enhance diagnostic precision and reliability [26]. These models can identify novel patterns across modalities, potentially resulting in new biomarkers and therapeutic targets [26]. Oncoimaging and oncopathology are developing AI algorithms for cancer screening, tumour characterisation, and clinical decision-making [27]. Radiomics, pathomics, and radiogenomics are nascent disciplines that use AI to derive significant insights from medical imaging and associate them with biochemical pathways [27].

5 The integration of image-based and molecular AI-driven biomarkers in cancer diagnosis and prognosis

The integration of image-based and molecular AI-driven biomarkers in cancer diagnosis and prognosis represents a promising frontier in precision oncology. By combining insights from medical imaging (such as radiology and pathology images) with molecular data (like genetic and genomic profiles), this integrated approach provides a comprehensive view of cancer, improving diagnostic accuracy and prognostic predictions [28].

Image-based AI Biomarkers: AI-driven image analysis, particularly through techniques like radiomics and deep learning, extracts valuable features from medical imaging (e.g., CT scans, MRIs, or histopathological slides). These biomarkers can reveal tumor heterogeneity, early signs of malignancy, and subtle patterns that may be missed by the human eye [29]. For instance, AI can identify tumor boundaries, estimate tumor size, and assess treatment response, which are crucial for staging and monitoring progression [17].

Molecular AI Biomarkers: Molecular biomarkers involve the analysis of genetic, epigenetic, and proteomic data to identify cancer-specific alterations. AI algorithms, especially those based on machine learning, can process vast datasets to identify genetic mutations, identify tumor subtypes, and predict responses to targeted therapies [30]. Molecular biomarkers are often used for personalized treatment, guiding clinicians to select the most appropriate therapies based on the tumor's genetic profile.

5.1 Integrated approach

By combining both image-based and molecular biomarkers, AI can correlate visual patterns with underlying genetic and molecular alterations. This fusion enables:

  i. *More accurate diagnoses*: AI can integrate both imaging and molecular data to detect early-stage cancers and identify hard-to-detect tumors [31].
  ii. *Improved prognosis prediction*: The integrated models predict cancer progression and metastasis more accurately, guiding treatment decisions and patient management [32].
 iii. *Personalized treatment*: The combination of imaging and molecular data helps tailor therapies to individual patients, improving efficacy and reducing unnecessary side effects [33].

5.2 Real-world case studies where AI-driven oncology biomarkers have proven efficient

AI-driven oncology biomarkers have shown improvements in clinical outcomes. Below are a few real-world case studies:

1. *AI-Driven Radiomics for Early Detection of Lung Cancer:* A study used AI algorithms to analyze radiomic features from CT scans to identify early-stage lung cancer biomarkers. The AI model demonstrated improved early detection rates compared to traditional imaging methods. It was able to identify subtle patterns in CT images that were overlooked by human clinicians, enabling earlier intervention and better survival rates [34].
2. *AI in Breast Cancer Diagnosis:* PathAI, a company specializing in AI-driven diagnostics, developed algorithms to analyze histopathological images for breast cancer detection. PathAI's AI models achieved higher accuracy than pathologists

in identifying breast cancer, especially in detecting smaller or less obvious malignancies. The AI system also helped reduce diagnostic errors, particularly in distinguishing between benign and malignant tissues. Improved diagnostic precision led to better treatment decisions, and faster, more accurate results were linked to enhanced patient care and outcomes [35].

3. *AI-Powered Genomic Profiling in Precision Oncology*: Tempus, a technology company, developed an AI-driven platform that integrates genomic data with clinical outcomes to help identify effective treatments for cancer patients. By analyzing tumor genomic profiles, Tempus' platform identified biomarkers that guided personalized treatment plans. In clinical trials, patients whose treatments were informed by AI-driven genomic profiling showed improved responses to targeted therapies and immune checkpoint inhibitors, leading to better response rates and survival outcomes for patients with various types of cancer [36].

4. *AI in Prostate Cancer*: IBM Watson for Oncology was used to analyze large volumes of medical data, including electronic health records (EHR) and clinical trials, to assist in the diagnosis and treatment of prostate cancer. In China, Watson for Oncology provided oncologists with evidence-based treatment recommendations that aligned with best practices [37].

5. *AI for Liquid Biopsy in Colorectal Cancer: Guardant Health*: Guardant Health developed an AI-powered liquid biopsy test (Guardant360) that analyzes tumor DNA in blood samples to detect actionable mutations in colorectal cancer. The Guardant360 test, powered by AI, helped identify specific mutations and guide treatment decisions for colorectal cancer patients. It was found to be highly effective in detecting mutations and monitoring treatment responses with minimal invasiveness [38].

6 Artificial intelligence (AI) and machine learning (ML) in clinical trials in oncology studies

AI and ML are progressively incorporated into oncology research and clinical practice. These technologies demonstrate potential in multiple facets of cancer care, encompassing early detection, diagnosis, classification, and therapy planning [39]. AI applications encompass several data modalities, including imaging, genomics, and medical records, and are being created for the four predominant cancer types. Notable AI subtypes that have been remarkably useful are neural networks and Bayesian methods. AI is the central domain, while neural networks and Bayesian methods are specific techniques within AI with distinct approaches. While neural networks are data-driven, Bayesian methods are probabilistic and leverage prior information. Recent research has concentrated on employing neural networks for cancer prediction utilising microarray data, encompassing gene expression filtering, cancer presence or type prediction, and unlabelled sample clustering [40]. Neural networks can integrate various biological scales, from genotype to phenotype, facilitating the comprehension of heterogeneity's role in cancer progression and the microenvironment's influence on evolutionary dynamics [41]. These networks have been utilised across multiple cancer types, including prostate, breast, lymphatic, and head and neck cancers [42]. Although neural networks exhibit potential in cancer research, challenges persist in identifying optimal network architectures and hyperparameters, often necessitating trial-and-error methodologies [40]. Bayesian methods have been prominent in cancer research because of their capacity to manage complicated statistical models and integrate prior information [43]. These strategies are very effective in the analysis of high-dimensional omics data for cancer prognosis and prediction, facilitating multi-omics integration and enhancing predictive performance [43]. Researchers have used Bayesian spatial–temporal models to study cancer incidence and death trends across time and geography, primarily using generalised linear mixed models [44]. AI approaches in clinical trials provide benefits over conventional statistics by enabling direct assertions regarding treatment success and effectively managing intricate trial designs [45]. Despite several constraints, including the necessity for prior information, AI methods offer significant resources for oncologists across multiple facets of cancer research, encompassing prognostic modeling and the design and analysis of clinical trials [46, 47]. Recent improvements concentrate on the incorporation of AI into clinical practice for the four predominant cancer types, involving tasks such as detection, diagnosis, and treatment planning across diverse data modalities [48]. Notwithstanding these encouraging advancements, difficulties persist, such as data transparency, interpretability, and potential biases, which must be resolved for the extensive implementation of AI in oncology [49]. Recent initiatives have highlighted AI's capability in forecasting emergency hospital admissions, and sudden death in clinical trials [50]. A review of randomized controlled trials indicated that AI and ML methodologies were predominantly investigated in prostate, colorectal, and lung cancers, with artificial neural networks being the most frequently employed algorithm [51]. Although AI demonstrates significant potential in oncology, its present applications

Discover

are predominantly centered on diagnostic and prognostic services, highlighting the necessity for additional research and development in treatment-related applications. Table 1 is a brief comparison of AI methodologies and traditional methods.

7 Hard computing-based algorithms in cancer management

Hard computing is a deterministic, rule-based algorithm intended for precise and logical problem solving. These algorithms rely on strict mathematical models and exact solutions, and so they are very powerful for problems where accuracy and reproducibility are of the essence [52]. While AI-based methods, particularly deep learning and machine learning methods, have revolutionized biomarker discovery, it's important to acknowledge the continued usefulness of hard computing-based algorithms for biomarker discovery in oncology. These deterministic, rule-based approaches are powerful and provide complementary benefits in clinical practice [53]. In cancer management, hard computing-based algorithms play a crucial role in the following areas:

i.   *Medical Imaging and Diagnosis*: Algorithms such as support vector machines (SVMs), decision trees, and linear regression models are used in analysing medical images such as magnetic resonance imaging, computerized tomography, and positron emission tomography scans to detect tumors, classify cancer types, and predict malignancy. These algorithms perform best at pattern recognition when the input data is well-structured and labeled [54].

ii. *Molecular Data Analysis*: Hard computing techniques are applied to analyse genetic and proteomic data, aiding identification of mutations and biomarkers linked with specific cancers. For example, logistic regression and principal component analysis are utilized to study gene expression profiles and stratify patients [55].

iii. *Treatment Optimization*: Rule-based algorithms assist in optimising radiation therapy and chemotherapy dosages by simulating treatment scenarios based on patient-specific parameters [56]. This ensures efficient tumor targeting while reducing damage to nearby healthy tissues.

iv. *Predictive Modeling*: Hard computing algorithms are utilized to create predictive models for cancer prognosis. For example, algorithms like Cox proportional hazards models assess survival rates based on clinical data, enabling personalized treatment planning [57].

v.   *Clinical Decision Support Systems (CDSS)*: Deterministic algorithms are implanted in CDSS to provide cancer specialists with evidence-based recommendations for diagnosis, treatment, and follow-up care, enhancing decision-making accuracy [58].

7.1 Advantages of hard computing in cancer applications

i.   *Efficiency and Resource Optimization*: Many hard computing algorithms (such as SVMs, and decision trees), and statistical models (such as linear regression and logistic regression) require significantly fewer computational resources compared to deep learning models. Their deterministic ability ensures speedy execution times, making them ideal for resource-constrained environments [59].

ii. *Data Requirements*: Unlike AI models that often demand large, high-quality datasets for training, hard computing-based methods can produce reliable results with minimal datasets, provided the quality of the data is high and well-structured [60]. This is particularly useful in situations where data availability is limited, such as rare cancer types.

iii. *Automation and Workflow Integration*: The automation of many hard computing techniques have streamlined their integration into clinical settings. For instance, rule-based systems and Bayesian methods are increasingly used in clinical decision support systems for cancer prognosis and treatment optimization [61].

iv. *Interpretability*: Hard computing algorithms offer a level of transparency that improves understanding and trust among clinicians and patients. Their results can often be traced back to specific rules or statistical principles, avoiding the "black-box" issues associated with AI-based approaches [62].

7.2 Limitations of hard computing in cancer management

The transformative potential of AI is confronted with challenges of large datasets, high computational power requirements and more complex architectures when it comes to real world clinical applications. This is addressed by complementing

Table 1 A brief comparison of AI methodologies and traditional methods

S/N	Feature	AI methodologies	Traditional methods
1	Nature	Probabilistic and data-driven	Empirical-based
2	Flexibility	Adaptive to changes and unstructured data	Limited flexibility; often static approaches
3	Handling of Uncertainty	Effective, leveraging probabilistic models	Typically not designed for uncertainty
4	Key Techniques	Machine learning, neural networks, deep learning	Observation-based, statistical or logical rules
5	Data Requirements	Large datasets; quality impacts performance	Varies, often smaller-scale datasets
6	Speed of Execution	High, depending on computational power	Moderate to slow
7	Scalability with Data	Highly scalable with increased data	Limited scalability with increasing data
8	Examples	Image recognition, natural language processing	Linear regression, traditional statistical tests
9	Error Tolerance	Can handle and learn from errors	Moderate, depending on the method
10	Real-World Adaptability	Excellent for dynamic, complex environments	Moderate, depending on the context
11	Innovation Potential	Enables discovery of new patterns and insights	Moderate, rooted in existing knowledge
12	Complexity of Implementation	High, often requiring advanced computational expertise	Moderate, depending on the domain
13	Human Intervention	Minimal after training; models often self-operate	Often requires significant human involvement
14	Adaptation to New Domains	Highly adaptable through retraining	Limited, often domain-specific
15	Reliance on Data Quality	Strong reliance; poor data leads to poor outcomes	Data quality is important but less critical

these tools with increasing frequency by soft computing techniques as neural networks and fuzzy logic, which are more suitable for dealing with complexities and ambiguity [63]. As such, hard computing approaches serve to encode balanced simplicity, interpretability and economic computability, and hence play an important role under certain circumstances.

8 Challenges and ethical considerations

Despite its potential for change, finding biomarkers through AI faces a variety of challenges.

i. *Data-related challenges:* An important challenge is the need for big, top-quality datasets. The quality of training datasets directly impacts the strength of AI models, and the presence of diverse, representative data in many regions is consistently limited [60]. Ethical hurdles appear in the context of biomarker discovery powered by AI, especially concerning data privacy and bias. The inclusion of large numbers of patient data points in AI development raises concerns about the security and privacy of important medical data [64]. In addition, when the data used to train AI models does not include enough diversity, the biomarkers revealed may not be applicable in all settings, which could worsen existing health disparities [65]. Statistical uncertainties in AI cancer diagnostics can be caused by data variability, sample sizes, and model assumptions. Management strategies include robust data collection, cross-validation, uncertainty quantification, explainable AI, and continuous learning. These practices improve diagnostic reliability and trust in AI systems.

ii. *Interpretability and Transparency:* The opaque "black-box" nature of AI models affecting clinical trust. The hard-to-decipher aspects of multiple AI algorithms, in particular, create difficulties in understanding and clinical implementation [66]. Oncology life-or-death choices require clinicians to have faith in AI model decision-making.

iii. *Regulatory and Validation Barriers:* there are challenges in transitioning AI-identified biomarkers into clinically validated tools. Validating technologies for AI-driven biomarker identification is critical to ensuring that these methods are reliable and suitable for clinical use. Even though several AI models have not gone through extensive experimental verification, the question of generalizability across various patient segments arises. Additionally, the integration of these technologies into established clinical workflows presents logistical challenges that require resolution to promote widespread adoption [67]. More so, navigating complex approval pathways for AI-driven methodologies delays its application [68].

iv. *Ethical Considerations*: The use of biomarker tools and AI technologies in healthcare raises ethical concerns such as patient consent, accountability for errors, and addressing disparities. These concerns aim to ensure that patients are fully informed about how their data will be used, establishing clear responsibility for any errors, and equity in healthcare by addressing disparities across socio-economic and geographic populations. These considerations ensure trust, fairness, and safety in the application of advanced medical technologies [69].

9 Procedure to address ethical and data privacy concerns

Recent studies have concentrated on resolving data privacy and bias concerns in cancer diagnosis. Tasci et al.[70] emphasise class imbalance as a critical issue in oncological datasets, suggesting methods to alleviate algorithmic bias. According to McGraw [71], to eradicate bias in radiology AI, there is need for different datasets and expert participation during the development as well as the implementation phase. Clinical informatics technologies, including real-world data analysis, natural language processing, and radiomics, are utilised to comprehend and mitigate cancer disparities among diverse demographic characteristics. Nonetheless, meticulous examination of algorithmic bias is essential to avert the exacerbation of preexisting inequalities [72]. Similarly, Malin and Goodman et al. [73] summarised research on reconciling data accessibility and privacy within medical informatics. According to them, addressing permission methodologies through consent strategy, privacy risk assessment, cryptography approaches for querying clinical data, and game theoretic method for publishing summary information, could mitigate data accessibility and privacy issues. These studies jointly underscore the necessity of resolving bias and privacy issues in cancer diagnostic data to guarantee equal and effective healthcare outcomes.

Recent studies emphasise measures to improve diversity in cancer clinical trials and research. Principal strategies encompass culturally sensitive outreach, education, and the dissemination of accessible research within local

communities [74]. Involving healthcare providers, community organisations, and other cancer survivors as stakeholders is essential [74]. Enhancing trial matching through the expansion of eligibility criteria, comprehensive patient screening, and diligent follow-up on matches can augment enrolment [75]. I recommend addressing barriers through site self-assessments, quality improvement plans, and tools like clinical trial checklists and patient navigators. Addressing the deficiency of diversity in cancer research necessitates enhancing funding accessibility for low- and middle-income countries, augmenting the involvement of under-represented scientists in editorial boards and conferences, and appropriately recognizing the contributions of low- and middle-income countries researchers [76]. These tactics aim to enhance representation and generalizability in cancer research.

10 Future directions

One can expect encouraging progress in biomarker discovery from AI within oncology, with several important advancements in the near future.

i. *Improving data privacy:* Developing global data-sharing frameworks while maintaining privacy. The precision of biomarker discovery will continue to rise as AI models become more sophisticated and access to larger, more diverse datasets increases.

ii. *Expanding explainable AI*: Developing explainable AI (XAI) models for better interpretability would foster trust among clinicians and patients.

iii. *Integration with other biomarkers*: The integration of AI with other pioneering technologies, such as CRISPR and single-cell sequencing, may trigger the discovery of earlier health issues in cancer biology, enabling a higher level of personalized cancer treatment [77].

iv. *Promoting collaboration among stakeholders to ensure equitable access to AI advancements*: Clinical practice can only use AI-derived biomarkers with confidence if there is collaboration between AI researchers, oncologists, and regulatory agencies. As these technologies develop, AI-driven biomarker discovery is very likely to become a key element of precision oncology, giving fresh opportunities for early detection, better prognostic predictions, and more tailored cancer treatments [78].

11 Conclusion

AI-driven innovation in biomarker identification signifies a pivotal advancement in precision oncology, presenting distinct opportunities to markedly improve cancer diagnosis and prognosis. Through the analysis of complex, paradoxical biomarker patterns derived from multi-omics data, AI can transform cancer treatment from a broad approach to a personalized one. Despite ongoing challenges, the application of AI in biomarker identification has the potential to transform oncology, facilitating progress towards the ultimate objective of creating tailored cancer medicines that enhance patient outcomes and survival rates. This suggests optimism that the anticipated increase in cancer incidence will not correlate with mortality rates.

Declarations

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J Clin. 2024;74:229–63. https://doi.org/10.3322/caac.21834.
2. Alum EU, Ugwu OPC, Obeagu EI, Ugwu CN. Beyond conventional therapies: exploring nutritional interventions for cervical cancer patients. J Cancer Res Cell Ther. 2024;8(1):1–6.
3. Ibiam UA, Uti DE, Ejeogo CC, Orji OU, Aja PM, Nwamaka EN, Alum EU, Chukwu C, Aloke C, Chinedum KE, Agu P, Nwobodo, V.. In Vivo and in silico assessment of ameliorative effects of Xylopia aethiopica on testosterone propionate-induced benign prostatic hyperplasia. Pharmac Front. 2023;05:e64–76. https://doi.org/10.1055/s-0043-1768477.
4. Alum EU, Tufail T, Uti DE, Aja PM, Offor CE, Ibiam UA, Ukaidi CUA, Alum BN. Utilizing indigenous flora in East Africa for breast cancer treatment: an overview. Anti Cancer Agent Med Chem. 2025;24:1–15. https://doi.org/10.2174/0118715206338557240909081833.
5. Farina E, Nabhen JJ, Dacoregio MI, Batalini F, Moraes FY. An overview of artificial intelligence in oncology. Futur Sci OA. 2022. https://doi.org/10.2144/fsoa-2021-0074.
6. Passaro A, Al Bakir M, Hamilton EG, Diehn M, André F, Roy-Chowdhuri S, Mountzios G, Witsuba I, Swanton C, Peters S. Cancer biomarkers—emerging trends and clinical implications for personalized treatment. Cell. 2024;187:1617–35. https://doi.org/10.1016/j.cell.2024.02.041.
7. Chakraborty C, Bhattacharya M, Pal S, Lee S-S. From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare. Curr Res Biotechnol. 2024;7: 100164. https://doi.org/10.1016/j.crbiot.2023.100164.
8. Barioni CTS, de Barros Wandresen RP, Pereira LF, Coimbra AF, de Araújo Oliveira Kubo BB, da Cunha RC. Artificial intelligence for the identification of biomarkers in cancer prevention and diagnosis: advances and perspectives. Revista Brasileira De Cancerol. 2024. https://doi.org/10.2635/2176-9745.RBC.2024v70n2.4692.
9. Mikdadi D, O'Connell KA, Meacham PJ, Dugan MA, Ojiere MO, Carlson TB, Klenk JA. Applications of artificial intelligence (AI) in ovarian cancer, pancreatic cancer, and image biomarker discovery. Cancer Biomark. 2022;33:173–84. https://doi.org/10.3233/CBM-210301.
10. Rodland KD, Webb-Robertson BJ, Srivastava S. Introduction to the special issue on applications of artificial intelligence in biomarker research. Cancer Biomark. 2022;33:171–2. https://doi.org/10.3233/CBM-229001.
11. Dai X, Shen L. Advances and trends in omics technology development. Front Med (Lausanne). 2022. https://doi.org/10.3389/fmed.2022.911861.
12. Fawaz A, Ferraresi A, Isidoro C. Systems biology in cancer diagnosis integrating omics technologies and artificial intelligence to support physician decision making. J Pers Med. 2023;13:1590. https://doi.org/10.3390/jpm13111590.
13. Kamya P, Ozerov IV, Pun FW, Tretina K, Fokina T, Chen S, Naumov V, Long X, Lin S, Korzinkin M, Polykovskiy D, Aliper A, Ren F, Zhavoronkov A. PandaOmics: an AI-driven platform for therapeutic target and biomarker discovery. J Chem Inf Model. 2024;64:3961–9. https://doi.org/10.1021/acs.jcim.3c01619.
14. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, Guidotti R, Del Ser J, Díaz-Rodríguez N, Herrera F. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inf Fusion. 2023;99: 101805. https://doi.org/10.1016/j.inffus.2023.101805.
15. Dwivedi K, Rajpal A, Rajpal S, Agarwal M, Kumar V, Kumar N. An explainable AI-driven biomarker discovery framework for non-small cell lung cancer classification. Comput Biol Med. 2023;153: 106544. https://doi.org/10.1016/j.compbiomed.2023.106544.
16. Das S, Dey MK, Devireddy R, Gartia MR. Biomarkers in cancer detection, diagnosis, and prognosis. Sensors. 2024;24:37. https://doi.org/10.3390/s24010037.
17. Chen MM, Terzic A, Becker AS, Johnson JM, Wu CC, Wintermark M, Wald C, Wu J. Artificial intelligence in oncologic imaging. Eur J Radiol Open. 2022;9: 100441. https://doi.org/10.1016/j.ejro.2022.100441.
18. Bio-Marker Cancer Prediction System Using Artificial Intelligence. https://ieeexplore.ieee.org/document/10430291.
19. Dlamini Z, Francies FZ, Hull R, Marima R. Artificial intelligence (AI) and big data in cancer and precision oncology. Comput Struct Biotechnol J. 2020;18:2300–11. https://doi.org/10.1016/j.csbj.2020.08.019.
20. Ozaki Y, Broughton P, Abdollahi H, Valafar H, Blenda AV. Integrating omics data and AI for cancer diagnosis and prognosis. Cancers. 2024;16:2448. https://doi.org/10.3390/cancers16132448.
21. Khanam N, Kumar R. Recent applications of artificial intelligence in early cancer detection. Curr Med Chem. 2022;29:4410–35. https://doi.org/10.2174/0929867329666220222154733.
22. Cellina M, Cè M, Irmici G, Ascenti V, Khenkina N, Toto-Brocchi M, Martinenghi C, Papa S, Carrafiello G. Artificial intelligence in lung cancer imaging: unfolding the future. Diagnostics (Basel). 2022. https://doi.org/10.3390/diagnostics12112644.
23. Xu Z, Wang X, Zeng S, Ren X, Yan Y, Gong Z. Applying artificial intelligence for cancer immunotherapy. Acta Pharm Sin B. 2021;11:3393–405. https://doi.org/10.1016/j.apsb.2021.02.007.
24. Tébar-Martínez R, Martín-Arana J, Gimeno-Valiente F, Tarazona N, Rentero-Garrido P, Cervantes A. Strategies for improving detection of circulating tumor DNA using next generation sequencing. Cancer Treat Rev. 2023;119: 102595. https://doi.org/10.1016/j.ctrv.2023.102595.
25. Al-Tashi Q, Saad MB, Muneer A, Qureshi R, Mirjalili S, Sheshadri A, Le X, Vokes NI, Zhang J, Wu J. Machine learning models for the identification of prognostic and predictive cancer biomarkers: a systematic review. Int J Mol Sci. 2023;24:7781. https://doi.org/10.3390/ijms24097781.
26. Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, Vaidya AJ, Chen C, Zhuang L, Williamson DFK, Shaban M, Chen TY, Mahmood F. Artificial intelligence for multimodal data integration in oncology. Cancer Cell. 2022;40:1095–110. https://doi.org/10.1016/j.ccell.2022.09.012.

27. Majumder A, Sen D. Artificial intelligence in cancer diagnostics and therapy: current perspectives. Indian J Cancer. 2021;58:481. https://doi.org/10.4103/ijc.IJC_399_20.

28. Wu J, Mayer AT, Li R. Integrated imaging and molecular analysis to decipher tumor microenvironment in the era of immunotherapy. Semin Cancer Biol. 2022;84:310–28. https://doi.org/10.1016/j.semcancer.2020.12.005.

29. Koh D-M, Papanikolaou N, Bick U, Illing R, Kahn CE, Kalpathi-Cramer J, Matos C, Martí-Bonmatí L, Miles A, Mun SK, Napel S, Rockall A, Sala E, Strickland N, Prior F. Artificial intelligence and machine learning in cancer imaging. Commun Med. 2022;2:1–14. https://doi.org/10.1038/s43856-022-00199-0.

30. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in cancer research and precision medicine. Cancer Discov. 2021;11:900–15. https://doi.org/10.1158/2159-8290.CD-21-0090.

31. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, Allison T, Arnaout O, Abbosh C, Dunn IF, Mak RH, Tamimi RM, Tempany CM, Swanton C, Hoffmann U, Schwartz LH, Gillies RJ, Huang RY, Aerts HJWL. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin. 2019;69:127–57. https://doi.org/10.3322/caac.21552.

32. Zhang B, Shi H, Wang H. Machine learning and AI in cancer prognosis, prediction, and treatment selection: a critical approach. J Multidiscip Healthc. 2023;16:1779–91. https://doi.org/10.2147/JMDH.S410301.

33. Salih S, Elliyanti A, Alkatheeri A, AlYafei F, Almarri B, Khan H. The role of molecular imaging in personalized medicine. J Pers Med. 2023;13:369. https://doi.org/10.3390/jpm13020369.

34. Tunali I, Gillies RJ, Schabath MB. Application of radiomics and artificial intelligence for lung cancer precision medicine. Cold Spring Harb Perspect Med. 2021;11: a039537. https://doi.org/10.1101/cshperspect.a039537.

35. Soliman A, Li Z, Parwani AV. Artificial intelligence's impact on breast cancer pathology: a literature review. Diagn Pathol. 2024;19:38. https://doi.org/10.1186/s13000-024-01453-w.

36. Beaubier N, Bontrager M, Huether R, Igartua C, Lau D, Tell R, Bobe AM, Bush S, Chang AL, Hoskinson DC, Khan AA, Kudalkar E, Leibowitz BD, Lozachmeur A, Michuda J, Parsons J, Perera JF, Salahudeen A, Shah KP, Taxter T, Zhu W, White KP. Integrated genomic profiling expands clinical options for patients with cancer. Nat Biotechnol. 2019;37:1351–60. https://doi.org/10.1038/s41587-019-0259-z.

37. Liu C, Liu X, Wu F, Xie M, Feng Y, Hu C. Using artificial intelligence (Watson for oncology) for treatment recommendations amongst Chinese patients with lung cancer: feasibility study. J Med Int Res. 2018;20: e11087. https://doi.org/10.2196/11087.

38. Gupta R, Othman T, Chen C, Sandhu J, Ouyang C, Fakih M. Guardant360 circulating tumor DNA assay is concordant with Foundation-One next-generation sequencing in detecting actionable driver mutations in anti-EGFR naive metastatic colorectal cancer. Oncologist. 2020;25:235–43. https://doi.org/10.1634/theoncologist.2019-0441.

39. Chen Z-H, Lin L, Wu C-F, Li C-F, Xu R-H, Sun Y. Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine. Cancer Commun. 2021;41:1100–15. https://doi.org/10.1002/cac2.12215.

40. Daoud M, Mayo M. A survey of neural network-based cancer prediction models from microarray data. Artif Intell Med. 2019;97:204–14. https://doi.org/10.1016/j.artmed.2019.01.006.

41. Gerlee P, Kim E, Anderson ARA. Bridging scales in cancer progression: mapping genotype to phenotype using neural networks. Semin Cancer Biol. 2015;30:30–41. https://doi.org/10.1016/j.semcancer.2014.04.013.

42. Mahmood M, Al-Khateeb B, Alwash WM. A review on neural networks approach on classifying cancers. IAES Int J Artif Intell (IJ-AI). 2020;9:317–26. https://doi.org/10.11591/ijai.v9.i2.pp317-326.

43. Chu J, Sun NA, Hu W, Chen X, Yi N, Shen Y. The application of Bayesian methods in cancer prognosis and prediction. Cancer Genom Proteom. 2022;19:1–11. https://doi.org/10.21873/cgp.20298.

44. Wah W, Ahern S, Earnest A. A systematic review of Bayesian spatial–temporal models on cancer incidence and mortality. Int J Public Health. 2020;65:673–82. https://doi.org/10.1007/s00038-020-01384-5.

45. Lee SY. Using Bayesian statistics in confirmatory clinical trials in the regulatory setting: a tutorial review. BMC Med Res Methodol. 2024;24:110. https://doi.org/10.1186/s12874-024-02235-0.

46. Bhattacharjee A. Application of Bayesian approach in cancer clinical trial. World J Oncol. 2014;5:109–12. https://doi.org/10.14740/wjon842e.

47. Zhou T, Zhang J. Types and progress of clinical trial design for breast cancer: a narrative review. Trans Breast Cancer Res. 2023. https://doi.org/10.21037/tbcr-23-22.

48. Lotter W, Hassett MJ, Schultz N, Kehl KL, Van Allen EM, Cerami E. Artificial intelligence in oncology: current landscape, challenges, and future directions. Cancer Discov. 2024;14:711–26. https://doi.org/10.1158/2159-8290.CD-23-1199.

49. Shreve JT, Khanani SA, Haddad TC. Artificial intelligence in oncology: current capabilities, future opportunities, and ethical considerations. Am Soc Clin Oncol Educ Book. 2022. https://doi.org/10.1200/EDBK_350652.

50. Kang J, Chowdhry AK, Pugh SL, Park JH. Integrating artificial intelligence and machine learning into cancer clinical trials. Semin Radiat Oncol. 2023;33:386–94. https://doi.org/10.1016/j.semradonc.2023.06.004.

51. Mangat G, Pilkhwal N, Sharma S, Singh B. PCN91 penetration of artificial intelligence (AI) and machine learning (ML) in the randomized controlled setting of oncology. Val Health Reg Issues. 2020;22:S22. https://doi.org/10.1016/j.vhri.2020.07.141.

52. Sebastian AM, Peter D. Artificial intelligence in cancer research: trends, challenges and future directions. Life (Basel). 2022;12:1991. https://doi.org/10.3390/life12121991.

53. Askr H, Elgeldawi E, Aboul Ella H, Elshaier YAMM, Gomaa MM, Hassanien AE. Deep learning in drug discovery: an integrative review and future challenges. Artif Intell Rev. 2023;56:5975–6037. https://doi.org/10.1007/s10462-022-10306-1.

54. Kaifi R. A review of recent advances in brain tumor diagnosis based on AI-based classification. Diagnostics (Basel). 2023;13:3007. https://doi.org/10.3390/diagnostics13183007.

55. Yaqoob A, Musheer Aziz R, Verma NK. Applications and techniques of machine learning in cancer classification: a systematic review. Hum-Cent Intell Syst. 2023. https://doi.org/10.1007/s44230-023-00041-3.

56. Field M, Hardcastle N, Jameson M, Aherne N, Holloway L. Machine learning applications in radiation oncology. Phys Imag Radiat Oncol. 2021;19:13–24. https://doi.org/10.1016/j.phro.2021.05.007.

57. Cao P, Dun Y, Xiang X, Wang D, Cheng W, Yan L, Li H. Machine learning-based individualized survival prediction model for prognosis in osteosarcoma: data from the SEER database. Medicine (Baltimore). 2024;103:e39582. https://doi.org/10.1097/MD.0000000000039582.

58. Mazo C, Kearns C, Mooney C, Gallagher WM. Clinical decision support systems in breast cancer: a systematic review. Cancers. 2020;12:369. https://doi.org/10.3390/cancers12020369.

59. Ahmed SF, Alam MdSB, Hassan M, Rozbu MR, Ishtiak T, Rafa N, Mofijur M, Shawkat Ali ABM, Gandomi AH. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif Intell Rev. 2023;56:13521–617. https://doi.org/10.1007/s10462-023-10466-8.

60. Aldoseri A, Al-Khalifa KN, Hamouda AM. Re-thinking data strategy and integration for artificial intelligence: concepts, opportunities, and challenges. Appl Sci. 2023;13:7082. https://doi.org/10.3390/app13127082.

61. Hak F, Guimarães T, Santos M. Towards effective clinical decision support systems: a systematic review. PLoS ONE. 2022;17: e0272846. https://doi.org/10.1371/journal.pone.0272846.

62. Sadeghi Z, Alizadehsani R, Cifci MA, Kausar S, Rehman R, Mahanta P, Bora PK, Almasri A, Alkhawaldeh RS, Hussain S, Alatas B, Shoeibi A, Moosaei H, Hladík M, Nahavandi S, Pardalos PM. A review of explainable artificial intelligence in healthcare. Comput Electr Eng. 2024;118: 109370. https://doi.org/10.1016/j.compeleceng.2024.109370.

63. Zhu S, Yu T, Xu T, Chen H, Dustdar S, Gigan S, Gunduz D, Hossain E, Jin Y, Lin F, Liu B, Wan Z, Zhang J, Zhao Z, Zhu W, Chen Z, Durrani T, Wang H, Wu J, Zhang T, Pan Y. Intelligent computing: the latest advances, challenges and future. Intell Comput. 2023. https://doi.org/10.34133/icomputing.0006.

64. Ugwu OPC, Alum EU, Ugwu JN, Eze VHU, Ugwu CN, Ogenyi FC, Okon MB. Harnessing technology for infectious disease response in conflict zones: challenges, innovations, and policy implications. Medicine (Baltimore). 2024;103: e38834. https://doi.org/10.1097/MD.0000000000038834.

65. Zou J, Schiebinger L. Ensuring that biomedical AI benefits diverse populations. EBioMedicine. 2021. https://doi.org/10.1016/j.ebiom.2021.103358.

66. Marey A, Arjmand P, Alerab ADS, Eslami MJ, Saad AM, Sanchez N, Umair M. Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. Egypt J Radiol Nucl Med. 2024;55:183. https://doi.org/10.1186/s43055-024-01356-2.

67. Karalis VD. The integration of artificial intelligence into clinical practice. Appl Biosci. 2024;3:14–44. https://doi.org/10.3390/applbiosci3010002.

68. Yazdi M, Zarei E, Adumene S, Beheshti A. Navigating the power of artificial intelligence in risk management: a comparative analysis. Safety. 2024;10:42. https://doi.org/10.3390/safety10020042.

69. Farhud DD, Zokaei S. Ethical issues of artificial intelligence in medicine and healthcare. Iran J Public Health. 2021. https://doi.org/10.18502/ijph.v50i11.7600.

70. Tasci E, Zhuge Y, Camphausen K, Krauze AV. Bias and class imbalance in oncologic data—towards inclusive and transferrable AI in large scale oncology data sets. Cancers. 2022;14:2897. https://doi.org/10.3390/cancers14122897.

71. McGraw M. Report offers strategies for eliminating radiology artificial intelligence bias. Oncol Time. 2022;44:13. https://doi.org/10.1097/01.COT.0000891804.61259.a2.

72. Chaunzwa TL, del Rey MQ, Bitterman DS. Clinical informatics approaches to understand and address cancer disparities. Yearb Med Inform. 2022;31:121–30. https://doi.org/10.1055/s-0042-1742511.

73. Malin B, Goodman K. Between access and privacy: challenges in sharing health data. Yearb Med Inf. 2018;27:055–9. https://doi.org/10.1055/s-0038-1641216.

74. Napoles A, Cook E, Ginossar T, Knight KD, Ford ME. Chapter four—applying a conceptual framework to maximize the participation of diverse populations in cancer clinical trials. In: Ford ME, Watson DK, editors. Advances in cancer research. Cambridge: Academic Press; 2017. p. 77–94.

75. Guerra CE, Fleury ME, Byatt LP, Lian T, Pierce L. Strategies to advance equity in cancer clinical trials. Am Soc Clin Oncol Educ Book. 2022. https://doi.org/10.1200/EDBK_350565.

76. Molina-Aguilar C, Robles-Espinoza CD. Tackling the lack of diversity in cancer research. Dis Model Mechan. 2023. https://doi.org/10.1242/dmm.050275.

77. Bhat AA, Nisar S, Mukherjee S, Saha N, Yarravarapu N, Lone SN, Masoodi T, Chauhan R, Maacha S, Bagga P, Dhawan P, Akil AA-S, El-Rifai W, Uddin S, Reddy R, Singh M, Macha MA, Haris M. Integration of CRISPR/Cas9 with artificial intelligence for improved cancer therapeutics. J Transl Med. 2022;20:534. https://doi.org/10.1186/s12967-022-03765-1.

78. Precision Medicine, AI, and the Future of Personalized Health Care—PMC, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7877825/.