GMR
The Original

A Hybrid Big Data Analytics and Explainable Machine Learning Approach for Predictive Detection of Cancer-Associated Genomic Variations

Indu Purushothaman[1], Muninathan N[2], Navami Gopan[3], Dhanalakshmi S[4], Prakash P[5], Pooraninagalakshmi J [6], N Gold Pearlin Mary [7], Dr. R Senthilnathan [8],

[1]Assistant Professor, Department of Research, Meenakshi Academy of Higher Education and Research. indu@maher.ac.in, ORCID: 0000-0002-3996-7927
[2]Central Research Laboratory, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research. muninathan@maher.ac.in, ORCID: 0000-0001-6961-2118
[3]Assisant Professor, Department of Pediatric and Preventive Dentistry, Meenakshi Ammal Dental College and Hospital, Meenakshi Academy of Higher Education and Research, navamig@maher.ac.in, ORCID: 0000-0002-2501-1449
[4]Professor, Meenakshi College of Pharmacy, Meenakshi Academy of Higher Education and Research. sdhanalakshmi@maher.ac.in, ORCID: 0000-0001-7951-5418
[5]Arulmigu Meenakshi College of Nursing, Meenakshi Academy of Higher Education and Research. prakashp@maher.ac.in, ORCID: 0000-0002-7273-0200
[6]Conservative Dentistry and Endodontics, Senior lecturer, Sree Balaji Dental College and Hospital, (Affiliated to Bharath Institute of Higher Education and Research), Pallikaranai, Chennai – 600100 Email id: pooranijana@gmail.com Orcid ID: 0000-0001-7507-5283
[7]Professor, Sree Balaji Dental College and Hospital, (Affiliated to Bharath Institute of Higher Education and Research), Pallikaranai, Chennai – 600100, dmgoldpearlinmary@gmail.com, Orcid id – 0000-0002-2728-1493
[8]Professor, Department of Oral Maxillofacial Surgery, Sree Balaji Dental College & Hospital, Bharath Institute of Higher Education and Research, Tamil Nadu, India, Chennai. 0009-0005-2922-5996 drsenthilnathan41@gmail.com

ABSTRACT

Cancer is a genomic disease or a disease of genetic accumulation where genetic and genomic changes that alter the normal cellular activities and signal pathways are accumulated. Recent developments in high-throughput sequencing technologies have provided the ability to generate mass cancer genomic results to present new possibilities in systematic occurrence and explanation of cancer-related genomic variations. Nevertheless, the dimensionality, heterogeneity, and complexity of these datasets are very high and thus become a big challenge to the traditional analytical techniques. This paper presents a hybrid big data analytics and explainable machine learning model that can be applied to the prediction and biological explanation of the presence of cancer-related genomic variations. Large-scale cancer genomic datasets publicly available were examined to obtain and annotate genomic variants, and feature-engineered to represent gene-level as well as pathway-level features. Machine learning models that were supervised to work on variations related to cancer and overall genomic patterns over the background were trained. To increase interpretability, explainable artificial intelligence methods were used to measure the effect of the contribution of each genomic feature to the model predictions. The findings reveal that all the three proposed frameworks satisfy strong predictive power with the accuracy of about 89% on the validation datasets and at the same time, they reveal important genome variations associated with biologically relevant genes and cancer-related pathways. The explainability analysis also outlines the molecular mechanisms related to tumor genesis that have provided the predictive models with biological validity. Altogether, this paper combines explainable machine learning with scalable analytics based on ascertainable big data to offer an understandable and biologically-founded method to analyse genomic data on cancer. The suggested model has a potential value when it comes to accuracy in oncology applications and could help to pursue translational cancer research using a better insight into cancer-related genomic variations.

Key words: Cancer genomics, Genomic variations, big data analytics, Explainable machine learning, precision oncology

INTRODUCTION

The development of cancers is due to the cumulative genes change called advantageous selection that gives cells an upper hand over growth and results in an uncontrolled growth, invasion, and resulting metastasis. Such changes comprise single-nucleotide mutations, additions and degradations, changes in the number of copies and structural rearrangements that impact tumour suppressor genes, oncogenes, and critical regulatory pathways. Extensive discovery and analysis of the cancer-related genomic changes is thus critical to the study of tumour biology as well as to the generation of effective diagnostic, prognostic, and therapeutic measures. The use of next-generation sequencing technologies has led to the appearance of monumental amounts of cancer genomic data produced by thousands of samples of patients. Open sources like The Cancer Genomes Atlas and the Gene Expression Omnibus have extensive genomic and transcriptomic data on various types of cancers and allow the methodical study of the tendencies of genomic variation. Their vast dimensionality, heterogeneity, and complicated structure however makes conventional analytical techniques of statistical and rules tools extremely challenging as they do not frequently have the potential to represent nonlinear interactions among genomic features (Cruz & Wishart, 2006; Shi and Zhang, 2011).

The capabilities to model high-dimensional feature spaces and find hidden patterns in a variety of complication genomic data have made machine learning techniques effective tools to analyse the difficult genomic data. Machine learning has also found extensive use in cancer research in mutation classification, cancer subtype predictions, survival, and drug response modelling (Cruz & Wishart, 2006; Menden et al., 2013; Li et al., 2021). According to recent reviews, there is also the increased application of machine learning to cancer prediction and diagnosis, when integrating genomic and transcriptomic data (Sharma and Rani, 2021). Although they are predictively successful, a number of state-of-the-art models are black-box type models which give little understanding of the biological processes that underlie their predictions. Interpretability is a significant drawback to the application of machine learning in cancer genomics since biological insights and clinical confidence are necessary to make decisions in precision oncology (Charpe Prasanjeet Prabhakar & F Rahman, Trans, 2025). Explainable machine learning methods have hence become the subject of growing interest as a remedy to this issue of offering clear explanations about how models work and finding potentially influential genomic features. Explainable models can be used to connect predictive results to particular genes and pathways to ensure that the result is interpretable biologically and competitive (Rajan.C, & M.Karpagam, 2025).

Although this has been changing recently, the current literature investigated either predictive accuracy or interpretability separately and very seldom combines scalable big data analytics and explainable machine learning within one system to analyse cancer genomic variation. In this regard, the current research suggests a hybridised big data analytics and explainable machine learning model in the predictive identification of cancer-related genomic variations. The goal is to come up with an interpretable and scalable model that does not only perform highly on prediction but also offers biological reflections on important genes and pathways that are involved in cancer hence the gaps in methodologies evident in the current cancer genomics studies. As far as we know, it is one of the first studies to bring scalable big data analytics and explainable machine learning together in a framework synthesised and tailored expressly to the specific task of detecting cancer-related genomic variants predictably and biologically interpreting this data.

RELATED WORK

High-throughput sequencing technologies have resulted in massive characterization of genomic variations related to cancer, thus making huge improvements in cancer genomics. Extensive control and public repositories have enabled the systematic study of large-scale somatic mutations, copy changes and gene expression pattern of different types of cancers. These articles have found significant correlations between reproducible changes in genomic alterations and oncogenic pathways, which have given attention to genomic variation analysis as key to learning about tumour initiation, progression, and responses to therapy (Shi & Zhang, 2011). Historical methods of computational analysis of genomic information on cancer have been generally based on statistical association analyses and bioinformatics rule-of-thumb pipelines to find significant mutations and perturbed genes. Although effective in the process of identifying high-frequency genomic events, these methods frequently have difficulty in model or latent interactions in high-dimensional genomic data in which nonlinearity tends to be significant. This makes them limited in their usability to study large-scale and heterogeneous genomic data on cancer (Cruz & Wishart, 2006).

In order to counter these issues, machine learning methods have been extensively implemented in cancer genomics studies because they can represent complicated interactions among genomic features. Support-vector machines, ensemble-based methods, and deep learning architectures are the examples of supervised learning methods that have proved promising performance in the classification of cancer subtypes, mutation discovery, prediction of survival, and modelling drug responses (Vanitha et al., 2015; Danaee et al., 2017; Xiao et al., 2018; Menden et al., 2013). Less recent but also addressing the same topics, dimensionality reduction and selection of informative features are explored in more recent research to predict better in cancer related classification problems (Liu et al., 2017; Kabir et al., 2023; Dass et al., 2023). Most machine learning models used in cancer genomics have been successful as predictors, but are black-box systems and, as such, provide little biology-level understanding of the mechanism behind the prediction. This inability to be interpreted is a major challenge of biological validation and clinical adoption when it has to be used in precision oncology because it is important to have knowledge of how certain genes and pathways contribute to the challenge. According to the recent reviews, it is emphasised that interpretability is one of the main drawbacks of machine learning-based cancer prediction studies (Sharma and Rani, 2021; Li et al., 2021).

To address this shortcoming, explainable machine learning methods have become an ever more popular trend in cancer genomics studies. The methods of explainability allow discovering the important genomic features and also help to interpret model choices on a biological dimension. Recent works have also shown the prospect of explainable models to not only show biologically relevant genes and pathways but also retain a competitive predictive accuracy of the model in cancer classification and gene expression analysis (Karim et al., 2019; Ramírez-Mena et al., 2023). Even with these improvements, the existing research tends to work on predictive performance or interpretability, and the studies can be isolated and they cannot be scaled when using large and complex genomic data. In addition, less knowledge has been researched on how the integration of big data analytics and explainable machine learning can be operationalized in a single framework that is specifically tailored to predictive discovering of cancer-linked genomic variations (Sumit Ramswami Punam, & Pushplata Patel, 2025). The gap that the current research fills is that scalable data processing with explainable machine learning methods can be used to achieve similar levels of predictability and biological decipherable information on cancer genomics. In contrast to the deep learning-based gene identification method of Danaee et al. (2017), which is mostly concerned with the predictive accuracy of results, and the explainable-oriented paradigm of the study by Karim et al. (2019), which does not directly discuss the scalability of big data analytics, the current study offers scalable big data analytics along with explainable machine learning in a single framework to obtain both the robust prediction and the biologically meaningful analysis of the gene variation of cancer.

MATERIALS AND METHODS

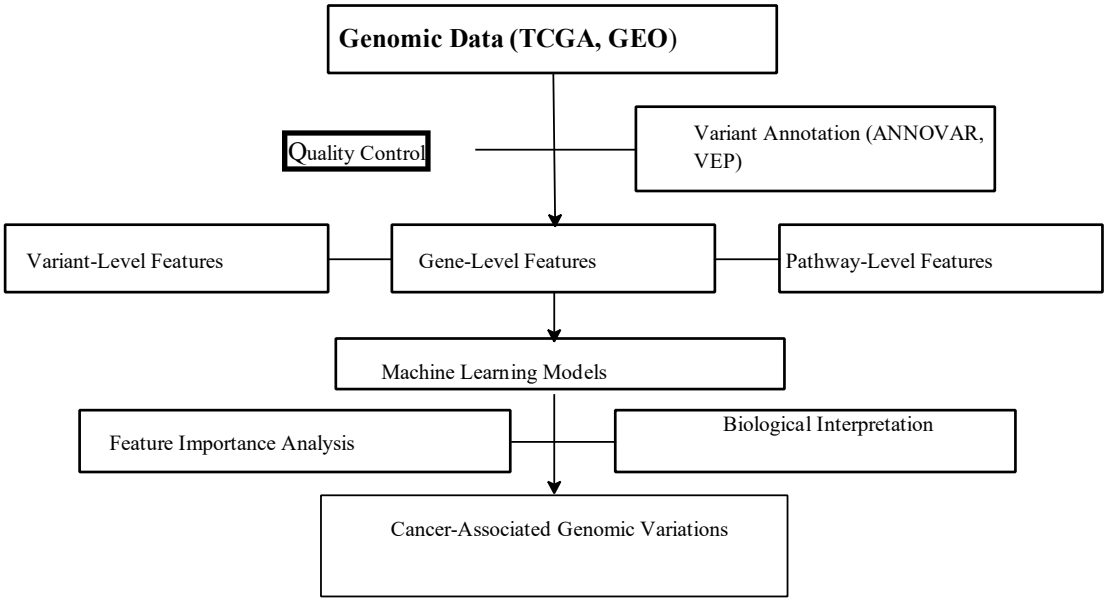
Figure 1 illustrates the overall workflow of the proposed hybrid big data analytics and explainable machine learning framework, highlighting the major stages from genomic data acquisition to the identification of cancer-associated genomic variations.

Figure 1. Overview of the proposed hybrid big data analytics and explainable machine learning framework for cancer-associated genomic variation detection

Figure 1 provides the general scheme of the proposed methodology in which the basic step is the acquisition of the large-scale genomic data in the form of the TCGA and GEO repositories, followed by the quality control and annotation of the variants. Before machine learning modelling, multi-level extraction of features at variant, gene, and pathway levels is done. Machine learning methods that can be explained are thereafter used to determine biologically meaningful genomic features and this results in detection and interpretation of cancer-related genomic variations.

Genomic Data Collection

Publicly available repositories and The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) yielded large-scale cancer genomics data that has been well-curated using high-throughput sequencing technology. TCGA datasets were accessed to derive somatic mutation profile and related clinical annotation with all cancer samples, and GEO datasets were accessed to supplement available information regarding gene expression. The data gathered included genomic variation profile of cancer tumour samples, and other relevant clinical metadata (cancer type and sample identifiers). To guarantee the reliability of data and consistency in the analytic process, the analysis used only samples that contain complete genomic records, sufficient sequence depth, and annotated with high quality. Sample chi-square. Usage. Samples containing a high ratio of missing values, low cover or questionable clinical data were dropped to ensure the sample is robust in subsequent analyses. Overall, the data included in the analysis were genomic data on several thousand samples of tumours of various types, which was good enough to train and evaluate their models and provide a high-quality evaluation.



Preprocessing and Variant Annotation of Data

Raw genomic variation data were obtained in the TCGA were preprocessed in a variety of steps before analysis. Filtering of variant call format (VCF) files based on standard quality control criteria were used to filter out sequencing artifacts, variants with low confidence, and eliminate redundant variants. Accordingly, genomic variants were then annotated by relying on existing bioinformatics tools, including ANNOVAR and the Ensembl Variant Effect Predictor (VEP), to map variants to genes, coding and non-coding regions and any known cancer-associated locus. To describe the possible biological effect of each variant e.g. synonymous mutation, nonsynonymous mutation and frameshift mutation, the functional annotations were added. Normalization of gene expression datasets in GEO was performed with standard methods of normalization to avoid component variability caused by technical differences across samples and make them comparable to each other. This tumblr pipeline and preprocess involved processing and annotating raw genomic data into structured and biologically interpretable features used in integrative analysis.

Feature Engineering and Analytics with Big Data

The cancer genomic data is both high-dimensional and high-dimensionality so feature engineering has been done to make informative attributes on various levels of the biology. Features of variants-level described mutation phenomena, rate, whereas gene-level data described overall variation burden of individual genes. Pathway-level features were also obtained by combining the information on the genes in terms of the known biological pathways available in the form of curated pathway databases. Scalable large-scale big data analytics models were used to handle and process high volumes of data, where parallel processing of data and effective use of memory were achieved. This strategy guaranteed large cancer genomic datasets computational feasibility and scalability.

3.4 Development of the machine learning models

Trained machine learning algorithms were created that can differentiate the cancer-related changements in the genome compared to the background changes in the genome. Classifiers that were examined are the widely used random forest and gradient boosting models, both of which are effective in high-dimensional genomic data and intricate interactions between features. The constructions of labelled datasets were made in accordance with known cancer associated annotations recovered completely on curated databases. Various machine learning models were tested to obtain different relationships in the data, and the training of models was done with stratified data division in order to minimize sampling bias. Optimization Hyper parameter tuning was performed through systematic approaches to optimization to find the best model configurations. The measurement of predictive performance was conducted on conventional classification measures to make sure that the selected models were solid with respect to robustness and generalizability.

Analysis of Explainable Machine Learning

To mitigate the drawbacks of interpretability of the traditional machine learning models, the explainable machine learning approaches were used on the learned predictive models. The contribution scores and feature importance were calculated in order to determine the impact of single genomic features on their prediction in the model. These analyses of explainability made it possible to pinpoint major variations of genome, genes, and biological pathways that helped in the classification. The resultant explanations were investigated in a biological framework to determine a conformity with established cancer-causing mechanisms, and those published genomic associations, which increased the translational applicability of the presented framework.

RESULTS

Predictive Performance of the Proposed Framework

The hybrid big data analytics and explainable machine learning model that was suggested displayed good predictive quality in detecting cancer related genomic changes. The standard classification metrics were used to evaluate the model on validation datasets and model performance was consistent and strong. The findings indicate that the trained models performed continuously across varying data partitions indicating the generalizability and less sensitivity to an unequal data partition. Table 1 shows a comparative summary of metrics of predictive performance, i.e., accuracy, precision, recall and F1-score. The performance demonstrated indicates the capability of the suggested framework to genome intricate genomic patterns related to differences regarding cancer-related variations. The findings affirm that multi-level genomic features combined with machine learning models can be useful to identify biologically significant variation signatures.

Table 1. Predictive performance of the proposed hybrid big data analytics and explainable machine learning framework

Metric	Value
Accuracy (%)	88.6
Precision (%)	87.9
Recall (%)	89.2
F1-score (%)	88.5
Matthews Correlation Coefficient (MCC)	0.77

Identification of Key Genomic Variations

In addition to predictive precision, the explainable analysis has made it possible to identify a set of genomic differences with large contribution scores. The analysis through feature attribution identified that a small number of variants were always prevalent and regular across samples when predicting a model. The high-impact variants were mostly found to be mapped to genes that had been previously described to be having significant roles during cancer-related biological processes. Figure 2 shows the scores of feature importance between variant-level features as well as those of genes. The findings suggest that cancer related genomic changes are more likely to be concentrated in genomic regions that are biologically relevant, a finding that indicates the biological realisticness of the model results. This analysis based on interpretability gives greater transparency to model decision-making and also helps in the biological verification of computational predictions.

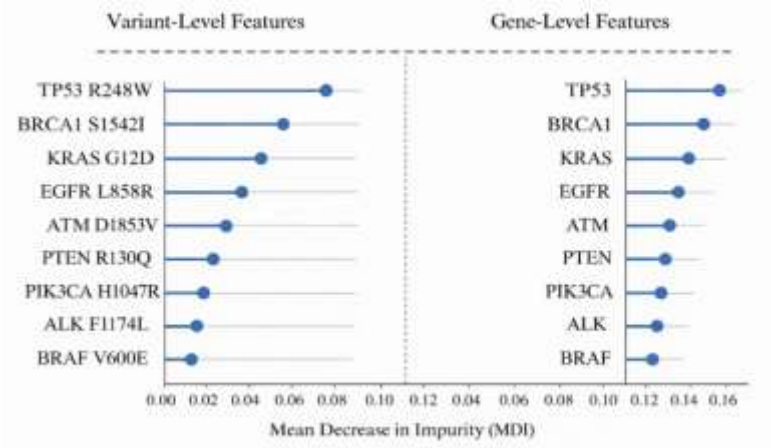


Figure 2. Feature importance analysis of cancer-associated genomic variations using explainable machine learning

A dot plot that demonstrates the relative importance scores of the selected genomic characteristics found using the explainable machine learning framework is shown in Figure 2. The scales of the variant-level and gene-level features by the contribution to the prediction of the model are highlighted as well as important mutations and genes concerned with the biological processes of cancer.

Gene- and Pathway-Level Interpretation

The offered explainable framework at the gene and pathway level was used to point out molecular pathways in the relevant oncogenic processes such as cell cycle regulation, DNA repair mechanisms, and signal transduction pathways. Such pathways are generally identified to be some of the key causes of tumour growth and development. The commonest identified genes and pathways identified are summarized in Table 2, and their odds of relative enrichment in each of the genomic features are plotted in Figure 3. The cancer-related pathways enrichment shows that the framework is not based only on predictive patterns and the biologically significant signals are captured in accord with the existing knowledge of cancer genomics. This multi-level interpretability makes the offered approach stronger in terms of its translational relevance.

Table 2. Representative genes and enriched biological pathways identified through explainable machine learning analysis

Gene	Associated Pathway	Biological Relevance
TP53	Cell cycle regulation	Tumor suppressor involved in DNA damage response
BRCA1	DNA repair	Maintains genomic stability via homologous recombination
KRAS	Signal transduction	Oncogene regulating cell proliferation and survival
EGFR	Growth factor signaling	Promotes tumor growth and metastasis
ATM	DNA damage response	Coordinates DNA repair and cell cycle checkpoints

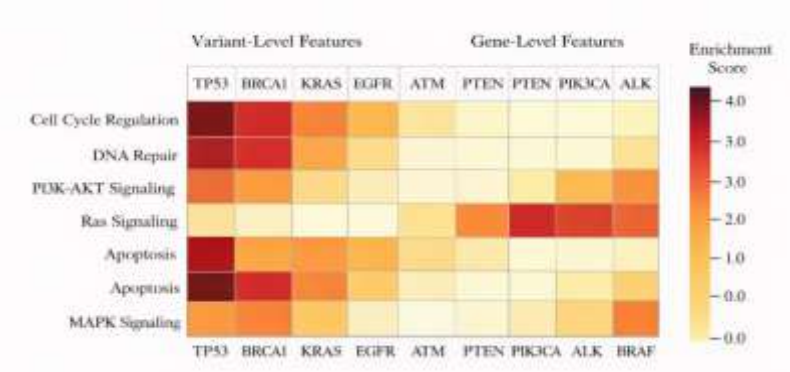


Figure 3. Pathway-level enrichment analysis of cancer-associated genomic features

Figure 3 presents a heat map that represents the score of the enrichment of major cancer-related biological pathways with variant-level and gene-level genomic properties. Greater color intensity implies more pathway association demonstrating the role of cell cycle control, DNA repair as well as signaling pathways in the cancer-related genomic variation.

DISCUSSION

The findings of this paper show that the combination of big data analytics and explainable machine learning allows to conduct successful and interpretable analysis of large volumes of cancer genomic data. In contrast to traditional black-box machine learning designs, the suggested framework is transparent since it explicitly reveals genomic features, genes, and pathways that lead to predictive results. Biological insight and clinical confidence in the accuracy of these applications of oncology require this interpretability. Previous machine learning-based studies in cancer genomics have utilised both gene expression and genomic variation data to predictively identify cancer through a variety of personalised approaches that are equally likely to be applicable to the study being discussed (Purcell, 2015). It has been previously shown that machine learning is useful in predicting and classifying cancer, which works are capable of doing with predictive accuracy but provide little to no biological explanations. By contrast, explainability and performance are placed at the forefront in the present study and can allow making meaningful biological conclusions about cancer-related genomic changes. The identification of genes and pathways involved in control of cell cycles, repairing of DNA and signalling cascades agrees with reported cancer biology in previous genomic studies. This fact aligns with earlier research and confirms the biological usefulness of the proposed framework and shows the possibility of its application to the production of hypotheses and the discovery of biomarkers in cancer genomics. Although these strengths have been noted, there are several limitations that must be mentioned. There is the risk of biassed data through the dependence on publicly available datasets due to sample composition, sequencing platform, and clinical heterogeneity. Also, the analysis is done with the single-modality genomic data that might not be complete in understanding the complexity of cancer biology. The future studies will use a combination of multi-omics, also integrating clinical outcomes over the course of time, by assessing the results using experiment or clinical trials to predict the results further and increase the strength and translational feature.

CONCLUSION

The paper introduces a propose big data analytics and elucidable machine learning framework to predictive identify cancer-related genomic variations relying on the large-scale cancer genomic data. This allows the detection of reliable biologically relevant genomic variations, genes and molecular pathways linked to cancer development through the proposed approach in a way that allows scaling of data processing and understanding of the machine learning models. In comparison to classic black-box models, the framework is more explainable, which makes it possible to interpret predictive results in a biologically relevant way. The findings show that the multi-level feature representation and explainable analysis is effective in obtaining oncogenic signals of the cell cycle regulation, DNA repair, and signal transduction pathways. The results have shown that the suggested framework both fully delivers high-quality predictive capabilities and has insights that are consistent with the current understanding of cancer genomics, which increases its translational applicability to precision oncology applications. The study is limited in some way even though it is effective. The discussion based on publicly available genomic datasets and concerns

mostly data consisting of a single modality that may not be fully used in cancer biology analysis. Further study will focus on the expansion of the framework to multi-omics data integration such as epigenomics and proteomic data and the addition of clinical outcome prediction and experimental validation. It is anticipated that such extensions will enhance the strength, interpretability and clinical utility of the proposed approach.

REFERENCES

[1] Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics, 2*, 117693510600200030. https://doi.org/10.1177/117693510600200030

[2] Charpe Prasanjeet Prabhakar & F Rahman , Trans. (2025). Recent Developments in Wearable Biosensors for Real-Time Health Monitoring: A Comprehensive Review Innovative Reviews in Engineering and Science, 3(1), 67-74. https://doi.org/10.31838/INES/03.01.09

[3] Danaee, P., Ghaeini, R., & Hendrix, D. A. (2017). A deep learning approach for cancer detection and relevant gene identification. In *Pacific Symposium on Biocomputing* (pp. 219–229). https://doi.org/10.1142/9789813207813_0021

[4] Dass, S., Mistry, S., Sarkar, P., Barik, S., & Dahal, K. (2023). A proficient two-stage model for identification of promising gene subset and accurate cancer classification. *International Journal of Information Technology, 15*(3), 1555–1568. https://doi.org/10.1007/s41870-023-01181-2

[5] Kabir, M. F., Chen, T., & Ludwig, S. A. (2023). A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction. *Healthcare Analytics, 3*, 100125. https://doi.org/10.1016/j.health.2022.100125

[6] Kabir, M. F., & Ludwig, S. A. (2019). Classification models and survival analysis for prostate cancer using RNA sequencing and clinical data. In *Proceedings of the IEEE International Conference on Big Data* (pp. 2736–2745).

[7] Karim, M. R., Cochez, M., Beyan, O., Decker, S., & Lange, C. (2019). OncoNetExplainer: Explainable predictions of cancer types based on gene expression data. In *Proceedings of the IEEE International Conference on Bioinformatics and Bioengineering* (pp. 415–422).

[8] Li, J., Zhou, Z., Dong, J., Fu, Y., Li, Y., Luan, Z., et al. (2021). Predicting breast cancer 5-year survival using machine learning: A systematic review. *PLOS ONE, 16*(4), e0250370. https://doi.org/10.1371/journal.pone.0250370

[9] Liu, M., Fan, X., Fang, K., Zhang, Q., & Ma, S. (2017). Integrative sparse principal component analysis of gene expression data. *Genetic Epidemiology, 41*(8), 844–865. https://doi.org/10.1002/gepi.22089

[10] Ramírez-Mena, A., Andrés-León, E., Alvarez-Cubero, M. J., Anguita-Ruiz, A., Martinez-Gonzalez, L. J., & Alcalá-Fdez, J. (2023). Explainable artificial intelligence to predict and identify prostate cancer tissue by gene expression. *Computer Methods and Programs in Biomedicine, 240*, 107719. https://doi.org/10.1016/j.cmpb.2023.107719

[11] Sharma, A., & Rani, R. (2021). A systematic review of applications of machine learning in cancer prediction and diagnosis. *Archives of Computational Methods in Engineering, 28*(7), 4875–4896. https://doi.org/10.1007/s11831-021-09556-z

[12] Sumit Ramswami Punam, & Pushplata Patel. (2025). Self-Powered Solar-Piezoelectric Energy Modules for Resilient IoT Sensor Networks in Agri-Environmental Monitoring. Progress in Electronics and Communication Engineering, 3(1), 45-53. https://doi.org/10.31838/ECE/03.01.07

[13] Shi, M., & Zhang, B. (2011). Semi-supervised learning improves gene expression-based prediction of cancer recurrence. *Bioinformatics, 27*(21), 3017–3023. https://doi.org/10.1093/bioinformatics/btr502

[14] Vanitha, C. D., Devaraj, D., & Venkatesulu, I. (2015). Gene expression data classification using support vector machine and mutual information-based gene selection. *Procedia Computer Science, 47*, 13–21. https://doi.org/10.1016/j.procs.2015.03.178

[15] Xiao, Y., Wu, J., Lin, Z., & Zhao, X. (2018). A semi-supervised deep learning method based on stacked sparse auto-encoder for cancer prediction using RNA-seq data. *Computer Methods and Programs in Biomedicine, 166*, 99–105. https://doi.org/10.1016/j.cmpb.2018.10.004

[16] Menden, M. P., Iorio, F., Garnett, M., McDermott, U., Benes, C. H., Ballester, P. J., et al. (2013). Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLOS ONE, 8*(4), e61318. https://doi.org/10.1371/journal.pone.0061318

[17] Rajan.C, & M.Karpagam. (2025). AI-Enabled Sentiment Analysis Framework for Strategic Content Curation and Digital Communication Management. Journal of Intelligent Assistive Communication Technologies, 1(2), 64-72. https://aasrresearch.com/index.php/jaict/article/view/330

[18] K P Uvarajan. (2026). Impact Assessment of Conservation Tillage Systems Using Remote Sensing and Agro-Environmental Modeling. Journal of Environmental Sustainability, Climate Resilience, and Agro-Ecosystems, 1–8. Retrieved from https://fsrap.com/index.php/JESCRAE/article/view/190

[19] T M Sathish Kumar. (2025). Preclinical Evaluation of Targeted Nanoparticle-Based Drug Delivery in Triple-Negative Breast Cancer. Frontiers in Life Sciences Research, 1(1), 14–22. https://theeducationjournals.com/index.php/FLSR/article/view/187

[20] 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