Advances in AI for Identifying Cancer Biomarkers

Intelligence Artificial para a Identificação de Biomarcadores na Prevenção e no Diagnóstico do Câncer: Avanços e Perspectivas

Carina Toledo Scoparo Barioni1; Renata Paes de Barros Wandresen2; Lucas Formicoli Pereira3; Amanda Franceschi Coimbra4; Bárbara Bruna de Araújo Oliveira Kubo5; Ricardo Corrêa da Cunha6

ABSTRACT

Introduction: The systematic analysis of cancer markers and the impact of artificial intelligence (AI) on early detection and therapeutic approach are crucial in today's medical field. Cancer represents a significant global burden of morbidity and mortality, making early identification of markers a priority for effective disease management. This study aims to explore recent advancements in the identification and characterization of cancer indicators, including genetic, molecular, protein, and imaging biomarkers. Objective: To analyze the latest advances in identifying and characterizing cancer indicators, covering a variety of biomarker types. Additionally, to investigate the role of AI in improving and applying methods for cancer detection, diagnosis, prognosis, and treatment, highlighting its significant contributions to enhancing the accuracy and efficiency of these approaches. Method: A systematic literature review was conducted, selecting relevant studies addressing the identification of cancer biomarkers and the use of AI in this context based on specific inclusion and exclusion criteria. Results: The results of this systematic analysis highlight recent advances in identifying and characterizing cancer indicators, as well as the impact of AI on enhancing detection, diagnosis, prognosis, and treatment approaches. Conclusion: This study offers valuable insights into the role of cancer indicators and AI in disease prevention and management, supporting evidence-based clinical practices and promoting the development of more efficient and personalized healthcare approaches.

Key words: Neoplasias/epidemiology; Artificial Intelligence; Early Detection of Cancer/methods; Early Detection; Biomarkers/analysis.

RESULTS:

The results of this systematic analysis highlight recent advances in identifying and characterizing cancer indicators, as well as the impact of AI on enhancing detection, diagnosis, prognosis, and treatment approaches. This study offers valuable insights into the role of cancer indicators and AI in disease prevention and management, supporting evidence-based clinical practices and promoting the development of more efficient and personalized healthcare approaches.
INTRODUCTION

Cancer is one of the leading global causes of death due to late detection and severity, with a reserved prognosis, limited therapeutic options, and often ineffective5. The lack of validated screening procedures contributes to low patient adherence and unnecessary complementary tests, increasing healthcare costs. Therefore, there is a great need for innovative, accurate, and minimally invasive tools for early cancer detection4. The increasing complexity of biomarkers in oncology leads to time-consuming and costly decisions in cancer care and treatment. With cancer being one of the leading causes of death, there is a growing focus on precision medicine to personalize treatment and improve patient outcome predictions2,3.

Digital image analysis in pathology, especially using deep learning (DL) platforms and artificial intelligence (AI), can identify features not visible to pathologists, improving disease modeling and stage and outcome predictions. Integrating image analysis with transcriptomic and clinical information using AI is crucial for advancing personalized medicine2,4.

Currently, diagnosing cancer after abnormal findings in screening procedures typically requires tissue biopsy for histopathological evaluation. However, this approach presents several disadvantages, such as invasiveness, difficulty accessing certain tumors, limitations in early detection and disease monitoring, and lack of tumor heterogeneity representation4.

To overcome these challenges, liquid biopsies have emerged, analyzing cancer-related markers in bodily fluids such as blood or urine. These biopsies comprise various analytes, including cell-free DNA (cfDNA), cell-free RNA (cfRNA), circulating tumor cells (CTCs), extracellular vesicles (EVs), tumor-educated platelets (TEPs), proteins, and metabolites, allowing the identification of tumor information and real-time tumor burden monitoring. Additionally, liquid biopsies have the potential to detect cancer even in early stages, when symptoms are not present or tumor masses are not detectable by imaging techniques5.

Considering the challenges of current screening methods, a blood test combined with AI, capable of simultaneously detecting various types of cancer in early stages, and even being applied in high-risk population screening, would be a clinically valuable tool5. This pan-cancer approach must have high sensitivity for early detection, high specificity to avoid false positives, and the ability to discriminate the tissue of origin of the detected cancer.

Diagnosing cancer in the early stage is crucial to increase the chances of effective treatment. This includes screening high-risk patients and quickly investigating symptoms. Machine learning, where computers learn patterns of complex data to make predictions, has the potential to revolutionize early cancer diagnosis5.

Within the perspective of machine learning, AI is being applied to early cancer detection, yet such practice is a complex and altruistic challenge in medicine. Its use can expedite and enhance diagnosis, reducing human errors and costs, especially in screening programs. AI algorithms can identify data imperceptible to the human eye in radiological and pathological images, leading to a deeper understanding of cancers2-5.

These advances have the potential to personalize oncological practice and develop new imaging biomarkers. AI in oncoimaging and oncopathology can be used for screening, tumor characterization, and clinical decision-making. However, it is not infallible and cannot completely replace humans, but it can complement insights and deeper understanding. Several issues, such as standardization, ethics, privacy, need to be addressed before the full potential of AI in cancer diagnosis can be achieved6.

The general objective of this systematic review is to investigate and critically synthesize the available evidence on cancer markers and the role of AI in cancer prevention and treatment. This includes the analysis of genetic, molecular, protein, and imaging biomarkers associated with different types of cancer, as well as the assessment of the impact of AI on disease detection, diagnosis, and prognosis. The aim is to provide a comprehensive and updated overview of the subject, identifying knowledge gaps, highlighting significant advances, and discussing challenges and perspectives for the implementation of these technologies in clinical practice.

METHOD

This systematic review was conducted following the PRISMA6 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure rigor and transparency in the review process. The aim of the study was to synthesize existing knowledge on the use of biomarkers in cancer diagnosis and prevention, with an emphasis on the application of AI, neural networks, and DL in this context.

The review has been recorded and published in the PROSPERO7 system (CRD42024531670).

The following descriptors were used for article retrieval, along with the combination of these terms and expressions: “biomarkers and cancer diagnosis and artificial intelligence”, “cancer prevention”, “neural network”, and “deep learning”. Searches were conducted...
in the following electronic databases: PubMed, Science Direct, and Web of Science. The search was limited to articles published in English from 2015 to April 2023. The search strategy combined terms using Boolean operators “AND” to link different domains of interest and “OR” to include synonyms or related terms within the same domain.

Original articles discussing the development or application of biomarkers for cancer diagnosis or prevention using AI, neural networks, and deep learning were included in the review. Review articles, commentaries, editorials, conference reports, and studies not directly related to the defined descriptors were excluded. Studies lacking empirical data or full accessibility were also excluded. Additionally, other articles were included for contextualization, enrichment, and justification of the topic addressed.

The identified studies were initially screened based on titles and abstracts. Team members independently reviewed the titles and abstracts to identify potentially relevant studies.

Following the initial screening, studies deemed potentially relevant were selected for full evaluation. The final selection of studies was based on comprehensive analysis of the articles and application of previously established inclusion and exclusion criteria.

Peer review was conducted, where two independent reviewers examined the titles and abstracts of articles identified in the initial search. Potentially relevant studies were selected for full-text review. Discrepancies were resolved through consensus or by a third reviewer.

For the selected articles, information was extracted regarding the type of biomarker investigated, cancer type, methodologies of AI applied (including specifications of deep learning techniques), key findings, and implications for cancer diagnosis and prevention. These data were qualitatively synthesized, highlighting emerging trends, challenges, and opportunities at the intersection of biomarkers, AI, and oncology.

All records related to the search and study selection were systematically documented. A bibliographic reference management system was utilized to organize the identified studies and facilitate the review process. This stage of the methodological process followed PRISMA guidelines to ensure transparency, rigor, and replicability in identifying relevant studies for the proposed systematic review.

The methodological quality of the included studies was assessed using a tool adapted for observational studies. Assessment criteria included clarity of biomarker definition, appropriateness of AI techniques employed, robustness of results, and applicability of conclusions in the context of cancer prevention and diagnosis.

RESULTS

Following the application of inclusion and exclusion criteria, a total of 29 articles were selected for review and analyzed in detail in Chart 1, which provides a description of the selected articles according to author, year of publication, study type, and findings. These studies encompass a wide range of cancer types, including breast, prostate, lung, and liver cancer, demonstrating the growing interest in the application of biomarkers and AI for cancer diagnosis and prevention. The majority of studies were published between 2015 and 2023, reflecting recent advancements in AI technology and its application in oncology.

Initially, search terms “biomarkers, cancer and artificial intelligence”, “biomarkers, cancer prevention and deep learning”, and “artificial intelligence, deep learning and cancer” identified 32,357 articles, with 23,483 articles in PubMed, 7,606 articles in Science Direct, and 1,268 in Web of Science. Subsequently, duplicate articles (325 articles) and studies not directly related to the topic (31,832) were excluded.

In this selection, the search revealed 200 articles related to the object of study of the present work, with 7 in Web of Science, 125 in PubMed, and 68 in Science Direct. After reading the introduction and conclusion of the articles, only 52 articles met the pre-specified eligibility criteria, representing 31 from PubMed, 3 from Web of Science, and 18 from Science Direct. Finally, after full-text reading of the articles, 29 were selected for review.

The reviewed studies highlighted the importance of biomarkers in early cancer detection, with several researches focusing on the development of AI algorithms to improve diagnostic accuracy. The application of deep learning techniques has shown potential in the analysis of biomedical images and genetic data to identify early signs of cancer. For example, studies reported the use of Convolutional Neural Networks for mammography analysis, achieving significantly higher accuracy in detecting early-stage tumors compared to conventional diagnostic methods.

The review identified a growing number of studies exploring the potential of biomarkers and AI in cancer prevention. For instance, 12 AI studies utilized deep learning algorithms to analyze genetic patterns and/or mutations in specific genes, providing accurate cancer risk predictions and recommending specific prevention strategies. Various AI techniques were studied, with an emphasis on Convolutional Neural Networks (CNNs) and deep learning. These techniques were applied in the analysis of histopathological images, detection of nodal
metastases, cellular nucleus segmentation, and biomarker expression quantification. Methods such as transfer learning, random forests, multiview Convolutional Neural Networks (MV-CNNs), and adversarial learning were also used\textsuperscript{12,13,25,27}.

Evaluation techniques included enriched pathway analysis using the IMPaLA tool, Spearman correlation to assess the relationship between metabolite intensity levels and enzyme expression, and the combination of immunohistochemistry (IHC) staining with AI to evaluate staining intensity and biomarker expression. These approaches showed high correlation with traditional methods and potential to improve therapeutic decision-making\textsuperscript{3,5,22,24}.

The articles analyzing the use of AI in the evaluation of breast cancer biomarkers addressed diagnostic, prognostic, and treatment response markers. The biomarkers evaluated in these studies include mammary artery calcifications, breast density, glutamate-to-glutamine ratio (GGR), isoleucine, taurine, and xanthine. These biomarkers are essential for diagnosis, prognosis, and treatment response prediction. For example, the GGR ratio was identified as a positive prognostic marker for breast cancer, while other biomarkers such as HER2, ER, PR, and Ki-67 were used to predict risk categories and therapy efficacy\textsuperscript{3,5,22,24}.

Regarding early detection, one study analyzes that conventional early detection methods for lung and breast cancer expose patients to high radiation doses and a high rate of false positives. Better alternatives for early diagnosis include mutation, methylation, gene expression/non-coding RNA, and circulating tumor cells\textsuperscript{1}. Other articles state that patient screening is the best way to diagnose leukemia early, allowing the identification of both symptomatic and asymptomatic patients, thus preventing major complications\textsuperscript{2,4}.

To evaluate prognostic markers in chimeric antigen receptor T-cell (CAR-T) therapies, studies used various advanced methodologies, such as single-cell genomics, multiplex immunofluorescence, multi-omics platforms, analysis of specific antigen expression in tumor cells, and AI algorithms for analyzing radiological and pathological data. These techniques allowed the identification of new targets for immunotherapy and the construction of data-driven medication schedules. To identify biomarkers, several techniques were employed, including single-cell sequencing to identify T-cell receptors specific to neoantigens, multiplex immunofluorescence to visualize multiple markers in a single sample, radiomics to create image biomarkers, and genetic engineering of T cells to express specific antigen receptors. Additionally, next generation sequencing (NGS), flow cytometry, real-time PCR (qPCR), IHC, and proteomic analyses were used for a detailed analysis of the genomic and proteomic profiles of tumors and CAR-T cells. These techniques enabled not only the identification and quantification of biomarkers but also the assessment of the interaction between CAR-T cells and tumors and the modulation of the tumor microenvironment (TME) to improve therapeutic efficacy\textsuperscript{28-31}.

Regarding treatment, some articles addressed lung cancer as the main topic, relating it to immunotherapies and tomography techniques\textsuperscript{21,26,28}. Prostate cancer, although less frequently mentioned, was discussed in other articles, analyzing treatments and ways to improve prognosis with the aid of AI\textsuperscript{1,27}. Furthermore, the present study demonstrated that neural networks can significantly assist pathologists, reducing interobserver variability and improving accuracy in mitosis detection, nucleus segmentation, and biomarker quantification. Studies showed that AI algorithms achieve a high degree of concordance with clinical gold standards, such as fluorescent in situ hybridization (FISH) for HER2, and can predict Oncotype DX risk categories from the automated detection of tubular nuclei\textsuperscript{3}.

The results are better expressed in Chart 1.
<table>
<thead>
<tr>
<th>Title</th>
<th>Author/ year of publication</th>
<th>Method</th>
<th>Goal</th>
<th>Results</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shifting the cancer screening paradigm: the rising potential of blood-based multi-cancer early detection tests</td>
<td>Tiago Brito Rocha et al. (2023)</td>
<td>Systematic literature review</td>
<td>Highlighting the variety of strategies currently under development for early detection of multiple cancers (MCED)</td>
<td>Although MCED tests have shown efficacy in cancer detection, they are not the best way due to the difficulty of identifying the tissue of origin (TOO)</td>
<td>MCED tests promise to reduce cancer mortality by shifting detection to earlier stages</td>
</tr>
<tr>
<td>Cancer early detection</td>
<td>National Cancer Institute José Alencar Gomes da Silva (2021)</td>
<td>Original guideline book</td>
<td>Forms of early cancer detection</td>
<td>Presenting the fundamentals of early cancer detection and its applicability to all health professionals interested in the topic</td>
<td>The recommendations are based on scientific evidence that underpins national guidelines and protocols</td>
</tr>
<tr>
<td>Artificial intelligence in digital breast pathology: techniques and applications</td>
<td>Asmaa Ibrahim et al. (2020)</td>
<td>Original research article</td>
<td>Addressing current uses of AI in digital pathology of breast cancer</td>
<td>Performance in diagnosis with scanned whole-slide imaging (WSIs) was almost equal to that achieved with traditional microscopy-based methods</td>
<td>AI applications in breast pathology are increasingly growing and are expected to be more than just a complementary tool</td>
</tr>
<tr>
<td>The role of artificial intelligence in early cancer diagnosis</td>
<td>Benjamin Hunter et al. (2022)</td>
<td>Systematic literature review</td>
<td>Identifying how artificial intelligence algorithms can assist physicians in tracking and diagnosing patients more effectively</td>
<td>AI is of paramount importance for early cancer detection. It has been observed that early-stage cancers can be detected with high accuracy using scanning images or biopsy</td>
<td>AI may enable more effective analysis of complex data from various modalities, including clinical text, metabolomic, genomic, and radiomic data</td>
</tr>
<tr>
<td>Future of biomarker evaluation in the realm of artificial intelligence algorithms: application in improved therapeutic stratification of patients with breast and prostate cancer</td>
<td>Jenny Fitzgerald et al. (2020)</td>
<td>Systematic literature review</td>
<td>Describing the application of integrative AI in the improved detection, classification, and prognosis of breast and prostate cancers</td>
<td>Biomarkers provide additional information independent of established clinical variables, enhancing predictive accuracy for diagnosis</td>
<td>AI and machine learning (ML) approaches offer significant new opportunities in precision oncology, with lower costs</td>
</tr>
</tbody>
</table>

*to be continued*
<table>
<thead>
<tr>
<th>Title</th>
<th>Author/ year of publication</th>
<th>Method</th>
<th>Goal</th>
<th>Results</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aging population and chronic diseases: reflections on the challenges for the system of public health</td>
<td>Mayckel Barreto et al. (2015)¹¹</td>
<td>Systematic literature review</td>
<td>Analyzing the challenges that the Public Health System faces in dealing with chronic diseases in the older population</td>
<td>Issues were observed in the Public Health System, including concerns in structuring Primary Care, assisting care networks, and availability of reliable information</td>
<td>The research was pivotal in gathering evidence that will contribute to better practices for the older population and their diseases</td>
</tr>
<tr>
<td>Cancer evolution: from biological insights to therapeutic opportunities</td>
<td>Andrew A. Davis et al. (2022)¹²</td>
<td>Systematic literature review</td>
<td>Discussing cancer evolution and the impact of genomic and epigenomic changes</td>
<td>Need for multi-omic biomarkers to track cancer evolution</td>
<td>Importance of studies on non-coding RNAs and RNA modifications in cancer</td>
</tr>
<tr>
<td>Machine learning in genetics and genomics</td>
<td>Maxwell W. Libbrecht, William Stafford Noble (2015)¹³</td>
<td>Systematic literature review</td>
<td>Discussing ML applications in genetic and genomic data</td>
<td>Identifies recurring challenges and provides guidelines for practical application</td>
<td>Highlights the importance of ML in genomics, with guidelines for overcoming challenges</td>
</tr>
<tr>
<td>Artificial intelligence with multifunctional machine learning platform development for better healthcare and precision medicine</td>
<td>Zeeshan Ahmed et al. (2020)¹⁴</td>
<td>Systematic literature review</td>
<td>Developing Machine Learning platforms for extraction, management, and analysis of clinical data, supporting clinicians in patient care</td>
<td>AI potential in healthcare to improve personalized and population medicine at lower costs</td>
<td>The importance of AI and ML in healthcare transformation through better data management and decision support</td>
</tr>
<tr>
<td>Dermatologist-level classification of skin cancer with deep neural networks</td>
<td>André Esteva et al. (2017)¹⁵</td>
<td>Original research article</td>
<td>Demonstrating the classification of skin lesions using only a Convolutional Neural Network, pixels, and disease labels as inputs</td>
<td>It was found that the Convolutional Neural Network (CNN) surpasses dermatologists in sensitivity and specificity, evidenced by most points lying below the CNN's blue curve</td>
<td>Deep Convolutional Neural Networks show potential for general and highly variable tasks in many categories of fine-grained objects</td>
</tr>
<tr>
<td>Deep learning in diverse intelligent sensor-based systems</td>
<td>Yanming Zhu et al. (2022)¹⁶</td>
<td>Systematic literature review</td>
<td>Reviewing the application of deep learning in sensor-based systems</td>
<td>Advances and successful applications of deep learning in various systems</td>
<td>Identification of challenges and future opportunities in the use of deep learning</td>
</tr>
<tr>
<td>The human pan-genome project: a global resource to map genomic diversity</td>
<td>Ting Wang et al. (2022)¹⁷</td>
<td>Experimental study</td>
<td>Creating a more accurate and diverse human reference genome</td>
<td>New reference for improving gene-disease association studies</td>
<td>Potential to advance research in genomics and precision medicine</td>
</tr>
</tbody>
</table>

Chart 1. continuation

---

¹¹ Barioni CTS, Wandresen RPB, Pereira LF, Coimbra AF, Kubo BBAO, Cunha RC

Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições, desde que o trabalho original seja corretamente citado.
<table>
<thead>
<tr>
<th>Title</th>
<th>Author/ year of publication</th>
<th>Method</th>
<th>Goal</th>
<th>Results</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advances in AI for identifying cancer biomarkers</td>
<td>Marcelo Alves-Ferreira, Ana Carolina Ramos Guimarães, Patricia Cuervo, (2014)</td>
<td>Systematic literature review</td>
<td>Advancing the molecular understanding of Leishmania spp. through genomic, transcriptomic, and proteomic technologies</td>
<td>Identification of Leishmania subgroups with distinct molecular profiles, advances in understanding protozoan biology</td>
<td>Importance of integrated approaches to parasitology research, contributing to Leishmaniasis control and treatment strategies</td>
</tr>
<tr>
<td>A comprehensive evaluation of metabolomics data preprocessing methods for deep learning</td>
<td>Krzysztof Jan Abram et al. (2022)</td>
<td>Original research article</td>
<td>Evaluating the performance of classification and reconstruction through omics datasets</td>
<td>A set of best practices has been identified as a starting point for researchers in classification and reconstruction using deep learning techniques</td>
<td>Experiments highlighted the unique data preprocessing requirements for classifying and reconstructing omics data using deep learning</td>
</tr>
<tr>
<td>Deep learning based multi-omics integration robustly predicts survival in liver cancer</td>
<td>Kumardeep Chaudhary et al. (2018)</td>
<td>Original research article</td>
<td>Presenting a deep learning study model that differentiates patient survival subpopulations</td>
<td>From the TCGA hepatocellular carcinoma project, approximately 360 cancer samples containing associated RNA-Seq, miRNA-Seq, and DNA methylation data were obtained</td>
<td>It is concluded that population heterogeneity affects the performance of the studied model</td>
</tr>
<tr>
<td>Integrated multiple ‘omics’ data reveal subtypes of hepatocellular carcinoma</td>
<td>Gang Liu, Chuangpeng Dong, Lei Liu (2016)</td>
<td>Experimental study</td>
<td>Classifying hepatocellular carcinoma into subgroups using omic data</td>
<td>Identification of five main subgroups with distinct molecular signatures</td>
<td>Classification aids in understanding the pathogenesis of hepatocellular carcinoma and may influence treatment strategies</td>
</tr>
<tr>
<td>Artificial intelligence to identify genetic alterations in conventional histopathology</td>
<td>Didem Cifci et al. (2022)</td>
<td>Systematic literature review</td>
<td>Summarizing the existing literature on predicting molecular alterations from H&amp;E using AI</td>
<td>AI methods showed genetic alterations in the FGFR, IDH, PIK3CA, BRAF, TP53 pathways</td>
<td>Deep learning could shed light on the viral etiology of tumors, which in some cases is relevant to patient management</td>
</tr>
<tr>
<td>The progress of multi-omics technologies: determining function in lactic acid bacteria using a systems level approach</td>
<td>Shane Thomas O’Donnell et al. (2020)</td>
<td>Systematic literature review</td>
<td>Exploring the progress of multi-omic technologies in relation to lactic acid bacteria</td>
<td>Discussion on the application and impact of multi-omic technologies</td>
<td>Multi-omic technologies offer valuable insights into understanding the complex functions of lactic acid bacteria</td>
</tr>
</tbody>
</table>

Chart 1. continuation
<table>
<thead>
<tr>
<th>Title</th>
<th>Author/ year of publication</th>
<th>Method</th>
<th>Goal</th>
<th>Results</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>The artificial intelligence and machine learning in lung cancer immunotherapy</td>
<td>Qing Gao et al. (2023)&lt;sup&gt;21&lt;/sup&gt;</td>
<td>Systematic literature review</td>
<td>Discussion of AI applications in predicting PD-L1/TMB, TME, and lung cancer immunotherapy</td>
<td>Both RESNET-50 and DENSENET 201 achieved good results</td>
<td>AI shows promise in the predictive application of cancer immunotherapy</td>
</tr>
<tr>
<td>Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data</td>
<td>Fadhl Alakwaa et al. (2018)&lt;sup&gt;22&lt;/sup&gt;</td>
<td>Original research article</td>
<td>Assessing the ability of deep learning frameworks to differentiate breast cancer patients based on their level of urgency</td>
<td>LDA and RPAT have the worst accuracy, probably due to their sensitivity to overfitting and being unsuitable for non-linear problems</td>
<td>Deep learning outperforms other machine learning algorithms for classifying ER status in breast cancer metabolomic data</td>
</tr>
<tr>
<td>Deep learning in medical image analysis</td>
<td>Dinggang Shen, Guorong Wu, Heung-II Suk (2017)&lt;sup&gt;23&lt;/sup&gt;</td>
<td>Systematic literature review</td>
<td>Reviewing the use of deep learning in medical image analysis</td>
<td>Significant advancements in various medical applications, such as tissue segmentation and computer-aided diagnosis</td>
<td>Deep learning is becoming the state of the art, improving performance in various medical applications</td>
</tr>
<tr>
<td>Convolutional neural networks for breast cancer detection in mammography: a survey</td>
<td>Leila Abdelrahman et al. (2021)&lt;sup&gt;24&lt;/sup&gt;</td>
<td>Systematic literature review</td>
<td>Examining applications of Convolutional Neural Networks for mammography and their effects on breast cancer</td>
<td>The article presented a review of research over the past 10 years where advances in neural networks have been applied to tomography as precision aids for radiologists</td>
<td>The presented material can serve as a guide for the development of CNN-based solutions to further enhance mammographic breast cancer detection</td>
</tr>
<tr>
<td>International evaluation of an AI system for breast cancer screening</td>
<td>Scott Mayer McKinney et al. (2020)&lt;sup&gt;25&lt;/sup&gt;</td>
<td>Experimental study</td>
<td>Developing an AI system to predict breast cancer outperforming specialists</td>
<td>Significant reductions in false positives and negatives, surpassing radiologists</td>
<td>Potential for clinical application in improving the accuracy and efficiency of breast cancer screening</td>
</tr>
<tr>
<td>End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography</td>
<td>Diego Ardila et al. (2019)&lt;sup&gt;26&lt;/sup&gt;</td>
<td>Experimental study</td>
<td>Enhancing the accuracy of lung cancer screening using deep learning</td>
<td>Deep learning model outperforms radiologists in predicting lung cancer risk</td>
<td>Potential of deep learning to enhance global lung cancer screening</td>
</tr>
<tr>
<td>Machine learning in prostate MRI for prostate cancer: current status and future opportunities</td>
<td>Huanye Li et al. (2022)&lt;sup&gt;27&lt;/sup&gt;</td>
<td>Systematic literature review</td>
<td>Evaluating ML applications in prostate magnetic resonance imaging for cancer</td>
<td>Highlights potential to improve diagnosis and treatment</td>
<td>Need for further research to overcome current limitations</td>
</tr>
</tbody>
</table>

**to be continued**
**DISCUSSION**

Changes in demographic patterns observed in various nations since the last decades of the 20th century, including increased life expectancy and the phenomenon of population aging, have emerged as significant challenges for the domain of public health, particularly in the context of public and universal healthcare systems, such as in Brazil\(^6\). In this context, non-communicable chronic diseases (NCDs), represented by various forms of cancer with their varied clinical manifestations and multifactorial etiologies, assume a prominent role, impacting the health and quality of life of the population. This is corroborated by data from the National Cancer Institute (INCA), which emphasize the relevance of these conditions in the current health scenario\(^2\).

According to INCA\(^3\), the incidence of cancer has been steadily increasing, reflecting changes in lifestyle, population aging, and advances in detection and diagnosis. The Institute estimated that for the 2023-2025 triennium, approximately 704 thousand new cases of cancer would occur annually in Brazil, with a slightly higher incidence in women (52%) compared to men (48%). Among the most frequent types, non-melanoma skin cancer is the most common, accounting for about 31.3% of the total new cases. Furthermore, breast cancer is the most common type among women, with around 66 thousand new cases expected per year, approximately 43.74 cases per 100...
thousand women. Prostate cancer, on the other hand, is the most prevalent among men, with an estimated 65 thousand new cases annually, approximately 66.12 cases per 100 thousand men. Additionally, lung, colon and rectum and stomach cancers are among the types with the highest incidence in both men and women.

Regarding geographical and demographic distribution, the South and Southeast regions concentrate about 70% of the cancer incidence, with the Southeast region accounting for more than half, due to the high population density and greater access to health services that improve detection. In the Northeast and North regions, incidences are lower but with rising rates, especially for stomach cancer and cervical cancer, associated with unfavorable socioeconomic conditions. The South and Central-West regions have intermediate incidences, with a highlight on lung and colorectal cancer.

In recent years, addressing these regional disparities in cancer incidence has been facilitated by advances in genomics and molecular biology, which have allowed for a deeper understanding of the mechanisms underlying tumor development and progression. In order to improve this process, AI has played a crucial role in transforming personalized medicine, especially in the identification of biomarkers for cancer prevention and diagnosis. This technological integration is pivotal in tailoring prevention and treatment strategies to specific regional and individual needs, thereby improving outcomes across different demographics. An example of this is machine learning, focused on the development of computational algorithms that improve through experience, offering the prospect of empowering computers to assist humans in interpreting vast and complex volumes of data, such as difficult-to-detect diagnostic reports and exams. This makes it possible to analyze comprehensive patient information in order to gain a better understanding of the biological indicators that may signal changes in health.

From this, deep learning algorithms have been used to analyze next-generation sequencing (NGS) data, identifying mutations and epigenetic alterations associated with different types of cancer. CNNs, a form of deep learning based on a biological data processing architecture, have been applied in the interpretation of medical images, such as mammograms and magnetic resonance imaging (MRI) to early detect signs of cancer. This type of network has greater utility in applications for detection, classification, and recognition in images and videos.

Based on the studies used for the development of this research, it is observed that AI plays a crucial role in cancer prevention and diagnosis. The analysis of tumor biomarkers, both genomic and proteomic, combined with advanced interpretation of medical images, offers a promising perspective for early disease detection. Therefore, this research focuses not only on early detection but also on personalized prevention strategies, leveraging the potential of AI to improve clinical outcomes and the quality of life of cancer patients.

Genomics represents the study of the genome, which comprises all the genetic material of an organism, including DNA, genes, and their sequences, structures, functions, and interactions. Transcriptomics, on the other hand, consists of the analysis of gene expression – through qualitative and quantitative study of different types of RNA – of a particular organism. AI has been applied to analyze genomic and transcriptomic data in search of mutations and gene expressions. Machine learning tools, such as neural networks and support vector machines, have been used to identify specific genetic signatures of various types of tumors.

A deep machine learning model has been developed that uses RNA-seq, miRNA-seq, and methylation data to predict survival in patients with hepatocellular carcinoma (HCC). This robust predictive model outperforms traditional models by integrating multiple types of genomic and clinical data, identifying molecular biomarkers and signaling pathways associated with the most aggressive subtype of HCC. These findings provide valuable insights into understanding disease progression.

In the same vein, they addressed the integration of multi-omics data through AI to identify subtypes of HCC, which is based on the integration of multi-omics data to predict survival in patients with HCC.

On the other hand, the text describes the application of AI in detecting genetic alterations through conventional histopathology, focusing on tumor markers. It highlights that AI is capable of predicting the probability of specific genetic alterations directly from conventionally stained tissue slides with hematoxylin and eosin (H&E), which can be used as a preliminary screening tool to reduce the workload of genetic analyses. Therefore, the potential of AI in predicting genomic and transcriptomic tumor markers is evident; however, there are current challenges and limitations that need to be overcome for its effective implementation in oncological clinical practice.

Proteome refers to the entirety of proteins expressed by a cell, tissue, or organism at a specific moment. Thus, proteomic analysis involves a broader set of variables compared to genomic and transcriptomic analysis due to the vast chemical diversity of proteins and their extensive interconnection in complexes and signaling networks. It is characterized by examining the proteome through separation and identification methods, including electrophoresis, chromatography, mass spectrometry,
and bioinformatics, being of great relevance for disease prognosis and therapy\textsuperscript{20}. AI-assisted proteomic analysis allows the identification of tumor proteins and post-translational modifications serving as biomarkers. Deep learning algorithms have been effective in interpreting mass spectrometry, facilitating the discovery of protein biomarkers with high specificity and sensitivity\textsuperscript{16}. An increasing number of studies have combined radiology, pathology, genomics, and proteomic data to predict the expression levels of biomarkers such as programmed death-ligand 1 (PD-L1), tumor mutation burden (TMB), and TME in cancer patients\textsuperscript{21}.

For example, a study reviewed the applications of AI in predicting PD-L1/TMB, TME, and immunotherapy in lung cancer and concluded that this tool shows great promise in the predictive application of cancer immunotherapy through data combination, disassembly, and analysis\textsuperscript{22}. Furthermore, the authors addressed the use of deep learning technique to accurately predict the estrogen receptor (ER) status in breast cancer patients\textsuperscript{20}. Moreover, they identified: a) specific metabolic markers, such as xanthines, which may be indicative of breast cancer metastasis; b) biosynthetic enzymes related to breast cancer metabolic pathways; c) glutathione biosynthesis, a critical component of cellular metabolism, as an innovative diagnostic opportunity in breast cancer. These points highlight the importance of integrating tumor markers, proteins, and AI, such as deep learning, in data analysis for a better understanding and prediction of specific cancer types.

Studies have also demonstrated the applicability of AI in identifying image-based biomarkers, especially in mammograms, MRI, and computed tomography scans (CT)\textsuperscript{23}. Some studies have used neural networks applied to mammography exams to increase efficiency and accuracy of the examination. The research presented and discussed the current literature on CNNs for four distinct mammography tasks: (1) breast density classification, (2) detection and classification of breast asymmetry, (3) detection and classification of calcifications, and (4) mass detection and classification, including presentation and comparison of results and the pros and cons of different CNN-based approaches\textsuperscript{24}. They also presented an AI model that outperformed human radiologists’ accuracy in breast cancer detection from mammograms. The model reduced false positive and false negative rates, enhancing breast cancer screening and diagnostic accuracy\textsuperscript{25}.

On the other hand, other studies performed automated classification of skin lesions through a single CNN, trained end-to-end directly from 129,450 images, using only pixels and disease labels as data. The system was able to identify basal cell carcinoma, melanoma, and squamous cell carcinoma with performance equivalent to dermatologists\textsuperscript{26}. Regarding lung cancer, they developed an AI model to detect this type of anomaly in CT images, outperforming six human radiologists in terms of sensitivity and specificity\textsuperscript{26}. The model, based on deep learning, identified malignant lung nodules with a lower rate of false positives. In the same way, similar results were observed with prostate cancer, through the analysis of MRI, achieving an accuracy that corresponds to or exceeds that of experienced radiologists\textsuperscript{27}.

Finally, the challenges faced in integrating biomarkers and AI in cancer diagnosis and prevention are highlighted. Among them are the need for large sets of well-annotated data to effectively train AI algorithms and the importance of validating AI models in diverse populations to ensure their universal applicability. However, the opportunities to advance in this field are significant, with studies indicating the potential to substantially improve patient outcomes through early detection and personalized preventive interventions based on AI.

In addition to the opportunities in integrating biomarkers and AI in cancer diagnosis and prevention, it is important to highlight the advancement in the study of therapies, such as CAR-T therapies, an innovative approach in cancer treatment, involves genetically modifying the patient’s T cells so they can identify and attack cancer cells. This is done by introducing a gene encoding a CAR-T, allowing them to recognize and bind to specific antigens present on cancer cells\textsuperscript{28}. Unlike conventional T cells, CAR-T cells can recognize antigens on the surface of cancer cells without relying on human major histocompatibility complex (MHC) molecules. This ability allows CAR-T cells to have a broader spectrum of targets compared to natural T cells\textsuperscript{29}.

Compared to blood cancer therapy, the application of CAR-T cells in solid tumors is challenged by the difficulty of penetration into the tumor environment due to the tumor’s immunosuppressive microenvironment and its physical barriers, such as the tumor stroma, which limits the penetration and mobility of CAR-T cells. One solution is local administration, avoiding the need for cell migration. Preclinical studies show efficacy in intraventricular and intrapleural injections in breast cancer, glioblastoma, and mesothelioma. This drives clinical trials to investigate these approaches in glioblastomas and brain metastases. However, local application may have limitations in single tumors or diseases with few metastases\textsuperscript{30}. The identification of new therapeutic targets and the improvement of CAR constructs are expanding the clinical scope of CAR-T cell therapy beyond hematologic diseases. However,
the rapid commercialization of this therapy presents significant challenges, such as toxicities and relapses, which require in-depth investigation of their underlying mechanisms to ensure additional benefits to patients. Current studies are exploring various combinations with CAR-T cell therapy, suggesting its potential as effective immunotherapy31. Despite these challenges, the present study underscores AI’s potential to personalize oncological care by pinpointing specific biomarkers linked to different cancer types. This personalized approach holds particular promise in Brazil, where genetic diversity is high, and cancer presentations can vary widely. Tailoring treatment plans to each patient’s genetic and molecular profile can enhance treatment outcomes and mitigate adverse effects. Moreover, in regions like the North and Northeast of Brazil, where healthcare resources are limited, AI can optimize resource utilization. AI algorithms can prioritize high-risk patients for further testing and treatment, ensuring that scarce healthcare resources are allocated where they’re most needed. This approach can help combat the increasing rates of stomach and cervical cancers in these regions, often associated with socio-economic disparities.

CONCLUSION

The present theoretical framework highlights the transformative impact of AI on the identification of tumor biomarkers, addressing everything from genomic and proteomic analysis to advanced interpretation of medical images. As technology advances, AI is expected to play an increasingly central role in personalized oncology, improving the diagnosis, prognosis, and selection of specific therapies for cancer patients.

However, it is important to note that the successful implementation of AI in clinical practice requires overcoming challenges such as large-scale clinical validation, ensuring data security, considering ethical and regulatory issues, and effectively integrating with existing healthcare systems. Despite these challenges, advances in AI offer promising prospects for enhancing cancer prevention and diagnosis, potentially resulting in better clinical outcomes and quality of life for patients.

In Brazil, where the incidence of cancer is increasing, particularly in regions with varying access to healthcare services, AI-driven tools can analyze complex data from clinical, genetic, and imaging sources, enabling earlier detection of cancers such as breast, prostate, lung, and colorectal cancers, which are prevalent in the country. Early detection is crucial for effective treatment and can significantly reduce mortality rates.

CONTRIBUTIONS

All the authors contributed to the study design, collection of information and wording of the manuscript. They approved the final version to be published.

DECLARATION OF CONFLICT OF INTERESTS

There is no conflict of interest to declare.

FUNDING SOURCES

None.

REFERENCES


Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições, desde que o trabalho original seja corretamente citado.


Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições, desde que o trabalho original seja corretamente citado.

Recebido em 22/4/2024
Aprovado em 11/6/2024