



Integrating Deep Learning for Mutation Detection in Cancer Genomes

Hiroshi Tanaka*

Department of Computational Biology Kyoto University, Kyoto, Japan

DESCRIPTON

The rapidly growing domain of cancer genomics continues to produce extensive and complex datasets, necessitating advanced computational methods for accurate interpretation. A central task in this field is mutation detection, which is essential for identifying the genetic alterations that drive cancer initiation and progression. Traditional methods, which typically rely on statistical comparisons between sequencing data and reference genomes, often face difficulties in differentiating true mutations from sequencing artifacts or noise. This limitation has spurred the adoption of deep learning techniques, which offer more precise and efficient mutation detection capabilities. Deep learning models are particularly effective at identifying intricate patterns within high-dimensional data. In the context of cancer genomics, these models can analyze raw sequencing outputs to detect subtle patterns linked to various genetic alterations, including insertions, deletions and Single Nucleotide Variants (SNVs). Convolutional Neural Networks (CNNs) are especially powerful in learning spatial and hierarchical features from sequence alignment data, while Recurrent Neural Networks (RNNs) excel at modeling the sequential nature of nucleotide sequences. By leveraging these architectures, deep learning-based mutation callers often achieve higher sensitivity and specificity than conventional bioinformatics pipelines.

A significant advantage of deep learning in genomics lies in its flexibility and adaptability across different cancer types. For instance, models trained on data from common cancers such as breast or lung can be fine-tuned using transfer learning to work effectively on rarer malignancies. This adaptability accelerates tool development and enables the broader application of AI models in various clinical and research settings. Furthermore, deep learning can integrate multiple layers of omics data such as transcriptomic, epigenomic and proteomic information leading to more holistic models that consider the broader molecular environment of a mutation. Validation remains a cornerstone in ensuring the robustness and clinical reliability of deep learning models. Independent test datasets and experimental

confirmation are essential for assessing generalizability and accuracy. Public databases such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC) offer large, well-annotated sequencing datasets that serve as benchmarks for evaluating model performance. The widespread use of these resources has also helped standardize practices and improve the reproducibility of computational approaches in mutation detection.

Clinically, the accurate identification of mutations has profound implications. Detecting driver mutations can inform the selection of targeted therapies, predict a patient's response to treatment and contribute to prognostic assessments. For example, identifying mutations in genes like *EGFR* in lung cancer or *BRCA1/2* in breast and ovarian cancers directly influences therapeutic strategies. As sequencing becomes more affordable and accessible, the integration of deep learning models into routine clinical genomics is expected to expand, empowering more patients to benefit from precision oncology. Despite these advancements, several challenges persist. One of the primary concerns is data heterogeneity. Sequencing outputs can vary significantly between different platforms or laboratories, impacting model performance and generalizability. Additionally, training deep learning models demands significant computational resources and specialized expertise, which may not be readily available in all clinical or research environments. To address these barriers, researchers are exploring techniques such as federated learning, which allows decentralized model training across multiple institutions without sharing sensitive data. Distributed computing and algorithmic optimization are also being pursued to improve the efficiency and scalability of model development.

Deep learning is increasingly becoming an integral part of clinical workflows, especially in genomic diagnostics. Neural network-based pipelines can process whole-genome or whole-exome sequencing data in a fraction of the time required by manual or traditional bioinformatics methods. This reduction in processing time is particularly valuable in high-stakes clinical

Correspondence to: Hiroshi Tanaka, Department of Computational Biology Kyoto University, Kyoto, Japan, E-mail: hiroshi.tanaka@kyoto-u.ac.jp

Received: 29-May-2025, Manuscript No. JDMGP-25-29763; **Editor assigned:** 31-May-2025, PreQC No. JDMGP-25-29763; **Reviewed:** 14-Jun-2025, QC No. JDMGP-25-29763; **Revised:** 20-Jun-2025, Manuscript No. JDMGP-25-29763; **Published:** 28-Jun-2025, DOI: 10.35248/2153-0602.25.16.382

Citation: Tanaka H (2025). Integrating Deep Learning for Mutation Detection in Cancer Genomes. Journal of Data Mining in Genomics & Proteomics. 16:382.

Copyright: © 2025 Tanaka H. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

contexts, such as late-stage or aggressive cancers, where timely decisions can be life-saving. Consequently, many hospitals and research centers are adopting AI-powered systems to support oncologists with data-driven, accurate mutation interpretation. One emerging application is the use of deep learning in liquid biopsies. These minimally invasive tests detect tumor-derived genetic material, such as Circulating Tumor DNA (ctDNA), in blood samples. Deep learning algorithms enhance the sensitivity and specificity of mutation detection from these samples, even when the signal is weak or masked by background noise. This capability supports early cancer detection, monitoring of treatment response and detection of minimal residual disease, making it a promising tool for longitudinal patient care.

As the integration of AI into genomics advances, ethical and regulatory issues become increasingly important. Concerns include data privacy, algorithmic bias and the "black box" nature of many deep learning models. If models are trained on non-representative datasets, they may underperform for minority or

underrepresented populations, potentially exacerbating healthcare disparities. Ensuring fairness in training data and building inclusive models are therefore critical priorities. Regulatory bodies such as the U.S. Food and Drug Administration (FDA) are beginning to assess AI-driven diagnostics for clinical use. This has led to a growing emphasis on explain ability and interpretability in model development. Tools like attention visualization and saliency maps are being employed to help clinicians understand which data features influenced a model's prediction, enhancing trust and facilitating clinical adoption.

Looking ahead, the integration of deep learning into cancer genomics holds immense promise. As computational tools evolve and become more accessible, their potential to transform diagnosis, treatment selection and long-term disease monitoring will only grow. This convergence of artificial intelligence and molecular biology is not only improving outcomes for cancer patients but also shaping the future of personalized medicine.