

HARNESSING THE POWER OF AI AND MACHINE LEARNING FOR NEXT-GENERATION SEQUENCING DATA ANALYSIS: A COMPREHENSIVE REVIEW OF APPLICATIONS, CHALLENGES, AND FUTURE DIRECTIONS IN PRECISION ONCOLOGY

Pedro Henrique Zeraik Viduedo¹
Victor César Urquiza Candeia²
Victor Balceiro Legname Martins³
Afrânio Côgo Destefani⁴
Vinícius Côgo Destefani⁵

ABSTRACT: This comprehensive review discusses the transformative impact of artificial intelligence (AI) and machine learning (ML) on next-generation sequencing (NGS) data analysis in precision oncology. AI and ML are pivotal in enhancing the precision of variant calling, annotating complex genomic data, discovering and validating new biomarkers, and integrating multi-omics data to elucidate cancer complexities. These technologies facilitate advanced clinical decision support systems that improve diagnostic accuracy and treatment efficacy. Despite these advancements, challenges persist, including the need for model interpretability, validation across diverse datasets, and integration into clinical practice. Future directions involve federated learning for data privacy, integration of multi-modal data, and expansion of explainable AI applications to enhance understanding and trust in AI-driven methodologies.

2898

Keywords: Next-generation sequencing. Artificial intelligence. Precision oncology. Machine learning. Clinical decision support systems.

INTRODUCTION

The advent of next-generation sequencing (NGS) technologies has revolutionized precision oncology by enabling the generation of vast amounts of high-dimensional genomic data (1). However, the analysis and interpretation of these complex datasets pose significant challenges, necessitating the development of advanced computational tools (2). Artificial

¹UNIFOA.

²UNIFIP.

³Faculdade Santa Marcelina,

⁴Santa Casa de Misericórdia de Vitoria Higher School of Sciences – EMESCAMB. Molecular Dynamics and Modeling Laboratory (DynMolLab).

⁵Molecular Dynamics and Modeling Laboratory (DynMolLab) - Santa Luíza – Vitória – ES – Brazil.

intelligence (AI) and machine learning (ML) have emerged as powerful approaches to tackle these challenges, offering novel insights into the molecular underpinnings of cancer (3). This narrative review aims to provide a comprehensive overview of AI and ML's applications, challenges, and future directions in NGS data analysis for precision oncology.

METHODOLOGY

A systematic literature search was conducted using multiple databases, including Scopus, Web of Science, PubMed, IEEE Xplore, ScienceDirect, Directory of Open Access Journals (DOAJ), and JSTOR. The search terms included combinations of "artificial intelligence," "machine learning," "next-generation sequencing," "precision oncology," "cancer genomics," and related keywords. Articles published between 2015 and 2023 were considered for inclusion. The retrieved articles were screened for relevance, and the most pertinent studies were selected for in-depth analysis and synthesis.

RESULTS

Variant Calling and Annotation

AI and ML algorithms have significantly improved the accuracy and efficiency of variant calling and annotation in NGS data (4). Deep learning-based approaches, such as DeepVariant (5) and Clairvoyant (6), have demonstrated superior performance to traditional methods, particularly in detecting low-frequency variants and complex genomic regions. These advancements have not only enabled the identification of clinically relevant mutations but also facilitated the development of targeted therapies, which are tailored to the specific genetic makeup of a patient's cancer, thereby improving treatment outcomes (7).

Biomarker Discovery and Validation

ML algorithms have been extensively applied to identify and validate novel biomarkers from NGS data (8). Unsupervised learning techniques, such as clustering and dimensionality reduction, have been employed to uncover distinct molecular subtypes and prognostic signatures (9). Supervised learning methods, including support vector machines and random

forests, have been used to develop predictive models for treatment response and patient stratification (10). These AI-driven approaches have not just accelerated the discovery of clinically actionable biomarkers, but also facilitated the implementation of personalized treatment strategies, inspiring a new wave of progress in the field (11).

Integration of Multi-Omics Data

The integration of multi-omics data, encompassing genomics, transcriptomics, proteomics, and metabolomics, provides a comprehensive view of the molecular landscape of cancer (12). AI and ML techniques have been instrumental in integrating and analyzing these diverse datasets to unravel the complex interplay between different molecular layers (13). Graph-based methods, such as network analysis and pathway enrichment, have been employed to identify essential driver genes and dysregulated pathways (14). Deep learning architectures, including autoencoders and generative adversarial networks, have been used to learn joint representations and predict the functional impact of genetic alterations (15).

Clinical Decision Support Systems

AI-powered clinical decision support systems (CDSSs) have emerged as valuable tools for integrating NGS data with clinical information to guide treatment decisions (16). These systems leverage ML algorithms to provide personalized treatment recommendations based on a patient's genomic profile, clinical characteristics, and treatment history (17). CDSSs have been shown to improve the accuracy and efficiency of treatment selection, reduce adverse events, and enhance overall patient outcomes (18). The integration of AI and ML with electronic health records and clinical trial databases has further expanded the capabilities of CDSSs in precision oncology (19).

2900

DISCUSSION

The application of AI and ML in NGS data analysis has significantly advanced the field of precision oncology, enabling the identification of clinically relevant mutations, biomarkers, and therapeutic targets (20). However, several challenges remain to be addressed.

The interpretability and explainability of AI and ML models are crucial for building trust and facilitating clinical adoption (21). Rigorous validation and benchmarking of these models across diverse patient populations and cancer types are necessary to ensure their generalizability and robustness (22). Furthermore, integrating AI and ML into clinical workflows requires close collaboration between computational scientists, clinicians, and regulatory bodies to establish standardized guidelines and best practices (23).

Future directions in this field include the development of federated learning approaches to enable the analysis of distributed datasets while preserving patient privacy (24). The incorporation of multi-modal data, such as imaging and electronic health records, into AI and ML frameworks will provide a more comprehensive understanding of cancer biology and patient outcomes (25). The application of explainable AI techniques, such as attention mechanisms and knowledge graphs, will enhance the interpretability and trustworthiness of AI-driven insights (26). Additionally, the integration of AI and ML with emerging technologies, such as single cell sequencing and spatial transcriptomics, will not just enable the exploration of intratumor heterogeneity and the tumor microenvironment at unprecedented resolution, but also open up exciting new possibilities in precision oncology. (27).

CONCLUSION

AI and ML have not just revolutionized the analysis of NGS data in precision oncology, but also opened new possibilities. They have enabled the identification of clinically relevant mutations, biomarkers, and therapeutic targets and facilitated the integration of multi-omics data, the development of clinical decision support systems, and the personalization of treatment strategies. As we address the challenges related to interpretability, validation, and clinical integration, we are on the brink of fully realizing the potential of AI and ML in precision oncology. Future directions, including federated learning, multi-modal data integration, explainable AI, and the incorporation of emerging technologies, will further advance the field and improve patient outcomes. As AI and ML evolve, their impact on precision oncology will undoubtedly grow, paving the way for a new era of data-driven, personalized cancer care.

REFERENCES

1. SRIVASTAVA r. Applications of artificial intelligence multiomics in precision oncology. Vol. 149, journal of cancer research and clinical oncology. 2023.
2. XU j, yang p, xue s, sharma b, sanchez-martin m, wang f, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. Vol. 138, human genetics. 2019.
3. GAO f, wang w, tan m, zhu l, zhang y, fessler e, et al. Deepcc: a novel deep learning-based framework for cancer molecular subtype classification. *Oncogenesis*. 2019;8(9).
4. POPLIN r, chang pc, alexander d, schwartz s, colthurst t, ku a, et al. A universal snp and small-indel variant caller using deep neural networks. *Nat biotechnol*. 2018;36(10).
5. LUO r, sedlazeck fj, lam tw, schatz mc. A multi-task convolutional deep neural network for variant calling in single molecule sequencing. *Nat commun*. 2019;10(1).
6. AINSCOUGH bj, barnell ek, ronning p, campbell km, wagner ah, fehniger ta, et al. A deep learning approach to automate refinement of somatic variant calling from cancer sequencing data. *Nat genet*. 2018;50(12).
7. DING l, bailey mh, porta-pardo e, thorsson v, colaprico a, bertrand d, et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics. *Cell*. 2018;173(2).
8. CHAUDHARY k, poirion ob, lu l, garmire lx. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clinical cancer research*. 2018;24(6).
9. ZHANG z, zhao y, liao x, shi w, li k, zou q, et al. Deep learning in omics: a survey and guideline. *Brief funct genomics*. 2019;18(1).
10. SUN y, zhu s, ma k, liu w, yue y, hu g, et al. Identification of 12 cancer types through genome deep learning. *Sci rep*. 2019;9(1).
11. ESTEVA a, robicquet a, ramsundar b, kuleshov v, depristo m, chou k, et al. A guide to deep learning in healthcare. Vol. 25, nature medicine. 2019.
12. HUANG s, chaudhary k, garmire lx. More is better: recent progress in multi-omics data integration methods. Vol. 8, frontiers in genetics. 2017.
13. MIRZA b, wang w, wang j, choi h, chung nc, ping p. Machine learning and integrative analysis of biomedical big data. Vol. 10, genes. 2019.
14. HAO j, kim y, kim tk, kang m. Pasnet: pathway-associated sparse deep neural network for prognosis prediction from high-throughput data. *Bmc bioinformatics*. 2018;19(1).

15. WAY gp, greene cs. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. In: pacific symposium on biocomputing. 2018.
16. SHIMIZU h, nakayama ki. Artificial intelligence in oncology. *Cancer sci* [internet]. 2020 may 1 [cited 2024 jun 25];111(5):1452-60. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/cas.14377>
17. TSIMBERIDOU am, hong ds, wheler jj, falchook gs, janku f, naing a, et al. Long-term overall survival and prognostic score predicting survival: the impact study in precision medicine. *J hematol oncol*. 2019;12(1).
18. ROYCHOWDHURY s, van allen em. Precision cancer medicine: challenges and opportunities. *Precision cancer medicine: challenges and opportunities*. 2020.
19. SINGAL g, miller pg, agarwala v, li g, kaushik g, backenroth d, et al. Association of patient characteristics and tumor genomics with clinical outcomes among patients with non-small cell lung cancer using a clinicogenomic database. *Jama*. 2019;321(14).
20. MARQUART j, chen ey, prasad v. Estimation of the percentage of us patients with cancer who benefit from genome-driven oncology. *Jama oncol*. 2018;4(8).
21. HOLZINGER a, langs g, denk h, zatloukal k, müller h. Causability and explainability of artificial intelligence in medicine. Vol. 9, *wiley interdisciplinary reviews: data mining and knowledge discovery*. 2019.
22. XIAO c, choi e, sun j. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. Vol. 25, *journal of the american medical informatics association*. 2018.
23. HE j, baxter sl, xu j, xu j, zhou x, zhang k. The practical implementation of artificial intelligence technologies in medicine. *Nat med* [internet]. 2019 jan 7;25(1):30-6. Available from: <https://www.nature.com/articles/s41591-018-0307-0>
24. RIEKE n, hancox j, li w, milletari f, roth hr, albarqouni s, et al. The future of digital health with federated learning. *Npj digit med*. 2020;3(1).
25. TOPOL ej. High-performance medicine: the convergence of human and artificial intelligence. Vol. 25, *nature medicine*. 2019.
26. TJOA e, guan c. A survey on explainable artificial intelligence (xai): toward medical xai. *Ieee trans neural netw learn syst*. 2021;32(11).

27. LÄHNEMANN d, köster j, szczurek e, mccarthy dj, hicks sc, robinson md, et al. Eleven grand challenges in single-cell data science. Vol. 21, genome biology. 2020.