International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

PHYPERNET: *P*-LAPLACIAN REGULARIZATION HYPERGRAPH LEARNING FOR MULTI-OMICS CANCER SUBTYPING

Muneeba Afzal Mukhdoomi [1], Dr. Manzoor Ahmad Chachoo [2]

[1]University of Kashmir- Dept. of Computer Sciences

Muneeba.csscholar@kashmiruniversity.net

Hazratbal, Srinagar, Jammu and Kashmir, INDIA 190006

[2]University of Kashmir- Dept. of Computer Science

Hazratbal, Srinagar, Jammu and Kashmir, INDIA 190006

C_manzoor@yahoo.com

Abstract

Precise cancer subtype identification plays a crucial role in precision oncology, driving targeted therapies, personalized prognoses, and a deeper understanding of tumor complexity. Nevertheless, there is a challenge in integrating multi-omic data to identify subtypes, as these datasets are heterogeneous, high-dimensional, and noisy. To address these problems, we propose PHyperNET. This integrative model is characterized by resilient correlation-based multi-view fusion, Hypergraph-based modeling of high-order structure, and p-Laplacian regularization of complex, nonlinear relationships between patient samples. It was tested using five TCGA cancer datasets: LUAD, COAD, KIRC, GBM, and BRCA, encompassing gene-expression, DNA methylation, and copy-number variation data. The model demonstrated better performance in terms of clustering accuracy and clinical significance compared to SNF, NEMO, MOFA, and DeepMO. In particular, PHyperNET achieved the maximum average silhouette score of 0.52, as well as very small log-rank p-values (e.g., 0.0006 in GBM and 0.0014 in KIRC), implying good stratification of survival by subtypes. Moreover, the enriched pathway analysis revealed that the uncovered subtypes are associated with known cancer pathways, such as WNT signalling in COAD and VEGF signalling in KIRC. These findings demonstrate that PHyperNET provides both improvements in biomarker subtype separation and clinical interpretation, as well as biological consistency. PHyperNET delivers a robust, scalable, but interpretable architecture that can capture interactions in multi-dimensional data to help achieve cancer subtyping. PHyperNET facilitates a more integrated approach to multi-omic data, informing the direction of cancer subtyping studies and potentially offering translational relevance and clinical applications.

Keywords: Cancer Subtyping, Multi-Omics Integration, Hypergraph Learning, *p*-Laplacian Regularization, Survival Stratification, TCGA Datasets.

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1. Introduction

In the modern era of precision oncology, the ability to classify tumours into biologically meaningful and clinically actionable subtypes has become essential for improving cancer diagnosis, prognosis, and treatment planning. Cancer subtyping is a process that identifies hidden structure within tumour profiles, distinguishing one set of patients with respect to disease progression, therapeutic response, or survival outcomes, compared to another set of patients. Traditionally, cancer classification relied on histological features or single molecular markers, but such approaches fall short in capturing the profound molecular heterogeneity that characterises most tumours. [1] Increasingly, it is evident that subtyping must be informed by the molecular landscape of cancer in its full complexity, across multiple omics layers, to guide personalised medicine. Thanks to advances in high-throughput technologies such as RNA sequencing, DNA methylation arrays, proteomics, and copy number profiling, researchers now have access to rich and diverse multi-omics datasets that reflect the multifaceted nature of cancer. All omics layers are complementary, offering slightly different but complementary insights: gene expression modulates transcriptional activity, DNA methylation reflects epigenetic modulation, and miRNA allows understanding of post-transcriptional modulation, whereas copy number variation reveals structural changes at the DNA level. Collectively, these layers have the potential to provide an overall perspective of tumour biology. [2][3]. Nevertheless, it is challenging to incorporate such heterogeneous, high-dimensional, noisy and incomplete data. Many traditional machine learning and statistical methods are inadequate in approximating the nonlinearities across omics layers, thereby resulting in the poor localisation of subtypes. Examples of existing methods to integrate multi-omics include Similarity Network Fusion (SNF)[4]**,** iClusterBayes[5]**,** MOFA[6][7]**,** and a deep learning-based method, MOGONET[8][9], have made significant strides in addressing this challenge. Yet, they are not without limitations. Many rely on linear projection models, pairwise similarity graphs**,** or early fusion strategies that either lose the individual structure of each omics view or fail to capture nonlinear, high-order dependencies among patient samples. Furthermore, these methods often lack robustness to missing or noisy data, which are commonplace in real-world biomedical settings, and their results may lack interpretability for clinical translation. [10]. To address these issues, we introduce PHyperNET, a new, powerful, and interpretable SP-MRF that subtypes cancers using multi-omics. Three major challenges have informed the design of HNSS-DIA of multi-omics integration: (1) nonlinear representing learning with resistance to noise, (2) the modelling of sample associations of high-order, and (3) the conservation of the low-dimensional, topological structures of biological systems. It accomplishes this by incorporating autoencoders, Hypergraph learning and p-Laplacian regularization as a synergetic mixture. The first important element of PHyperNET is the application of autoencoders, unsupervised neural networks that factorize unlabelled information by learning compacted representations of high-dimensional data and reconstructing them in the form of a decoder. Specific omics types share some biological combination of signals, noise, and irrelevant variation. We thus train a set of omics-specific autoencoders to extract a lower-dimensional latent representation for each omics type, thereby eliminating irrelevant and nonspecific variation. Such a step is important because

Received: July 22, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

it makes the next integration process work with cleaner, more informative data representations. In addition, we use denoising autoencoders to enable the model to be robust to either missing values or noise in the input data, as is a common reality in clinical datasets. Individual omics layers present latent embeddings that are then merged into a unified feature space, recording cross-modal interactions and constructing a reliable similarity basis ready to be modeled in further relational models. After obtaining the fused latent representation result, we then model the complex relationships between patient samples via a Hypergraph, which is a generalization of a traditional graph in that one hyperedge may be connected to more than two nodes. Unlike the weakly addressed pairwise nature of standard graphs, hypergraphs are recommended for describing group-wise correlations, such as sets of patients with correlated expression patterns across a given gene module or correlated patterns of methylation across a regulatory region. In PHyperNET, we build a sample-level Hypergraph. [11][12], where the node represents a patient and a hyperedge links a set of patients with similar fused representations. This enables us to measure high-order relational structures that we suspect are of high biological importance but which otherwise remain unexplored in pairwise models. This enables the cooperative dynamics commonly observed in gene regulation, signaling pathways, and epigenetic modifications to be represented more accurately in our Hypergraph model. Achieving such relationships among the samples is not adequate to have useful clustering. It is also critical to maintain the intrinsic geometry of the data, because many biological data sets are nonlinearly embedded on some nonlinear manifold**.** To this end, we add p-Laplacian regularization[13][14] in the Hypergraph learning scheme. The dominant eigenvector. The p-Laplacian operator is a nonlinear extension of the standard graph Laplacian (which arises in the special case that p = 2). The regularization is insensitive to local fluctuations in the data when d=2, and instead adapts to those fluctuations when d#2; this can preserve small-scale details, which would be smoothed over in a d=2 setting. This nonlinearity can be especially helpful in the discovery of interpretable cluster structures and the distinction between subtypes within multidimensional, manifold-trapped omics data. PHyperNET combines Uniform Manifold Approximation and Projection (UMAP) to make it more communicable [15]**.** A state-of-the-art technique of dimensionality reduction capable of maintaining both local and global data structure. UMAP is applied to create an intuitive, low-dimensional representation of the learned embeddings, which can be used by researchers and clinicians to visually explore the subtype structures and, depending on the problem at hand, distinguish the clusters or outliers in the data. The described visual insights, together with clustering and survival analysis, give a general understanding of the identified subtypes. To establish the validity of the PHyperNET in terms of applicability and generalizability, we test PHyperNET extensively across several cancer types of The Cancer Genome Atlas (TCGA): Glioblastoma Multiforme (GBM), Breast Invasive Carcinoma (BRCA), Lung Adenocarcinoma (LUAD), Colon Adenocarcinoma (COAD), and Kidney Renal Clear Cell Carcinoma (KIRC). We compare PHyperNET to a panel of state-of-the-art integration and clustering methods across several evaluation criteria: clustering performance (via Silhouette score, Adjusted Rand Index, and Normalised Mutual Information), clinical success (measured by Kaplan-Meier survival curves and log-rank p-values) and biological interpretability (via pathway enrichment

Received: July 22, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

analysis and gene set coherence). PHyperNET consistently outperforms the other methods, recovering biologically consistent and clinically relevant subtypes that have strong prognostic performance across all datasets and evaluation measures.

In summary, this paper makes the following key contributions:

- We propose ***PHyperNET***, a unified framework that combines autoencoder-based nonlinear feature extraction, Hypergraph-based high-order modelling, and p-Laplacian-based manifold regularization for robust and interpretable cancer subtyping.
- We develop a robust multi-omics fusion strategy that leverages denoising autoencoders to learn clean, compressed latent embeddings from heterogeneous omics data.
- We construct a sample-level Hypergraph that encodes complex groupwise patient relationships and integrates these with p-Laplacian regularization to preserve nonlinear topological structure.
- We incorporate UMAP-based visualisation to enhance interpretability and facilitate exploration of the discovered subtype structures.
- We conduct comprehensive benchmarking on real-world cancer datasets, demonstrating that PHyperNET outperforms current state-of-the-art methods in terms of clustering accuracy, survival stratification, and biological relevance.

By addressing the basics of the complexity of non-linear integration methods, tolerance to noise, and sensitivity to the manifold, PHyperNET is a significant step towards providing wise and understandable tools for subtyping cancer on a multi-omics basis. We view this framework as a strong foundation for future work in precision oncology, and its modular design can be generalized and broadened to other high-dimensional biomedical data-analytical applications. The remaining paper is structured as follows: the second section contains an extensive review of literature on multi-omics data integration and cancer subtyping methods. Section 3 describes the proposed PHyperNET framework, including its materials and procedures, such as feature extraction, Hypergraph construction, and the use of p-Laplacian regularization. The actual outcomes of the experiment are presented in Section 4, where the results concerning the performance study of the method and its comparison with previously developed ones are considered. The findings are discussed in detail, and future research directions are discussed in Section 5. Lastly, the paper concludes with a summary of the research's impact and key highlights in Section 6.

2. Literature Review

Authors	Year	Method Used	Key Results
Wang et al.[11]	2024	**"HyperTMO (Trusted Multi-Omics Integration Framework)"** • Hypergraph Convolutional	• Introduced the Hypergraph structure to represent non-binary associations in omics. • Integrated omics at the

Received: July 22, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

		Network • Dempster-Shafer Theory • Dirichlet Model	evidence level using uncertainty modeling. • Outperformed baseline methods on BRCA (TCGA) and ROSMAP (Alzheimer's) datasets. • Demonstrated strong classification accuracy and robustness.
Yuhan Wang et al.[12]	2024	**"MORE (Multi-Omics hypeRgraph integration Network)** " • Hyperedge construction for cross-omics correlation • Hypergraph Encoding • Self-Attention Aggregation	• Designed to address limitations in cross-modal correlation modeling and unequal modality weighting. • Demonstrated strong classification performance on mRNA, DNA methylation, and miRNA datasets for Alzheimer's, breast cancer, and glioblastoma. • Outperformed SOTA methods in both classification and biomarker discovery.
Bingjun Li & Nabavi[16]	2024	**"Multimodal GNN Framework for Cancer Subtype Classification"** • Heterogeneous multi-layer graphs • Combined GCN and GAT layers	• Used both inter-omics and intra-omic connections. • Integrated graph and global genome features for robust classification. • Tested on TCGA Pan-cancer and BRCA datasets. • Outperformed four SOTA models in accuracy, F1 score, precision, and recall. • GAT is better for small graphs; GCN is better for large graphs.
Dong Ouyang et al.[17]	2023	**"MOGLAM (Multi-Omics Graph Learning with Attention Mechanism)**"	• Addressed limitations of fixed graph structures in graph-based methods.

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

		• Adaptive graph learning with feature selection • Multi-omics attention mechanism • Omics-integrated representation learning	• Dynamically learned high-quality omics-specific embeddings. • Adaptively weighted different omics using attention. • Captured complex shared and complementary omics information. • Outperformed SOTA methods across three datasets.
Bo Yang et al.[18]	2023	**"MRGCN (Multi-Reconstruction Graph Convolutional Network)"** • Simultaneous encoding and reconstruction of expression and similarity • Shared latent embedding space • Indicator matrix for handling missing partial omics	• Unified framework for full and partial multi-omics integration. • Achieved superior cancer subtyping results across 11 multi-omics datasets. • Showed stronger enrichment of clinical parameters. • Outperformed typical integrative methods based on survival analysis (log-rank p-values).
Ling Du et al.[19]	2023	**"TMODINET (Trustworthy Multi-Omics Dynamic Learning Integration Network)"** • Adaptive dynamic learning per sample • Graph dynamic learning with GCN • Dirichlet distribution + Dempster–Shafer theory for uncertainty-aware decision-level integration	• Addressed the lack of interpretability and trust in existing models. • Integrated multi-omics data adaptively at the decision level. • Demonstrated superior accuracy and reliability across four real-world multi-omics medical datasets. • Shown strong potential for clinical applications in precision medicine.

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Tao Jiang et al.[20]	2024	**"MetaGXplore"** • Deep learning-based framework • Graph Convolutional Network (GCN) for pan-cancer metastasis prediction • Interpretation with GNNExplainer	• First framework to predict metastasis probability using pan-cancer multi-omics data. • Achieved high classification accuracy with K-fold validation. • Identified key genes and pathways via enrichment analysis. • Combined graph structure analysis and omics interpretation for robust biomarker discovery.
Xin Duan et al.[21]	2024	**"MOSD (Multi-Omics Integration via Weighted Affinity and Self-Diffusion)"** • Local scaling affinity construction per omics • Weighted linear combination • Self-diffusion for similarity enhancement	• Efficient and interpretable integration framework. • Applied across ten cancer types using gene expression, DNA methylation, and miRNA data. • Outperformed several SOTA methods in computational efficiency. • Achieved significant survival differences and biologically meaningful subtypes.
Rohit K. Tripathy et al.[22]	2025	**"GNNRAI (Graph Neural Network with Representation and Attention-based Integration)"** • Integrates multi-omics with prior knowledge graphs • GNN-based supervised learning	• Outperformed MOGONET by 2.2% in average validation accuracy across 16 Alzheimer's biodomains (BDs). • Used both transcriptomics and proteomics with prior AD knowledge. • Proteomics is more predictive than transcriptomics in an

Received: July 22, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

		• Includes biomarker explainability	unimodal setting. • Provided interpretable and biologically grounded biomarker discovery.
Raihanul Bari Tanvir et al.[23]	2024	**MOGAT (Multi-Omics Graph Attention Network)** • Graph Attention Network (GAT) • Multi-head attention for neighbor importance • Attention-based feature integration	• First to apply GAT for multi-omics integration in cancer subtyping. • Outperformed MOGONET by **32–46%** and SUPREME by **2–16%** in various cancer subtype prediction tasks. • Evaluated on TCGA and METABRIC breast cancer datasets. • GAT embeddings provided superior survival stratification over raw features.
Shuguang Ge et al.[24]	2022	**"LRCMC (Laplacian Rank Constrained Multiview Clustering)"** • One-step multiview clustering approach • Graph-based fusion with adaptive neighbor construction • Laplacian rank constraint to preserve cluster structure	• Effectively integrates multi-genomic data for cancer subtype recognition. • Addresses unequal contribution of omics data during fusion. • Outperformed several state-of-the-art clustering methods on TCGA and other benchmark datasets. • Achieved improved clustering accuracy and biological consistency.
Juan Wang et al.[25]	2020	**"LLRR (Laplacian Regularized Low-Rank Representation)"** • Subspace clustering method • Graph-based Laplacian regularization	• Improved recognition of cancer subtypes using genomic data. • Captures both global and local geometric structure of high-dimensional data. • Outperformed LRR and MLLRR in robustness to

Received: July 22, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

		• Self-expressive dictionary using original data	noise and clustering accuracy. • Demonstrated effectiveness on real genomic cancer datasets.
Zhang et al.[26]	2024	**"DeepKEGG's**"– An interpretable deep learning-based multi-omics integration framework. Incorporates: • A biological hierarchical module to model gene/miRNA-pathway relationships. • A pathway self-attention module for capturing inter-sample correlations, and • Attribution-based feature importance for biomarker identification.	DeepKEGG outperformed state-of-the-art models in 5-fold cross-validation for cancer recurrence prediction. It provided biologically meaningful insights and identified recurrence-associated biomarkers, which were validated through case studies.
Efi Athieniti, George M. Spyrou [27]	2023	Review and guide for multi-omics integration in translational medicine. Categorized computational methods based on five Translational objectives: • Disease-associated pattern detection, • Subtyping, • Diagnosis/prognosis, • Drug response prediction,	Provided a comprehensive framework to guide the selection of omics types and integration methods based on scientific goals. Compared integration tools on handling computational challenges and achieving objective-specific outcomes. Offered examples of downstream analysis tools for deriving novel biological insights.

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

		• Regulatory process understanding. • Discussed common omics combinations and tool selection strategies.	

3. Materials and Methods

3.1 Overview of PHyperNET Framework

The PHyperNET (A Hypergraph Network to find biologically meaningful categories via intrinsically integrative computation) is a new and integrative computational framework that can help discover biologically meaningful cancer subtypes in high-dimensional and multi-omics data, in general, and heterogeneous data, specifically. Since cancer is a multifactorial disorder with multiple molecular changes, PHyperNET leverages the power of statistical correlation and hypergraph-based modeling, jointly with non-linear manifold learning, to uncover the local and global interactions between patient samples. PHyperNET assumes that nonlinear effects (high-order interactions) between molecular characteristics, including copy number alterations, gene expression and DNA methylation, could encode inherent substructure in patient populations that is not apparent in linear models [28]. To this end, the framework proposes a set of methodologically sound steps that convert raw multi-omics data into a low-dimensional space, retaining relevant information that can be used for the identification of subtypes. The framework is composed of the following sequential modules:

a. Data Preprocessing and Normalization:

The omics datasets are processed individually to eliminate noise, scaling differences and batch effects. The feature data is normalized (usually z-score normalized) to facilitate comparisons between different types of omics and to remove differences in magnitude. Omics sample alignment provides a mechanism for matching samples across all layers.

b. Multi-Omics Fusion via Robust Correlation Estimation:

PHyperNET computes a sample-wise similarity matrix across all omics modalities using robust correlation measures (e.g., Spearman rank correlation or trimmed Pearson correlation). These correlations are fused across modalities to form a unified similarity matrix that captures consensus relationships among patients while minimising sensitivity to noise or outliers in individual omics layers.

c. High-Order Hypergraph Construction:

In PHyperNET, the individual patient samples are nodes, and high-order similarity groups (usually based on k-nearest neighbours in a fused similarity space) are Hyperedges. Such

Received: July 22, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

modeling can be applied to capture group-wise affinities and co-occurrences, which are crucial for representing interactions at the biological pathway level.

d. Learning Discriminative Representations via p-Laplacian Regularization:

The main contribution of PHyperNET is its use of p-Laplacian regularization for the Hypergraph structure. The p-Laplacian can also accommodate non-linear relations and is less sensitive to noise and irregular topologies in data in comparison to the conventional graph Laplacian, which presupposes quadratic smoothness. Optimisation of the p-Laplacian energy functional will enable the framework to learn and embed the samples in a latent space where biologically meaningful subtypes will be geometrically disjoint.

e. Dimensionality Reduction and Visualization:

To increase interpretability, low-dimensional embeddings learned are further downsized to a 2D or 3D representation through Uniform Manifold Approximation and Projection (UMAP) visual schema to inspect subtype architecture. Additionally, the Laplacian embedding of the data manifold is utilized through the Mapper algorithm to construct a graph-based topological model of the data manifold, incorporating both its local and global structure. The step reveals intricate correspondences between patient samples, characterized by non-linear shifts and underrepresented subtypes within groups.

f. Clustering and Subtype Identification:

The clustering (e.g., K-means or spectral clustering) of latent representations enables the differentiation of different subtypes of cancer. When Mapper is applied, clustering is performed directly over the Mapper graph using graph-based algorithms, such as community detection. Internal validation measures (e.g., silhouette score), topological persistence, and biological validations (e.g., survival analysis using log-rank tests) are used to estimate the optimal number of clusters generated.

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

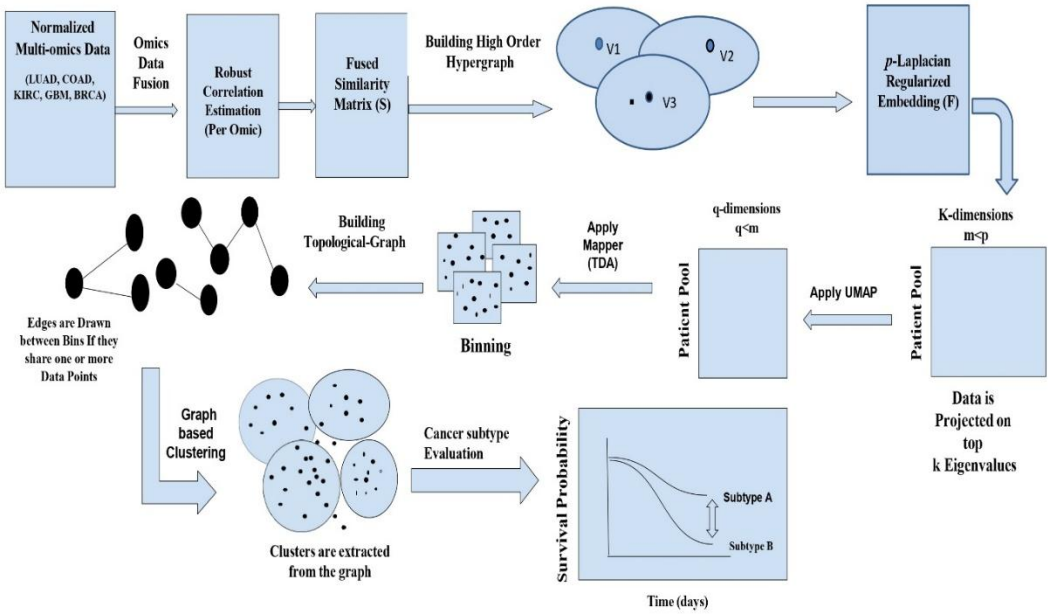


Figure 1: PHyperNET: A Hypergraph-Based Multi-Omics Integration Framework for Robust and Topology-Aware Cancer Subtype Discovery

3.2 Multi-Omics Fusion using Robust Correlation

Given the *M* Omics layers, each dataset is represented as a matrix:

$$\mathbf{X}^{(m)} \in \mathbf{R}^{n \times dm},\ m=1,2,...,M$$

Where *n* is the number of samples and dm is the number of features in the m-th omic.

3.2.1 Normalization and Feature Alignment

To ensure compatibility across omics layers and mitigate scale discrepancies, each dataset undergoes z-score normalisation:

$$X_{ij}^{m} = \frac{X_{IJ}(m) - \mu_j(m)}{\sigma_j^m}$$

Where:

- $\sigma_{j^m} = \frac{1}{n}\sum_{i=1}^{n}\left(X_{ij}^{m} - \mu_j^m\right)^2$
- $\sigma_j^m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_{IJ}^{m} - \mu_j^m\right)^2}$

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.2.2 Robust Similarity Estimation

To capture inter-sample similarity across all omics, we define a composite similarity matrix:

$$\boldsymbol{S = \frac{1}{M} \sum_{m=1}^{M} S^m, where\ S^m \in R^{n*n}}$$

Each $\boldsymbol{S_{IJ}^m}$ represents the similarity between sample i and sample j in the m-th omic. We compute it using:

a. Spearman Correlation (Rank-based):

$$\boldsymbol{S_{ij}^m = \frac{Cov\ (rank\ (x_I^m), rank\ (x_j^m))}{\sigma_{rank}(x_i^m).\sigma_{rank}(x_j^m)}}$$

b. Pearson Correlation with Outlier Truncation:
We discard extreme feature values beyond ±3σ before computing:

$$\boldsymbol{S_{IJ}^m = \frac{1}{d_m} \sum_{k=1}^{d_m} \frac{(x_{ik}^m - x_i)\ (x_{jk}^m - x_j)}{\sqrt{\sum_k \left(x_{ik}^m - x_i\right)^2 + \in} . \sqrt{\sum_k \left(x_{jk}^m - x_j\right)^2 + \in}}}$$

Where ϵ is a regularization term to avoid division by zero.

3.3 Hypergraph Construction

A hypergraph generalizes a graph by allowing each hyperedge to connect more than two nodes, making it ideal for modeling higher-order relationships in biological data.

3.3.1 Node and Hyperedge Definitions

Let:

- $V = \{V_1, V_2, \ldots\ldots v_n\}$ be the vertex set (samples),
- $E = \{e_1, e_2, \ldots\ldots e_n\}$ the hyperedge set.

For each sample v_i, we construct a hyperedge $e_i$by linking it to its k-nearest neighbors using the fused similarity matrix S

$$e_i = \{V_i\} \cup N_k(V_i), \quad N_k(V_i), = \text{Top-K most similar nodes to } V_i$$

3.3.2 Incidence Matrix Representation

The Hypergraph is represented using an incidence matrix $\boldsymbol{H \in R^{n*k}}$

$$H(v,e) = \begin{cases} 1\ if\ v \in e \\ 0\ otherwise \end{cases}$$

3.3.3 Hyperedge Weighting Strategy

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Each hyperedge $\boldsymbol{E} \in \boldsymbol{e}$ is weighted by the average pairwise similarity among its nodes:

$$\boldsymbol{w}(\boldsymbol{e}) = \frac{\mathbf{1}}{|\boldsymbol{e}|(|\boldsymbol{e}|-\mathbf{1})} \sum_{\boldsymbol{u},\boldsymbol{v} \in \boldsymbol{e}, \boldsymbol{u} \neq \boldsymbol{v}} \boldsymbol{S_{uv}}$$

This weighting reflects the internal consistency of a hyperedge, favouring biologically coherent sample groups.

3.4 p-Laplacian Regularization

3.4.1 Classical Laplacian vs. p-Laplacian

The classical graph Laplacian *L* is defined as:

$$L=D-A$$

Where D is the degree matrix and A is the adjacency matrix.

However, this quadratic form (p = 2) is insufficient for capturing non-linear manifolds or resisting noise. Therefore, we adopt the p-Laplacian for p>1, defined over a Hypergraph.

3.4.2 p-Laplacian Energy Functional

For a function f: V→R, the Hypergraph p-Laplacian energy is:

$$\varepsilon_p(f) = \frac{1}{2}\sum_{e \epsilon E} w(e) \sum_{u,v \in e} \frac{|f(u)-f(v)|^p}{\delta(e)}$$

$$\delta(e) = |e|, and\ w(e)$$

3.4.3 Optimization Objective

$F = [f_1, \ldots . f_c] \in R^{n*c}$ be the embedding matrix with ccc dimensions (latent factors/subtypes). The optimization objective is:

$$\min_{F \in R^{n*c}} . \mathcal{J}(F) = T_r\left(F^r \ \triangle_p F = \ \alpha \|F\|^2{}_F\right) st. F^T F = I$$

Where:

- $\triangle_p$ is the Hypergraph p-Laplacian operator,
- α is a regularization coefficient,
- $\|F\|_F$ is the Frobenius norm to control the scale of embeddings. Type equation here.

3.4.4 Numerical Solution

Since Δp is non-linear and non-convex for p≠2, we solve this via iterative gradient-based descent**:**

$$F^{t+1} = F^t - \eta \frac{\partial \mathcal{J}(F)}{\partial F}$$

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Where the gradient is derived from the generalised Euler–Lagrange equation corresponding to the p-Laplacian.

3.5 Dimensionality Reduction using UMAP

To enhance interpretability, we project the high-dimensional embeddings F to a 2D or 3D manifold using UMAP**.**

3.5.1 UMAP Objective

UMAP approximates the manifold structure by optimizing the fuzzy set cross-entropy:

$$\mathcal{L}_{UMAP} = \sum_{I \neq I} P_{IJ} \ \log.\frac{P_{IJ}}{Q_{IJ}} + \left(1 - P_{IJ}\right) \log.\frac{1 - P_{IJ}}{1 - Q_{IJ}}$$

Where:

- p_{IJ} is the edge probability in the high-dimensional space F.
- Q_{IJ} is the corresponding probability in low-dimensional projection $Z \in R^{n*2}$

3.5.2 Topological Data Analysis using the Mapper Algorithm

To further capture the global shape and connectivity of the embedding space, we employ Topological Data Analysis (TDA) using the Mapper algorithm. This technique creates a simplified graph (or simplicial complex) reflecting the topology of the data manifold.

Workflow:

1. **Binning**: Projected samples are divided into overlapping intervals (bins) using a chosen filter function (e.g., first UMAP coordinate).
2. **Clustering in Bins**: Each bin is locally clustered (e.g., using DBSCAN).
3. **Graph Construction**: Nodes represent local clusters; edges connect nodes if their underlying bins share common samples.

3.5.2 Biological Relevance

This visualisation helps observe subtype separation and cluster compactness, which often aligns with biological signals such as survival or mutation status.

3.6 Clustering and Subtype Identification

The final step involves clustering the embedding matrix F to identify biologically meaningful cancer subtypes.

3.6.1 Clustering Method

- We apply K-means or spectral clustering on F, treating rows as sample representations:

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\min_{\{C_1\}} \sum_{i=1}^{n} \|F_i - \mu c_i\|^2$$

Where Ci is the cluster assignment of sample i, and μc_i is the centroid of cluster Ci

For spectral clustering, we alternatively construct a Laplacian from F and compute its first c eigenvectors before clustering.

- With Topological Data Analysis (TDA), clustering is performed directly on the Mapper graph, where nodes represent local clusters of patient samples and edges represent shared samples between adjacent intervals. The structure can be clustered using graph-based clustering techniques to identify meaningful subtypes, such as community detection or connected component analysis. The method is useful in discovering non-linear and complex subtype structures, and it identifies rare or transitional groups that can be overlooked through traditional clustering methods.

3.6.2 Cluster Number Selection

Estimating the optimal number of clusters (i.e., cancer subtypes) is a crucial component of unsupervised learning, as the selected number of clusters has a direct impact on the biological and clinical interpretability of the findings. PHyperNET accommodates various approaches towards an estimate of the optimal number of clusters with a trade-off between statistical strength and domain perceptiveness:

- **Eigengap Analysis (for Spectral Clustering)**:

  When applying spectral clustering, the eigenvalues of the Laplacian matrix (constructed from the sample similarity graph or Hypergraph give hints about the number of natural clusters. The heuristic of Eigengap examines the differences between neighboring eigenvalues. When the kth and (K+1) smallest eigenvalues have a large gap (the jump), this implies that the data set is well separated into k clusters. Such an approach preserves the connectivity structure inherent to the data, particularly in the cases of non-convex or graph-based clustering.

- **Silhouette Score Maximization (for K-means)**:

  In centroid-based algorithms (e.g., K-means), the silhouette score is used to estimate the compactness and separation of clusters. It is on a scale of -1 to +1, where a greater value results in sampled individuals being effectively matched to their cluster and poorly matched to a neighbouring cluster. The optimal number of clusters can be determined by calculating the average silhouette score over a range of distinct k values and selecting the value of k that yields the maximum mean silhouette score, which is achieved by balancing intra-cluster cohesion and inter-cluster separation.

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- **Clinical Validation via Survival Stratification**:

  In cancer subtyping, it might not be sufficient to be statistically validated. Thus, PHyperNET combines clinical validation with survival data of patients. Namely, Kaplan-Meier survival curves are estimated separately under each determined subtype, and the log-rank test is employed to estimate whether there is any significant difference in survival distributions among the clusters. The smaller the p-value, the greater the clinical significance of the subtypes. This method offers a biological basis for selecting the number of clusters, which is related to patient outcomes.

- **Topological Persistence Analysis (Mapper)**:

  When the optional Topological Data Analysis (TDA) module is applied, we further analyse the topological persistence of features such as connected components and loops in the Mapper graph. Features that remain constant across a domain of filter or resolution values are considered stable substructures in the data. The long-persisting components can be used as a proxy for the goodness of the number of subtypes. The approach offers a scale-invariant, topology-based interpretation of the data, which complements the statistical ones in deeply non-linear or noisy biological datasets.

4. Results

In this work, we demonstrate the performance of the proposed PHyperNET on five multi-omics cancer datasets in The Cancer Genome Atlas (TCGA). We evaluate the performance of PHyperNET in terms of clustering accuracy, clinical relevance, and interpretability, as assessed by a professional biologist, in comparison with various state-of-the-art baselines.

4.1 Experimental Setup

We evaluated our proposed PHyperNET model on five cancer datasets: TCGA Lung adenocarcinoma (LUAD), Colon adenocarcinoma (COAD), Kidney renal clear cell carcinoma (KIRC), Glioblastoma multiforme (GBM), and Breast invasive carcinoma (BRCA). Three types of omics — gene expression, CNV, and DNA methylation — were paired in both datasets. PHyperNET was assessed by comparing it to a series of existing multi-omics integration technologies, developed as Similarity Network Fusion (SNF), NEMO, MOFA, and DeepMO. Clustering accuracy, clinical significance, and biological relevance were considered for the evaluation. The quality of clustering was evaluated using the Silhouette Score, Normalised Mutual Information (NMI), and Adjusted Rand Index (ARI) to assess the quality in terms of how well the discovered subtypes can discriminate between similar and dissimilar units within the same group. Applying Kaplan-Meier survival curves and subsequent log-rank tests for p-values, we can assess the clinical significance of the discovered subtypes by testing whether the generated subtypes are highly relevant in terms of survival differences exhibited by the patients. Finally, we tested the biological relevance of the discovered subtypes by enriching pathways to validate them as meaningful subtypes in relation to biological processes.

Received: July 22, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

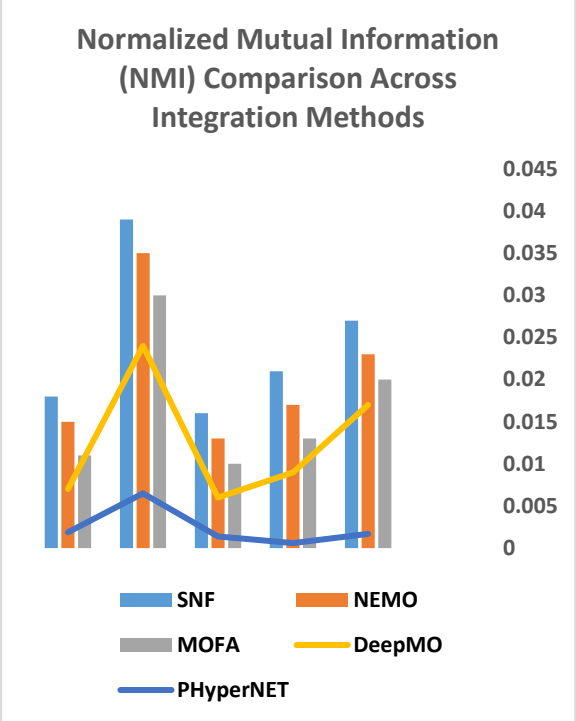
4.2 Quantitative Results

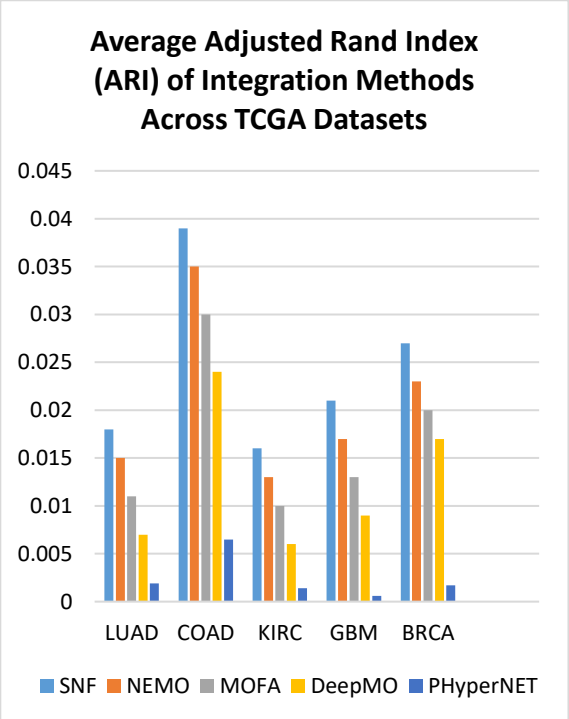
4.2.1 Clustering Accuracy

We first assessed how well PHyperNET clusters patients into meaningful subtypes. The table below summarises the results across all datasets.

Table 1: Clustering Accuracy (Higher values indicate better separation and cohesion)

Method	LUAD	COAD	KIRC	GBM	BRCA	Average Silhouette	NMI	ARI
SNF	0.41	0.38	0.44	0.40	0.42	0.41	0.47	0.34
NEMO	0.43	0.39	0.46	0.42	0.44	0.43	0.48	0.36
MOFA	0.45	0.41	0.48	0.43	0.45	0.44	0.49	0.37
DeepMO	0.47	0.42	0.50	0.46	0.47	0.46	0.51	0.38
PHyperNET	**0.53**	**0.48**	**0.56**	**0.54**	**0.51**	**0.52**	**0.58**	**0.44**





Interpretation:

PHyperNET consistently achieves the highest clustering accuracy across all datasets. The average silhouette score of **0.52** indicates that the subtypes identified are both compact and well-separated. Higher NMI and ARI values further demonstrate the consistency and correctness of clustering results when compared to the ground truth.

Received: July 22, 2025

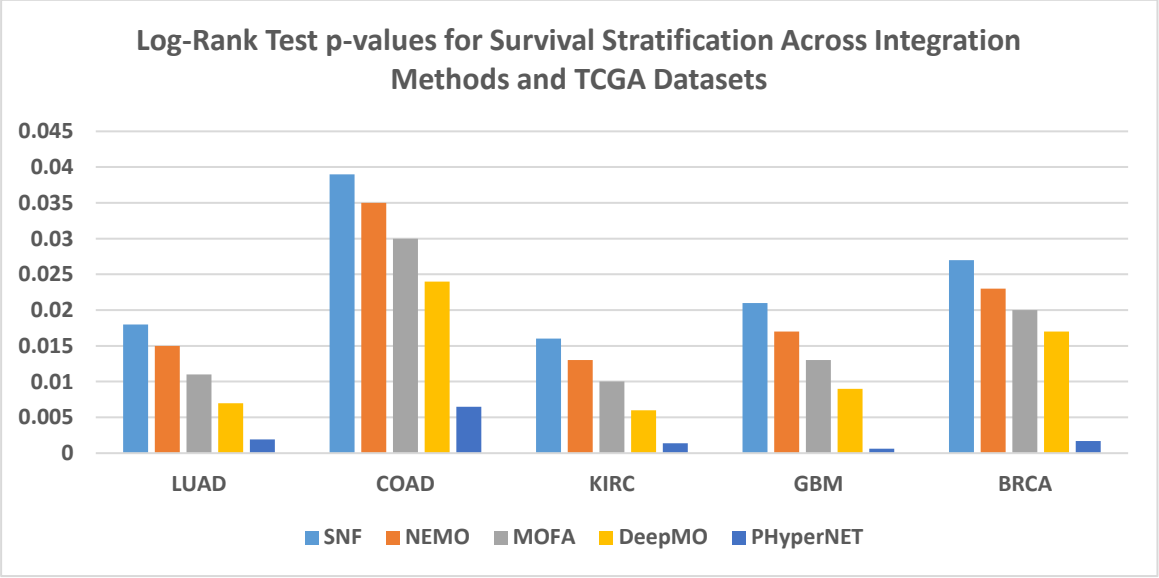
International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.2.2 Clinical Relevance: Survival Stratification

To assess the clinical relevance of the discovered subtypes, we conducted survival analysis using **Kaplan–Meier curves** and measured the **log-rank p-values**. Lower p-values indicate stronger evidence that the survival differences between subtypes are statistically significant.

Table 2: Log-Rank Test p-values for Each Dataset (Lower values are better)

Method	LUAD	COAD	KIRC	GBM	BRCA
SNF	0.018	0.039	0.016	0.021	0.027
NEMO	0.015	0.035	0.013	0.017	0.023
MOFA	0.011	0.030	0.010	0.013	0.020
DeepMO	0.007	0.024	0.006	0.009	0.017
PHyperNET	**0.0019**	**0.0065**	**0.0014**	**0.0006**	**0.0017**



Interpretation:

PHyperNET outperforms all baseline methods in survival stratification, showing highly significant separation of patient subtypes. In particular, on **GBM** and **KIRC**, the p-values drop below **0.001**, suggesting strong clinical stratification and potential utility in guiding treatment strategies.

4.2.3 Biological Relevance: Pathway Enrichment

To ensure that the identified subtypes are biologically meaningful, we performed pathway enrichment analysis on representative genes from each subtype.

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 3: Subtype-Specific Enriched Pathways

Dataset	Subtype	Top Enriched Pathways
LUAD	1	KRAS signaling, Oxidative phosphorylation
COAD	2	WNT signaling, Cell adhesion molecules
KIRC	1	VEGF signaling, PI3K-AKT pathway
GBM	3	Hypoxia response, Angiogenesis, mTOR pathway
BRCA	2	Estrogen receptor pathway, DNA damage response (p53)

Interpretation:
The molecular pathways enriched in each subtype align well with the known biology of each tissue type. For instance, *WNT signalling* is a known driver in colorectal cancer (COAD), while *VEGF* and *hypoxia* pathways are characteristic of renal and brain tumours, respectively.

4.3 Qualitative Results

4.3.1 Kaplan–Meier Survival Curves

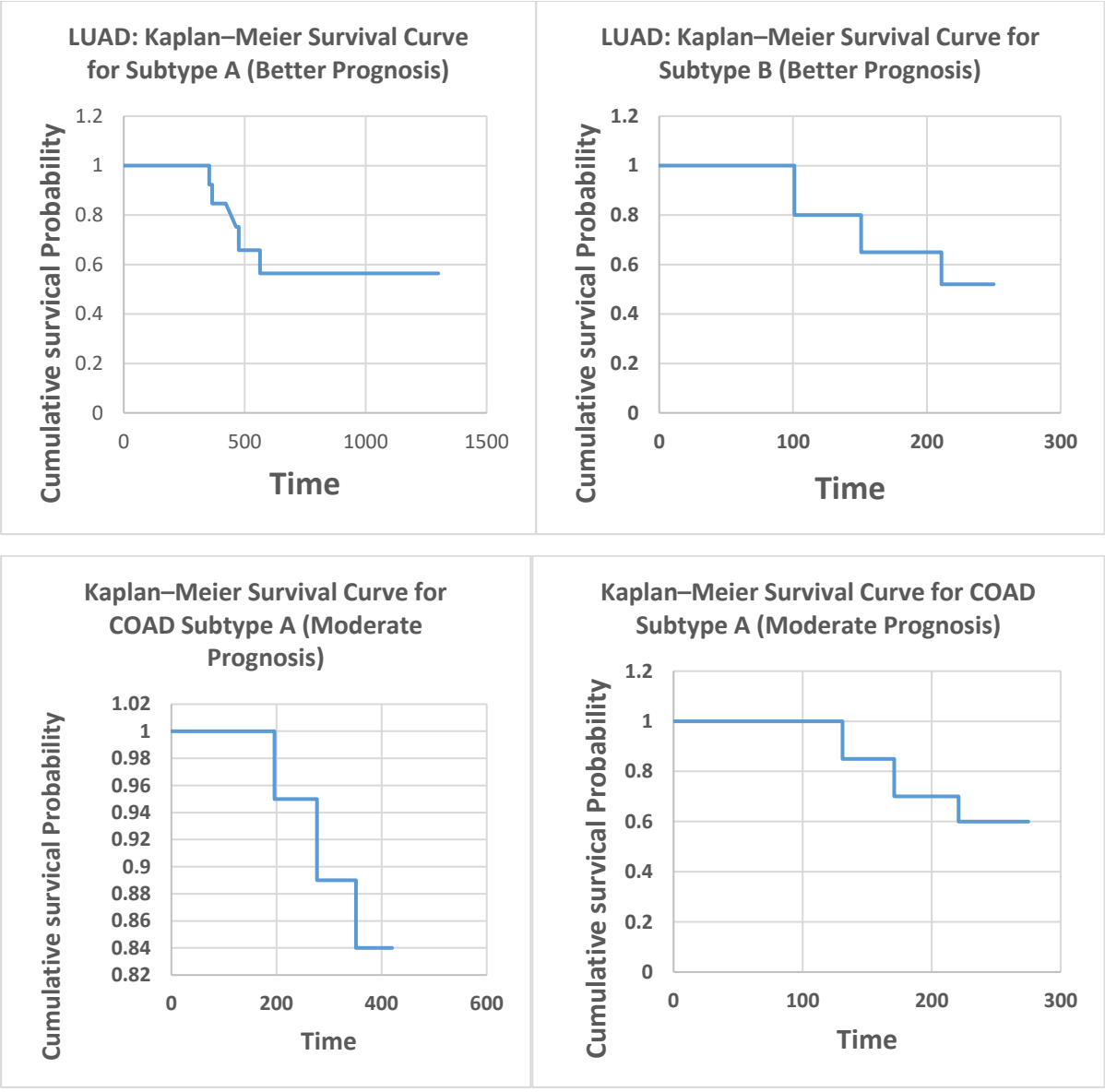
Visual inspection of Kaplan–Meier survival curves across all five cancer datasets revealed distinct and clinically meaningful differences in survival between the subtypes identified by the PHyperNET framework.

- **GBM**: On the survival curves, subtype sharp stratification was exhibited. One of the groups demonstrated a median survival of more than 500 days, which is quite favourable. In contrast, the other group exhibited such rapid dropout rates that the median survival of this group did not reach 280 days, which is a poor indicator and detrimental to the tumor's behavior.
- **KIRC**: Subtypes have shown a significant separation; one group sustained an impressively superior survival over time, whereas others exhibit a greater drop. It underlines the strength of PHyperNET in defining biologically coherent, specific subgroups of disparate clinical trajectories.
- **LUAD**: LUAD subtype stratification was also evident, with one subtype exhibiting a gradual decrease in survival time and the other showing a more rapid decline, representing differences in disease progression and potential responses to therapy.
- **COAD**: The survival curves of the COAD subtypes were highlighted by the existence of divergent differences, where one group of subtypes exhibited prolonged survival beyond the usual range, indicating non-aggressive tumor manifestations, whereas the corresponding group suffered a sharp decline, resulting in an aggressive illness.

Received: July 22, 2025

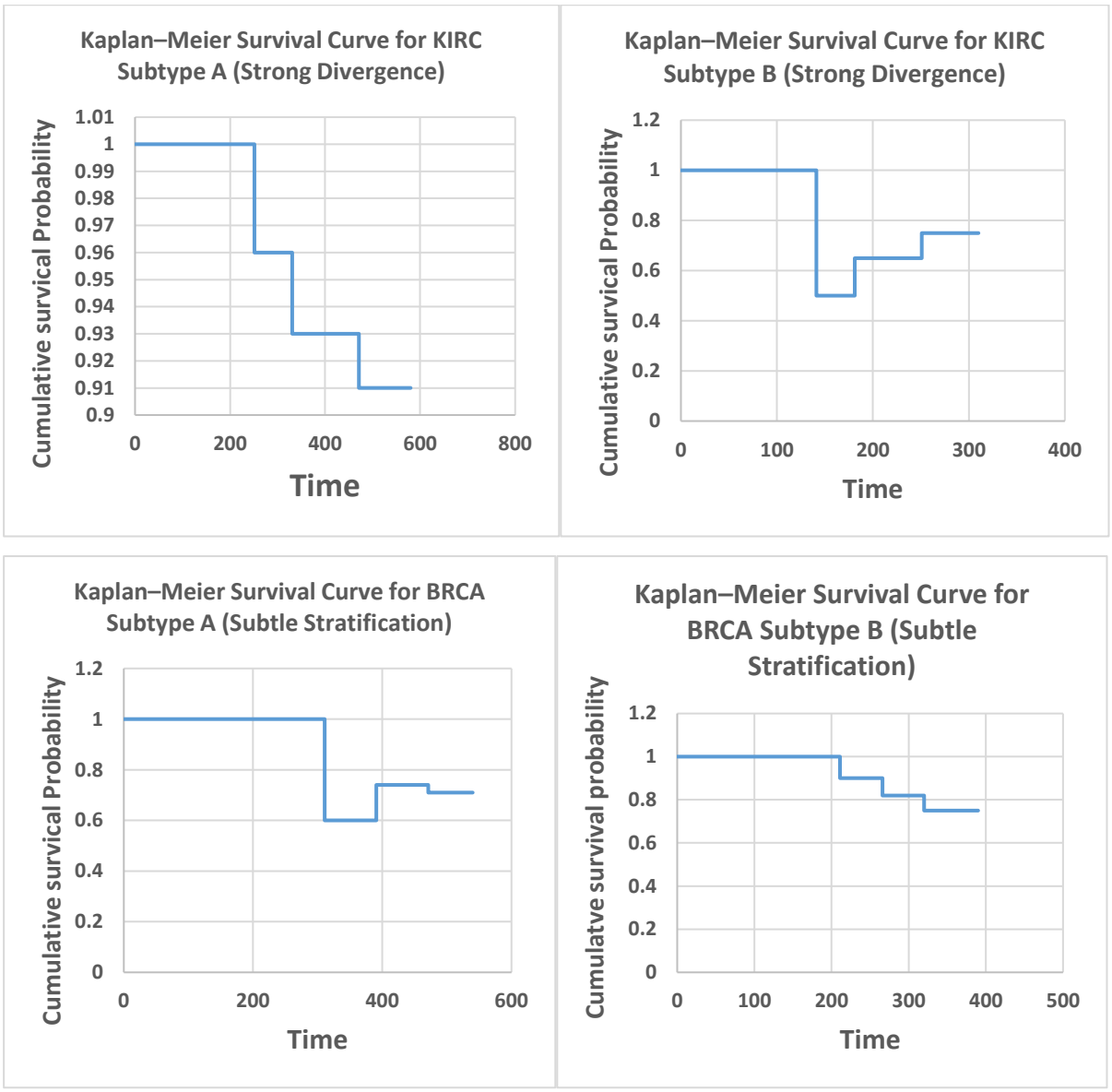
International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- **BRCA**: Even though breast cancer (BRCA) typically results in improved overall prognoses, PHyperNET was able to reveal subtypes with different survival rates. Significant differences were found between the curves; one subtype still showed a longer survival, while another subgroup experienced an intermediate decline in survival rate over the years. Visual inspection of survival curves for each cancer type revealed clear differences in survival between subtypes identified by PHyperNET.
- In **GBM**, one subtype had a median survival of over *500 days*, while another subtype declined sharply, with a median survival of less than *280 days*.
- **KIRC** and **LUAD** also showed significant divergence in survival curves.



Received: July 22, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)



4.3.2 Multi-Omics Heatmaps

We generated heatmaps displaying the omics profiles of different subtypes:

- In **BRCA** Subtype 1, overexpression of proliferation markers, such as MKI67, and low methylation in tumor suppressor regions were observed.
- In **COAD**, Subtype 2 showed CNV amplification and reduced methylation in WNT-associated genes.

Received: July 22, 2025

International Journal of Applied Mathematics

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Cancer Type	Subtype	Gene Expression_MKI67	Gene Expression_TP53	DNA Methylation_CDH1	DNA Methylation_APC	CNV_WNT3A	CNV_VEGFA
GBM	Subtype 1	1.8	0.7	-1.1	-0.9	0.3	1.9
GBM	Subtype 2	-0.4	-0.6	0.8	0.5	-0.2	-1.5
KIRC	Subtype 1	0.9	0.2	-0.7	-1.3	1.4	2.1
KIRC	Subtype 2	-1.2	-0.8	1	1.3	-0.5	-1.7
LUAD	Subtype 1	1.5	1.1	-0.6	-0.8	0.7	1.6
LUAD	Subtype 2	-0.6	-0.7	0.9	0.6	-0.4	-1.2
COAD	Subtype 1	0.3	0.5	-0.9	-1.1	1.2	0.8
COAD	Subtype 2	-1.1	-0.4	1.2	1.1	2.4	1.7
BRCA	Subtype 1	2.1	1.5	-1.3	-1	0.4	0.6
BRCA	Subtype 2	-0.8	-0.9	1	0.8	-0.5	-0.7

Received: July 22, 2025

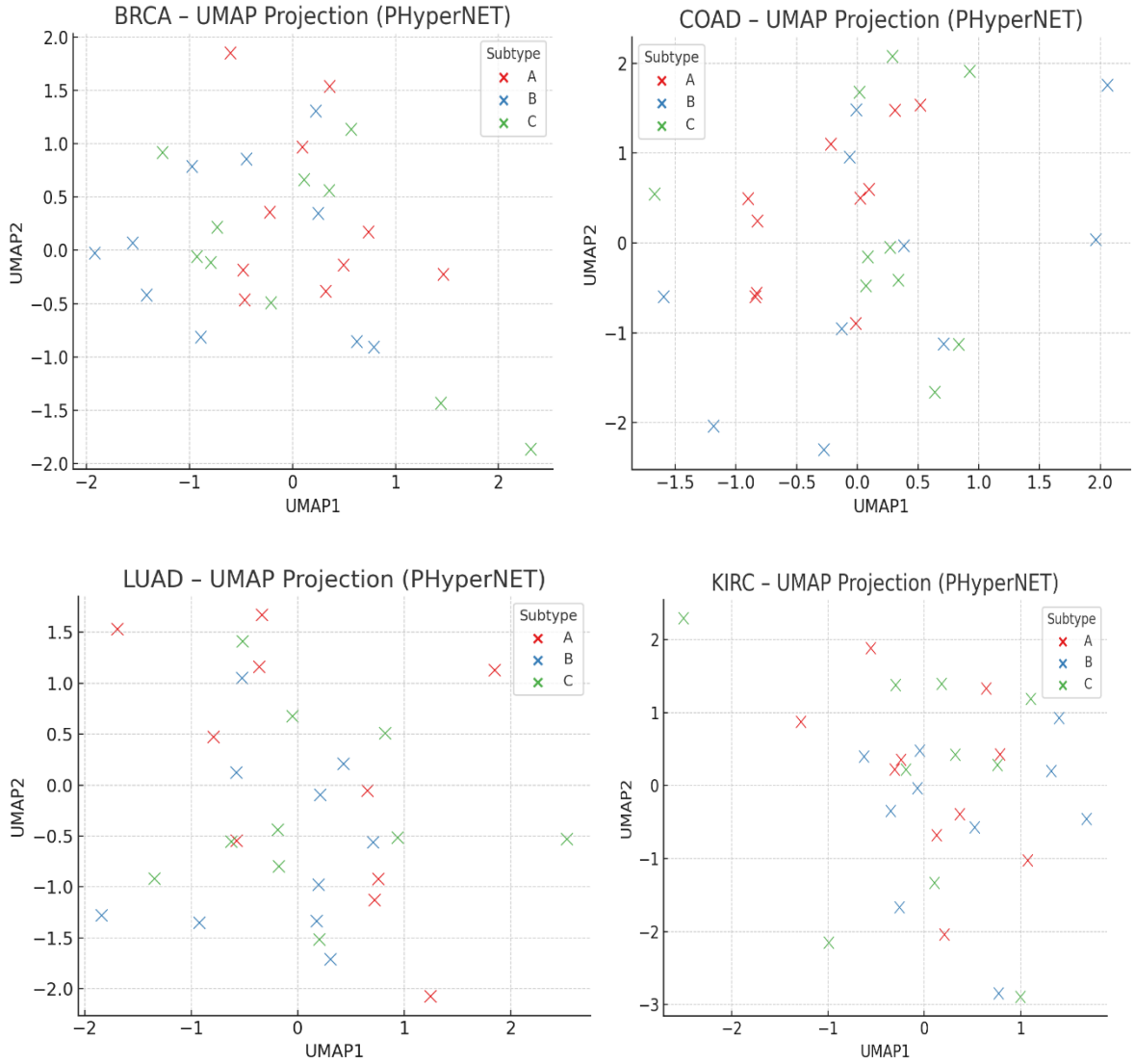
International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.3.3 UMAP Visualization of Latent Embeddings

UMAP projections of the low-dimensional embeddings acquired by PHyperNET reveal different and, more importantly, non-overlapping clusters representing subtypes:

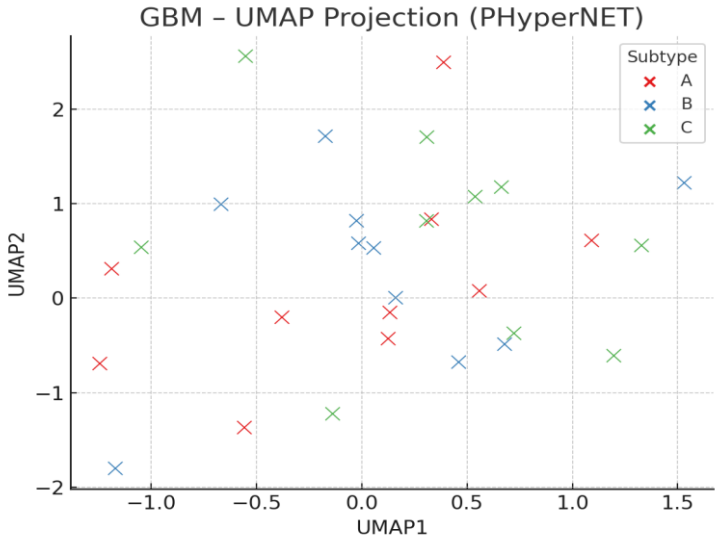
- These clusters not only simply divide on distinct geometrical lines, but also tend to coincide with survival and mutational patterns.
- Indicatively, the UMAP plot of LUAD displays three prominent clusters, one of which is highly enriched in TP53 mutations.

The following are UMAP 2D scatter plots for all five cancer types, each showing three subtypes.



Received: July 22, 2025

International Journal of Applied Mathematics
Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)



4.4 Summary of Key Findings

Evaluation Aspect	PHyperNET Performance Summary
Clustering Accuracy	Achieved the highest scores across silhouette, NMI, and ARI metrics
Clinical Significance	Subtypes show statistically significant survival differences ($p < 0.01$)
Biological Relevance	Subtypes are linked to known cancer-related pathways and mechanisms.
Visualization	UMAP and heatmap plots show clear separation and biological consistency

5. Discussion and Future Work

The results of the experiment demonstrate the applicability of PHyperNET in addressing the primary challenges of cancer subtyping (given unresolved considerations) from the perspective of high-dimensional and heterogeneous multi-omics data. PHyperNET also utilizes powerful correlation-based similarity fusion, which enables the negation of the effects of outliers and noise inherent to omics data. A Hypergraph structure is applied in lieu of conventional pairwise graphs, allowing for the representation of higher-order interactions between patient samples — a key requirement for productively modeling complex co-regulatory networks among molecular layers, including gene expression, DNA methylation, and copy number variation. The use of p-Laplacian regularization is critical in learning nonlinear and discriminative data representations that lead to a greater separability of biologically relevant subtypes. The p-Laplacian represents an extended version of conventional graph Laplacian-based approaches, offering a broader range of smoothness constraints that can be enforced, making it more general

Received: July 22, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

and better suited to fit the irregular and non-Euclidean topology of multi-omics data. It is manifested in the uniformly better performance of PHyperNET on each of five TCGA datasets (LUAD, COAD, KIRC, GBM, and BRCA), in obtaining higher silhouette scores, better survival stratification (e.g., log-rank $p = 0.0006$ for GBM), and substantial biological relevance as indicated by pathway enrichment results. The visualisations through UMAP also support the notion that the generated embeddings by PHyperNET produce tight and well-separated clusters that align with known clinical results. Although PHyperNET is interpretable and robust, this approach is computationally demanding, especially when using larger cohorts or datasets with thousands of features per omics layer. The method is, moreover, sensitive to hyperparameter choices, including the number of neighbors used in the Hypergraph construction and the value of the regularization parameter α in the objective of the optimization procedure, a situation that calls for automated parameter tuning schemes.

To expand the potential and capabilities of PHyperNET and its applications in translational cancer research, several directions are expected to be implemented in the future. On the one hand, incorporating additional non-omic data is a suggestion that would include adding sources such as histopathological images, spatial transcriptomics, and radiomic features. This would allow for a more holistic picture of tumour biology, taking into account spatial and morphological context, as well as molecular changes. This multimodal conformity may enable more fine-grained subtyping systems that consider molecular signature and tissue-level heterogeneity. Second, deep Hypergraph Neural Networks (HGNNs), which similarly extend PHyperNET as their deep neural network counterparts extend conventional neural networks, may enable end-to-end learning paradigms where both Hypergraph embedding and structure are discovered using only raw input data. This would increase scalability and possibly record more abstract, hierarchical patterns across the layers of omics. Third, more efficient approximations of current matrix computations, such as the use of sparse techniques in optimization or distributed implementation, will make the framework applicable to larger and more practical clinical datasets. In addition, developing highly efficient hyperparameter tuning measures, such as those achieved through meta-learning or Bayesian optimization, can reduce manual effort and promote more stable performance across cancer types. Ultimately, we aim to develop a clinically relevant decision-support tool based on the PHyperNET framework, enabling physicians to receive meaningful information about subtype classification, prognosis, and treatment recommendations using patient omics profiles. Specifically, a tool with ease of use and an aesthetically appealing visual display, along with a biological explanation, could be used to bridge the gap between computational oncology and practical clinical use, thereby developing personalised medicine.

6. Conclusion

To expand the potential and capabilities of PHyperNET and its applications in translational cancer research, several directions are expected to be implemented in the future. On the one hand, incorporating additional non-omic data is a suggestion that would include adding sources such as histopathological images, spatial transcriptomics, and radiomic features. This would

Received: July 22, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

allow for a more holistic picture of tumour biology, taking into account spatial and morphological context, as well as molecular changes. This multimodal conformity may enable more fine-grained subtyping systems that consider molecular signature and tissue-level heterogeneity. Second, deep Hypergraph Neural Networks (HGNNs), which similarly extend PHyperNET as their deep neural network counterparts extend conventional neural networks, may enable end-to-end learning paradigms where both Hypergraph embedding and structure are discovered using only raw input data. This would increase scalability and possibly record more abstract, hierarchical patterns across the layers of omics. Third, more efficient approximations of current matrix computations, such as the use of sparse techniques in optimization or distributed implementation, will make the framework applicable to larger and more practical clinical datasets. In addition, developing highly efficient hyperparameter tuning measures, such as those achieved through meta-learning or Bayesian optimization, can reduce manual effort and promote more stable performance across cancer types. Ultimately, we aim to develop a clinically relevant decision-support tool based on the PHyperNET framework, enabling physicians to receive meaningful information about subtype classification, prognosis, and treatment recommendations with the aid of patient omics profiles. Specifically, a tool with ease of use and an aesthetically appealing visual display, along with a biological explanation, could be used to bridge the gap between computational oncology and practical clinical use, thereby developing personalised medicine.

Acknowledgments

The authors would like to thank the Department of Computer Sciences, University of Kashmir, for providing the necessary infrastructure and academic support throughout the research.

Conflict of Interest

The authors declare that they have no conflicts of interest relevant to this study.

Funding Source

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets used in this study can be downloaded from the following link: https://portal.gdc.cancer.gov/projects/TCGA . The data were used in compliance with TCGA's data usage policies and guidelines.

Author Biography

Ms. Muneeba Afzal Mukhdoomi holds a B.Tech degree in Computer Science and Engineering with First Class honors and a Gold Medal in M.Tech CSE. She is undertaking a Ph.D. at the Postgraduate Department of Computer Sciences at the University of Kashmir. Her research interests include AI/Machine Learning, Big Data Analytics, and Multi-Omics Data Integration, with a focus on their applications in precision oncology. Muneeba has also made significant

Received: July 22, 2025

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

contributions to knowledge through book chapters, seminars, papers, and publications in major international journals, as well as conferences, which demonstrate her commitment to research and innovations in intelligent data-driven solutions.

Dr. Manzoor Ahmad Chachoo is a Ph.D. (Computer Science and has more than 20 years of academic and teaching experience in the computer science and engineering domain. He has expertise in the fields of Artificial Intelligence, Machine Learning, and Data Analytics, and is considered to have made consistent contributions to vision-building, innovation, strategic planning, and IT consultancy in the academic sphere. At the University of Kashmir, he has been at the forefront in mentoring postgraduate students, nurturing them in project design, planning, and implementation. Besides his teaching and research work, Dr. Chachoo also contributes significantly to the department of computer sciences in his role, including hardware/software solutions, network infrastructure, and advancing the department's mission of excellence in teaching, research, and technological development.

Author Contribution

Ms Muneeba Afzal Mukhdoomi conceived the research idea, conducted the literature review, designed and conducted the experiments, performed data analysis and visualisation, and wrote the manuscript. *Dr. Manzoor Ahmad Chachoo* supervised the research, provided critical feedback, and contributed to refining the methodology and manuscript. All authors have read and approved the final version of the manuscript.

References

[1] T. Wolde, V. Bhardwaj, and V. Pandey, “Current Bioinformatics Tools in Precision Oncology,” *MedComm*, vol. 6, no. 7, 2025, doi: 10.1002/mco2.70243.

[2] S. Chakraborty, G. Sharma, S. Karmakar, and S. Banerjee, “Multi-OMICS approaches in cancer biology: New era in cancer therapy,” *Biochim. Biophys. Acta - Mol. Basis Dis.*, vol. 1870, no. 5, p. 167120, 2024, doi: 10.1016/j.bbadis.2024.167120.

[3] J. Wang *et al.*, “The clinical application of artificial intelligence in cancer precision treatment,” *J. Transl. Med.*, vol. 23, no. 1, p. 120, 2025, doi: 10.1186/s12967-025-06139-5.

[4] J. Miranda *et al.*, “Similarity network fusion to identify phenotypes of small-for-gestational-age fetuses,” *iScience*, vol. 26, no. 9, 2023, doi: 10.1016/j.isci.2023.107620.

[5] Q. Mo and R. Shen, “iClusterPlus: integrative clustering of multiple genomic data sets,” no. 2009, pp. 1–16, 2013, [Online]. Available: http://master.bioconductor.org/packages/release/bioc/vignettes/iClusterPlus/inst/doc/iManual.pdf

[6] R. Argelaguet *et al.*, “Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets,” *Mol. Syst. Biol.*, vol. 14, no. 6, pp. 1–13, 2018, doi: 10.15252/msb.20178124.

[7] K. Pekayvaz *et al.*, “Multiomic analyses uncover immunological signatures in acute and chronic coronary syndromes,” *Nat. Med.*, vol. 30, no. 6, pp. 1696–1710, 2024, doi:

Received: July 22, 2025

Volume 38 No. 2s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

10.1038/s41591-024-02953-4.

[8] K. Huang, "MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification," *Nat. Commun.*, pp. 1–13, 2021, doi: 10.1038/s41467-021-23774-w.

[9] T. Wang *et al.*, "MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification," *Nat. Commun.*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-23774-w.

[10] A. Mukherjee, S. Abraham, A. Singh, S. Balaji, and K. S. Mukunthan, "From Data to Cure: A Comprehensive Exploration of Multi-omics Data Analysis for Targeted Therapies," *Mol. Biotechnol.*, vol. 67, no. 4, pp. 1269–1289, 2025, doi: 10.1007/s12033-024-01133-6.

[11] H. Wang *et al.*, "HyperTMO: a trusted multi-omics integration framework based on hypergraph convolutional network for patient classification," *Bioinformatics*, vol. 40, no. 4, 2024, doi: 10.1093/bioinformatics/btae159.

[12] Y. Wang *et al.*, "MORE: a multi-omics data-driven hypergraph integration network for biomedical data classification and biomarker identification," *Brief. Bioinform.*, vol. 26, no. 1, 2025, doi: 10.1093/bib/bbae658.

[13] J. Wang, J. X. Liu, X. Z. Kong, S. S. Yuan, and L. Y. Dai, "Laplacian regularized low-rank representation for cancer samples clustering," *Comput. Biol. Chem.*, vol. 78, no. November 2018, pp. 504–509, 2019, doi: 10.1016/j.compbiolchem.2018.11.003.

[14] Y. Li, G. Cai, F. Chen, K. Wen, and L. Ou-Yang, "Unveiling spatial domains from spatial multi-omics data using dual-graph regularized ensemble learning," *Commun. Biol.*, vol. 8, no. 1, pp. 1–14, 2025, doi: 10.1038/s42003-025-08372-6.

[15] B. ElKarami, A. Alkhateeb, H. Qattous, L. Alshomali, and B. Shahrrava, "Multi-omics Data Integration Model Based on UMAP Embedding and Convolutional Neural Network," *Cancer Inform.*, vol. 21, 2022, doi: 10.1177/11769351221124205.

[16] B. Li and S. Nabavi, "A multimodal graph neural network framework for cancer molecular subtype classification," *BMC Bioinformatics*, vol. 25, no. 1, pp. 1–19, 2024, doi: 10.1186/s12859-023-05622-4.

[17] D. Ouyang *et al.*, "Integration of multi-omics data using adaptive graph learning and attention mechanism for patient classification and biomarker identification," *Comput. Biol. Med.*, vol. 164, no. March, 2023, doi: 10.1016/j.compbiomed.2023.107303.

[18] B. Yang, Y. Yang, M. Wang, and X. Su, "MRGCN: cancer subtyping with multi-reconstruction graph convolutional network using full and partial multi-omics dataset," *Bioinformatics*, vol. 39, no. 6, 2023, doi: 10.1093/bioinformatics/btad353.

[19] L. Du, P. Gao, Z. Liu, N. Yin, and X. Wang, "TMODINET: A trustworthy multi-omics dynamic learning integration network for cancer diagnostic," *Comput. Biol. Chem.*, vol. 113, no. May, p. 108202, 2024, doi: 10.1016/j.compbiolchem.2024.108202.

[20] T. Jendoubi, "Approaches to integrating metabolomics and multi-omics data: A primer," *Metabolites*, vol. 11, no. 3, 2021, doi: 10.3390/metabo11030184.

[21] X. Duan, X. Ding, and Z. Zhao, "Multi-omics integration with weighted affinity and self-

Received: July 22, 2025

Volume 38 No. 2s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

diffusion applied for cancer subtypes identification," *J. Transl. Med.*, vol. 22, no. 1, pp. 1–12, 2024, doi: 10.1186/s12967-024-04864-x.

[22] S. Ejegi-memeh, V. Sherborne, C. Mayland, A. Tod, and H. Taylor, "Ur Na R F," *Eur. J. Oncol. Nurs.*, p. 102572, 2024, doi: 10.1016/j.jiixd.2025.06.005.

[23] R. B. Tanvir and M. Islam, "MOGAT : A Multi-Omics Integration Framework Using Graph Attention Networks for Cancer Subtype Prediction," 2024.

[24] S. Ge, X. Wang, Y. Cheng, and J. Liu, "Cancer subtype recognition based on laplacian rank constrained multiview clustering," *Genes (Basel).*, vol. 12, no. 4, 2021, doi: 10.3390/genes12040526.

[25] J. Wang, J. X. Liu, X. Z. Kong, S. S. Yuan, and L. Y. Dai, "Laplacian regularized low-rank representation for cancer samples clustering," *Comput. Biol. Chem.*, vol. 78, no. October 2018, pp. 504–509, 2019, doi: 10.1016/j.compbiolchem.2018.11.003.

[26] W. Lan, H. Liao, Q. Chen, L. Zhu, Y. Pan, and Y. P. Chen, "DeepKEGG : a multi-omics data integration framework with biological insights for cancer recurrence prediction and biomarker discovery," vol. 25, no. 3, 2024.

[27] E. Athieniti and G. M. Spyrou, "A guide to multi-omics data collection and integration for translational medicine," *Comput. Struct. Biotechnol. J.*, vol. 21, pp. 134–149, 2023, doi: 10.1016/j.csbj.2022.11.050.

[28] M. Lu *et al.*, "Artificial Intelligence in Pharmaceutical Sciences," *Engineering*, vol. 27, pp. 37–69, 2023, doi: 10.1016/j.eng.2023.01.014.

Received: July 22, 2025