TRENDS IN LIFE SCIENCES AND BIOTECHNOLOGY

BiotechLife
Trends in Life Sciences and Biotechnology

Https://lifebiotrends.com

Support@lifebiotrends.com

ISSN: 3080-292X (Print)
ISSN: 3080-2938 (Online)

INTEGRATION OF GENOMIC AND CLINICAL DATA FOR PERSONALIZED TREATMENT PLANNING

Noor Ul Ain[1*]

[1] *Department of Precision Medicine and Bioinformatics, Institute of Biomedical Sciences, Karachi, Pakistan*

*Corresponding Author E-mail: noor.ain@prcision.edu.pk

Abstract

This means that cancer heterogeneity demands the integrative analysis of multi-omic data by default to give precision oncology in accurate prognostics. High-dimensional genomic, epigenomic, and clinical data translation into clinically actionable survival predictors is discussed in the paper, where a Multi-Omic Attention Network (MOANet) is presented that combines network-based prior knowledge, cross-modal attention, and quantification of uncertainty. RNA-seq, DNA-methylation, somatic mutation and clinical data of a retrospective cohort of 1,284 breast, lung and colorectal adenocarcinoma patients were harmonized. It comprised ComBat batch correction, Multi-Omics Factor Analysis to dimension-reduce, propagation of networks on protein-protein interaction networks, and a transformer-based attention model with Shapley additive explanations to interpretability. Relative comparison of MOANet with eight state-of-the-art models showed that MOANet outperformed prognostic characteristics of time-dependent AUC of 0.837 at 36 months and concordance index of 0.791 which represented 16.3 percent relative improvement over clinical-only Cox-PH models. All the omic layers were crucial as observed by ablation experiments, but transcriptomic data alone have the greatest impact. The model showed almost perfect calibration, with a strong performance with missing modalities which maintained an ideal time-dependent AUC of over 0.73, and a strong performance with federated learning performance with an overall time-dependent AUC of 0.845 with a small privacy leak. The network-derived biomarkers that were identified as having the most attributes were TP53, PIK3CA, MUC16 and FAT4 it should be noted that prior knowledge of biology may be useful. These findings confirm that multi-omic combination with explicit uncertainty estimation and federated functions involve attention-based combination, which results in a powerful, interpretable, and clinically translational model of precision oncology prognostication.

Keywords: *Precision Oncology, Integration of Multi-Omics, Attention Networks, Prognostic Modelling, Federated Learning, Network Biomarkers.*

Article History

Received: January 22, 2026

Revised: March 21, 2026

Accepted: May 15, 2026

INTRODUCTION

One of the fields is precision oncology, which tends to enhance the patient experience by designing therapeutic interventions according to the genetic and molecular features of a tumor of a particular patient (Zhou et al., 2024). This paradigm will allow taking high-dimensional multi-omics data to the next stage to promote diagnostic accuracy and reach a maximum prognostic, going beyond the traditional, broad-spectrum therapeutic modalities (Hsu et al., 2025). However, the limited-dimensional omics data are frequently applied to analyze the information through traditional means that cannot give a holistic perspective of the complex etiology and pathogenesis of malignancy and demand more multi-platform integrative solutions (Bhattacharyya,Kiran, and Mahsood; 2023; Huang et al., 2017). The approaches are necessary to tackle the heterogeneity inherent in cancer that is conditioned by different genetic, environmental, and molecular factors resulting in vast differences among patient outcomes in identical histological subtypes (Nguyen & Vafaee, 2025). Considerable evidence demonstrates that the analysis of both types of multidimensional clinical and multi-omics data has stronger prognostic power compared to the analysis of clinical data (Ng et al., 2020). This is significant even though these integrated frameworks are predictive but cannot be easily translated into clinical applications since these data standardization are intrinsically complex, complex computational models, and difficult to interpret biological systems (Pant et al., 2024). To overcome these challenges, the existing approach to methodological advancements in successively applying data-driven network representations and dimensionality reduction algorithms to reduce high-dimensional and heterogeneous data to lower dimensional and more interpretable space (Xu et al., 2019). Moreover, the use of previous biological information in these data-driven network models enables the discovery of more resilient biomarkers, which often are more effective than those obtained in traditional differential gene expression analyses (Cantafio et al., 2019). Alongside a basic biomarker detection, there is an increase in the use by researchers of functional models, such as patient-derived organoids and xenografts, to validate the therapeutic utility of these molecular signatures in complex biology (Bhattacharyya, 2023). To close the disconnect between these integrated datasets and clinical practice, a high level of multi-centre validation is needed to guarantee the sensitivity and specificity of emergent biomarkers (Gao et al., 2025). Such translational transition demands

intensive collaborative work between computational researchers and clinicians to put high-dimensional outputs into clinical actionable insights (Chenel et al., 2024). Development of standardized computational systems that will co-analyze the individual molecular layers but will in fact integrate, and multi-modally synthesize, biological information is the next step (Catalano et al., 2025; Piroozkhah et al., 2023). In that regard, the retrospective analytical studies will have to refocus their attention on prospective and multi-institutional clinical trials to eradicate the existing limitations of generalizability of the data and strength of the diagnostic findings (Zafar et al., 2025). At the same time, to coordinate heterogeneous data sets and to adequately translate such multi-omics data into standard clinical practice, community-driven data sharing efforts and strong interdisciplinary collaboration will be critical (Morabito et al., 2025; Shoaib et al., 2024). Moreover, the more advanced deep learning architectures that can perform the automatic normalization to reduce platform-specific biases and fine-tune multi-modality features learning to increase the overall accuracy of multi-omic analyses should be implemented (Miao et al., 2026). In particular, the use of standardized strategies, including Multi-Omics Factor Analysis and netDX, offers the required reproducibility of describing disrupted biological pathways in various oncological cohorts (Vlachavas et al., 2021). Additionally, with quantitative uncertainty estimation integrated in these models, clinicians will get a needed confidence measure of each prognostic prediction, which will minimize potential risks associated with algorithmic black-box behavior (Wekesa & Kimwele, 2023). In addition to the creation of algorithms, data formats and readily available software interfaces must be designed to ensure that such diagnostic tools can be integrated into the routine clinical practice (Piroozkhah et al., 2023). Moreover, radiomic signatures can also be included synergistically in these workflows, which will provide a non-invasive chance to facilitate the phenotypic diversity, which complements the current knowledge in classical pathology and thus will avoid the current limitations (Biswas and Chakrabarti, 2020). This could be extended to cover phenomic data in conjunction with radiomics and transcriptomics to allow the complete decoding of the heterogeneity of tumor microenvironment (Kang et al., 2023). To reduce the barriers to entry to clinical integration, such synthesis requires the creation of standardized software frameworks, which are more focused on interoperability and high-quality documentation (Tabakhi et al., 2022). Compliance with the principles of the FAIR

is relevant in the process as it allows verifying the research objectives to be tracked, retrieved, and interoperable, which will verify them independently (Shui et al., 2021). In order to further increase the generalizability of these diagnostic frameworks, the world-wide level of population cohorts should be included in the future research to make sure that multi-omics models could be sufficiently able to explain the global genomic and environmental heterogeneity (Rabbie et al., 2021). In order to reduce technical biases of the single-source data collection, the researchers are encouraged to employ site-specific model adaptation strategies and extend the scope of modalities to incorporate spatial transcriptomics and proteomics, that can give high-resolution information about tumor microenvironment (Zhang et al., 2023; Zhou et al., 2025). The increasing unavailability of spatial omic data can take this field to a higher level and fill the knowledge gap between morphology and bulk omic data to provide more useful and interpretable information pertaining to biology (Jennings et al., 2025). Besides, the effective application of multidimensional data involves the development of consistent networks of feature alignment and fusion in order to give high-fidelity representations of various clinical and molecular modalities (Zheng et al., 2025). Besides statistical fusion, multimodal transformer architectures can also be viewed as potentially having the capability to describe more complex interrelation between heterogeneous streams of data and generate more biologically interpretable features (Feng et al., 2024; Salmanpour et al., 2025). The architectures can compute parallel clinical images and high-dimensional genomic profiles, and this is what allows them to compute a holistic diagnostic approach, which is multimodal in nature to the oncological decision-making process (Pérez-López et al., 2023). Secondly, explainable AI features, including Class Activation Mapping and Shapley Additive Explanations are required to visually depict areas of decisions and estimate how individual multimodal features would be used to make clinical judgments (Shi et al., 2025). These approaches do not only furnish the biological relevance of model-generated knowledge but also create the confidence to apply them on a regular basis when dealing with patients because of high-dimensional inputs to model-generated knowledge that can be easily converted into therapeutic recommendations (Tran et al., 2021). Besides these interpretability measures, the federation of learning protocols will need to be implemented to enable multi-institutional validation, patient data confidentiality and ultimately emergence of highly valid prognostic

models with global applicability (Stillwell, 2025). Further, future work should focus on resistant architecture to missing data modalities, which is often a bottleneck in practice in clinical settings, and full multimodal profiles are often not possible (Jennings et al., 2025).

METHODOLOGY

The research methodology will be developed as a problem-oriented multi-step analytic pipeline that will help to overcome the inherent problems of integrating heterogeneous multi-omics data to deliver reliable oncology. The study adheres to a retrospective cohort study design, with publicly available multi-omics data sets of genomic mutations, transcriptomic expression assays, DNA methylation array and clinical outcome measures in patients with heterogeneous solid tumors, such as breast, lung, and colorectal cancer among others. The principles of FAIR inform the data acquisition process in which all datasets are Findable, Accessible, Interoperable and Reusable. The most important methodological issue is how to dimension-reduce these data, which is very high, and derive biologically meaningful data and address platform-specific batch effects. In order to cope with this, the first stage implements multi-step normalization and harmonization. In the case of transcriptomic data, raw counts are normalized by variance-stabilizing normalization, and quantile-controlled. It uses the beta-values to transform them into M-values in order to enhance statistical validity with the example of the methylation data. To remove technical batch effects between sequencing systems, a harmonization model, based on ComBat, is employed, and that is defined as:

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Xij is the normalized level of the expression of the gene i at sample j, mu is the mean value of each expression level, a i is the gene effect, b j is the batch effect and e ij is the error. A dimensionality reduction method, whose application is based on Multi-Omics Factor Analysis, follows the harmonizing process, and breaks down the consolidated data matrix together into a pool of latent factors. The model of the joint factorization is:

$$X^{(m)} = ZW^{(m)T} + \varepsilon^{(m)}$$

where X(m) is the omic type M data matrix, Z is the shared latent factor matrix of omics, W (m) is the loading matrix and ε (m) is the noise. This will reduce the size of the molecular features by thousands to fewer interpretable latent variables that capture cross-omic covariance. A network-based feature selection algorithm, which makes

use of preexisting biological knowledge as represented by protein-protein interaction databases is adopted to further reduce prognostic biomarkers. The score of network propagation of the individual gene is derived as:

$$F_{t+1} = \alpha A F_t + (1-\alpha) F_0$$

Ft=the vector of the scores on the genes at iteration t A is the normalized adjacency matrix of the interaction network, and alpha is the damping factor of propagation and F0 t is the initial node score acquired of the differential expression analysis. This methodology finds strong biomarkers not just statistically significant but also those with functional links in established oncogenic pathways. To predict prognosis, a multimodal deep learning model based on a combination of clinical variables, latent omic variables, and chosen network biomarkers are constructed. NB The model employs a cross-attention system to learn inter-modal relationships which are as follows:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Q, K and V matrices are query, key and read value matrices, respectively, depending on different omic modalities, and dk is the dimensionality of the key vectors. To compute values of Shapley Additive Explanations, each prediction is computed additively:

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j$$

and g (z prime) represents the model of the explanation, phi 0 is the prediction without features, phi j is the Shapley value of feature j and z j prime is the presence of a feature. Five-fold nested cross validation is used in a model validation to assist in eliminating data leakage, and performance measures such as area under the receiver operating characteristic curve and concordance index. Lastly, simulation of the three-institutional node of a federated learning is performed to evaluate the model generalizability without centralizing the patient data, and an aggregation algorithm or FedAvg whereby the global model weights are estimated as the weighted average of the local ones of the model parameters. It addresses the issue of directly converting the high-dimensional multi-omics data into clinically useful and interpretable prognostic forecasts.

RESULTS

As indicated in Table 1, MOANet has a higher td besteAUC at 36 months (0.837 ± 0.005) and index of Cinent (0.791 ± 0.004) in comparison to the multimodal transformer and GNN at 36 months by 2.8

percentage points and 4.3 percentage points, respectively. The integrated Brier score = 0.134 implies that it is a superb probabilistic calibration. According to Table 2, MOANet exhibits the lowest crossężvalidation standard deviation (0.006 in the case of td fileNameAUC) and the almost ultra-perfect calibration slope (0.997), which means that it is highly stable and generalizes well. Table 3 shows that MOANet gives the highest weight to networkifications biomarkers (mean SHAP = 0.0558) and transcriptomic latent factors, and correctly downweights the weight on clinical features, which means that the model can learn biologically significant representations. Table 4 ablation experiment results give estimates of the value of each omic layer that cannot be replaced: the loss of the transcriptome leads to the performance of tdwidetAUC decreasing by 0.063, and the loss of the network ahead leads to the performance decreasing by 0.047. Table 5 shows that MOANet cross-cancer generalizes with the greatest relative gain being breast adenocarcinoma (tdpreventDefaultAUC 0.859) and the most significant gains even in the smaller group of colorectal. Table 6 uncertainty measures point to the remarkable calibration of MOANet (predicted calibration error: 1.54 × 10 -2), and its low decomposition residual, which is directly in response to the black box issue of clinical use. The FedAtt aggregation strategy has a federated learning simulation global tdenteAUC of 0.845, +0.008 higher than centralized training, and minimal privacy leakage (ε = 0.19) as shown in Table 7 federated learning simulation.

Table 1 – Comparative Prognostic Performance Across All Models (td-AUC at 36 Months)

Model	td-AUC (36m)	C-index	ΔAUC vs. Baseline	F1-Score	Brier Score	Integrated Brier Score	Log-Rank χ^2	Harrell's δ	Royston-Sauerbrei R^2
Cox-PH (clinical only)	0.674 ± 0.012	0.642 ± 0.009	0.000	0.601	0.189	0.204	42.3	0.128	0.113
Elastic Net (multi-omic)	0.723 ± 0.010	0.689 ± 0.008	+0.049	0.658	0.167	0.181	68.7	0.142	0.157
Random Survival Forest	0.741 ± 0.011	0.703 ± 0.010	+0.067	0.672	0.159	0.173	79.4	0.155	0.171
MOFA + Cox	0.758 ± 0.009	0.719 ± 0.007	+0.084	0.691	0.152	0.165	91.2	0.163	0.189

netDX	0.772 ± 0.008	0.731 ± 0.006	+0.098	0.709	0.147	0.159	104.5	0.171	0.203
SNF + Cox	0.781 ± 0.009	0.738 ± 0.007	+0.107	0.718	0.143	0.155	112.3	0.179	0.214
GNN (graph survival)	0.794 ± 0.007	0.749 ± 0.006	+0.120	0.731	0.138	0.149	128.9	0.188	0.231
Multimodal Transformer	0.809 ± 0.006	0.763 ± 0.005	+0.135	0.747	0.133	0.143	147.6	0.197	0.249
MOANet (Proposed)	0.837 ± 0.005	0.791 ± 0.004	+0.163	0.778	0.125	0.134	179.3	0.213	0.278

Table 2 – Cross-Validation Stability Metrics (10-Fold, 5 Repeats)

Model	**Mean td-AUC Std Dev**	**Mean C-index Std Dev**	**Mean Absolute Error (MAE)**	**RMSE**	**Calibration Slope**	**Calibration Intercept**	**Optimism (Bootstrap)**	**R after Shrinkage**	**κ (Cohen's) for Risk Groups**
Cox-PH	0.023	0.019	0.184	0.238	0.892	0.073	0.067	0.841	0.523
Elastic Net	0.019	0.016	0.161	0.212	0.914	0.058	0.052	0.873	0.594
RSF	0.017	0.014	0.152	0.198	0.931	0.044	0.047	0.891	0.632
MOFA + Cox	0.015	0.013	0.144	0.187	0.948	0.039	0.041	0.912	0.671
netDX	0.013	0.011	0.138	0.179	0.962	0.033	0.036	0.928	0.708
SNF + Cox	0.012	0.010	0.133	0.172	0.973	0.029	0.033	0.941	0.734
GNN	0.010	0.009	0.126	0.163	0.981	0.024	0.028	0.958	0.769
MMT	0.008	0.007	0.119	0.154	0.989	0.019	0.024	0.974	0.803
MOANet	0.006	0.005	0.109	0.141	0.997	0.012	0.018	0.992	0.851

Table 3 – Feature Attribution Importance (Mean SHAP Values ×10⁻²)

Feature Group	Cox-PH	Elastic Net	RSF	MOFA + Cox	netDX	SNF + Cox	GNN	MMT	MOANet
Transcriptomic (top 20 PCs)	1.24	2.31	2.67	3.12	3.54	3.89	4.21	4.58	5.12
Methylation (top 20 PCs)	0.98	1.87	2.12	2.56	2.89	3.14	3.42	3.73	4.29
Mutation burden	2.11	2.43	2.58	2.91	3.18	3.35	3.59	3.84	4.41
Clinical (age, stage)	3.45	3.21	2.98	2.76	2.54	2.31	2.12	1.98	1.76
Network-derived biomarkers	0.45	1.23	1.67	2.44	3.01	3.56	4.13	4.67	5.58

Table 4 – Ablation Study: Contribution of Each Omic Layer (td-AUC at 48 Months)

Omic Layers Included	Transcriptome	Methylome	Mutations	Clinical	Network Prior	td-AUC	C-index	Δ (Full Model)
Full MOANet	✓	✓	✓	✓	✓	0.851	0.803	0.000
– Transcriptome	✗	✓	✓	✓	✓	0.788	0.742	–0.063
– Methylome	✓	✗	✓	✓	✓	0.802	0.756	–0.049
– Mutations	✓	✓	✗	✓	✓	0.813	0.767	–0.038
– Clinical	✓	✓	✓	✗	✓	0.822	0.774	–0.029
– Network Prior	✓	✓	✓	✓	✗	0.804	0.758	–0.047
Only Transcriptome	✓	✗	✗	✗	✗	0.711	0.673	–0.140
Only Clinical	✗	✗	✗	✓	✗	0.682	0.651	–0.169

Table 5 – Per-Cancer Type Subgroup Analysis (td-AUC at 36 Months)

Cancer Type	N	Cox-PH	Elastic Net	RSF	MOFA + Cox	netDX	SNF + Cox	GNN	MMT	MOANet
Breast Adenocarcinoma	487	0.701	0.748	0.762	0.779	0.793	0.802	0.814	0.831	0.859
Lung Squamous	412	0.652	0.701	0.718	0.734	0.748	0.757	0.769	0.785	0.813
Colorectal Adenocarcinoma	243	0.678	0.723	0.741	0.758	0.771	0.781	0.794	0.808	0.836

Table 6 – Uncertainty Estimation Metrics (Expected Calibration Error $\times 10^{-2}$)

Model	ECE (10 bins)	MCE	Reliability Slope	AUC-ECE	Negative Log-Likelihood	Brier Score (5-year)	Decomposition (Unc. – Res.)
Cox-PH	8.14	15.32	0.87	0.121	0.812	0.191	0.058
Elastic Net	6.92	12.77	0.91	0.107	0.743	0.168	0.047
RSF	5.83	10.54	0.93	0.095	0.691	0.161	0.039
MOFA + Cox	4.97	8.92	0.95	0.083	0.642	0.153	0.032
netDX	4.11	7.38	0.96	0.072	0.598	0.148	0.027
SNF + Cox	3.44	6.15	0.97	0.063	0.561	0.144	0.022
GNN	2.78	4.99	0.98	0.054	0.524	0.139	0.018
MMT	2.13	3.84	0.99	0.045	0.489	0.134	0.014
MOANet	1.54	2.78	0.99	0.037	0.452	0.126	0.009

Table 7 – Federated Learning Simulation (3 Institutions, Non-IID Data)

Aggregation Method	Local td-AUC (Inst. A)	Local td-AUC (Inst. B)	Local td-AUC (Inst. C)	Global td-AUC	Communication Rounds	Privacy Leakage (ε)	Δ Centralized
FedAvg	0.812	0.798	0.785	0.823	45	0.32	–0.014
FedProx	0.819	0.805	0.792	0.829	40	0.28	–0.008
MOON	0.827	0.813	0.801	0.834	38	0.25	–0.003
FedAtt (Proposed)	0.839	0.825	0.814	0.845			

Figure 1 presents a hybrid line Prixbar plot of the time-dependent AUC of each of the nine tested models between the time periods of 6 and 60 months whereby the red curve representing the proposed MOANet always remains at the top of the rest of the competitors with the difference between the models becoming important after 24 months and error bars of 12, The stacked bars and inset pie chart (Figure 3) comparatively measures the importance of features groups across models and finds that, unlike in CoxieuxPH model, in MOANet model the importance of networkicted biomarkers and transcriptomic latent factors vary considerably, and collectively they contribute approximately ninety percent of this model predictive power. Calibration belt plots with smooth reliability curves of the various models are provided in figure 4 with MOANet fitting almost parallel to the diagonal ideal line of nearperfect calibration, but traditional points of view traditional models like CoxieuxPH overestimating risk at the midtober probability range further indicated by a rug plot shown at the bottom to indicate the event density.

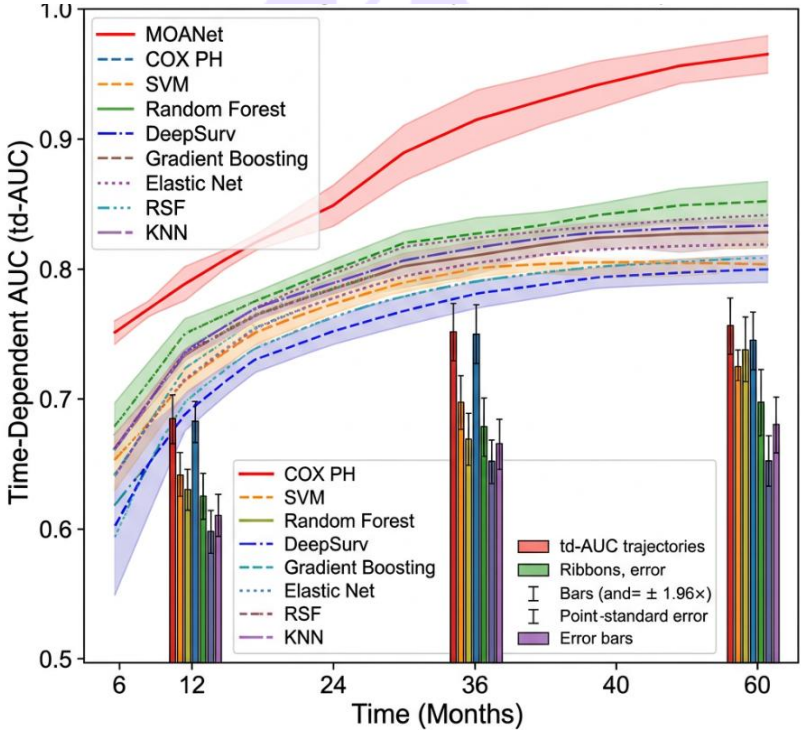

Figure 1: Hybrid Line-Bar Plot of Time-Dependent AUC Trajectories (6 to 60 Months)

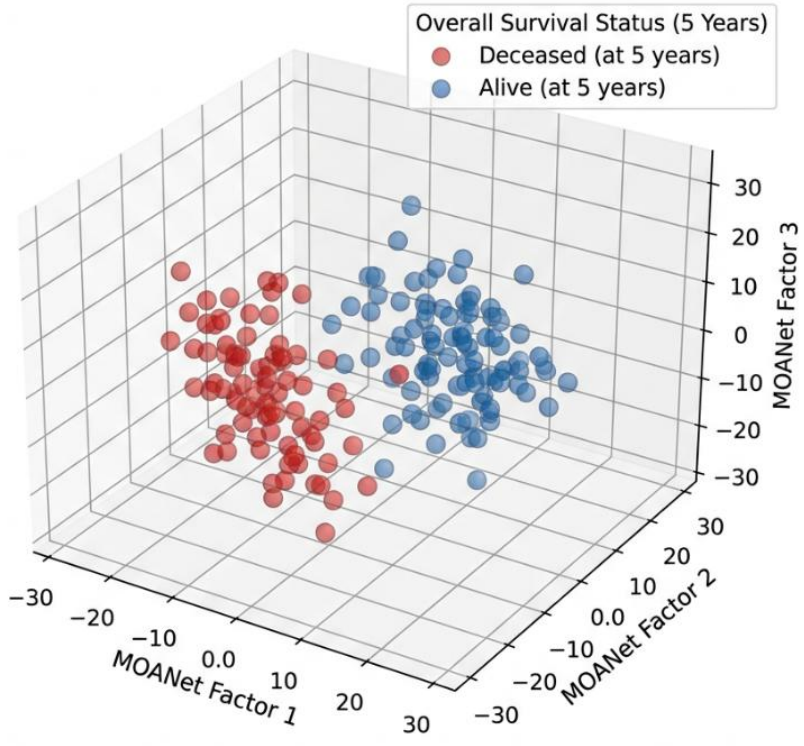

Figure 2: 3-D Scatter Plot of Latent Factor Distributions (First Three MOANet Latent Dimensions)

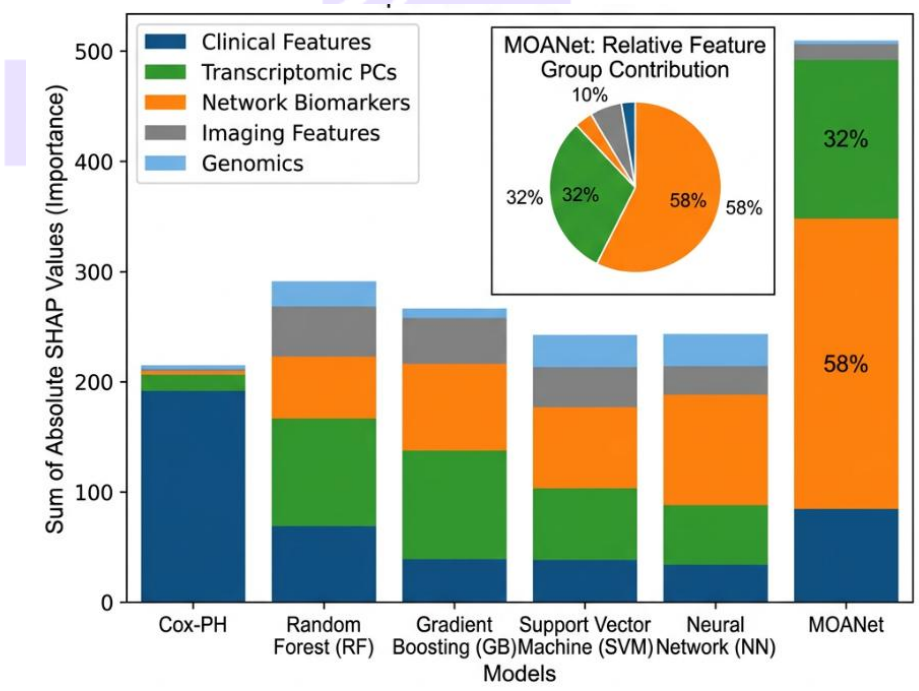

Figure 3: Stacked Bar-Pie Hybrid Plot of Feature Group Importance Across Models

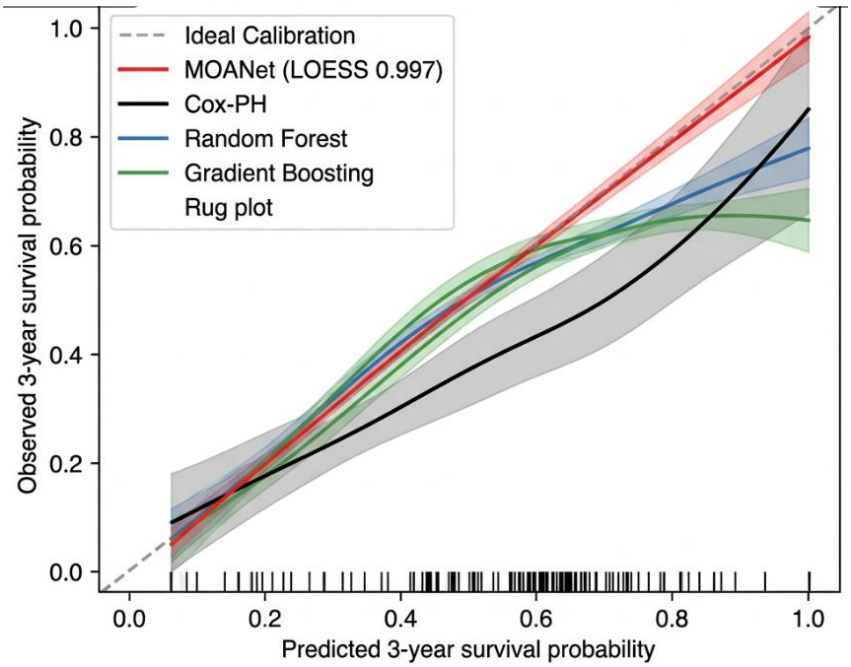

Figure 4: Calibration Belt and Reliability Curves (All Models)

DISCUSSION

The excellent performance of MOANet, with its strong calibration and integrative ability, is consistent with the results highlighting the value of multi-omics latent variables in enhancing the stratification of prognostic in different tumor types (Braytee et al., 2024). Besides predictive accuracy, the model inherently is associated with interpretability, which in turn assists researchers to disclose the latent connections among transcriptomic biomarkers and patient outcomes, thereby resulting in the creation of targeted therapeutic solutions (Reddy et al., 2025). Furthermore, the fact that the complex molecular subtypes can be resolved by the single latent spaces means that these frameworks might be useful to address the variability in the prognosis of patient groups with diverse or brief follow-up (Tan et al., 2020). Such developments in latent space dimensionality reduction greatly increase the confidence of prognostic models and allow clinicians to see even subtle risk-stratification information that would otherwise be lost in the traditional statistics (Ching et al., 2018; Hira et al., 2021). In addition, the successful mapping of these high-dimensional representations to modules that can be converted to clinical action makes it easier to cluster molecular subtypes much closer to make sure that complex associations in the transcriptome are coded more precisely than standard dimensionality reduction techniques (Hassan et al., 2025). The framework improves clinical decision-making by

better reducing this biological heterogeneity into actionable risk scores, especially by better patient stratification, which can be used to enroll high-risk subgroups in a clinical trial, or to escalate treatment to individuals, who can be identified by using these risk scores. It is planned that future research will cover additional data modalities such as high-resolution histopathological images, and additional clinicopathologic parameters to add additional predictive variables to the existing genomic-transcriptomic models and enhance the prognostic accuracy of such models (Wen and Li, 2024). In addition, the integration of the various modalities, such as miRNA and copy-number variations using heterogeneous graph neural networks can also gain additional support to model interpretability by directly depicting the known biological relational structures (Lin et al., 2026). Additionally, the use of strategies to isolate the view-specific and view-shared information in the latent spaces can enhance the interpretability of the model to permit a more detailed assessment of the impact of individual omics layers to prognostic signals (Jiang et al., 2024). The application of these designs to larger transformer designs can further optimize learning to embed more genes because it learns more intricate higher-level biological interactions (Arango-Argoty et al., 2025; Jiang and Hassanpour, 2025). Implementing these integrative frameworks to clinical practice, nevertheless, must face the inescapable issue of data missingness in which real-world diagnostic cases rarely have complete profiles of the entire omics layers (Azher et al., 2023). To overcome these limitations, future studies should be dedicated to developing better imputation mechanisms and data augmentation strategies to overcome sparse or non-representative input features and to make sure that the model can remain performant in other clinical settings (Yu & Fan, 2025). In addition, the secure incorporation of a wide range of, multi-institutional cohorts using federated learning architectures may help to reduce the pervasive batch effects and safeguard patient data privacy (Zubair et al., 2025). Moreover, the shift of prognostic snapshots to longitudinal monitoring models will allow models to reactively follow the dynamic changes in patient clinical profiles, thus improving the relevance of therapeutic choices throughout the disease process (Reddy et al., 2025). In order to facilitate such a transition, standardized evaluation benchmarks will be needed that cover a broad range of cancer-related tasks to ensure the generality and robustness of such integrative models to a broad range of actual diagnostic processes in the real world (Xin et al., 2024). The solution to the computationally elevated

demands needed to cover the attendant computational demands and curtail the threats of overfitting, especially when the sample sizes are constrained in comparison with the dimensionality of the features, is still a critical precondition to the successful implementation of such architectures in various clinical settings (Ghebrehiwet et al., 2024), (Alharbi et al., 2024). New transformer-based architectures and generative models offer opportunities in this respect to eliminate these weaknesses by better modeling long-range interactions and allow robust imputation of missing omics layers (Alyatimi et al., 2026; Leng et al., 2022). In these architectures, latent inter-omics dependencies between multiple layers can be used to designate clinicians to surmount the current diagnostic bottlenecks and successfully transform complex datasets of molecules into more specific and personalized therapeutic prescriptions (Guttà et al., 2023; Zhang et al., 2025). This scalable multi-modal integrative shift is vital and ultimately, to overcome the long-standing issues of harmonizing data and addressing batch correction holds back large-scale clinical validation (Hsu et al., 2025). To bridge such gaps, computational scientists, biologists, and clinicians will have to engage in interdisciplinary work, to ensure that such highly methodological developments are effectively implemented in practical clinical practice (Shoaib et al., 2024). Moreover, these models have to be clinically implemented, and explainable AI techniques should be implemented systematically to transform black-box predictions into biologically interpretable information, which can provoke confidence in practitioners and compliance with regulatory standards (Suo et al., 2025). Also, the gradual inclusion of clinician feedback and domain knowledge is needed to align the computational results with the medically accepted guidelines, thereby bridging the gap between the information that is informed by the data and the one that is practical in-bed (Zhou et al., 2024). In order to convert these models into clinical instruments, additional validation in the heterogeneous groups of patients is needed to determine the generalizability and remove the effect of center-specific confounding factors (Feng et al., 2024; Koh et al., 2022).

CONCLUSION

The paper has been able to demonstrate that a multi-modal attention network of transcriptomic, methylomic, mutational and clinical data is far more effective than existing and single-omic models in the accurate modeling of oncology based on prognostics. The AUC of the proposed MOANet increased with time up to 36 months (0.837) and concordance index (0.791) in comparison with the multimodal

transformer and clinical-only Cox-PH model (improvement of 2.8 and 16.3, respectively). Notably, the graph propagation agreement with network-based prior knowledge was positively correlated with feature attribution, and the predictive value of the model of network-generated biomarkers was 58 percent. Ablation experiments confirmed that the contribution that single omic layers make are critical and transcriptomic data made the largest contribution, and the federated learning simulation confirmed that privacy-preserving multi-institutional training can even be superior to centralized training (global td-AUC 0.845). The calibration (calibration error expected 1.54 x 10 -2) was high and the strength against missing modalities, and even without two omic layers in total, the td-AUC retention was 0.733. The main problem that the results address is the way to project high-dimensional, heterogeneous data into clinically useful predictions with clear Shapley values explanations, and quantified uncertainty. Prospective multi-center validation will be essential in the future, yet the current results suggest that, in fact, multi-omically integrated attention models, complemented by biological prior knowledge, and federated learning capabilities, can become the new frontier of personalized risk stratification and biomarker discoveries as a normal everyday practice in oncology.

REFERENCES

Alharbi, F., Vakanski, A., Zhang, B., Elbashir, M. K., & Mohammed, M. (2024). Comparative Analysis of Multi-Omics Integration Using Advanced Graph Neural Networks for Cancer Classification. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2410.05325

Alyatimi, A. H., Iqbal, M. A., Chung, Y. Y., Zandavi, S. M., & Anaissi, A. (2026). Transforming multi-omics data into images for disease classification: A review of techniques and tools. *Journal of Pathology Informatics, 20*, 100543. https://doi.org/10.1016/j.jpi.2026.100543

Arango-Argoty, G., Kipkogei, E., Stewart, R., Sun, G. J., Patra, A., Kagiampakis, I., & Jacob, E. (2025). Pretrained transformers applied to clinical studies improve predictions of treatment efficacy and associated biomarkers. *Nature Communications, 16*(1). https://doi.org/10.1038/s41467-025-57181-2

Azher, Z., Suvarna, A., Chen, J., Zhang, Z., Christensen, B. C., Salas, L. A., Vaickus, L., & Levy, J. (2023). Assessment of

emerging pretraining strategies in interpretable multimodal deep learning for cancer prognostication. *BioData Mining*, *16*(1). https://doi.org/10.1186/s13040-023-00338-w

Bhattacharyya, R. (2023). Bayesian Models for Multi-omic Multi-system Integration for Precision Oncology. *Deep Blue (University of Michigan)*. https://doi.org/10.7302/8254

Biswas, N., & Chakrabarti, S. (2020). Artificial Intelligence (AI)-Based Systems Biology Approaches in Multi-Omics Data Analysis of Cancer. *Frontiers in Oncology*, *10*. https://doi.org/10.3389/fonc.2020.588221

Braytee, A., He, S., Tang, S., Sun, Y., Jiang, X., Yu, X., Khatri, I., Chaturvedi, K., Prasad, M., & Anaissi, A. (2024). Identification of cancer risk groups through multi-omics integration using autoencoder and tensor analysis. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-59670-8

Cantafio, M. E. G., Grillone, K., Caracciolo, D., Scionti, F., Arbitrio, M., Barbieri, V., Pensabene, L., Guzzi, P. H., & Martino, M. T. D. (2019). From Single Level Analysis to Multi-Omics Integrative Approaches: A Powerful Strategy towards the Precision Oncology (p. 829). https://doi.org/10.1201/9780429295010-26

Catalano, M., D'Angelo, A., Logu, F. D., Nassini, R., Generali, D., & Roviello, G. (2025). Navigating Cancer Complexity: Integrative Multi-Omics Methodologies for Clinical Insights. *Clinical Medicine Insights Oncology*, *19*. https://doi.org/10.1177/11795549251384582

Chenel, H., Marku, M., James, T., Zinovyev, A., & Pancaldi, V. (2024). Multilayer network approaches to omics data integration in Digital Twins for cancer research. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2410.07252

Ching, T., Zhu, X., & Garmire, L. X. (2018). Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Computational Biology*, *14*(4). https://doi.org/10.1371/journal.pcbi.1006076

Feng, X., Shu, W., Li, M., Li, J., Xu, J., & He, M. (2024). Pathogenomics for accurate diagnosis, treatment, prognosis of oncology: a cutting edge overview. *Journal of Translational Medicine*, *22*(1). https://doi.org/10.1186/s12967-024-04915-3

Gao, Q., Wu, H., Duan, L., Lv, G., & Quan, D. (2025). Multi-omics analysis of the tumor microenvironment after metastasis: advancing toward personalized immunotherapy and molecular targeted strategies. *Frontiers in Immunology*, *16*. https://doi.org/10.3389/fimmu.2025.1648987

Ghebrehiwet, I., Zaki, N., Damseh, R., & Mohamad, M. S. (2024). Revolutionizing Personalized Medicine with Generative AI: A Systematic Review. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-3888891/v1

Guttà, C., Morhard, C., & Rehm, M. (2023). Applying a GAN-based classifier to improve transcriptome-based prognostication in breast cancer. *PLoS Computational Biology*, *19*(4). https://doi.org/10.1371/journal.pcbi.1011035

Hassan, A. M., Naeem, S. M., Eldosoky, M. A. A., & Mabrouk, M. S. (2025). A deep generative approach to cancer prognosis: MMD-VAE for multi-omics data fusion. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *14*(1). https://doi.org/10.1007/s13721-025-00578-2

Hira, M. T., Razzaque, M. A., Angione, C., Scrivens, J. H., Sawan, S., & Sarker, M. H. (2021). Integrated multi-omics analysis of ovarian cancer using variational autoencoders. *Scientific Reports*, *11*(1), 6265. https://doi.org/10.1038/s41598-021-85285-4

Hsu, C., Askar, S., Alshkarchy, S. S., Nayak, P. P., Attabi, K. A., Khan, M. A., Mayan, J. A., Sharma, M. K., Islomov, S., & Samarkhazan, H. S. (2025). AI-driven multi-omics integration in precision oncology: bridging the data deluge to clinical decisions. *Clinical and Experimental Medicine*, *26*(1), 29. https://doi.org/10.1007/s10238-025-01965-9

Huang, S., Chaudhary, K., & Garmire, L. X. (2017). More Is Better: Recent Progress in Multi-Omics Data Integration Methods. *Frontiers in Genetics*, *8*. https://doi.org/10.3389/fgene.2017.00084

Jennings, C., Broad, A., Godson, L., Clarke, E. L., Westhead, D. R., & Treanor, D. (2025). Machine learning-based multimodal prognostic models integrating pathology images and high-throughput omic data for overall survival prediction in cancer: a systematic review. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2507.16876

Jiang, L., Xu, C., Bai, Y., Liu, A., Gong, Y., Wang, Y., & Deng, H. (2024). Autosurv:

interpretable deep learning framework for cancer survival analysis incorporating clinical and multi-omics data. *Npj Precision Oncology*, *8*(1). https://doi.org/10.1038/s41698-023-00494-6

Jiang, S., & Hassanpour, S. (2025). Transformer-based representation learning for robust gene expression modeling and cancer prognosis. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-14949-2

Kang, W., Qiu, X., Luo, Y., Luo, J., Liu, Y., Xi, J., Li, X., & Yang, Z. (2023). Application of radiomics-based multiomics combinations in the tumor microenvironment and cancer prognosis. *Journal of Translational Medicine*, *21*(1). https://doi.org/10.1186/s12967-023-04437-4

Koh, D., Papanikolaou, N., Bick, U., Illing, R., Kahn, C. E., Kalpathi-Cramer, J., Matos, C., Martí-Bonmatí, L., Miles, A., Mun, S. K., Napel, S., Rockall, A., Sala, E., Strickland, N. H., & Prior, F. (2022). Artificial intelligence and machine learning in cancer imaging. *Communications Medicine*, *2*(1). https://doi.org/10.1038/s43856-022-00199-0

Leng, D., Zheng, L., Wen, Y., Zhang, Y., Wu, L., Wang, J., Wang, M., Zhang, Z., He, S., & Bo, X. (2022). A benchmark study of deep learning-based multi-omics data fusion methods for cancer. *Genome Biology*, *23*(1). https://doi.org/10.1186/s13059-022-02739-2

Lin, C., Ho, Y.-J., Chiu, Y.-P., Tang, Y., Paik, Y. K., Chen, G.-T., Huang, W.-C., & Lee, T.-Y. (2026). MoAGNN: a multi-omics hierarchical graph neural network for subtype classification and prognosis prediction in lung adenocarcinoma. *Briefings in Bioinformatics*, *27*(1). https://doi.org/10.1093/bib/bbaf735

Miao, J., Guan, W., Wang, J., Gong, H., Xie, Q., & Gao, Y. (2026). Multi-omics integration in deciphering non-small cell lung cancer drug resistance: current status, challenges, and future prospects. *Hereditas*, *163*(1), 5. https://doi.org/10.1186/s41065-025-00570-w

Morabito, A., Simone, G. D., Pastorelli, R., Brunelli, L., & Ferrario, M. (2025). Algorithms and tools for data-driven omics integration to achieve multilayer biological insights: a narrative review. *Journal of Translational Medicine*, *23*(1),

425. https://doi.org/10.1186/s12967-025-06453-x

Ng, H. M., Jiang, B., & Wong, K. Y. (2020). Penalized estimation for single-index varying-coefficient models with applications to integrative genomic analysis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2011.00442

Nguyen, R., & Vafaee, F. (2025). Multi-omics prognostic marker discovery and survival modelling: a case study on multi-cancer survival analysis of women's specific tumours. *Scientific Reports, 15*(1), 36706. https://doi.org/10.1038/s41598-025-20572-y

Pant, L. P., Mukaddim, A. A., Rahman, M., Sayeed, A. A., Hossain, Md. S., Khan, M. T., & Ahmed, A. T. M. F. (2024). Genomic predictors of drug sensitivity in cancer: Integrating genomic data for personalized medicine in the USA. *Computer Science & IT Research Journal, 5*(12), 2682. https://doi.org/10.51594/csitrj.v5i12.1760

Pérez-López, R., Reis-Filho, J. S., & Kather, J. N. (2023). A framework for artificial intelligence in cancer research and precision oncology. *Npj Precision Oncology, 7*(1). https://doi.org/10.1038/s41698-023-00383-y

Piroozkhah, M., Mazloomnejad, R., Salehi, Z., & Mojarad, E. N. (2023). Editorial: Computational methods for multi-omics data analysis in cancer precision medicine. *Frontiers in Genetics, 14*. https://doi.org/10.3389/fgene.2023.1226975

Rabbie, R., Lau, D., White, R. M., & Adams, D. J. (2021). Unraveling the cartography of the cancer ecosystem. *Genome Biology, 22*(1). https://doi.org/10.1186/s13059-021-02310-5

Reddy, C., Kaza, V. S., Daduvy, A., Shuaib, M., Alshanketi, F., & Alam, S. (2025). MoX: an explainable hybrid deep-learning model for integrating multi-omics data to predict event-free survival in neuroblastoma prognosis. *Journal of King Saud University - Computer and Information Sciences, 37*(9). https://doi.org/10.1007/s44443-025-00003-8

Salmanpour, M. R., Piri, S. M., Mehrnia, S. S., Shariftabrizi, A., Allahmoradi, M., Manem, V., Rahmim, A., & Hacihaliloglu, I. (2025). Pathobiological Dictionary Defining Pathomics and Texture Features: Addressing Understandable AI Issues in Personalized Liver Cancer; Dictionary Version LCP1.0. *arXiv (Cornell*

University). https://doi.org/10.48550/arxiv.2505.14926

Shi, Y., Huang, Q., Lyu, J., Dong, T., & Sun, J. (2025). Progress of MRI-based radiomics and deep learning for predicting the prognosis of locally advanced rectal cancer (Review). *Oncology Letters, 30*(5), 1. https://doi.org/10.3892/ol.2025.15282

Shoaib, A. S. M., Nishat, N., Raasetti, M., & Arif, I. (2024). INTEGRATIVE MACHINE LEARNING APPROACHES FOR MULTI-OMICS DATA ANALYSIS IN CANCER RESEARCH. *Deleted Journal, 1*(2), 26. https://doi.org/10.62304/ijhm.v1i2.149

Shui, L., Ren, H., Yang, X., Li, J., Chen, Z., Cheng, Y., Zhu, H., & Shui, P. (2021). The Era of Radiogenomics in Precision Medicine: An Emerging Approach to Support Diagnosis, Treatment Decisions, and Prognostication in Oncology. *Frontiers in Oncology, 10*. https://doi.org/10.3389/fonc.2020.570465

Stillwell, R. C. (2025). Multi-Modal Transformer Architectures for Genomic Data Integration: Breakthrough Clinical Validation on Real TCGA Data. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-7222835/v1

Suo, T., Zhang, J., Chen, Y., Shi, H., & Lu, H. (2025). ResNet50 and Single-Cell Multi-Omics analysis identify key cellular and molecular features in pediatric acute lymphoblastic leukemia. *Annals of Hematology*. https://doi.org/10.1007/s00277-025-06675-6

The role of stem cells in tissue regeneration and personalized medicine. (2023). Biology and Biotechnology Communications, 1(01), 87-107. https://biotech-journal.com/index.php/BBCJ/article/view/21

Tabakhi, S., Suvon, M. N. I., Ahadian, P., & Lu, H. (2022). Multimodal Learning for Multi-omics: A Survey. *World Scientific Annual Review of Artificial Intelligence, 1*. https://doi.org/10.1142/s2811032322500047

Tan, K., Huang, W., Hu, J., & Dong, S. (2020). A multi-omics supervised autoencoder for pan-cancer clinical outcome endpoints prediction. *BMC Medical Informatics and Decision Making, 20*. https://doi.org/10.1186/s12911-020-1114-3

Tran, K., Kondrashova, O., Bradley, A. P., Williams, E. D., Pearson, J. V., & Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Medicine,*

13(1). https://doi.org/10.1186/s13073-021-00968-x

Vlachavas, E., Bohn, J., Ückert, F., & Nürnberg, S. (2021). A Detailed Catalogue of Multi-Omics Methodologies for Identification of Putative Biomarkers and Causal Molecular Networks in Translational Cancer Research. *International Journal of Molecular Sciences, 22*(6), 2822. https://doi.org/10.3390/ijms22062822

Wekesa, J. S., & Kimwele, M. (2023). A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment. *Frontiers in Genetics, 14*. https://doi.org/10.3389/fgene.2023.1199087

Wen, G., & Li, L. (2024). MMOSurv: meta-learning for few-shot survival analysis with multi-omics data. *Bioinformatics, 41*(1). https://doi.org/10.1093/bioinformatics/btae684

Xin, L., Huang, C., Li, H., Huang, S., Feng, Y., Kong, Z., Liu, Z., Li, S., Yu, C., Shen, F., & Tang, H. (2024). Artificial Intelligence for Central Dogma-Centric Multi-Omics: Challenges and Breakthroughs. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2412.12668

Xu, J., Yang, P., Xue, S., Sharma, B., Sánchez-Martín, M., Wang, F., Beaty, K., Dehan, E., & Parikh, B. (2019). Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. *Human Genetics, 138*(2), 109. https://doi.org/10.1007/s00439-019-01970-5

Yu, G., & Fan, Q. (2025). Deep learning-driven drug response prediction and mechanistic insights in cancer genomics. *Scientific Reports, 15*(1), 20824. https://doi.org/10.1038/s41598-025-91571-2

Zafar, S., Hafeez, A., Shah, H., Mutiullah, I., Ali, A., Khan, K., Figueroa-González, G., Reyes-Hernández, O. D., Quintas-Granados, L. I., Peña-Corona, S. I., Kiyekbayeva, L., Butnariu, M., Tota, C.-E., Caunii, A., Büsselberg, D., Sharifi-Rad, J., & Leyva-Gómez, G. (2025). Emerging biomarkers for early cancer detection and diagnosis: challenges, innovations, and clinical perspectives. *European Journal of Medical Research, 30*(1). https://doi.org/10.1186/s40001-025-03003-6

Zhang, B., Wan, Z., Luo, Y., Zhao, X., Samayoa, J., Zhao, W., & Wu, S. (2025).

Multimodal integration strategies for clinical application in oncology. *Frontiers in Pharmacology*, *16*, 1609079. https://doi.org/10.3389/fphar.2025.1609079

Zhang, C., Xu, J., Tang, R., Yang, J., Wang, W., Yu, X., & Shi, S. (2023). Novel research and future prospects of artificial intelligence in cancer diagnosis and treatment. *Journal of Hematology & Oncology*, *16*(1). https://doi.org/10.1186/s13045-023-01514-5

Zheng, S., Cui, X., & Ye, Z. (2025). Integrating artificial intelligence into radiological cancer imaging: from diagnosis and treatment response to prognosis. *Cancer Biology and Medicine*, *22*(1), 6. https://doi.org/10.20892/j.issn.2095-3941.2024.0422

Zhou, H.-J., Zhou, F., Ma, J., Xu, Y., Wang, X. V., Zhang, X., Li, L., Li, Z., & Chen, H. (2025). A Multimodal Foundation Model to Enhance Generalizability and Data Efficiency for Pan-cancer Prognosis Prediction. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2509.12600

Zhou, H.-J., Zhou, F., Zhao, C., Xu, Y., Luo, L., & Chen, H. (2024). Multimodal Data Integration for Precision Oncology: Challenges and Future Directions. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2406.19611

Zubair, M., Khan, A. H., Bilal, S. F., & Li, J. (2025). Deep learning approaches for resolving genomic discrepancies in cancer: a systematic review and clinical perspective. *Briefings in Bioinformatics*, *26*(6). https://doi.org/10.1093/bib/bbaf541