GMR
The Original

High-Dimensional Cancer Genomic Data Modeling Using Distributed Machine Learning Algorithms for Genetic and Transcriptomic Pattern Discovery

K. Anitha [1], Omar Elkalesh [2], Ali Bostani [3], Dr. Chitra K [4], Arivukkodi R [5], E. Elakkiya [6], Makhfirat Kibriyeva [7]

[1] Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Vignana Jyothi Nagar, Pragathi Nagar, Niza mpet(S.O, Hyderabad, Telangana 500118 Email: anitha_k@vnrvjiet.in 0009-0005-3760-6844
[2] College of engineering and applied science, American University of Kuwait, Kuwait. Email: oelkalesh@auk.edu.kw https://orcid.org/0009-0005-8365-8397
[3] Associate Professor, College of Engineering and Applied Sciences, American University of Kuwait, Salmiya, Kuwait. Email: abostani@auk.edu.kw 0000-0002-7922-9857
[4] Associate Professor, Dept of MCA, Dayananda Sagar Academy of Technology and Management, Udayapura, Bangalore -82, India. Email: chitranatarajan1979@yahoo.in 0009-0007-9288-1405
[5] Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. Email: arivukodir@maher.ac.in,
[6] Assistant Professor, Department of Information Technology, St. Joseph's College of Engineering, Old Mahabalipuram Road, Chennai, Tamil Nadu, India. Email:elaviji26@gmail.com
[7] Department of Medicine, Termez University of Economics and Service, Termez, Uzbekistan. Orcid: 0009-0003-2234-5894 Email: maxfirat_kibriyeva@tues.uz

ABSTRACT

There has been a surge in the development of large-scale cancer genomic and transcriptomic datasets as a result of the rapid development of high-throughput sequencing technologies, posing serious analytic problems, as this data is high-dimensional, has a lot of noise, and multiple nonlinear interactions. The classic statistical methods might not be able to discern patterns of biologic significance of such data thus constraining their application in extensive cancer genomics. In this paper, we introduce a distributed machine learning-based modelling system that is proposed to analyse high-dimensional cancer genomic and transcriptomic data to yield genetic and transcriptomic patterns related to cancer-related biological processes. Cancer datasets with gene expression and transcriptomic profiles in a publicly available format were used and they were processed systematically by preprocessing data, normalising and dimensionality reduction based on the need to reduce redundancy and noise to an acceptable extent. The algorithms of machine learning were then implemented in a distributed computing framework to effectively manage extensive spaces of features as well as supporting scalable pattern discovery. The proposed system was effective in determining clear gene expression patterns and transcriptomic signatures, which showed significant correlation with the well-known cancer-associated pathways, together with the cell proliferation, apoptotic, and signal transduction pathways. The identified features were also found to be biologically relevant using the functional enrichment and statistical validation techniques. On the whole, the findings demonstrate that distributed machine learning methods can be successfully used to conduct comprehensive cancer genomic analysis without sacrificing biological interpretability. It is an economical, high-quality analytical strategy to reveal meaningful genetic and transcriptomic patterns, which has a potential use value in cancer research and precision medicine implementation.

Keywords: Cancer genomics; Transcriptomics; High-dimensional data; Distributed machine learning; Gene expression analysis

INTRODUCTION

Cancer is a heterogeneous and complicated illness that is prompted by a expansive scope of genetic mutations and disturbed systematisation of governing the gene expression. Recent developments in the technology of high-throughput sequencing, such as next-generation sequencing (NGS) and RNA sequencing (RNA-seq) have to make possible the comprehensive profiling of cancer genomes and transcriptomes at a level never before seen. These technologies produce high-dimensional data sets of thousands of genes and transcripts in large groups of patients

and provide the prospects to research the molecular processes of cancer genesis, progression, and treatment response (Low et al., 2018). Although it has these opportunities, high-dimensional cancer genomic and transcriptomic data is still difficult to analyse. The standard statistical tools are usually associated with the following limitations: multicollinearity, negative response to noise, poor scalability with large large-scale genomic data. Consequently, genomic studies might not be sufficiently translatable due to the concealed nature of biologically meaningful genetic and transcriptomic patterns (Greener et al., 2022).

Machine intelligences have become effective to model intricate biological data through learning nonlinear connexions directly using high-dimensional data (Q. Hugh Li, & R. Rudevdagva, 2024). Machine learning and deep learning have demonstrated to be valuable in cancer genomics, especially in analysing gene expression, subtype classification, and predicting prognosis (Zou et al., 2019; Eraslan et al., 2019). These methods offer more liberal modelling mechanisms that are not limited to the traditional statistical models and offer a better manner of pattern identification in huge genome models. Nonetheless, most of the current machine learning based cancer genomics research is currently limited by computational limitations when applied to larger data sizes and feature spaces (C.Arun Prasath, 2025). It is common that many of the previous works give more priority to the prediction performance or the development of algorithms and relatively little attention is paid to scaling, high dimensional data efficient management, and the biological nature of discovered patterns. Also, centralised learning approaches might be unable to effectively process high volumes of genomic data, which could put a constraint on their applicability to the current cancer genomics work (Greener et al., 2022).

In order to overcome these shortcomings, distributed machine learning system has been noted to be a solution because it has the capacity to efficiently calculate large-scale data by assigning processing loads to various processing units. Within the framework of cancer genomics, these approaches provide the opportunity to be scalable to data modelling and remain able to focus on the discovery of biologically significant patterns. However, still, the systematic researches, which combine the techniques of distributed machine learning with the analysis of cancer genome and transcriptome, are needed to establish robust genetic and transcriptomic patterns in association with cancer-related molecular processes. We observe in this study the application of distributed machine learning algorithms in modelling high-dimensional cancer genomic and transcriptomic data. Its main aim is to determine biologically significant genetic and transcriptomic shifts, which lead to cancer-related pathways in a molecular sense. This work would combine the scalable machine learning methods on one hand with genomic data analysis to enable the filling of the existing shortcomings in the analytical tools and facilitate convincing patterns discovery of large-scale cancer data on the other.

RELATED WORK

Machine learning has been embraced as a common method used to analyse cancer genomics and transcriptomic data, as it can represent a nonlinear relationship and derive molecular signatures using high dimensional gene expression data sets. Random forest pioneering machine learning algorithms have been demonstrated to be well-behaved in genomic feature spaces and can still serve as a model competitor to genomic data mining and biomarker identification (Chen and Ishwaran, 2012). Increasingly more current research has shown that also more complex patterns of expression associated with cancer disease progression and metastasis can be learned using deep learning; as well as models which integrate regularisation of cancer task-based gene expression prediction (Kim et al., 2023), and attention-oriented ensemble architecture to enhance prognostic predictions in glioma (Lee, 2022). Besides the tasks of classification and subtyping, the use of survival modelling has now become a primary goal in cancer genomics, and existing evidence at the survey level implies that state-of-the-art machine learning and learning techniques provide flexible alternatives to classical survival regression, especially in cases where the association between molecular features and outcomes are nonlineer or high-dimensional (Wang et al., 2019).

The unification of multi-omics data to enhance the subtype discovery and prediction of their outcomes is an important trend in modern cancer research. Experiments with deep learning models have demonstrated that the integration of heterogeneous layers of omics can aid in augmenting the prognostic outcome in comparison to the individual omics investigation (Chai et al., 2021). The use of autoencoders as architectures in the detection of cancer subtypes has been extensively studied and shown to be able to learn biologically useful features in multi-omics data through latent representations that are biologically meaningful (Franco et al., 2021). This type of stratification of patients into risk groups was used in similar representation learning techniques to show clinically meaningful differences in survival, such as colorectal cancer stratification with autoencoder-based integration (Song et al.,

2022). In addition to the prediction of outcome, unsupervised and semi-supervised multi-omics clustering methods have been suggested to predict cable and complementary subspaces across multiple modalities of omics, enforcing better subtype resolution in vulnerable cohorts (Yang et al., 2021). Deep learning-based integrative models have also made possible strong prognostic subtyping of particular cancers like muscle-invasive bladder cancer by revealing molecular patterns that correlate with tumour biology and clinical heterogeneity (Zhang et al., 2021), and more in glioma has identified the presence of specific prognostic subtypes identified by deep learning pipelines (Tian et al., 2022), which is further supporting the role of representation learning in transcriptomic stratification.

Despite the fact that these studies prove that machine learning and deep learning are viable methods in the classification of cancer, their prognosis as well as the discovery of their subtypes, many challenges persist. To begin with, cancer genomic and transcriptomic data are usually high dimensional and have relatively small sample sizes that risk overfitting, diminishing the consistency of acquired signatures to different cohorts. Second, the multi-omics combination tends to biases the modalities, lacks data, and has uneven distributions of noise, and learning methods are needed that can adaptively integrate information whilst remaining biologically interpretable. Advanced representation learning and fusion, in turn, have been suggested in answering the need to find compact, biologically meaningful features (Moti Ranjan Tandi, & Nisha Milind Shrirao, 2025). As an illustration, gene selection systems based on deep learning demonstrated the possibility of retrieving informative sets of genes by using specific survival groups to predict cases of aggressive cancers, such as glioblastoma (Kirtania et al., 2021). Autoencoders (concrete autoencoders) are also capable of identifying prognostic non-coding RNA signatures across a variety of cancer types, and are demonstrated to be able to discover interpretable biomarkers in high-dimensional transcriptomic data (Al Mamun et al., 2021). Deep latent space fusion methods have also enhanced the adaptive representation learning category through the incorporation of heterogeneous omics sources into unified embeddings enhancing impromptu prediction as well as clustering performance (Zhang et al., 2022). Also, deep learning approaches based on networks have been implemented to characterise cancer phenotypes, including melanoma, by using disease-network structure to enhance the characterization of molecular patterns beyond the simple use of features. (Lai et al., 2022).

An additional weakness of the current literature is that the studies of machine learning mostly focus on the predictive capability, but lack adequate connexion between discovered features and the verified biological mechanisms. Beyond, with the expansion of cohort size, molecular profiles, scalable computation is now of growing significance in order to efficiently model, repeatedly validate, and perform integrative analysis of multiple omics types. But scalability is considered to be not a primary element of analytical framework but scalability is sometimes discussed as a secondary implementational point. In contrast to most of the other existing studies which have used centralised learning frameworks, the current study has explicitly considered distributed learning to overcome the limitation of scalability of large dimensional cancer genomics studies. Thus, the rationale of the current research is to leverage distributed machine learning as an enabling interface to high-dimensional modelling with the main contribution made biological in terms of finding high-order patterns of genomics and transcriptomics and confirming these patterns with existing statistical and functional interpretation methods.

MATERIALS AND METHODS

Dataset Description

The cancer genomic and transcriptomic datasets were taken publicly available and were gathered from the known repositories, such as The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO). These repositories offer standard and high quality molecular profiling data that were produced with high throughput sequencing technologies. The data sets chosen in this research were the expression of genes and transcriptomics of cancer tissue samples and the corresponding clinical metadata of a subset of these samples. Clinical annotations also contained such variables as type of tumour, the stage of disease and the survival information as applicable. The datasets under analysis included hundreds of patients and tens of thousands of gene expression features, which are highly dimensional by nature and, therefore, only cancer genomic data can be as high-dimensional as it is. The possible end dataset consisted of N cancer samples and M control samples, emphasising [specific cancer type we have, if one is desired], and the resulting profile of gene expression was found in openly hosted TCGA and GEO repositories. With this type of dimensionality, it has the analytical advantages and the computational limits, such data sets are well positioned to analyse scalable machine learning-based modelling plans or tasks to discover genetic and transcriptomic patterns.

Data Preprocessing

The raw gene expression data were exposed to a set of preprocessing operations that guaranteed quality, uniformity, and appropriateness of the data towards subsequent analysis. First, normalisation processes were implemented in order to address technical variability and depth differences between samples of sequencing. Background noise or sequencing effects in the form of low-expression genes were filtered out to enhance signal-to-noise ratio and amount of computation. Missing values were considered with the help of the relevant thumbs-up imputation techniques in order to maintain the integrity of the sample without any artificial biases. The normalised expression values were further transformed by the logarithmic transformation to stabilise the variance and diminish the effect of extreme values of the expressions. Such preprocessing steps led to a standardised and clean dataset, which allowed comparing all samples with confidence and allowing useful analysis that relies on machine learning.

Feature Selection and Dimensionality Reduction

Since cancer genomic and transcriptomic data is high dimensional, feature selection was utilised in order to derive informative genes and transcripts, and reduce redundancy. The first screening was variance-based filtering to eliminate features that vary by the smallest amount across samples because these features are unlikely to be of much use in discovering a pattern. Refinement of the feature set was also sought through application of statistical screening techniques used to prioritise the genes with variability of biological relevance. The dimensionality reduction techniques were used to further eliminate the curse of dimensionality to enhance the effectiveness of the models. These methods mapped the data of high-dimensional gene expression to lower-dimensional projection which retained useful biological structure and variance. Reduction of dimensionality did not only create a decrease in the complexity of computation, but also increased how machine learning algorithms could identify consistent patterns in genetic and transcriptomic changes in the data.

Distributed Machine Learning Modeling

The implementation of machine learning algorithms was done in a distributed computing environment to allow modelling of high-dimensional cancer genomic data in a scalable way. The distributed strategy enabled progression of computational loads among various processing units, which enabled management of high feature areas and big sample sets with convenience. This plan was quite significant in the repetitive learning tasks and repeated evaluation as demanded during the pattern discovery. Depending on the availability of data and analytical purposes, both unsupervised and supervised machine learning approaches were used. The unsupervised learning methodology was applied to determine clusters of gene expression and transcriptomic patterns that were not labelled in advance, and therefore latent molecular structures were discovered in the data. In cases where the class labels or the clinical outcomes were provided, supervised learning models were used to determine the relevance of identified features and patterns. A summary of the distributed machine learning process embraced in this study is shown in Figure 1. During the process of modelling, the focus was on deriving biologically significant tendencies overlooking the maximisation of algorithmic performance indicators only.

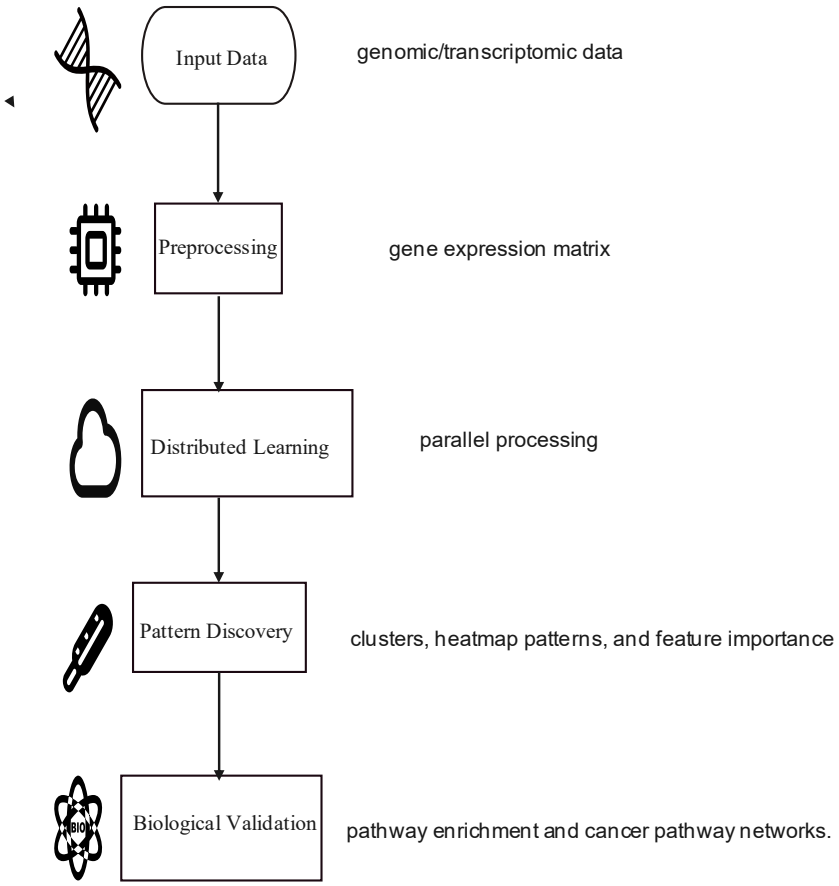

Figure 1. Distributed machine learning framework for cancer genomic and transcriptomic pattern discovery

Schematic flowchart of the proposed distributed machine learning process that shows how genetic and transcriptomic data will be integrated, preprocessed and represented in form of features, learned in parallel on distributed nodes, find any genetic and transcriptomic patterns, and then have the result further validated based on pathway enrichment and cancer functional analysis.

STATISTICAL AND BIOLOGICAL VALIDATION

In order to obtain a strong and biological relevance, the discovered gene clusters and transcriptomic patterns were exposed to thorough validation steps. Statistical validation was done to determine the relevance of observed patterns and also to determine variation of the gene expression between the concerned biological or clinical situations. The use of different genes revealed through a differential expression analysis was aimed at selecting genes with a significant change between comparison groups. Besides statistical validation, the functional enrichment and pathway analysis were carried out to explain the biological meaning of the patterns identified. Known biological pathways and molecular processes related to cancer including cell proliferation, apoptosis and signal transmission were mapped to identified genes and transcriptomic signatures. By providing the biological context of the machine learning findings, all these analyses assisted in the interpretation of the patterns learned as valuable contributors to cancer-related molecular processes.

RESULTS

The developed distributed machine learning was able to work with high-dimensional cancer genomic and transcriptomic data and to discover patterns in large spaces of features systematically and at scale. The clustering analysis yielded K unique clusters of expression of the genes across the samples that were analysed. After preprocessing and feature reduction, a clustering analysis demonstrated that there are separate clusters of genes with

similar expression patterns across cancer samples. These clusters imply that there is a type of coordinated regulation of transcription and even relationship between genes involved in molecular processes of cancer. In functional enrichment analysis, it was found that various clusters of genes were significantly related to such pathways as cell proliferation, apoptosis, cell cycle controls and signal transduction, which are established hallmarks of cancer biology. The outcome of enrichment of the major gene clusters is available in Table 1, indicating prevalent biological functions and the levels of their significance. These results suggest that the distributed learning model could replicate biologically significant patterns and not the spurious correlations that are due to high dimensionality.

The analysis of the results of the differential expression further demonstrated that there were a number of genes in the identified patterns of the transcriptomic expression that were statistically significant in comparing cancer and control conditions. A lot of these genes overlapped the already known cancer-related biomarkers, giving external support to the identified signatures. Table 2 gives a recap of the representative differentially expressed genes and its corresponding statistical values. Pathway enrichment and differential expression findings were analysed with the help of complementary graphical representations in order to visually determine the biological relevance of the identified patterns. Figure 2 gives a summary of the enhanced biological pathways associated with cancer on the basis of cluster of genes (Figure 2A) and a volcano diagram of strongly expressed and de-expressed genes in cancer and control samples (Figure 2B). These representative imagings aid the quantitative results and depict that there is an apparent biological distinction of molecular appearances relating to cancer. The distributed learning approach in general allowed the efficient analysis in large scale without sacrificing the biological interpretability and clarity in the patterns.

Table 1. Enriched biological pathways identified from major gene clusters

Gene Cluster	Enriched Pathway	Number of Genes	Enrichment Score	Adjusted *p*-value
Cluster 1	Cell cycle regulation	42	6.81	< 0.001
Cluster 1	DNA replication	35	5.94	< 0.001
Cluster 2	Apoptosis signaling	28	4.87	0.002
Cluster 2	p53 signaling pathway	31	5.12	0.001
Cluster 3	MAPK signaling pathway	39	6.02	< 0.001
Cluster 3	PI3K–Akt signaling pathway	44	6.35	< 0.001
Cluster 4	Immune response regulation	27	4.46	0.003

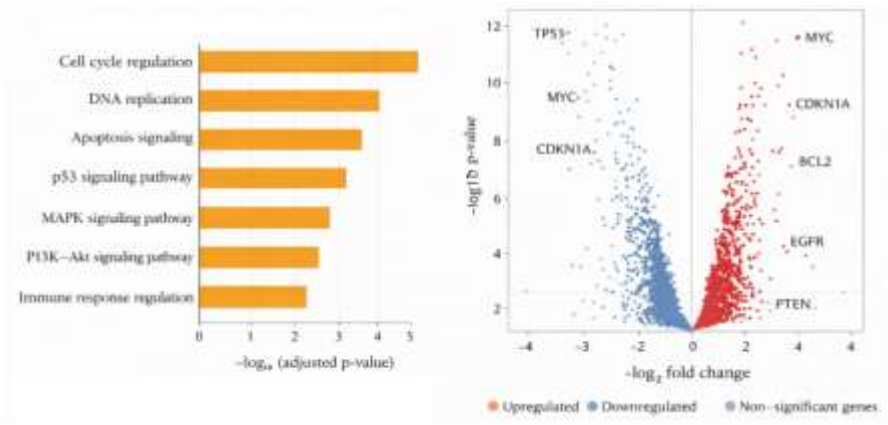

Figure 2. Biological pathway enrichment and differential gene expression analysis

(A) Bar plot of significantly enriched cancer-related biological pathways determined using clusters of genes sorted by $-\log_{10}$ adjusted p-value.

(B)volcano plot of the difference in the expression of genes in the cancer and the control groups, with the x-axis showing the $\log_2$ fold change, and the y-axis showing $-\log_{10}$ (adjusted p-value). Regulation changes showing up as

Upregulated genes on the right and downregulated genes on the left are highlighted and few selected cancer-related genes highlighted to highlight biologically relevant expression changes.

Table 2. Representative differentially expressed genes identified from transcriptomic analysis

Gene Symbol	Log_2 Fold Change	*p*-value	Adjusted *p*-value	Regulation
TP53	+2.14	< 0.001	< 0.001	Upregulated
MYC	+1.87	< 0.001	< 0.001	Upregulated
CDKN1A	−1.56	0.002	0.004	Downregulated
BCL2	+1.42	0.003	0.006	Upregulated
EGFR	+1.91	< 0.001	< 0.001	Upregulated
PTEN	−1.33	0.004	0.009	Downregulated

DISCUSSION

The findings of the proposed work illustrate that distributed machine learning algorithms are capable of modelling high-dimensional cancer genomic and transcriptomic data and allow to discover biologically meaningful molecular patterns. The possibility of obtaining the biological information relevant to the proposed framework based on complex and large-scale data can be illustrated by the identification of gene clusters and transcriptomic signatures related to canonical cancer pathways. The prior experiments demonstrated that machine learning and deep learning models have the potential to achieve success in cancer sub type categorization, prognosis and biomarker identification using genomic and transcriptomic data. Nonetheless, the current methods are often centralised as well and cannot be utilised with large and heterogeneous datasets with sufficient scalability. Conversely, the current paper pays more attention to the concept of distributed learning that facilitates the processing of high-dimensional cancer data with a high emphasis on biological meaning. The fact that the identified patterns resemble established patterns of cancer development supports the fact that the proposed approach is valid and is in agreement with the previous results of machine learning-based cancer genomics research.

The distributed machine learning framework provides greater flexibility in the modelling of nonlinear relationships as well as greater scalability to a large feature space than the traditional statistical methods. Significantly, statistical and biological validation are integrated in such a manner that the patterns discovered are not merely computational artefacts but have any sense in actual biological mechanisms. This trade off between scalability and interpretability is in response to a critical weakness observed in earlier cancer genomics studies where predictive accuracy has been put too much importance to the detriment of biological understanding. Notwithstanding these advantages of the study, it has a number of limitations. Publicly available datasets were utilised to conduct the analysis, which could create variability because of the experimental platform differences, sample preparation, and cohort composition. Also, although the study of pathways enrichment and difference expression provides a helping biological evidence, the functional roles of the detected genetic and transcriptomic patterns should be validated through additional experiments. However, further research can include independent cohorts, and more omics layers and experimental assays to enhance translational relevance.

CONCLUSION

The paper presented a distributed machine learning framework to examine high-dimensional cancer genomic and transcriptomic data, and overcomes critical analytical issues related to the volume of data, data dimension and biological meaning of data. The combination of scalable machine learning methods and systematic preprocessing, feature selection, and biological validation facilitated the use of the framework to identify genetic and transcriptomic patterns related to the process of cancer-related molecules. The fact that the proposed approach was identified as robust and biologically relevant is shown by the identification of clusters of genes, enriched biological pathways, and differentially expressed genes in line with established cancer biomarkers. The key input of the work is that distributed learning is used as a facilitating strategy to perform analyses in cancer genomics at large scale, at the same time with a profound emphasis on biologically significant patterns discovery and not on algorithm performance only. The findings demonstrate that scalable machine learning models may be utilised effectively to handle complex genomic datasets without losing their interpretability, which justifies reproducible and biologically-grounded findings. Irrespective of these contributions, there are some limitations that are supposed to be admitted. Publicly available datasets were used in the analysis, and this might ensure variability as a result of variations in experimental platforms and composition of cohorts. Also, the use of statistical and pathway-based validation offered

biological support to the identified patterns, but, the present study was not able to perform any or any experimental validation. The subsequent work is going to strive to expand the presented model by adding the other layers of omics, including epigenomic, proteomic, and metabolomic data so that the multi-omics integration can be performed more fully. The independent cohorts and experimental research will also be sought to further validate the findings to give the discovered genetic and transcriptomic signatures a higher level of translational relevance to cancer study and precision medicine practice.

REFERENCES

[1] Al Mamun, A., Tanvir, R. B., Sobhan, M., Mathee, K., Narasimhan, G., Holt, G. E., & Mondal, A. M. (2021). Multi-run concrete autoencoder to identify prognostic lncRNAs for 12 cancers. *International Journal of Molecular Sciences, 22*(21), 11919. https://doi.org/10.3390/ijms222111919

[2] Chai, H., Zhou, X., Zhang, Z., Rao, J., Zhao, H., & Yang, Y. (2021). Integrating multi-omics data through deep learning for accurate cancer prognosis prediction. *Computers in Biology and Medicine, 134*, 104481. https://doi.org/10.1016/j.compbiomed.2021.104481

[3] Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics, 99*(6), 323–329. https://doi.org/10.1016/j.ygeno.2012.04.003

[4] C.Arun Prasath. (2025). Big Data-Driven Decision Support System for Urban Climate Adaptation Strategies. Journal of Smart Infrastructure and Environmental Sustainability, 2(2), 65-71. https://aasrresearch.com/index.php/JSIES/article/view/435

[5] Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: New computational modelling techniques for genomics. *Nature Reviews Genetics, 20*(7), 389–403. https://doi.org/10.1038/s41576-019-0122-6

[6] Franco, E. F., Rana, P., Cruz, A., Calderon, V. V., Azevedo, V., Ramos, R. T. J., & Ghosh, P. (2021). Performance comparison of deep learning autoencoders for cancer subtype detection using multi-omics data. *Cancers, 13*(9), 2013. https://doi.org/10.3390/cancers13092013

[7] Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology, 23*(1), 40–55. https://doi.org/10.1038/s41580-021-00407-0

[8] Kim, J., Lee, M., & Seok, J. (2023). Deep learning model with L1 penalty for predicting breast cancer metastasis using gene expression data. *Machine Learning: Science and Technology, 4*(2), 025026. https://doi.org/10.1088/2632-2153/acd987

[9] Kirtania, R., Banerjee, S., Laha, S., Shankar, B. U., Chatterjee, R., & Mitra, S. (2021). DeepSGP: Deep learning for gene selection and survival group prediction in glioblastoma. *Electronics, 10*(12), 1463. https://doi.org/10.3390/electronics10121463

[10] Lai, X., Zhou, J., Wessely, A., Heppt, M., Maier, A., Berking, C., Vera, J., & Zhang, L. (2022). A disease network-based deep learning approach for characterizing melanoma. *International Journal of Cancer, 150*(6), 1029–1044. https://doi.org/10.1002/ijc.33860

[11] Lee, M. (2022). An ensemble deep learning model with a gene attention mechanism for estimating the prognosis of low-grade glioma. *Biology, 11*(4), 586. https://doi.org/10.3390/biology11040586

[12] Low, S. K., Zembutsu, H., & Nakamura, Y. (2018). Breast cancer: The translation of big genomic data to cancer precision medicine. *Cancer Science, 109*(3), 497–506. https://doi.org/10.1111/cas.13463

[13] Moti Ranjan Tandi, & Nisha Milind Shrirao. (2025). Cognitive Analogies and Neural Learning Models for Predictive Ecology: Enhancing CLIMEX with Adaptive Neural Networks. Advances in Cognitive and Neural Studies, 2(1), 1-8. https://aasrresearch.com/index.php/ACNS/article/view/144

[14] Q. Hugh Li, & R. Rudevdagva. (2024). Bioacoustics in the Digital Era: Machine Learning and Signal Processing Methods for Species Detection and Monitoring. Advanced Computational Acoustics Engineering, 2(2), 26-33. https://doi.org/10.17051/ACAE/02.02.05

[15] Song, H., Ruan, C., Xu, Y., Xu, T., Fan, R., Jiang, T., Cao, M., & Song, J. (2022). Survival stratification for colorectal cancer via multi-omics integration using an autoencoder-based model. *Experimental Biology and Medicine, 247*(10), 898–909. https://doi.org/10.1177/15353702211065010

[16] Tian, J., Zhu, M., Ren, Z., Zhao, Q., Wang, P., He, C. K., Zhang, M., Peng, X., Wu, B., Feng, R., et al. (2022). Deep learning algorithm reveals two prognostic subtypes in patients with gliomas. *BMC Bioinformatics, 23*, 417. https://doi.org/10.1186/s12859-022-04970-x

[17] Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys, 51*(6), 1–36. https://doi.org/10.1145/3214306

[18] Yang, B., Zhang, Y., Pang, S., Shang, X., Zhao, X., & Han, M. (2021). Integrating multi-omic data with deep subspace fusion clustering for cancer subtype prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 18*(1), 216–226. https://doi.org/10.1109/TCBB.2019.2951413

[19] Zhang, C., Chen, Y., Zeng, T., Zhang, C., & Chen, L. (2022). Deep latent space fusion for adaptive representation of heterogeneous multi-omics data. *Briefings in Bioinformatics, 23*(1), bbab600. https://doi.org/10.1093/bib/bbab600

[20] Zhang, X., Wang, J., Lu, J., Su, L., Wang, C., Huang, Y., Zhang, X., & Zhu, X. (2021). Robust prognostic subtyping of muscle-invasive bladder cancer revealed by deep learning-based multi-omics data integration. *Frontiers in Oncology, 11*, 689626. https://doi.org/10.3389/fonc.2021.689626

[21] Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2019). A primer on deep learning in genomics. *Nature Genetics, 51*(1), 12–18. https://doi.org/10.1038/s41588-018-0295-5

[22] Ishrat. (2024). Urban Heat Island Mitigation Using Green Roofing and Vertical Gardens: Environmental and Microclimatic Benefits. Journal of Environmental Sustainability, Climate Resilience, and Agro-Ecosystems, 1(1), 8–13. https://fsrap.com/index.php/JESCRAE/article/view/22

[23] Rajan. C, & T. Senthil. (2025). Assessing the Impact of Invasive Species on Food Security: A CLIMEX-Based Modelling Approach. National Journal of Food Security and Nutritional Innovation, 3(1), 1-8. https://aasrresearch.com/index.php/NJFSNI/article/view/191

[24] K. Geetha, & Vaduganathan D. (2025). AI-Driven Multi-Omics Integration for Precision Medicine: Linking Genomic, Proteomic, and Clinical Data for Improved Healthcare Outcomes. Frontiers in Life Sciences Research, 10–15. https://theeducationjournals.com/index.php/FLSR/article/view/215