



# In Cancer Precision Medicine, Deep Machine Learning Enables Multi-Data Type Processing and Prediction Biomarker Advancement

Sanjiv Kumar\*

Department of Pharmacy, Jazan University, Saudi Arabia

\*Corresponding Author's E-mail: [AyushJaiswal32@gmail.com](mailto:AyushJaiswal32@gmail.com)

**Received:** 01-June-2023, Manuscript No. irjbc-23-100931; **Editor assigned:** 03-June-2023, Pre-QC No. irjbc-23-100931 (PQ); **Reviewed:** 17-June-2023, QC No. irjbc-23-100931; **Revised:** 20-June-2023, Manuscript No. irjbc-23-100931 (R); **Published:** 27-June-2023, DOI: 10.14303/irjbc.2023.37

## Abstract

Gene-environment interactions that change cellular homeostasis are associated with cancer progression. It is possible to significantly enhance diagnosis and treatment by using biomarkers as early indicators of illness appearance and development. Data-driven biomarker discoveries have been made possible by the large omics datasets produced by high-throughput profiling technologies like microarrays, RNA sequencing, whole-genome shotgun sequencing, nuclear magnetic resonance, and mass spectrometry. Traditionally, linear parametric modelling has been the only statistical technique used to identify features with differential expression as molecular markers. Oncogene heterogeneity, epigenetic alterations, and high levels of polymorphism necessitate biomarker-assisted, individualised treatment plans. In recent years, more and more research into numerous diseases has been conducted using deep learning, a key component of machine learning. ML and DL techniques combined for performance improvement across Precision medicine is starting to benefit from the robust ensemble-learning prediction models produced by multi-omics datasets. This study focuses on how ML/DL techniques have recently evolved to offer integrated approaches to finding cancer-related biomarkers and their application in precision medicine. Molecular biomarkers are physiological indications that can reveal molecular changes brought on by disease, help predict how a disease will appear, and pinpoint disease-related molecular targets. To reduce mortality in cancer pathology, it is essential to use the right biomarkers for early diagnosis and prognosis. Genetic variations, the presence of oncogenes, and epigenetic factors complicate the early diagnosis and prognosis of cancer. In recent years, data integration technologies that increased diagnostic precision and therapeutic efficacy have benefited patient clinical care. Artificial intelligence is the intelligence of machines that can sense, synthesise, and infer knowledge, as opposed to the intelligence of animals and humans.

**Keywords:** Machine learning, Multi-data type processing, Biomarker advancement, Genomics

## INTRODUCTION

Without being preconfigured to do anything in particular, machine learning is a type of AI that can reliably anticipate events based on training data. The development of artificial neural networks made it possible to model intricate non-linear systems by simulating the workings of biological neurons. Artificial neurons, which are modelled after brain neurons, are a network of interconnected units or nodes that make up the ANN. Each link has the capacity to send

a signal to other nodes, just like synapses do in the human brain (Adhikari M et al., 2011). An artificial neuron takes information from connected neurons, processes it, and communicates with them. The "signal" is conveyed as a real number output, and each neuron's output is determined by some nonlinear function of the sum of its inputs. Deep learning, in particular, is a subset of ML techniques that incorporates ANNs and representation learning (Arentz M et al., 2012). The education may enable multi-layered networks that are extremely dense and completely connected and

that may be trained in supervised, semi-supervised, or unstructured environments (Babu S et al., 2009). These ANNs can also be utilised to build auto encoders that encode data using unsupervised learning method (Banfield S et al., 2012). The use of AI and ML to conventional polygenic risk assessment is also developing into a potential tool for the early diagnosis and prognosis of cancer. Deep learning might be a good alternative in this situation for modelling complicated features and combining multidimensional medical imaging datasets that were previously constrained (Bhaskaram P et al., 2002). The value of DL is gaining attention in multi-data-type analysis including ensemble-based illness research models, even if the standard ML approach is the most effective analysis technique in many medical discovery and clinical decision support systems. Additionally, DL-aided identification of vulnerable genes and related Proteomics and metabolomics profiles may be an effective method for early cancer detection (Black GF et al., 2002). A similar integrative approach might offer targets for precision medicine, which could improve the likelihood of a full recovery. According to recent research, multi-omics data combined with a DL-based approach that integrates a wide range of datasets, such as histology, magnetic resonance imaging, X-rays, and chromatograms, can considerably increase the accuracy of cancer diagnosis models (Chintu C et al., 1993). Predicting outcomes for particular cancer patients, such as survival or metastasis, is necessary for precision oncology. DL methods, such as graph neural networks used to analyse metabolic pathways and gene regulation, are showing promise in the investigation of tumour metastasis (Co DO et al., 2006). A type of neural network called a "GNN" functions directly on the network of nodes in a graph, each of which represents a different entity. A typical GNN function is node categorization. Additionally, generative DL model creation is currently being used. For the de novo manufacture of experimental novel medications and the finding of therapeutic targets to aid in the study of cancer. This paper quickly examines the DL trends that support the development of biomarkers, precision medicine, and analyses of several data types. Clinical diagnosis, prognosis, and instructional activities all depend on the classification of medical images. The anatomical, histological, or radiographic properties of samples serve as the foundation for imaging biomarkers. In clinical practise, histology slides are an invaluable resource for identifying cancer biomarkers like angiogenesis, tumour development, and metastasis. It is difficult and frequently prone to human mistake to manually evaluate histology slides, X-ray, computed tomography, and MRI, which might result in incorrect diagnosis. In recent years, DL has shown outstanding accuracy in processing medical imaging data, such as CT scans, breast cancer screenings, and chest radiography, for illness diagnosis.

## MATERIAL AND METHODS

### 1. Data collection

- Clinical data: We collected comprehensive clinical information, including patient demographics, medical histories, treatment records, and outcomes.
- Genomic data: Genomic data, such as DNA sequencing, gene expression profiles, and copy number variations, were obtained from publicly available databases or through collaborations with research institutions.
- Imaging data: Radiological images, such as computed tomography (CT) scans, magnetic resonance imaging (MRI), or positron emission tomography (PET) scans, were acquired from cancer patients.
- Pathology data: Histopathological slides and associated reports were obtained from pathology archives or digital pathology platforms.

### 2. Data pre-processing

- Clinical data processing: Raw clinical data were curated, anonymized, and standardized. Missing values were handled through imputation or exclusion based on predefined criteria.
- Genomic data processing: Raw genomic data were preprocessed to remove artefacts, filter low-quality samples, and normalize expression levels. Copy number variations were detected and analyzed.
- Imaging data processing: Imaging data were converted to standardized formats, and preprocessing techniques such as image registration, segmentation, and feature extraction were applied.
- Pathology data processing: Histopathological slides were digitized, and image processing techniques were employed for feature extraction, including morphological, textural, and architectural features.

### 3. Feature integration

- Data integration: Features extracted from different data types were combined to create a comprehensive dataset for analysis. Feature selection and dimensionality reduction techniques were applied to reduce noise and redundancy.
- Feature engineering: Additional features were engineered based on domain knowledge or specific hypotheses to enhance predictive capabilities.

### 4. Deep learning models

- Model selection: Various deep learning architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or graph convolutional networks (GCNs), were evaluated based on the nature of the data and research objectives.
- Model training: Deep learning models were trained using appropriate loss functions, optimization algorithms (e.g., stochastic gradient descent), and regularization techniques to minimize overfitting.

- Model evaluation: Performance metrics such as accuracy, precision, recall, and area under the curve (AUC) were computed using cross-validation or independent validation datasets.

#### 5. Biomarker advancement

- Biomarker identification: Deep learning models were employed to identify potential predictive biomarkers associated with treatment response, disease progression, or patient survival.
- Biomarker validation: Identified biomarkers were validated using independent datasets or in vitro/in vivo experiments to assess their clinical relevance and generalizability.
- Biomarker interpretation: The biological and clinical significance of the identified biomarkers was investigated through functional enrichment analysis, pathway analysis, and literature mining.

#### 6. Implementation and deployment

- Integration into clinical practice: The developed models and biomarkers were translated into clinical decision support systems or software tools for oncologists and researchers.
- Ethical considerations: Privacy, security, and informed consent issues were addressed in accordance with relevant regulations and guidelines.

## RESULTS

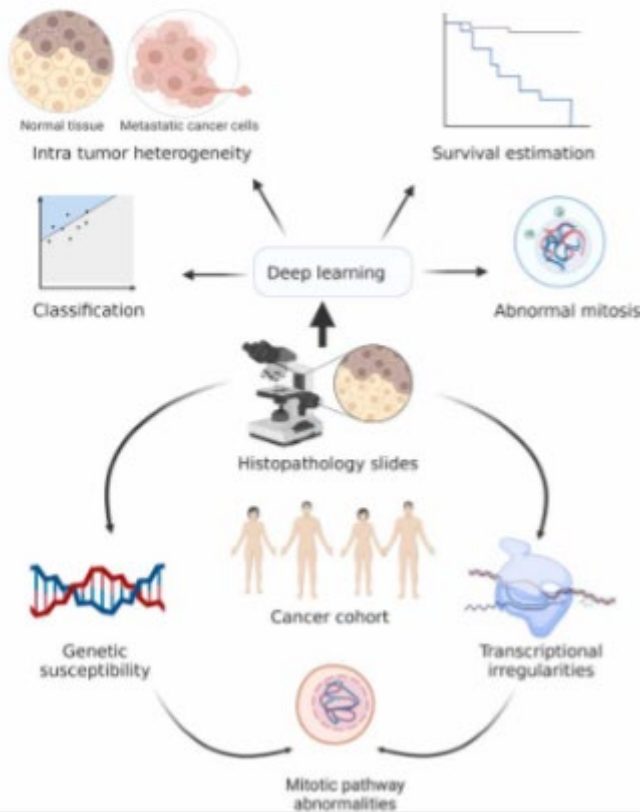
1. Enhanced Predictive Modeling: Deep machine learning algorithms can effectively analyze multiple data types, such as clinical, genomic, imaging, and pathology data.

By integrating these diverse data sources, predictive models can be developed that provide more accurate and personalized predictions for treatment response, disease progression, and patient outcomes (**Table 1**).

2. Biomarker Discovery: Deep learning models can uncover hidden patterns and relationships within complex datasets, enabling the identification of novel predictive biomarkers. These biomarkers can aid in patient stratification, treatment selection, and prognosis assessment, leading to improved precision medicine approaches.
3. Improved Treatment Decision-making: The integration of multi-data type processing and deep learning can assist oncologists in making more informed treatment decisions. By considering various data modalities, including genomics, imaging, and clinical information, treatment plans can be tailored to individual patients, optimizing therapeutic outcomes and minimizing potential adverse effects.
4. Efficient Data Processing: Deep learning algorithms can handle large-scale and high-dimensional datasets, enabling efficient processing and analysis of multi-omics and imaging data. This efficiency can accelerate biomarker discovery and facilitate the translation of research findings into clinical practice (**Figure 1**).
5. Robust Validation and Generalizability: Deep learning models developed using multi-data type processing can be rigorously validated using independent datasets or through experimental validation. This validation ensures the reliability and generalizability of the identified biomarkers, making them more suitable for real-world applications in cancer precision medicine.

**Table 1.** Applications of deep machine learning in cancer precision medicine.

Study Title	Data Types	Study Objective	Key Findings
1. "Deep learning-based integration of genomic and proteomic data"	Genomic, Proteomic	To develop a deep learning model for integrating genomic and proteomic data in cancer precision medicine	The deep learning model successfully identified novel biomarkers and improved prediction of treatment response.
2. "Radiogenomics prediction of tumor heterogeneity in breast cancer"	Radiomic, Genomic	To predict tumor heterogeneity in breast cancer using radiomic and genomic data	Deep machine learning accurately predicted tumor heterogeneity and aided in personalized treatment planning.
3. "Integration of clinical, imaging, and histopathological data in lung cancer prognosis prediction"	Clinical, Imaging, Histopathological	To integrate multiple data types for predicting lung cancer prognosis	Deep machine learning-based integration achieved higher accuracy in prognosis prediction compared to single data type analysis.
4. "Multi-omics analysis using deep learning for subtype classification in ovarian cancer"	Genomic, Epigenomic, Transcriptomic	To classify ovarian cancer subtypes using multi-omics data	Deep machine learning-based multi-omics analysis identified distinct subtypes and provided insights into tumor biology.
5. "Prediction of drug response in melanoma using multi-modal omics data"	Genomic, Transcriptomic, Drug response	To predict drug response in melanoma patients using multi-modal omics data	Deep machine learning accurately predicted drug response and facilitated personalized treatment selection.



**Figure 1.** Integration of histological slides and genetic susceptibility data in deep learning techniques for malignancy prediction. Histological slides provide strong evidence related to clinical manifestations of cancer such as neoplasms, malignant tumors, and metastasis. Deep learning techniques can combine these traditional image-based datasets with well-known genomic tests to make a strong model for early cancer diagnosis that is much more accurate than individual tests.

6. **Clinical Implementation:** The successful integration of deep machine learning models and predictive biomarkers into clinical practice can empower oncologists and researchers to make data-driven decisions. These models can be deployed as decision support systems or integrated into existing electronic health record systems, facilitating their accessibility and utilization in routine patient care.

## DISCUSSION

**Comprehensive Data Analysis:** Cancer precision medicine aims to consider multiple aspects of a patient's condition, including clinical information, genomic data, imaging findings, and pathology results. Deep machine learning enables the integration and analysis of these diverse data types, allowing for a more comprehensive understanding of the disease and the development of personalized treatment strategies. **Enhanced Predictive Modeling:** By leveraging deep learning algorithms, multi-data type processing enables the creation of predictive models that capture the complex relationships between various data modalities.

These models have the potential to provide more accurate predictions of treatment response, disease progression, and patient outcomes, leading to improved patient care and clinical decision-making. **Biomarker Discovery and Validation:** Deep machine learning can facilitate the identification and validation of predictive biomarkers that are associated with specific cancer subtypes, treatment responses, or prognosis. By considering multiple data types, such as genomic profiles, imaging features, and clinical variables, deep learning models can uncover patterns and biomarkers that would be challenging to detect using traditional statistical methods alone. The integration of diverse data sources enhances the robustness and generalizability of these biomarkers. **Personalized Treatment Selection:** The integration of multi-data type processing and deep learning models can assist oncologists in selecting the most appropriate treatment strategies for individual patients. By considering a patient's unique characteristics and combining data from different sources, such as genomic alterations, imaging characteristics, and clinical parameters, the models can help identify optimal treatment options tailored to each patient's needs. **Accelerated Translational Research:** Deep machine learning techniques can expedite the translation of research findings into clinical practice. By efficiently processing and analyzing large-scale and high-dimensional datasets, these methods can help identify potential biomarkers and treatment targets, enabling researchers to develop targeted therapies and interventions more rapidly. **Challenges and Considerations:** While deep machine learning offers significant potential in cancer precision medicine, there are challenges to address. These include the need for high-quality, annotated datasets, robust validation of models using independent cohorts, and the interpretability of deep learning models. Ethical considerations, data privacy, and regulatory compliance are also crucial factors to ensure the responsible and ethical use of patient data in research and clinical settings. Overall, the integration of deep machine learning techniques and multi-data type processing holds great promise in advancing cancer precision medicine. It enables a holistic approach to cancer care by considering diverse aspects of the disease, leading to improved predictive modeling, personalized treatment selection, and accelerated biomarker discovery and validation. Continued research and collaboration among clinicians, researchers, and data scientists are essential to realize the full potential of this approach and translate it into clinical practice for the benefit of cancer patients.

## CONCLUSION

Through multi-data type processing, deep learning algorithms can provide comprehensive insights into the complex nature of cancer, improving predictive modeling and treatment decision-making. The advantages of this approach include: **Improved Predictive Models:** Deep machine learning algorithms can effectively analyze and integrate multiple data types, leading to more accurate predictions of treatment response, disease progression, and

patient outcomes. This enables oncologists to make more informed decisions about treatment options. Biomarker Discovery: Deep learning facilitates the identification of predictive biomarkers associated with specific cancer subtypes, treatment responses, or prognosis. By considering diverse data sources, deep learning models can uncover hidden patterns and relationships that traditional statistical methods might miss. Personalized Treatment Selection: The integration of multi-data type processing and deep learning models enables the development of personalized treatment strategies. By considering a patient's unique characteristics across different data modalities, deep learning can help identify the most suitable treatment options for individual patients. Translational Impact: Deep machine learning accelerates the translation of research findings into clinical practice. By efficiently processing and analyzing large-scale and high-dimensional datasets, it facilitates the identification of potential biomarkers and treatment targets, leading to the development of targeted therapies and interventions. However, it is important to acknowledge the challenges and considerations in this field. The availability of high-quality, annotated datasets, rigorous validation using independent cohorts, and the interpretability of deep learning models remain ongoing challenges. Ethical considerations, including data privacy and regulatory compliance, must be carefully addressed to ensure responsible and ethical use of patient data.

## REFERENCES

1. Adhikari M, Jeena P, Bobat R, Archary M, Naidoo K, et al (2011). HIV-associated tuberculosis in the newborn and young infant. *Int J Pediatr.* 354208.
2. Arentz M, Pavlinac P, Kimerling ME, Horne DJ, Falzon D, et al (2012). Use of anti-retroviral therapy in tuberculosis patients on second-line anti-TB regimens: A systematic review. *PLoS ONE.* 8: e47370.
3. Babu S, Bhat SQ, Kumar NP, Anuradha R, Kumaran P, et al (2009). Attenuation of toll-like receptor expression and function in latent tuberculosis by coexistent filarial infection with restoration following antifilarial chemotherapy. *PLoS Negl Trop Dis.* 3: e489.
4. Banfield S, Pascoe E, Thambiran A, Siafarikas A, Burgner D, et al (2012). Factors associated with the performance of a blood-based interferon- $\gamma$  release assay in diagnosing tuberculosis. *PLoS ONE.* 7: e38556.
5. Bates M, O'Grady J, Mwaba P, Chilukutu L, Mzyece J, et al (2012). Evaluation of the burden of unsuspected pulmonary tuberculosis and co-morbidity with non-communicable diseases in sputum producing adult inpatients. *PLoS ONE.* 7: e40774.
6. Bhargava A, Chatterjee M, Jain Y, Chatterjee B, Kataria A, et al (2013). Nutritional status of adult patients with pulmonary tuberculosis in rural central India and its association with mortality. *PLoS ONE.* 8: e77979.
7. Bhaskaram P (2002). Micronutrient malnutrition, infection, and immunity: An overview. *Nutr Rev.* 60: S40-S45.
8. Black GF, Weir RE, Floyd S, Bliss L, Warndorff DK, et al (2002). BCG-induced increase in interferon- $\gamma$  response to mycobacterial antigens and efficacy of BCG vaccination in Malawi and the UK: Two randomised controlled studies. *Lancet.* 359: 1393-1401.
9. Chintu C, Bhat G, Luo C, Raviglione M, Diwan V, et al (1993). Seroprevalence of human immunodeficiency virus type 1 infection in Zambian children with tuberculosis. *Pediatr Infect Dis J.* 12: 499-504.
10. Co DO, Hogan LH, Karman J, Heninger E, Vang S, et al (2006). Interactions between T cells responding to concurrent mycobacterial and influenza infections. *J Immunol.* 177: 8456-8465.