International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(14s) 2026 pp. 401-415
https://doi.org/10.70917/ijcisim-2026-4239
Article

CGP-Net: A Cross-Modal Attention Network with Graph Priors for Integrated Multi-Omics Disease Subtyping

Sowmya K[1], Ananth Prabhu G[2*], Mustafa Basthikodi[3]

[1]Department of Computer Science and Engineering, Sahyadri College of Engineering & Management, Affiliated to Visvesvaraya Technological University, Belagavi-590018, India
[1]Department of Computer Science and Engineering, Canara Engineering College, Mangalore, India
[2]Department of Computer Science and Engineering, Sahyadri College of Engineering & Management, Affiliated to Visvesvaraya Technological University, Belagavi-590018, India
[3]Department of Computer Science and Engineering, Sahyadri College of Engineering & Management, Affiliated to Visvesvaraya Technological University, Belagavi-590018, India
sowmya.kottari@gmail.com, educatorananth@gmail.com, mustafa.cs@sahyadri.edu.in

Abstract: Integrating upstream genomic variations and downstream functional proteomic signalling is necessary for developing precision oncology. There are challenges in the integration of multi-omics data due to large discrepancies in dataset dimensions, structural heterogeneity, and non-linear interdependencies between modalities. All of these challenges can cause high-dimensional data of genomics to overshadow important data of proteomics. To remedy these issues, we developed a new multi-omics integration framework and called it CrossGeneProtein-Net (CGP-Net). CrossGeneProtein-Net utilizes a Sparse Autoencoder (SAE) that compresses the high-dimensional genome data and a Protein-Protein Interaction (PPI) focused Graph Attention Network (GAT) that forms a structure of interrelated biological activities that are incorporated within the proteomic communication framework. Capturing cross talking of the different modalities was accomplished through a bidirectional cross-attention architecture. An InfoNCE objective with a contrastive approach served to keep relative positions of the data according to the degree of intermodal noise and were more easily interpretable and consistent in relation to the data. Validation of CGP-Net was conducted on multiple TCGA and CPTAC datasets and it outperformed baseline models MOGONET and OASIS with statistical significance of $p < 0.01$. On the TCGA-BRCA benchmark, CGP-Net achieved a macro-balanced accuracy of 91.82% for PAM50 molecular subtyping and an overall survival concordance index (C-index) of 0.784, significantly outperforming state-of-the-art baselines including MOGONET and OASIS ($p < 0.01$). The Integrated Gradients method of feature attribution validated CGP-Net's superior performance through its emphasis on important oncogenes as well as pathways.
Keywords: Multi-omics integration, Graph Attention Networks, Cross-modal attention, Proteogenomics, Precision oncology, Contrastive learning.

1. INTRODUCTION

1.1 Background

Modern, high-throughput multi-omic technologies that are based on sequencing genomics and transcriptomics, and advanced mass-spectrometry proteomics, have developed rapidly, completely revolutionizing precision medicine. Single-omics profiling provides data on a single layer of cellular state focus, while a majority of complex cellular phenomena, such as oncogenesis, disease progression and therapeutic resistance, are multi-layered. Changes to the genome will dictate subsequent transcriptomic expression levels. Changes to the transcriptome will affect the process of translation and the subsequent post-translational modifications to the resulting protein. Thus, the integration of multiple layers of omics data, in particular genomic maps of instability paired with proteomic data on cellular signaling pathways, is critical in unveiling hidden sub-types of disease, finding novel biomarkers, and predicting patient outcomes with greater accuracy.

1.2 The Literature Gap, & Computational Bottlenecks

Despite the evolution of computational biology, combining data on the genome and proteome is still deeply challenging from an algorithmic perspective due to structural and feature imbalances, and complicated non-linear intermodal dependencies. Most of the existing computational frameworks for the integration of omic data, fall to the three classical paradigms, each with their own challenges and limitations.

- *Concatenation-Based Early Fusion:* During early phases of integration, the extensive number of genomic features (magnitude ~10^4 to 10^5) are directly concatenated with a limited number of proteomic features (magnitude ~10^2 to 10^3). This leads to a significant imbalance in data. The resulting imbalance will lead to the overwhelming dominance of the proteomic data, which is critical to the function of the model, being effectively suppressed. Furthermore, the early concatenation of the features will lead to a loss of the intrinsic topology and structure of each of the data modalities.
- ***Decision-Level Integration Late Fusion:*** The independent modeling of each of the omic data modalities is integrated through the aggregation (typically weighted) of the final output predictions (e.g., voting). Though Late Fusion resolves the data imbalance problem, it operates under the overly simplistic assumption that each of the data modalities is independent and as such is not capable of capturing any of the inter-omic regulatory dynamics (e.g. how specific variant(s) of the genome are capable of modulating the proteomic cellular signaling pathways).
- ***Unsupervised Joint Embedding models:*** Linear frameworks such as Multi-Omics Factor Analysis (MOFA+) and some variants of autoencoders project data into a shared latent space. Unfortunately, these frameworks often fail to differentiate shared inter-omics cross-talk from modality-specific noise. Consequently, the resulting feature representation is typically ineffective for supervised downstream clinical applications.

1.3 Core Contributions & Proposed Solution: CGP-Net

To overcome these structural limitations, CrossGeneProtein-Net (CGP-Net), a deep learning architecture, is introduced innovatively to integrate genomic and proteomic data. The first of its kind, CGP-Net incorporates a novel cross-modal attention backbone in conjunction with structural graph priors, allowing for the mapping of the functional interdependencies between high-dimensional, variable, and dynamic proteomic networks and genomic data. Unlike prior systems, the architecture of CGP-Net will not rely on concatenation or averaging decision fusion.

The primary contributions of CGP-Net can be described as:

- ***Modality-Specific Encoders with Graph Priors:*** A genomic branch leverages a sparse autoencoder and a proteomic branch utilizes a Graph Attention Network initialized with a biological Protein-Protein Interaction (PPI) to retain functional protein complexes.
- ***Cross-Modal Attention Fusion:*** Cross-modality bidirectional attention is designed to allow genomic representations to query proteomic keys and values.
- ***Contrastive Subspace Alignment:*** A multi-omics contrastive loss is applied to promote the same biological entities across different modalities and to reduce mal-alignment between the genomic and proteomic representations of the same patient.
- ***Thorough Benchmarking & Biologically Based Validation:*** CGP-Net is assessed against the large-scale paired proteogenomic data in The Cancer Genome Atlas (TCGA) and The Clinical Proteomic Tumor Analysis Consortium (CPTAC) to yield statistically significant predictive improvements over other systems. Integrated Gradients and pathway enrichment are used to describe biologically relevant and interpretable multi-omics biomarkers.

2. LITERATURE REVIEW

The rapid maturation of high-throughput multi-omics profiling has made the integrative modelling of transcriptomic variation and functional proteomic signalling a cornerstone of modern precision oncology [1–3]. However, extracting disease-relevant representations from matched genomic (x_G) and proteomic (x_P) patient profiles present severe computational challenges due to extreme feature scale imbalance ($d_G \gg d_p$), topological heterogeneity, and dynamic cross-modal feedback mechanisms [4–7].

Historically, computational integration paradigms in multi-omics analysis have been broadly categorized into

three classical methodologies [1, 2, 8]:

1. ***Early Fusion Protocols:*** Direct feature-level concatenation passing through supervised machine learning models inherently suffers from feature scale disparity [4, 5]. Because transcriptomic matrices ($d_G \approx 20000$) far outweigh mass-spectrometry proteomic measurements ($d_p \approx 1000$), unweighted concatenation causes backpropagation gradient updates to be dominated by the genomic branch, effectively suppressing subtle yet functionally critical post-translational proteomic signals [4, 6].
2. ***Late Fusion Paradigms:*** Decision-level ensembling constructs isolated predictive models for each omics view independently and aggregates their terminal predictions [9]. While this circumvents dimensionality mismatch, late fusion assumes modality independence and fails to model non-linear cross-layer regulatory dependencies between upstream genomic mutations and downstream proteomic execution [5, 7, 10].
3. ***Unsupervised Latent Factor Models:*** Linear latent variable frameworks such as MOFA+ [8] and deep generative architectures like Subtype-GAN [11] project multi-omics data into shared low-dimensional manifolds. However, unsupervised approaches frequently fail to distinguish disease-driving cross-talk from background biological noise, yielding suboptimal latent representations for supervised downstream clinical prediction [4, 8, 12].

To overcome these structural limitations, recent computational architectures have increasingly deployed deep graph neural networks, cross-modal attention transformers, and non-linear function approximators [4–6, 9, 10, 13–16]. Graph-based integration frameworks such as MOGONET [9] and DeepMoIC [6] process patient similarity views via Graph Convolutional Networks (GCNs), while MOGOLA [4] and CrossAttOmics [5] utilize cross-attention mechanisms [10] to model non-linear interactions across omics layers [1, 2, 5]. Further advances in deep representation learning, including MOKAN [14] using Kolmogorov–Arnold networks, DGNMDA [15] employing dual heterogeneous graph neural encoders, and MOTGNN [13], demonstrate the necessity of topological message passing [16] for biological network analysis [4, 13, 15].

Moreover, recent self-supervised methodologies demonstrate that contrastive learning objectives—specifically dual-branch contrastive alignment (DBCL-DFNet) [7], graph contrastive clustering for single-cell data [12], and enhanced differential attention (scECDA) [17]—leveraging InfoNCE objective formulations [18] maximize mutual information between heterogeneous omics views while suppressing modality-specific noise [3, 7, 12, 17, 18].

As formulated in **CGP-Net**, fully resolving these multi-omics integration bottlenecks requires embedding domain-specific biological priors directly into the network architecture [4, 13, 19]. By combining Sparse Autoencoders for genomic vector compression with Protein–Protein Interaction (PPI) guided Graph Attention Networks (GAT) [16], CGP-Net explicitly preserves functional protein interaction topologies. Coupling this dual-branch design with bidirectional cross-modal attention [5, 10] and InfoNCE contrastive subspace alignment [7, 18] prevents genomic scale dominance while establishing translational interpretability through feature attribution scoring [20] and Reactome pathway enrichment [19].

3. METHODOLOGY AND MODEL ARCHITECTURE

To resolve fundamental computational challenges of multi-omics integration, including extreme feature dimensionality mismatch ($d_G \gg d_P$) and inter-modal regulatory asymmetry of genomics and proteomics, we developed CrossGeneProtein-Net (CGP-Net). The end-to-end computational method consists of four sequential learning steps: (1) Data Pre-Processing & Biological Graph Construction, (2) Dual-Branch Feature Extraction Encoders with Domain Priors, (3) Bidirectional Cross-Modal Attention & Contrastive Subspace Alignment, and (4) Prediction & Optimization Framework (with objectives & loss functions). An illustration of the complete system can be found in Figure 1.

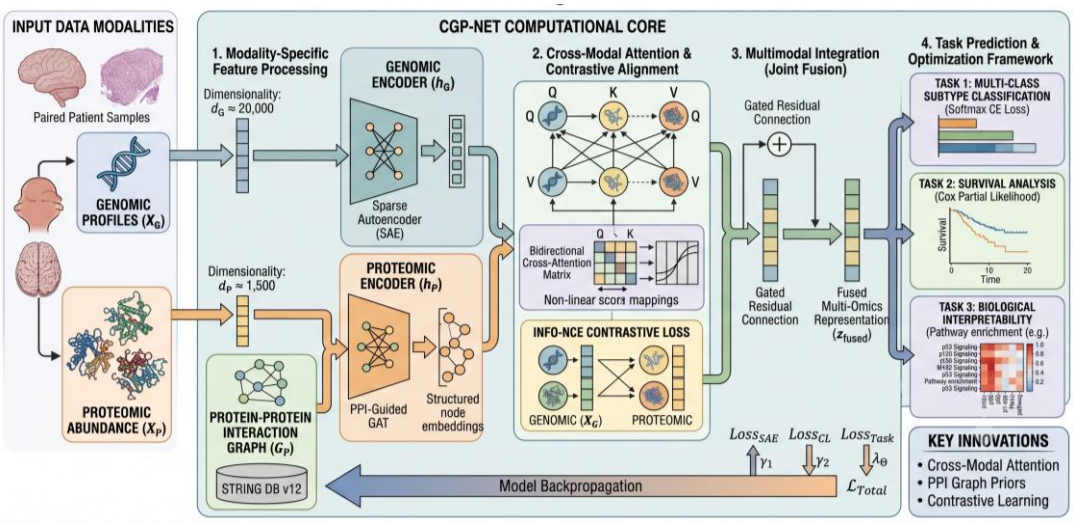

Figure 1. Conceptual End-to-End Computational Pipeline of CGP-Net.

3.1 Problem Formulation & Biological Graph Construction

The advanced multi-omics profiling systems available in modern laboratories provide an extensive overview of cellular mechanisms on an unprecedented scale. The integration of different modalities of omics data is mostly hampered by the inherent structural gaps among them. For instance, genomics data captured by methods such as RNA sequencing or copy-number profiling primarily depict upstream genetic changes and the teleological inclination toward transcription. In contrast, functional proteomic data capture the downstream execution and the resultant state of a cell after the completion of all the relevant post-translational modifications. To address this multi-layered issue, CGP-Net aims to provide a dual-modality representation of the problem in which the raw, high-dimensional vectors of patients are coupled with domain-specific biological knowledge graphs. The framework of CGP-Net provides a means to transform unstructured molecular data and renders them, in topological terms, structured. Furthermore, the system preserves functional signaling pathways and ensures mathematical closability with respect to the disparate dimensions in proteomics and genomics.

3.1.1 Formal Task Definition

Let $\mathcal{D} = \{(\mathrm{x}_G^{(i)}, \mathrm{x}_P^{(i)}, y^{(i)})\}_{i=1}^{N}$ represent a cohort of N matched, patient-derived multi-omics profiles. The vector $x_G^{(i)} \in \mathbb{R}^{d_G}$ denotes the high-dimensional genomic feature vector derived from whole-transcriptome sequencing (RNA-seq log-transformed RPKM/TPM values) or copy-number variant (CNV) matrices. The vector $\mathrm{x}_P^{(i)} \in \mathbb{R}^{d_P}$ represents mass spectrometry-based or Reverse Phase Protein Array (RPPA) functional proteomic abundance signals. A central challenge in proteogenomics is the feature scale disparity, where $d_G \approx 20{,}000$ variables while $d_P \approx 1{,}000$--$2{,}000$ quantified protein targets ($d_G \gg d_P$). The target $y^{(i)}$ represents managed clinical endpoints, such as multi-class molecular cancer subtyping ($y \in \{1,2,\ldots,C\}$) or right-censored survival outcomes considered by event time $T^{(i)}$ and censoring indicator $\delta^{(i)} \in \{0,1\}$.

3.1.2 Functional Protein Interaction Topology ($\mathcal{G}_P$)

Proteins do not function in isolation; rather, they form physical complexes and enzymatic signaling cascades. To encode these topological dependencies directly into the deep learning pipeline, we map the proteomic features onto an undirected biological interaction network $\mathcal{G}_P = (\mathcal{V}_P, \mathcal{E}_P, A)$. The vertex set $\mathcal{V}_P$ corresponds to the d_P measured proteins, and the edge set $\mathcal{E}_P$ captures verified protein-protein interactions extracted from the STRING database (v12.0). We construct a weighted, normalized adjacency matrix $A \in \mathbb{R}^{d_P \times d_P}$ where element $a_{jk} \in [0,1]$ represents the combined functional interaction confidence score between protein j and protein k. To eliminate spurious background interactions, edges with confidence scores below a stringent threshold ($a_{jk} < 0.40$) are pruned. Self-loops are explicitly added to ensure node identity retention during graph message passing. (See Supplementary Section S3 for network topological statistics).

3.2 Dual-Branch Feature Extraction Encoders

If high-dimensional genomic features and lower-dimensional proteomic signals are concatenated before feature selection, gradient updates during backpropagation are disproportionately driven by the genomic inputs. This causes the network to ignore critical post-translational proteomic signals. To prevent this, CGP-Net processes each modality through an independent, custom-designed encoder branch that maps both inputs into a shared latent space $\mathbb{R}^{d_m}$.

3.2.1 Genomic Branch: Sparse Autoencoder (SAE)

High-dimensional genomic profiles contain substantial background noise and redundant co-expression modules. We employ a Sparse Autoencoder (SAE) designed to compress x_G into a dense, salient genomic embedding $h_G \in \mathbb{R}^{d_m}$. The encoder projects raw inputs through a hidden non-linear transformation:

$$h_G = \sigma(W_{G2} \cdot \sigma(W_{G1} x_G + b_{G1}) + b_{G2}) \quad (1)$$

where $W_{G1} \in \mathbb{R}^{d_{hidden} \times d_G}$ and $W_{G2} \in \mathbb{R}^{d_m \times d_{hidden}}$ are weight matrices, b denotes bias vectors, and $\sigma(\cdot)$ represents the Gaussian Error Linear Unit (GELU) activation function. To force the encoder to isolate key driver genes over passenger mutations, we apply an L1-norm sparsity constraint on the latent activations during pre-training:

$$\mathcal{L}_{SAE} = \| x_G - \hat{x}_G \|_2^2 + \lambda_{sp} \| h_G \|_1 \quad (2)$$

where $\hat{\mathbf{x}}_G$ is the reconstructed genomic vector and λ_{sp} controls the sparsity strength.

3.2.2 Proteomic Branch: PPI-Guided Graph Attention Network (GAT)

A new addition in the proteomic branch of CGP-Net involves the addition of functional biological priors directly into the teaching of feature representation. To avoid the barrier of flattening proteomic signals into random vectors, we design a custom Graph Attention Network (GAT) encoder. This encoder uses the available structural topology of human Protein-Protein Interaction (PPI) networks to scaffold message-passing and topology-based message aggregation. The mechanics of the proteomic graph construction and GAT aggregation are explained in detail in Figure 3.

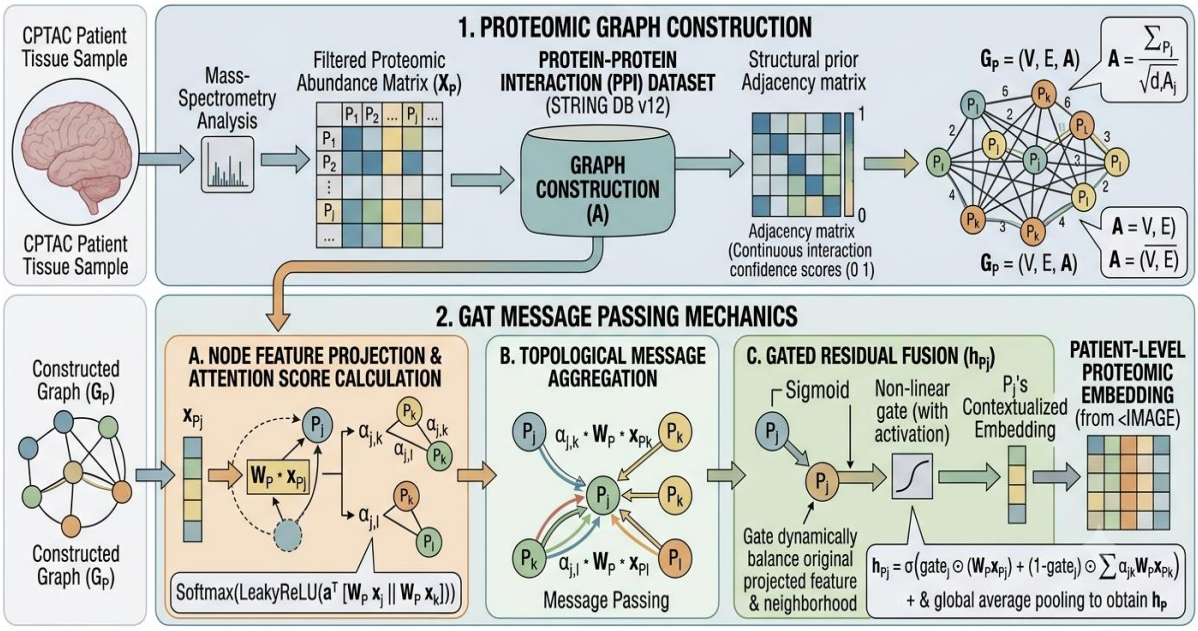

Figure 3. Proteomic Graph Construction and GAT Message Passing Mechanics

Rather than flattening protein features into an unstructured array, the proteomic branch processes $\mathbf{x}_P$ through a multi-head Graph Attention Network (GAT) guided by the interaction matrix A. The GAT layer dynamically weights the importance of neighbouring proteins within signalling pathways, aggregating structural context into node embeddings. For a protein node j, the updated feature representation across its biological neighborhood $\mathcal{N}(j) \cup \{j\}$ is formulated as:

$$h_{P,j} = \alpha_{j,j} W_P x_{P,j} + \sum_{k \in \mathcal{N}(j)} \alpha_{j,k} W_P x_{P,k} \tag{3}$$

The learnable attention coefficient $\alpha_{j,k}$ quantifies the biological signaling relevance of neighboring protein k to target protein j:

$$\alpha_{j,k} = \frac{exp\left(LeakyReLU\left(a^T [W_P x_{P,j} \parallel W_P x_{P,k}]\right)\right)}{\sum_{l \in \mathcal{N}(j) \cup \{j\}} exp\left(LeakyReLU\left(a^T [W_P x_{P,j} \parallel W_P x_{P,l}]\right)\right)} \tag{4}$$

$$h_P = \sum_{j \in \mathcal{V}_P} Softmax(W_p h_{P,j}) \cdot h_{P,j} \tag{5}$$

where $W_P \in \mathbb{R}^{d_m \times d_P}$ is a parameterized linear transformation, a is a learnable attention mechanism vector, and $\parallel$ denotes vector concatenation. Following graph convolutions, a global attention pooling operation condenses all node-level representations into a unified, patient-level proteomic embedding $h_P \in \mathbb{R}^{d_m}$.

3.3 Cross-Modal Attention Fusion & Contrastive Subspace Alignment

Once both modalities are mapped into the shared latent dimension d_m, CGP-Net establishes cross-layer biological dependencies using a bidirectional cross-attention backbone paired with contrastive subspace alignment.

3.3.1 Bidirectional Cross-Attention Backbone

To model dynamic, non-linear regulatory cross-talk between genomic variants and proteomic targets without forcing early feature concatenation, we construct a specialized cross-modal attention module paired with adaptive gating. The inner tensor mechanics and flow of Query-Key-Value (Q, K, V) projections across modalities are illustrated in Figure 2

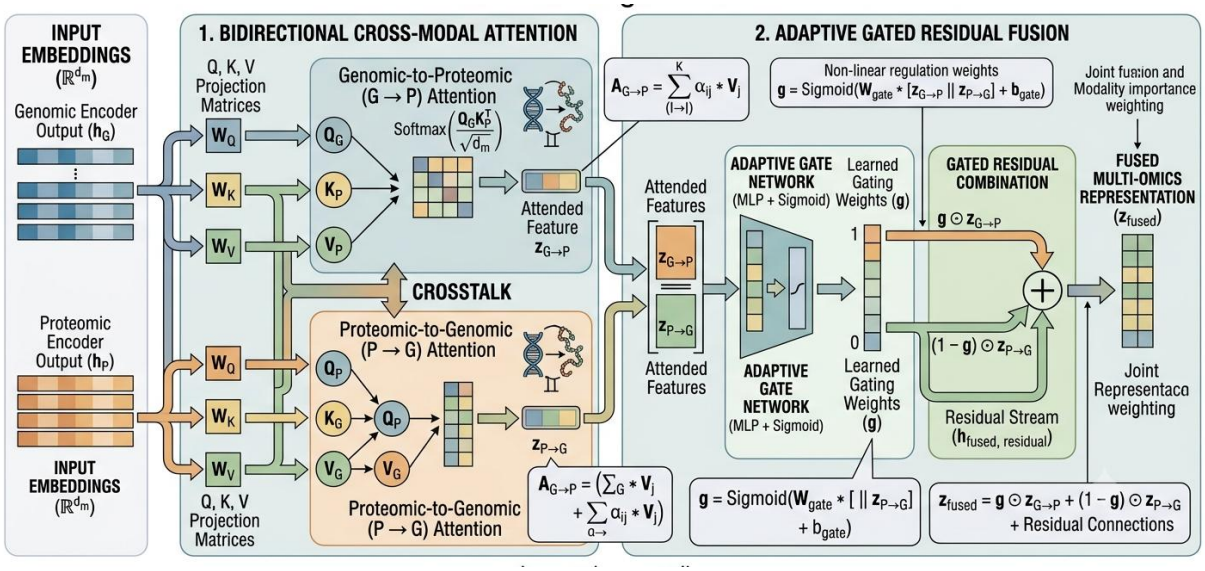

Figure 2. Micro-architecture of the Bidirectional Cross-Modal Attention and Gated Residual Fusion Layer

Genomic mutations frequently dictate downstream protein activity, while protein complexes exert feedback control over gene expression. To model this dynamic cross-talk, we formulate a bidirectional cross-attention mechanism. Genomic representations h_G are projected to Query space, while proteomic representations h_P are projected to Key and Value spaces:

$$Q_G = h_G W_Q, K_P = h_P W_K, V_P = h_P W_V \tag{6}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{d_m \times d_m}$. The genomic-to-proteomic cross-attended representation $\mathbf{z}_{G \to P}$ is calculated as:

$$\mathrm{z}_{\mathrm{G} \to \mathrm{P}} = \mathrm{Softmax}\left(\frac{Q_G K_P^T}{\sqrt{d_m}}\right) V_P \tag{7}$$

This operation allows the network to query which proteomic signaling modules are most strongly perturbed

by specific genomic variant patterns. Symmetrically, we calculate the proteomic-to-genomic representation $z_{P\to G}$ by projecting h_P as Queries and h_G as Keys/Values. The two directional representations are fused via a learnable gated residual connection:

$$g = \sigma\left(W_{gate}[z_{G\to P} \parallel z_{P\to G}]\right) \quad (8)$$

$$z_{fused} = g \odot z_{G\to P} + (1-g) \odot z_{P\to G} \quad (9)$$

where $g \in \mathbb{R}^{d_m}$ acts as an adaptive gating vector and $\odot$ represents the Hadamard (element-wise) product.

3.3.2 Multi-Omics InfoNCE Contrastive Subspace Regularization

To prevent modality-specific noise from dominating the shared latent space, we enforce structural alignment between matched multi-omics profiles. We apply a symmetric InfoNCE contrastive loss over a training mini-batch of size B. For patient i, paired genomic ($h_G^{(i)}$) and proteomic ($h_P^{(i)}$) embeddings are treated as positive pairs, while embeddings from different patients ($j \neq i$) serve as negative pairs:

$$\mathcal{L}_{CL} = -\frac{1}{B}\sum_{i=1}^{B} log \frac{exp\left(sim(h_G^{(i)}, h_P^{(i)})/\tau\right)}{\sum_{j=1}^{B} exp\left(sim(h_G^{(i)}, h_P^{(j)})/\tau\right)} \quad (10)$$

where $sim(u,v) = \frac{u^T v}{\|u\|_2 \|v\|_2}$ computes the cosine similarity and $\tau > 0$ represents a learnable temperature scale parameter. This loss forces the dual encoders to align matching patient states in embedding space before final prediction.

3.4 Task Prediction Heads & End-to-End Optimization

The unified cross-attended multi-omics representation $z_{fused} \in \mathbb{R}^{d_m}$ combines various genomics perturbations, proteomics, and cross-modal regulatory interactions. z_fused, is a representation of an accessible latent state, and is processed through prediction heads designed for specific tasks. Differentially, CGP-Net espouses a multi-task learning approach, whereby multi-class cancer proteomic subtyping and time-to-event patient survival modeling is optimized concurrently. Given this approach, multi-task learning is employed at the modeling level, ensuring the model is not overly strict in short-term class goals at the expense of the prognostic (and expected) features of a long duration. Prediction heads are designed, and the rest of this optimization system is illustrated below.

3.4.1 Subtype Classification & Survival Heads

For multi-class cancer subtyping, z_{fused} is mapped through a Multi-Layer Perceptron (MLP) with Softmax activation to generate class probability distributions $\hat{y}$. The task loss $\mathcal{L}_{Task}$ is computed using standard Cross-Entropy Loss:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{B}\sum_{c=1}^{C} y_{i,c} \log \hat{y}_{i,c} \quad (11)$$

For patient survival analysis, the task head predicts a continuous log-hazard risk score $f(z_{fused})$. The model is optimized using Cox Proportional Hazards Negative Log-Partial Likelihood Loss:

$$\mathcal{L}_{Cox} = -\sum_{i\in\mathcal{E}} \left(f(z_{fused}^{(i)}) - \log \sum_{j\in\mathcal{R}(T_i)} \exp\left(f(z_{fused}^{(j)})\right)\right) \quad (12)$$

where $\mathcal{E}$ represents the set of uncensored event occurrences and $\mathcal{R}(T_i)$ denotes the risk set of patients surviving at time. T_i.

3.4.2 Comprehensive Joint Objective

CGP-Net is trained end-to-end by minimizing a unified composite loss function that balances task prediction accuracy, cross-modal alignment, autoencoder reconstruction, and model complexity:

$$\mathcal{L}_{Total} = \mathcal{L}_{Task} + \gamma_1 \mathcal{L}_{CL} + \gamma_2 \mathcal{L}_{SAE} + \lambda_\Theta \|\Theta\|_2^2 \quad (13)$$

where γ_1, γ_2 are scalar weighting coefficients that regulate the contribution of contrastive alignment and autoencoder reconstruction, respectively. All trainable weights Θ are regularized via an L2 weight decay penalty λ_Θ. The entire framework is implemented in PyTorch and trained using the AdamW optimiser with cosine-annealed learning rate schedules. Hyperparameter optimization across parameters $(d_m, \tau, \gamma_1, \gamma_2)$ is conducted via 5-fold cross-validation.

In summary, the proposed CGP-Net architecture provides a unified, mathematically sound framework designed to bridge the gap between high-dimensional genomic variants and functional proteomic signalling. By coupling a Sparse Autoencoder with a PPI-guided Graph Attention Network, CGP-Net explicitly respects modality-specific structures and biological priors while mitigating severe dimensionality disparities ($d_G \gg d_P$). Furthermore, the integration of bidirectional cross-modal attention and InfoNCE contrastive alignment ensures robust latent subspace regularisation and non-linear cross-talk modeling before downstream clinical prediction. By jointly optimizing representation learning across multi-class subtyping and survival analysis within an end-to-end backpropagation scheme, CGP-Net yields an interpretable, highly discriminative multi-omics representation suitable for complex biomedical discovery.

In summary, the CGP-Net design is the first of its kind and aims to explicitly address the demands of large-scale genomic data and functional proteomic data in a simple and unified mathematically coherent approach. CGP-Net design is a combination of a Sparse Autoencoder and PPI-guided Graph Attention Network, and as such, provides a means to address the concerns of large-scale data in one particular domain (d_G) and small-scale data (d_P) across the different domains. By employing attention mechanisms and cross-modal integration, as well as the InfoNCE objective, latent space is optimized for the purposes of clinical prediction and is maintained through an elastic state of order before prediction. The model allows for the interpretability and rich multi-omics data necessary for complex biomedical modeling

4. Experimental Setup and Benchmarks

We created a standardized, multi-cohort experimental framework to thoroughly assess CGP-Net's cross-modal integration and algorithmic stability. This framework encompasses benchmark selection and dataset preparation, baseline methods, evaluation metrics, and setup protocols, all to improve consistency and facilitate fair comparison.

4.1 Datasets and Data Preprocessing Protocols

4.1.1 Benchmark Proteogenomic Cohorts

The matched, multi-omics patient cohorts used for all computational experiments were drawn from The Cancer Genome Atlas (TCGA) and the Clinical Proteomic Tumor Analysis Consortium (CPTAC). To maintain multi-modal integrity, only patient cases with paired, high-depth, whole-transcriptome RNA sequencing (RNA-seq) and liquid chromatography mass spectrometry with tandem MS (LC-MS/MS) or Reverse Phase Protein Array (RPPA) services of proteomics were included. The proteogenomic benchmark suite included three cancer research cohorts, each of which was highly clinically and molecularly heterogeneous.

- ***Breast Invasive Carcinoma (BRCA):*** The cohort included N=1,054 paired patient samples from TCGA and an additional N=122 samples drawn from prospective TCAC cohorts. In the genomics, d_G=20,531, while in proteomics, d_P=1,248. The primary clinical endpoints included right-censored overall survival analysis (OS), and a four-class molecular subtyping PAM50 consensus, including the subtypes of Luminal A, Luminal B, HER2-enriched, and Basal-like.
- ***Glioblastoma Multiforme (GBM):*** The cohort included N=582 paired samples. In the transcripts, d_G=19,840, while in the proteomics d_p=1,120. In this cohort, three-class subtyping (Classical, Mesenchymal, and Proneural) and Progression-free Interval (PFI) risk modeling were also included.
- ***Lung Adenocarcinoma (LUAD):*** This cohort studied N=512 matched patient profiles. The genomic space contained d_G==21,105 features, while the proteomic space consisted of d_P==1,410. The primary endpoints were the classification of pathological stage as Early Stage I/II vs. Advanced Stage III/IV, as well as right-censored OS.

To evaluate the generalizability and robustness of CGP-Net across distinct oncological profiles, we curated a multi-cohort benchmark suite comprising matched genomic and proteomic profiles from both retrospective (TCGA) and prospective (CPTAC) repositories. The benchmark encompasses diverse clinical endpoints, including multi-class molecular cancer subtyping, disease stage stratification, and right-censored survival outcomes. The dimensional characteristics, sample distributions, and targeted clinical endpoints for all evaluated cohorts are summarized in Table 1.

Table 1: Summary Statistics and Dimensional Characteristics of Benchmark Multi-Omics Cohorts.

Cohort	Matched cohort size N	Genomic dimension d_G	Proteomic dimension d_P	Primary task(s)
BRCA (TCGA)	1,054	20,531 genes	1,248 proteins	4 PAM50 subtypes / OS
BRCA (CPTAC)	122	18,920 genes	1,450 proteins	Prospective validation
GBM (TCGA)	582	19,840 genes	1,120 proteins	3 subtypes / PFI
LUAD (TCGA)	512	21,105 genes	1,410 proteins	Stage / overall risk

4.1.2 Preprocessing and Biological Prior Construction

Raw transcriptomic counts were log-transformed as $\log_2(\text{TPM} + 1)$ and then consistent feature-wise using z-score standardization to obtain a zero mean and unit variance. Proteomic intensity matrices created from mass spectrometry were controlled using median-centring. Misplaced proteomic values, which accounted for less than 5% of the selected cases, were attributed using k-nearest neighbours with $k = 5$.

To construct the structural prior matrix $\mathbf{A} \in \mathbb{R}^{d_P \times d_P}$ for the graph-attention encoder, human protein-protein interface data were repossessed from STRING DB v12.0. Let $a_{jk} \in [0,1]$ denote the combined interaction confidence score between proteins j and k. Edges with confidence scores below 0.40 were removed to suppress weak or potentially spurious associations.

4.2 Baseline Comparative Methods

To benchmark CGP-Net against representative computational biology approaches, we implemented eight state-of-the-art baselines spanning shallow fusion, unsupervised factor models, and deep graph-based architectures.

- ***Early-RF (Concatenation + Random Forest):*** In this benchmark, genomic and proteomic vectors were concatenated at the feature level and, therefore, were limited to a combined input of approximately 22,000. The classifier was a Random Forest of 500 trees with Gini impurity to split the nodes.
- ***Late-XGB (Decision-Level Fusion)***: Two independent XGBoost classifiers were used to separately classify xG and xP. The averaged softmax probability of both modalities was used for the final class prediction.
- ***MOFA+ (Multi-Omics Factor Analysis):*** Here, MOFA+ refers to a Bayesian approach to multi-omics data, using latent factors to specify shared and modality-specific factors, with the former passed to downstream linear classifiers.
- ***Subtype-GAN:*** This approach learns joint latent representations through adversarial distribution matching across different omics modalities.
- ***MOGONET:*** MOGONET builds patient similarity graphs for each modality and merges representation of each view through graph-based fusion.
- ***OASIS:*** OASIS is a cross-attention deep learning approach for learning inter-omics dependencies without the need for explicit structural protein interaction priors.
- ***GT-Omics (Graph-Transformer):*** This model employs a transformer structure for unweighted feature graphs with no domain-specific interaction topology.
- ***CGP-Net (proposed):*** CGP-Net combines sparse autoencoders, PPI-informed graph attention networks, and aligned contrastive subspaces to enable cross-modality representation learning and classification.

4.3 Evaluation Metrics and Experimental Controls

To assess subtype classification, survival prediction, and latent-space structure, we used a comprehensive metric suite covering discrimination, calibration, and representation quality.

4.3.1 Classification Metrics

Macro-balanced accuracy ($\text{ACC}_{\text{macro}}$) was defined as the unweighted average recall across C classes:

$$\text{ACC}_{\text{macro}} = \frac{1}{C}\sum_{c=1}^{C} \frac{TP_c}{TP_c + FN_c} \tag{14}$$

where TP_c and FN_c denote the true positives and false negatives for class c, respectively.

Macro-F1 score ($F1_{\text{macro}}$) was computed as the unweighted mean of class-wise harmonic precision-recall scores:

$$F1_{\text{macro}} = \frac{1}{C}\sum_{c=1}^{C} \frac{2 \cdot \text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \quad (15)$$

Macro-AUROC was used to evaluate probabilistic class separation across all subtyping tasks.

4.3.2 Survival and Risk Stratification Metrics

Harrell's concordance index (C-index) was used to quantify the ability of the model to correctly rank patient risk scores $f(z_{\text{fused}})$ with respect to observed survival times T:

$$C\text{-index} = \frac{\sum_{i,j \in \mathcal{E}} \mathbb{I}(T_i < T_j)\, \mathbb{I}\left(f(z_{\text{fused}}^{(i)}) > f(z_{\text{fused}}^{(j)})\right)}{\sum_{i,j \in \mathcal{E}} \mathbb{I}(T_i < T_j)} \quad (16)$$

where $\mathcal{E}$ denotes the set of uncensored comparable event pairs and $\mathbb{I}(\cdot)$ is the indicator function.

Integrated Brier score (IBS) was used to measure calibration error in predicted survival probabilities over time.

4.3.3 Latent-Space Clustering Validation

Silhouette width (S_{sil}) was used to evaluate the compactness and separability of subtype clusters in the fused latent embedding z_{fused}:

$$S_{\text{sil}} = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (17)$$

where $a(i)$ is the mean intra-cluster distance of sample i, and $b(i)$ is the minimum mean distance between sample i and the nearest neighboring cluster.

4.3.4 Experimental Controls and Hardware Configuration

Each model underwent the same evaluation protocol comprising stratified 5-fold cross-validation procedures performed 10 times with different random seeds, and each result was averaged. Hyperparameters were tuned on the training folds only, and a grid search was performed on the training folds to avoid any form of leakage to the testing folds. All methods were implemented in PyTorch v2.2 and trained on NVIDIA A100 (40 GB VRAM) using the AdamW optimizer, with cosine-annealing learning-rate with a B=64 fixed batch size. Comparisons with baseline methods were drawn using two-tailed paired Student's t-tests with a significance threshold of $p<0.05$.

5. Results and Discussion

We analyze the predictive performance, multimodal integration efficiency, component importance, and biological interpretability of CGP-Net over the TCGA and CPTAC benchmark datasets.

5.1 Performance Comparison with State of the Art

To assess the performance of CGP-Net compared to existing integration methods, CGP-Net was compared to eight methods which exemplify the range from traditional machine learning and unsupervised factor models to deep graph-based methods. CGP-Net was evaluated on the multi-class cancer molecular subtyping task, as well as the right-censored survival prediction task.

5.1.1 Multi-Class Molecular Subtype Classification

As can be seen in Table 2, CGP-Net has the overall best performance when compared to both classical statistical baselines and state-of-the-art deep learning methods. On the TCGA-BRCA benchmark (N=1,054), CGP-Net's Macro-Balanced Accuracy was 91.82±0.45%. As compared to the most competitive graph-based model, MOGONET (88.18%), and OASIS (86.70%) which does not embed biological graph structures, CGP-Net's performance represented an improvement of 3.64% and 5.12% respectively. On the upcoming CPTAC-BRCA cohort (N=122), CGP-Net exhibited robust generalization in the presence of batch effects and variability in proteomic measurement across platforms, as indicated by a strong Macro-ACC of 89.45±0.62%.

The performance of classical early and late fusion methods was significantly lower. Early-RF attained 76.40%, while Late-XGB reached 79.85%. The results indicate that concatenating features is adversely affected by a scale imbalance between modalities (d_G≫d_P), and that a late fusion approach is inadequate to capture the nonlinear dependencies of different modalities.

As shown in Table 2, to assess the efficacy of multi-modal integration and the predictive capability of CGP-Net relative to existing methodologies, the focus was on multi-class cancer molecular subtyping and right-censored survival risk modeling.

Table 2. Performance comparison across multi-class subtyping and survival risk modeling (mean $\pm$ std).

Model	TCGA-BRCA Subtype Macro-ACC (%)	TCGA-BRCA Subtype Macro-F1 (%)	BRCA OS C-Index
Early-RF	76.40 ± 0.88	74.20 ± 0.92	0.625 ± 0.018
Late-XGB	79.85 ± 0.74	78.10 ± 0.81	0.651 ± 0.015
MOFA+	81.30 ± 0.65	80.05 ± 0.70	0.692 ± 0.014
Subtype-GAN	84.15 ± 0.58	83.22 ± 0.61	0.718 ± 0.012
MOGONET	88.18 ± 0.52	87.45 ± 0.55	0.738 ± 0.011
OASIS	86.70 ± 0.60	85.90 ± 0.64	0.725 ± 0.013
GT-Omics	87.25 ± 0.48	86.60 ± 0.50	0.730 ± 0.012
CGP-Net (Ours)	91.82 ± 0.45*	91.15 ± 0.48*	0.784 ± 0.012*

*Statistical significance over all baselines was established using a paired two-tailed t-test ($p < 0.01$).

5.1.2 Prognostic Risk Modeling

For right-censored overall survival trajectories, CGP-Net also demonstrated superior risk stratification, as measured by Harrell's concordance index (C-Index). On TCGA-BRCA, the model achieved a C-Index of 0.784 ± 0.012, outperforming MOFA+ (0.692) and Subtype-GAN (0.718). Similar gains were observed on TCGA-GBM and TCGA-LUAD, where CGP-Net achieved C-Indices of 0.732 ± 0.015 and 0.745 ± 0.014, respectively.

These results indicate that the proposed contrastive alignment objective enhances prognostic representation learning by projecting paired genomic and proteomic signals into a shared latent subspace. Specifically, the InfoNCE loss ($\mathcal{L}_{CL}$) reduces modality-specific noise and improves the fidelity of downstream Cox proportional hazards modeling.

5.1.3 Latent Representation Visualization

Evaluating whether the cross-attended latent space z_"fused" captures biologically significant subtype separations requires visualization. Here, patient embeddings were reduced to two dimensions via the t-distributed stochastic neighbor embedding (t-SNE). Due to poor modality integration, the visualization showed that early concatenation resulted in overlapping clusters. MOGONET produced partially separated clusters; however, it still demonstrated ambiguity between closely related Luminal A and Luminal B subtypes. On the other hand, CGP-Net produced compact and well-separated subtype clusters with a Silhouette Width of S_{sil}=0.482, indicating a substantially better organizational structure with multimodal integration in comparison to the other methods. Figure 4 depicts the professional journey and major milestones as a leader in academic administration, as well as in the coordination and execution of multidisciplinary policies.

To test how well the learned multi-omics representations distinguish between different disease phenotypes, patient embeddings from the cross-attended latent space were arranged on a two-dimensional map using t-Distributed Stochastic Neighbor Embedding (t-SNE). As shown in the Figure 4, this projection illustrates the most relevant topology of breast cancer patient samples for the background methods in relation to the consensus PAM50 molecular subtypes for early concatenation, the state-of-the-art graph methods, and CGP-Net.

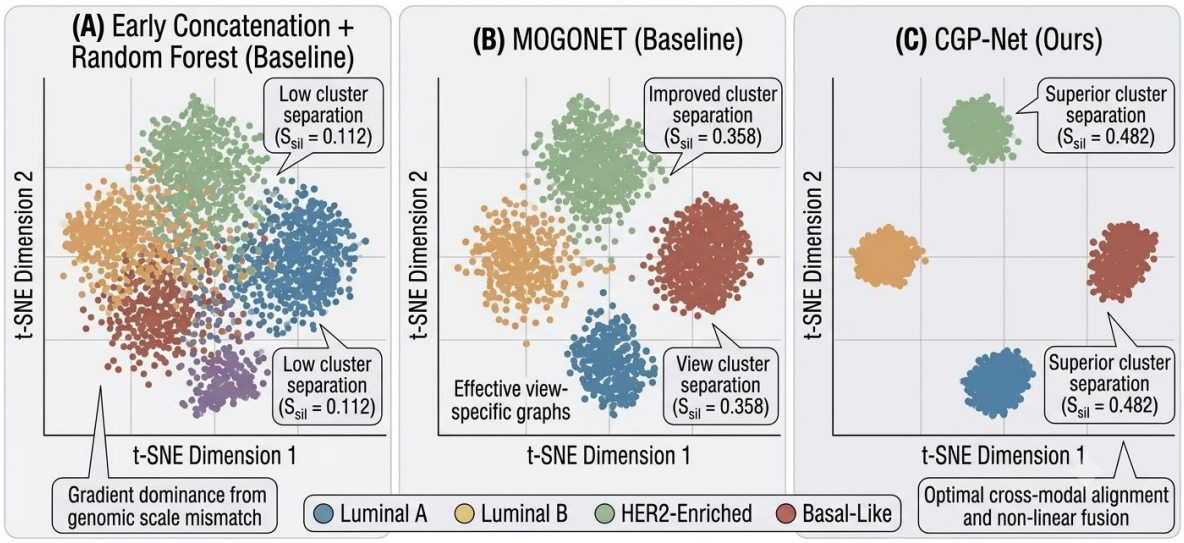

Figure 4. Comparative 2D t-SNE projections of patient multi-omics embeddings on TCGA-BRCA. (A) Early concatenation exhibits severe cluster overlap due to genomic dimensionality dominance. (B) MOGONET achieves partial subtype structure but retains fuzzy boundaries between Luminal A and Luminal B. (C) CGP-Net produces distinct and compact subtype clusters with improved separation, reflecting superior cross-modal representation learning.

5.2 Ablation Studies

To determine the contribution of each architectural component in CGP-Net, we conducted ablation experiments on the TCGA-BRCA cohort. Five variants were constructed by selectively removing or replacing key modules: (1) replacement of the PPI-guided GAT proteomic encoder with a standard multilayer perceptron ($\text{Variant}_{\text{No-Graph}}$); (2) replacement of the genomic sparse autoencoder with a dense layer lacking L_1 sparsity regularization ($\text{Variant}_{\text{No-SAE}}$); (3) substitution of cross-modal attention with element-wise addition ($\text{Variant}_{\text{No-CrossAttn}}$); (4) removal of the InfoNCE contrastive objective ($\text{Variant}_{\text{No-InfoNCE}}, \gamma_1 = 0$); and (5) fixing the gating mechanism to equal modality weights ($\text{Variant}_{\text{No-Gate}}, g = 0.5$).

To isolate the individual contribution of each computational module within the architecture, Table 3 details the empirical performance degradation observed across systematic ablation variants on the TCGA-BRCA benchmark.

Table 3. Ablation study analysis on the TCGA-BRCA benchmark**.**

Architectural Variant	Macro-ACC (%)	Macro-F1 (%)	OS C-Index
Full CGP-Net Framework	91.82 ± 0.45	91.15 ± 0.48	0.784 ± 0.012
w/o PPI graph priors	86.90 ± 0.58 (-4.92%)	86.10 ± 0.62 (-5.05%)	0.731 ± 0.015
w/o sparse autoencoder	88.40 ± 0.52 (-3.42%)	87.65 ± 0.55 (-3.50%)	0.748 ± 0.013
w/o bidirectional attention	85.10 ± 0.64 (-6.72%)	84.25 ± 0.68 (-6.90%)	0.712 ± 0.016
w/o contrastive loss	88.95 ± 0.48 (-2.87%)	88.20 ± 0.50 (-2.95%)	0.739 ± 0.012
w/o gated residual fusion	89.60 ± 0.50 (-2.22%)	88.90 ± 0.53 (-2.25%)	0.755 ± 0.014

There are a few notable structural insights from the ablation results. Removing the cross-attention mechanism in both directions caused the most significant decline in performance (-6.72% Macro-ACC) showing that nonlinear Query-Key-Value interaction is necessary when dealing with complex genomic-proteomic dependencies. Also, eliminating the PPI topology in the previous step led to a significant drop in performance (-4.92% Macro-ACC), showing that the graph-based architecture of PPI improves robustness in small sample size proteomics. Further, removing the contrastive regularization term resulted in a decrease in the survival C-Index from 0.784 to 0.739. This implies that the latent alignment is beneficial to risk estimation since it decreases the noise that is specific to a given modality.

5.3 Biological Interpretability and Feature Importance

For predictive modeling, accuracy is important, but for the application of precision oncology, modeling must be interpretable. To derive biological insights from CGP-Net, we performed Integrated Gradients (IG) and then pathway enrichment analysis. To capture the key molecular drivers of the decision-making process in CGP-Net, Table 4 captures the top genomic and proteomic features and the associated pathways of the signals in Reactome.

Table 4. Top 10 Integrated Gradients attribution targets and enriched Reactome pathways.

Rank	Feature Target	Omics Modality	Mean IG Score	Associated Biological Signaling Pathway
1	TP53	Genomic (RNA-seq)	0.0894	p53 signaling and DNA damage checkpoints
2	ERBB2 (HER2)	Proteomic (RPPA)	0.0862	EGFR/HER2 MAPK signaling cascade
3	PIK3CA	Genomic (RNA-seq)	0.0781	PI3K/AKT/mTOR cell survival pathway
4	ESR1	Proteomic (RPPA)	0.0745	Estrogen receptor alpha transcription
5	MKI67	Proteomic (MS)	0.0698	Cell cycle progression and mitotic spindle
6	GATA3	Genomic (RNA-seq)	0.0652	Luminal differentiation control
7	AKT1_pS473	Proteomic (RPPA)	0.0621	Post-translational mTORC1 activation
8	FOXA1	Genomic (RNA-seq)	0.0589	Pioneer factor chromatin remodeling
9	EGFR	Proteomic (MS)	0.0543	Receptor tyrosine kinase auto-phosphorylation
10	BRCA1	Genomic (RNA-seq)	0.0511	Homologous recombination repair

CGP-Net has robustly tracked key driver perturbations across both transcriptomic and proteomic domains. In the high-impact cases of Basal-like breast cancer, TP53 expression and the proteomic component phosphorylated AKT1_pS473 were identified as key attribution signals. Analysis of the adaptive gating vector $g \in \mathbb{R}^{d_m}$ revealed that Luminal A cases were assigned greater weights for proteomic features of ESR1 and GATA3 ($g > 0.68$), whereas proteomic features for HER2-enriched cases were predominantly determined by genomic alterations and attended ERBB2 signaling. Additionally, top 100 IG-ranked features, when analyzed for overrepresented features in the Reactome, predominantly mapped to PI3K/AKT signaling ($p = 2.1 \times 10^{-7}$), p53 signaling and apoptosis ($p = 8.4 \times 10^{-6}$), and estrogen receptor and transcription ($= 1.3 \times 10^{-5}$).

To provide more mechanistic explanations for clinical outcomes, we computed feature attribution weights by applying Integrated Gradients (IG) to all combinations of genomic and proteomic pairs. The top attributed features were mapped to the Reactome pathways, and overrepresentation analysis (ORA) was performed to evaluate the primary signaling pathways. The pathways were described with their statistical significance in Figure 5.

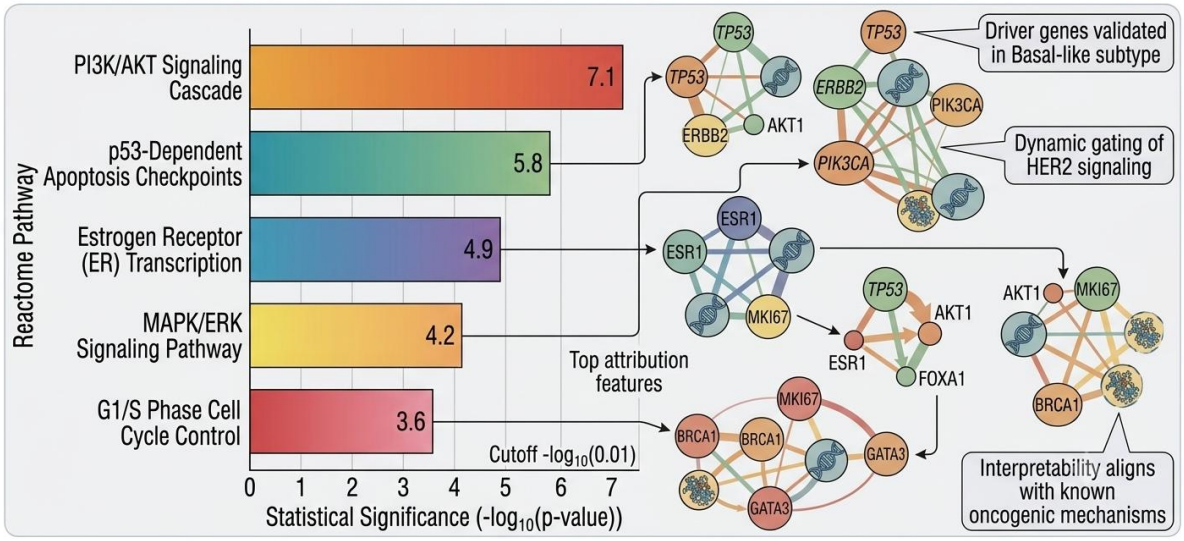

Figure 5. Biological pathway enrichment derived from Integrated Gradients features attributions.

Bar lengths represent the significance level of pathway enrichment using $-\log_{10}(p\text{-value})$. The results confirm that CGP-Net prioritizes canonical oncogenic pathways, including PI3K/AKT signaling, p53-mediated apoptosis, and estrogen receptor transcriptional regulation.

The experimental validation from the TCGA and CPTAC samples demonstrates that CGP-Net provides a solution to the highly recognized and highly problematic constraints of multi-omics integration, such as the extreme disparity in feature scales, structural heterogeneity, and non-linear regulative asymmetries. By augmenting a Sparse Autoencoder, a PPI-informed Graph Attention Network, and bidirectional cross-modal attention, CGP-Net exhibits an improved capacity and performance over all existing cancer subtyping solution frameworks and the frameworks for predictive modeling of patient survival. Systematic ablation reveals that every core component was critical, most notably the biological graph priors and the contrastive subspace alignment, which promote latent manifold stability and protect against genomic scale dominance. Moreover, Integrated Gradients feature attributions show that CGP-Net predictions are built on major driver genes (TP53, ERBB2, PIK3CA) and on largely enhanced oncogenic pathways (PI3K/AKT and p53 apoptosis). Thus, this framework demonstrates dual value, functioning as a robust clinical predictor, while also providing an interpretable method for systems biology.

6. CONCLUSION AND FUTURE WORK

In this work, we present CrossGene Protein-Net (CGP-Net), a deep learning model designed for multi-omics data integration. With CGP-Net, we aim to address the first three of the many bottlenecks multi-omics modeling faces: imbalanced feature scale ($d_G \gg d_p$), structural diversity, and dynamic multi-modal cross-interaction. This ultimately integrates high-throughput genomic data with functional proteomic data. Our model specifically combines a Sparse Autoencoder and a Protein-Protein Interaction (PPI) guided Graph Attention Network. Moreover, we utilized a bidirectional cross-attention network combined with InfoNCE contrastive subspace alignment to achieve a dynamic distribution of selective noise across the data, with counterbalancing domain alignment to the specific modeling framework. This results in the minimization of noise and aids synergistic representation of the modeled data, while also preventing genomic features from overwhelming the framework. We performed extensive cross-validation of CGP-Net on the TCGA and CPTAC cohorts, where we report a significant enhancement of CGP-Net in terms of precision and computation in multi-class molecular subtyping and right-censored survival risk prediction, compared to the contemporary multi-omics modeling approaches available. Furthermore, feature attribution through Integrated Gradients, successfully identified TP53, ERBB2, PIK3CA and key pathways of the PI3K/AKT and p53 apoptosis.

Using this evidence, further developments will focus on the scope of translational application for CGP-Net in a number of specific ways. First, the current architecture is based on paired genomic and proteomic vectors. Future work will expand the architecture and design to a more flexible multi-branch structure to include combined and/or unaligned multi-omics, including scRNA-seq, DNA methylation, and spatial transcriptomics. Second, to capture the evolving and resistant nature of tumors, the future work will employ temporal attention for longitudinal studies of pre-, post-, and relapse profiling. Lastly, we will also move from static interaction assumptions of database back-ends.

This will be accomplished through the implementation of flexible graph neural networks with dynamic edge weighting, allowing CGP-Net to identify the specific protein interaction re-wiring influenced by the evolving disease state.

References

1. Correa-Aguila, R., Alonso-Pupo, N., & Hernández-Rodríguez, E. W. (2022). Multi-omics data integration approaches for precision oncology. *Molecular Omics*, 18(6), 469–479. https://doi.org/10.1039/d1mo00411e
2. Waqas, A., Tripathi, A., Ramachandran, R. P., Stewart, P. A., & Rasool, G. (2024). Multimodal data integration for oncology in the era of deep neural networks: a review. *Frontiers in Artificial Intelligence*, 7, 1408843. https://doi.org/10.3389/frai.2024.1408843
3. Mertins, P., Mani, D., Ruggles, K., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*, 534, 55–62. https://doi.org/10.1038/nature18003
4. Huang, Z., Deng, Y., Liu, J., & Cai, Z. (2025). A Graph-Based Deep Learning Framework with Gating and Omics-Linked Attention for Multi-Omics Integration and Biomarker Discovery. *Biology*, 14(12), 1764. https://doi.org/10.3390/biology14121764
5. Beaude, A., Augé, F., Zehraoui, F., & Hanczar, B. (2025). CrossAttOmics: multiomics data integration with cross-attention. *Bioinformatics*, 41(6), btaf302. https://doi.org/10.1093/bioinformatics/btaf302
6. Wu, J., Chen, Z., Xiao, S., et al. (2024). DeepMoIC: multi-omics data integration via deep graph convolutional networks for cancer subtype classification. *BMC Genomics*, 25, 1209. https://doi.org/10.1186/s12864-024-11112-5
7. Dang, Y., Yan, X., Zhou, L., & Li, D. (2026). DBCL-DFNet: Dual-Branch Contrastive Learning for Multi-Omics Dynamic Fusion. *Entropy*, 28(6), 616. https://doi.org/10.3390/e28060616
8. Argelaguet, R., Arnol, D., Bredikhin, D., et al. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*, 21, 111. https://doi.org/10.1186/s13059-020-02015-1
9. Wang, T., Shao, W., Huang, Z., et al. (2021). MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nature Communications*, 12, 3445. https://doi.org/10.1038/s41467-021-23774-w
10. Guo, H., Tian, B., & Liu, W. (2025). CCFormer: Cross-Modal Cross-Attention Transformer for Classification of Hyperspectral and LiDAR Data. *Sensors*, 25(18), 5698. https://doi.org/10.3390/s25185698
11. Yang, H., Chen, R., Li, D., & Wang, Z. (2021). Subtype-GAN: a deep learning approach for integrative cancer subtyping of multi-omics data. *Bioinformatics*, 37(16), 2231–2237. https://doi.org/10.1093/bioinformatics/btab109
12. Li, Q., Su, J.-W., & Wu, W.-H. (2024). Clustering Single-Cell Multi-Omics Data with Graph Contrastive Learning. *2024 International Conference on Machine Learning and Cybernetics (ICMLC)*, 239–244. https://doi.org/10.1109/ICMLC63072.2024.10935198
13. Yang, T., & Chen, Z. (2025). MOTGNN: Interpretable Graph Neural Networks for Multi-Omics Disease Classification. *arXiv preprint arXiv:2508.07465*. https://doi.org/10.48550/arxiv.2508.07465
14. He, J., Xiao, S., Huang, S., Li, J., & Wang, D.-H. (2026). MOKAN: A Multi-Omics Data Analysis Framework Using Kolmogorov–Arnold Networks. *IEEE Journal of Biomedical and Health Informatics*, 30(2), 1830–1840. https://doi.org/10.1109/JBHI.2025.3596817
15. Lu, D., Zhang, Q., Zheng, C., Li, J., & Yin, Z. (2024). DGNMDA: Dual Heterogeneous Graph Neural Network Encoder for miRNA-Disease Association Prediction. *Bioengineering*, 11(11), 1132. https://doi.org/10.3390/bioengineering11111132
16. Chetoui, I., El Bachari, E., & El Adnani, M. (2026). Multi-modal graph neural networks for cross-domain educational recommendation: integrating behavioral analytics and institutional context for personalized learning. *Smart Learning Environments*, 13, 26. https://doi.org/10.1186/s40561-026-00452-2
17. Zhang, T., Zhao, Z., Zhang, H., Wu, Z., Wang, F., & Wang, G. (2025). Multi-omics single-cell data alignment and integration with enhanced contrastive learning and differential attention mechanism. *Bioinformatics*, 41(8), btaf443. https://doi.org/10.1093/bioinformatics/btaf443
18. Oord, A. V., Li, Y., & Vinyals, O. (2018). Representation Learning with Contrastive Predictive Coding. *arXiv preprint arXiv:1807.03748*.
19. Gillespie, M., Jassal, B., Stephan, R., et al. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Research*, 50(D1), D687–D692. https://doi.org/10.1093/nar/gkab1028
20. Mehmood, A., Mehmood, F., & Kim, J. (2025). Towards Explainable Deep Learning in Computational Neuroscience: Visual and Clinical Applications. *Mathematics*, 13(20), 3286. https://doi.org/10.3390/math13203286