

Review article

Pathway-guided architectures for interpretable AI in biological research

Qi Zhou^a, Naga Sekhar Madala^b, Chen Huang^{a,c,*} ^a Department of Genetics, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL 35233, USA^b Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL 35233, USA^c O'Neal Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL 35233, USA

ARTICLE INFO

Keywords:

Knowledge-guided deep learning

Interpretable AI

Pathway databases

Multiomics data

ABSTRACT

Understanding the dysregulation of complex biological pathways is essential for uncovering molecular mechanisms and identifying novel therapeutic opportunities for complex diseases. In recent years, deep learning (DL) models have shown great potential in modeling biological multi-omics data; however, their “black box” nature limits their application in biological and clinical translation. Knowledge-guided deep learning, particularly methods based on Pathway-Guided Interpretable Deep Learning Architectures (PGI-DLA), aims to improve model performance and interpretability by integrating prior pathway knowledge into the model structure. Here, we review the current progress in PGI-DLA, focusing on omics compatibility, architectural design, feature interpretation, and biological and clinical applications. For widely used pathway databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Reactome, and MSigDB, we summarize their differences in knowledge scope, hierarchical structure, level of detail and curation focus. We discuss how the choice of database impacts model design, performance, and interpretability. This review provides valuable guidance for selecting and optimizing pathway databases to implement PGI-DLA to translate omics data into actionable biological and clinical insights.

1. Introduction

Complex diseases like cancer exhibit significant molecular complexity, with genetic abnormalities, dysregulated signaling pathways, and intricate molecular interactions [1]. To understand the underlying biological mechanisms for various disease phenotypes, traditional analytical methods, such as pathway enrichment, have been widely used to explain the dysregulation of specific functional modules in disease states. However, with the rapid increase in the scale and dimensionality of omics data, traditional analytical methods face limitations in performing multi-source data integration and global modeling

due to their limited capabilities in processing complex data modalities and handling nonlinear data relationships within the data.

In recent years, deep learning (DL) has played an increasingly vital role in analyzing large-scale biological data [2,3]. Unlike traditional machine learning (ML), DL automatically extracts key features from raw data, reducing reliance on subjective manual processes and minimizing information loss from human bias (Fig. 1A). It also effectively captures complex nonlinear dynamics in biological systems, enabling integration of diverse molecular profiles, such as genes, proteins, and metabolites. However, DL models, with their multilayer neural networks and numerous parameters, lack a transparent causal chain [4]. Despite

Abbreviations: DL, Deep learning; PGI-DLA, Pathway-Guided Interpretable Deep Learning Architectures; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; AI, Artificial intelligence; GSEA, Gene set enrichment analysis; ML, Machine learning; GNN, Graph neural network; CNVs, Copy number variations; SNPs, Single nucleotide polymorphisms; PPIs, Protein-protein interactions; ScRNA-seq, Single-cell RNA-seq; NGS, Next-generation sequencing; MS, Mass spectrometry; TMT, Tandem Mass Tag; KNN, K-nearest neighbors; TAM, Transcriptional activity matrix; TF, Transcription factor; CCRES, cis-chromatin regulatory elements; LLM, Large language model; ViT, Vision Transformer; BP, Biological processes; MF, Molecular functions; CC, Cellular components; DAG, Directed acyclic graph; ORA, Over-representation analysis; SPIA, Signaling Pathway Impact Analysis; TPEA, Topology-based Pathway Enrichment Analysis; TopologyGSA, Topology-based Gene Set Analysis; DNN, Deep Neural Network; VNN, Visible Neural Network; CNN, Convolutional neural network; MLP, Multilayer perceptron; HGNN, Hypergraph neural network; IG, Integrated Gradients; LRP, Layer-wise Relevance Propagation; RLIPP, Relative Local Improvement in Predictive Power; C-index, Concordance Index; RMSE, Root Mean Squared Error; MAE, Mean Absolute Error; MSE, Mean Squared Error; R², Coefficient of Determination; AUPR, Area Under the Precision–Recall Curve.

* Correspondence to: University of Alabama at Birmingham, 210D 705 20th St S, Birmingham, AL 35233-2031, USA.

E-mail address: huangc@uab.edu (C. Huang).

<https://doi.org/10.1016/j.csbj.2025.10.063>

Received 23 August 2025; Received in revised form 29 October 2025; Accepted 30 October 2025

Available online 1 November 2025

2001-0370/© 2025 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

demonstrating superior performance in predicting biological phenotypes or clinical outcomes, their 'black box' nature hinders identification of key molecules or signaling pathways driving decisions [5].

To enhance the potential of DL in biological applications, the Pathway-Guided Interpretable Deep Learning Architectures (PGI-DLA)

have been proposed to improve interpretability while retaining DL's robust capacity to process multi-omics data. This approach directly integrates biological pathway knowledge, such as that from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [6], Gene Ontology (GO) [7], Reactome [8], and MSigDB [9], into model design. The network

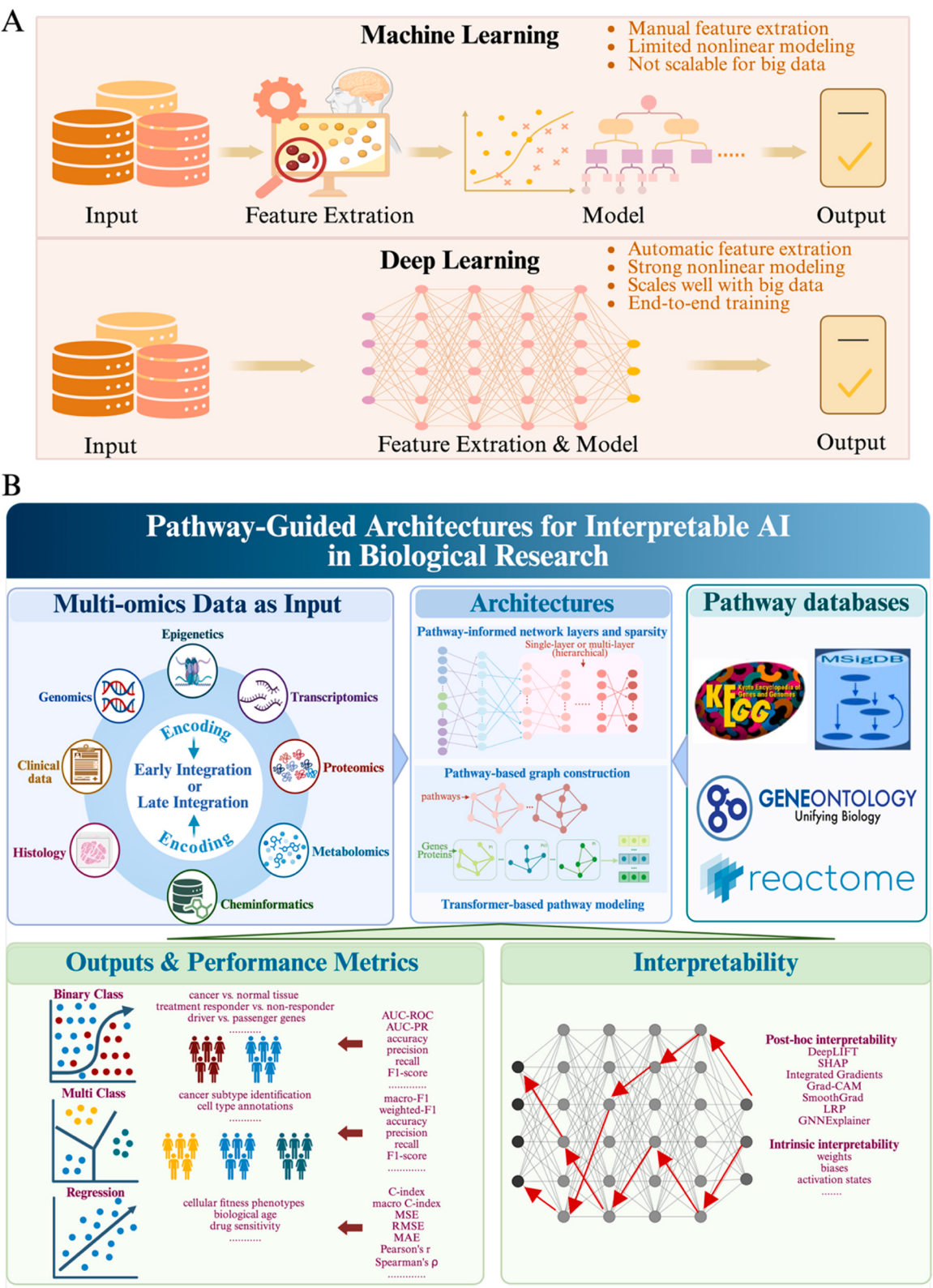


Fig. 1. Schematic workflow of PGI-DLA in biological research. (A) Comparison of machine learning and deep learning paradigms. (B) End-to-end PGI-DLA workflow from data input to biological interpretation.

architecture is based on known biological interaction relationships, ensuring intrinsic consistency between the model's decision-making logic and biological mechanisms. Unlike traditional approaches that use pathways solely for input feature preprocessing (e.g., gene set enrichment analysis (GSEA) to quantify pathway activity) [10], this biologically informed DL architecture embeds domain knowledge into the model structure to guide the learning process by mimicking the flow of biological information. This design not only enables biological priors to guide predictions but also provides interpretable knowledge units for feature interpretation and experimental or clinical follow-up.

Since the emergence of the first notable model, DCell [11], numerous innovative PGI-DLA designs, optimizations, and applications have followed (Fig. 1B, Table 1, Supplementary Table 1). Compared to previous reviews focusing primarily on PGI-DLA model architectures [12–15], this review aims to provide an end-to-end framework of the modeling process, emphasizing model evaluation and applications from the perspective of model users. Specifically, this review covers: (1) systematic encoding and integration of diverse omics data types (genomics, epigenomics, transcriptomics, proteomics, metabolomics); (2) the first in-depth comparison of pathway databases (KEGG, GO, Reactome, MSigDB) as architectural blueprints, analyzing how their structural and annotation differences fundamentally shape model design, performance, and interpretability — a critical dimension overlooked in prior reviews; (3) a comprehensive classification of PGI-DLA architectural paradigms and their design principles; and (4) a systematic evaluation of task-specific outputs, performance metrics, and interpretability methods. We particularly highlight PGI-DLA's application in cancer research, where the complex molecular landscape demonstrates its utility, though its principles and applications extend broadly to other complex diseases.

2. Multi-omics data as input for PGI-DLA

PGI-DLA models can process large-scale multimodal data as input to construct the molecular map of disease, including genomic (e.g., mutations, copy number variations (CNVs), and SNPs), epigenetic (e.g., DNA methylation), transcriptomic, proteomic, and metabolomic data. These models can also integrate other high throughput data for specific predictive or modeling purposes, such as chemoinformatic data (e.g., drug chemical fingerprints), clinical data with physiological context (e.g., patient survival), and digital pathology histology capturing tissue spatial heterogeneity (e.g., whole-slide images). Each omics data type is processed in specific ways to feed into the model. (Fig. 2, Table 1)

Genomic variations with binary or categorical measurements are commonly encoded into integers. For instance, mutations are encoded as '0/1' [11,16,17], allele copy numbers are encoded as '0/1/2' [18], and CNVs are encoded as integers ranging from -2 to 2 [19,20]. These encodings can be concatenated to represent multiple types of genomic variation [21–25]. For continuous data (epigenomics, transcriptomics, proteomics, and metabolomics), the standard workflow involves data cleaning, normalization, and transformation. For example, after normalization, DNA methylation levels are represented by beta values ranging from 0 (unmethylated) to 1 (completely methylated) [26,27].

Depending on specific high-throughput techniques, customized preprocessing might be applied to the data to achieve optimal model performance. For example, while bulk transcriptomic data are commonly normalized to control the sequencing depth, single-cell RNA-seq (scRNA-seq) data undergo more rigorous quality control, normalization, data transformation, and variable gene selection. Additionally, as exemplified by scGO [28], scRNA-seq data require dedicated batch-effect correction when integrating different datasets. Beyond standardized preprocessing steps, some models may construct more complex input features tailored to specific analytical objectives. For example, the TSGPN model [29] takes in differentially expressed mRNA, miRNA, and lncRNA as input to identify regulatory mechanisms among different types of RNA molecules.

Besides next-generation sequencing (NGS), molecular profiling by other techniques is increasingly available. Mass spectrometry (MS)-based proteomics data, for instance, have been widely used to elucidate molecular mechanisms less explored by genomics and transcriptomics [30]. Compared to NGS data, MS proteomic data require more preprocessing due to variable data formats. In cancer research, large-cohort MS proteomics data have been generated by various MS techniques, such as iTRAQ- or Tandem Mass Tag- (TMT) labeled and label-free MS, each with distinct data distribution and linearity [31]. Moreover, due to lower sensitivity, MS proteomics data often contain missing values, which need to be either zero-filled (e.g., BINN [32]) or imputed using methods like k-nearest neighbors imputation [33] to meet neural network input requirements. Similarly, metabolomics data, another understudied omics data type, face data processing challenges in missing values and heterogeneity of MS platforms [34]. Preprocessing of these high-throughput omics data types requires further improvement to maximize their utility in PGI-DLA models.

3. Data integration as input for PGI-DLA

The complexity of disease biology stems from abnormalities across multiple molecular levels. To build a comprehensive disease molecular profile, PGI-DLA models often integrate omics data using early integration and late integration strategies.

In early integration, feature vectors from different omics are concatenated at the input layer to form an extended vector at the gene or sample level. For example, GPC-Net [35] combines mutation, copy number alteration, and DNA methylation data; c-Diadem [36], BioVNN [37], and MPVNN [38] integrate gene expression and mutation data; MULGONET [39], DeepOmix [40], MCDHGN [41], PGLCN [42] and Pathformer [43] concatenate gene expression, mutation, and DNA methylation data. Besides direct concatenation, early integration can use more customized combining strategies. For instance, regX [44], designed to predict drivers of cell state transitions, constructs a transcriptional activity matrix (TAM) as input by integrating transcription factor (TF) expression, accessible chromatin regions (cCREs), and TF-cCRE interactions. For each gene, activation signals from all these resources are weighted and summed to generate TAM entries. Early integration is straightforward and can capture low-level interactions but may struggle with heterogeneous data, as it is sensitive to missing values and can be dominated by modalities with high feature dimensions.

In late integration, independent network branches process and extract features from each omics type before combining them through concatenation or weighted fusion [45]. For example, the c-Triadem [46] model integrates SNP data and microarray-based expression data by designing independent input layers to process each type of data before concatenating them. ViLoN [47] constructs similarity maps from gene expression, mutation, and DNA methylation data, then uses variant information distance for standardization and weighted fusion to construct a unified weight map for the model input. Late integration improves the model's adaptability and stability when handling incomplete overlap or missing data across omics datasets. However, the independent branch structure increases model complexity, resulting in higher parameter counts, computational costs, and training time.

Omics data are frequently integrated with clinical data to predict patient outcomes. Depending on the feature type, clinical data can be standardized as numerical values (e.g., z-scored age) or encoded as binary or one-hot variables (e.g., sex, tumor stage). These features are concatenated with omics data at various layers. For instance, VNNSurv [48] encodes clinical variables such as age and treatment for input-layer integration, while Cox-PASNet [49], PathGNN [50], c-Triadem [46], Cox-Path [51], and Multilevel-GNN [52], integrate encoded clinical features at the final hidden layer. Missing clinical data are handled by removing incomplete samples (e.g., PathGNN [50]) or imputing values using KNN (e.g., in c-Triadem [46]). In addition to tabular clinical data, clinical text data can also serve as input. For example, Lee, et al. [53]

Table 1

Representative PGI-DLA tools across input types, pathway knowledge bases, model architectures, and interpretability strategies.

Input	Pathway database	Models	Interpretability	Tools
Genomics	GO	VNN	RLIPP	DCell [11]
	KEGG	Sparse DNNs	DeepLIFT	KP-NET [21]
	KEGG, PathwayCommons	Sparse DNNs	Integrated Gradients	IntNet [16]
	KEGG	VNN	Intrinsic Interpretability	GenNet [18]
	KEGG	GNN	Intrinsic Interpretability	GraphPath [22]
	MSigDB	GNN	Intrinsic Interpretability	DISHyper [17]
Epigenomics	Reactome	Sparse DNNs	DeepLIFT	P-NET [23], GCS-Net [19], PR-NET [120]
	Reactome	Sparse DNNs	Intrinsic Interpretability	BiGMLVQ [24]
	Reactome	GNN	DeepLIFT	IBPGNET [20], PNETwa-MLP [25]
	KEGG	Transformer	Intrinsic Interpretability	PathMethy [26]
	Reactome	VNN	DeepLIFT	XAI-AGE [27]
	GO	VNN	DeepLIFT	scGO [28], Liu Y [83]
Transcriptomics	GO(BP)	VNN	Intrinsic Interpretability	GraphGONet [88], KDDSL [87]
	KEGG	Sparse DNNs	SHAP	SPIN [89]
	KEGG	Sparse DNNs	LRP	XModNN [85]
	KEGG	Sparse DNNs	Intrinsic Interpretability	PathExpSurv [84], SigPrimedNet [78]
	KEGG	GNN	Intrinsic Interpretability	IRnet [90], Lee, S. [121]
	MSigDB (Hallmark)	Sparse DNNs	Intrinsic Interpretability	Holzschek, N. [122]
Proteomics	MSigDB (Reactome)	Sparse VAE	Intrinsic Interpretability	pmVAE [123]
	MSigDB (Reactome)	Sparse DNNs	Intrinsic Interpretability	PathDeep [124], PASNet [109]
	MSigDB (Reactome)	GNN	Intrinsic Interpretability	TSGPN [29]
	MSigDB, Reactome	GNN	GNNExplainer, Grad-CAM	Cell Decoder [93]
	MSigDB, Reactome	Sparse VAE	Intrinsic Interpretability	VEGA [112]
	MSigDB, Reactome	Sparse VAE	Integrated Gradients	PAUSE [113]
Metabolomics	Reactome	Sparse DNNs	Intrinsic Interpretability	d-sclGM [125], CellTICS [126]
	Reactome	Sparse CNNs	Integrated Gradients	ReGeNNe [127]
	Reactome	Sparse DNNs	SHAP	BDKANN [128]
	Reactome	GNN	Integrated Gradients	Burkhart JG [91]
	MSigDB (Reactome)	Sparse DNNs	Intrinsic Interpretability	DeepBINN [110]
	MSigDB (Reactome)	Sparse DNNs	DeepSHAP	BINN [32]
Chemoinformatics	KEGG	Sparse DNNs	SHAP	PiDeeL [79]
	Reactome	VNN	LRP	DTox [58]
	Reactome	VNN	Intrinsic Interpretability	AIDTox [57]
	KEGG	Sparse DNNs	DeepLIFT	GPC-Net [35]
	MSigDB (KEGG)	Sparse DNNs	SHAP	c-Diadem [36]
	Reactome	VNN	Intrinsic Interpretability	BioVNN [37]
Genomics & Epigenomics	KEGG	VNN	Intrinsic Interpretability	MPVNN [38]
	GO(BP)	VNN	RLIPP	DrugCell [54]
	GO	VNN	Intrinsic Interpretability	ParsVNN [55]
	Reactome	VNN, GNN	Intrinsic Interpretability	XMR [59]
	Reactome	VNN	SHAP	VNNSurv [48]
	GO	GNN	Intrinsic Interpretability	HHAN-DSI [129]
Genomics & Transcriptomics	GO	GNN	Intrinsic Interpretability	regX [44]
	KEGG	VNN	Intrinsic Interpretability	van Hilten, A.[130]
	GO	VNN	Intrinsic Interpretability	CancerIDP [131]
	KEGG	GNN, Transformer	Grad-CAM	DRPreter [92]
	MSigDB	CNNs	Integrated Gradients	PathSynergy [60]
	Reactome	VNN, GCN	SHAP, Integrated Gradients	DrugVNN [61]
Genomics & Clinical data	MSigDB (KEGG, Reactome)	Sparse DNNs	Intrinsic Interpretability	Cox-PASNet [49]
	Reactome	GNN	Integrated Gradients	PathGNN [50]
	Reactome, MSigDB (Hallmark)	Transformer	Integrated Gradients	SURVPATH [66]
	MSigDB (Hallmark)	Transformer	Intrinsic Interpretability	MMP [64]
	KEGG	Sparse DNNs	Intrinsic Interpretability	PathDNN [80]
	GO	Sparse DNNs	Integrated Gradients	MULGONET [39]
Epigenomics & Transcriptomics	KEGG	GNN	Intrinsic Interpretability	GGNN [132]
	MSigDB (KEGG, Reactome)	Sparse DNNs	Intrinsic Interpretability	DeepOmix [40]
	MSigDB(GO, KEGG)	Sparse multi-layer network	Intrinsic Interpretability	ViLoN [47]
	MSigDB	GNN	Intrinsic Interpretability	MCDHGN [41]
	Reactome	GNN	GNNExplainer	PGLCN [42]
	KEGG, Reactome, BioCarta, PID	Transformer	SHAP	Pathformer [43]
Transcriptomics & Chemoinformatics	Reactome	Sparse DNNs	Integrated Gradients	PONET [65]
	KEGG	Sparse DNNs	SmoothGrad	DeepSigSurvNet [81]
	Reactome	Sparse DNNs	SHAP	c-Triadem [46]
	Reactome	Sparse DNNs	DeepLIFT	BioXNet [133]
	KEGG	Sparse DNNs	SmoothGrad	consDeepSignaling [82]
	Reactome	Sparse DNNs	LRP	xNNDriver [68]
Transcriptomics & Clinical data	KEGG	GNN	Intrinsic Interpretability	Cox-Path [51]
	Reactome	Transformer	LRP, SHAP	DeePathNet [95]
	KEGG	GNN	Intrinsic Interpretability	
	Reactome	Transformer	LRP, SHAP	
	KEGG	GNN	Intrinsic Interpretability	
	Reactome	Transformer	LRP, SHAP	

(continued on next page)

Table 1 (continued)

Input	Pathway database	Models	Interpretability	Tools
Genomics & Transcriptomics & Chemoinformatics	GO	VNN	DeepLIFT	MOViDA [86]
Genomics & Epigenomics & Transcriptomics & Clinical data	KEGG	GNN	GNNExplainer, Integrated Gradients	Multilevel-GNN [52]
Epigenomics & Transcriptomics & Proteomics & Clinical data	Reactome	GNN	Intrinsic Interpretability	Lee, C. [53]

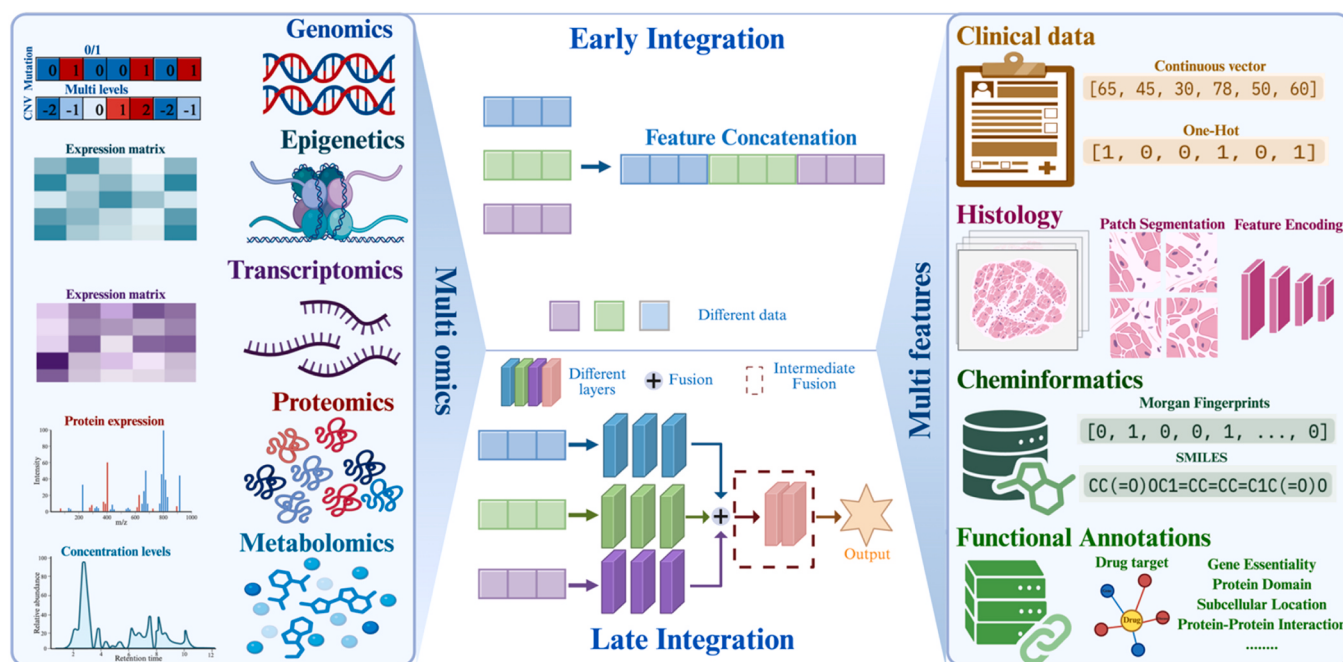


Fig. 2. Input data types and integration strategies for PGI-DLA. The left and right sections depict the diverse data modalities accepted by the model. The central framework compares two multimodal integration strategies: early integration and late integration.

processed the clinical records of Parkinson's disease patients using a large language model (LLM) to generate dense clinical phenotype embeddings.

To predict drug response or enable drug repurposing, omics data are commonly integrated with drug compound descriptors, such as molecular weight, chemical bond types, functional groups, topological structure, and physicochemical properties. These compound features are digitized for computational analysis. For example, DrugCell [54] and ParsVNN [55] convert chemical structures into fixed-length vectors such as Morgan fingerprints [56]. To integrate compounds with target molecules, AIDTox [57] uses a binary chemical-gene matrix derived from a knowledge graph, while DTTox [58] converts chemical fingerprints into target-binding probabilities. Similar to late integration, independent branches can be designed to process drug and gene-related features separately before fusing them. For example, models like XMR [59], PathSynergy [60], and DrugVNN [61] utilize graph neural networks (GNNs) to learn compounds' graphical representations and generate chemical embeddings that can be fused with gene embeddings.

Omics data can also be combined with histological images through late integration. In the histological branch, images are divided into segments using sliding windows or grids, and these segments are encoded into high-dimensional vectors using visual encoders like Vision Transformer (ViT) [62] or Swin Transformer [63]. ViT captures global tissue features (e.g., MMP [64], PONET [65]), while Swin Transformer builds local-to-global context (e.g., SURVPATH [66]). Segment-level features are combined via concatenation or attention mechanisms to generate slide-level representations.

Finally, high-throughput perturbations and knowledge graphs can be

integrated into the model as input to guide its operation. For example, DepMap's CRISPR screening data [67], which quantify the importance of each gene for cell survival, can be integrated with mutation data to predict cancer driver genes, as demonstrated in xNNDriver [68].

In summary, PGI-DLA models can integrate multiomics, clinical, drug, histological, and knowledge graph data, through feature concatenation (early integration) or independent network branches (late integration). This enables PGI-DLA models to learn biological processes at various molecular levels and be tailored to handle different prediction tasks.

4. Annotated Pathways as the Blueprint for Model Design

Biological pathways are ordered sets of gene products and small molecules that drive specific biological processes. Pathway databases describe and organize diverse pathways, serving as blueprints for PGI-DLA model network structures. Due to differences in design principles and historical factors, pathway annotations and component genes vary significantly across these databases [69], which must be considered for database selection to optimize the performance and interpretability of PGI-DLA models.

Our survey highlights four widely used public databases: Reactome, KEGG, MSigDB, and GO (Tables 1 and 2). KEGG links genes and small molecules to diseases and presents molecular interaction networks through expert-curated pathway diagrams. The KEGG pathway database is organized in a three-layer structure (core categories, functional subcategories, and specific pathways), with 367 human pathways in its current release. Reactome depicts complex biological pathways in a

Table 2
Systematic comparison of the four major pathway databases.

	KEGG	GO	Reactome	MSigDB
Core Purpose	Manually drawn pathways showing molecular interactions	Structured vocabulary for gene and gene product function (BP, MF, CC)	Reaction-based process maps	Collections of gene sets for functional genomics
Structure	3-layer system (BRITE)	DAG structure	Hierarchically organized from reactions to pathways to systemic processes	Grouped sets and subsets
Source	Expert curated	Literature-based curation combined with inferring methods	Expert-written and peer-reviewed	Varies by collection; mixed sources
Update Frequency	Major release quarterly; annotation daily	Ontology monthly; annotation daily	Quarterly release	Irregular, typically annual
Human Data (2025)	Release 114, 367 pathways, 20,535 proteins, and 4150 RNAs.	Release 2025–03–16, 20,580 genes - MF: 11,243 (603 terms), BP: 12,434 (2120 terms), CC: 12,918 (487 terms)	Release 92, 2769 pathways, 11,356 proteins	v2025.1.Hs, 35,134 gene sets, nine within main collections and subsets

highly structured biological reaction network centered on reactions. Each reaction specifies molecular participants, biochemical events, and hierarchical relationships, enabling detailed analysis of specific steps and causal relationships in disease biology. The current version of Reactome includes 2769 human pathways. GO provides hierarchical ontology terms describing gene functions in biological processes (BP), molecular functions (MF), and cellular components (CC). Its directed acyclic graph (DAG) structure is well-suited for hierarchical clustering of genes by function, with over 3000 human terms in the current version. MSigDB serves as an integration hub for gene sets, collecting pathway information from the aforementioned databases and functional gene sets derived from experiments, such as perturbations mimicking disease or biological phenotypes.

Before the advent of DL models, architectural differences across pathway databases were known to impact their application in pathway analysis. For instance, while pathway annotations from these databases can be used for over-representation analysis (ORA) or GSEA, treating all component genes as equivalent, the detailed pathway structures in KEGG and Reactome enable more informative topology-based analysis. Tools like SPIA [70], TPEA [71], and TopologyGSA [72] consider gene positions and interaction directions (i.e., inhibition or activation) within pathways when inferring pathway activity. Conversely, GO’s DAG structure suits functional similarity analysis, enabling quantification of gene functional closeness to predict protein interactions and functional modules. GO’s hierarchy also supports advanced algorithms (e.g., parent-child, Elim) to reduce redundancy, thereby enhancing the accuracy and interpretability of enrichment analyses [73].

Beyond architecture, variations in component genes significantly influence database selection for PGI-DLA models. For example, in the apoptosis pathway, only 17 genes (~10 %) overlap across KEGG (hsa04210, 137 genes), Reactome (R-HSA-109581, 168 genes), GO (GO:0006915, 1075 genes), and MSigDB (HALLMARK_APOPTOSIS, 161 genes) (Fig. 3). These 17 genes are involved in core apoptosis execution, such as caspase cascades and death receptor signaling. KEGG includes unique genes involved in signal transduction and inflammation (e.g.,

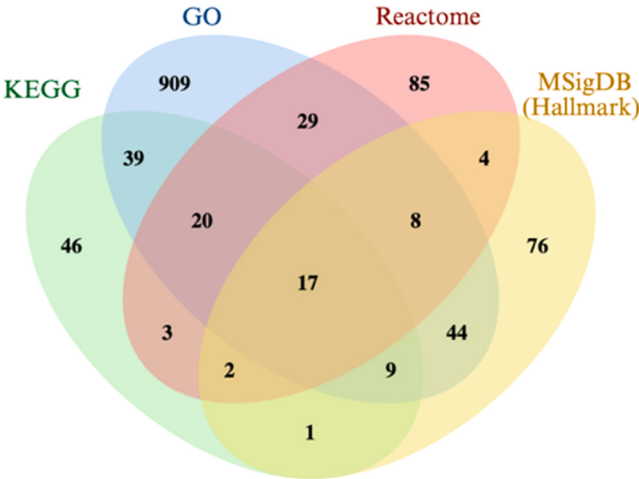


Fig. 3. Venn diagram comparing apoptosis pathway genes across pathway databases.

PARP3, *CASP12*) but misses non-caspase genes like *GSDMD* (present in Reactome). Reactome covers inflammatory and degradation genes (e.g., *TLR4*, *PSMD8*) but omits signal transduction genes like *AKT3* (present in KEGG). GO includes significantly more genes than other databases, with broad coverage of stress responses and non-caspase-dependent apoptosis genes (e.g., *HSP90AA1*, *AIFM3*); however, its extensive scope may reduce specificity. MSigDB’s HALLMARK gene set encompasses diverse biological events involved in apoptosis, including oxidative stress (e.g., *HMOX1*), immune inflammation (e.g., *IL6*), and transcriptional regulation (e.g., *JUN*, *RELA*). However, due to redundancy removal, this gene set lacks some key apoptosis-relevant genes (e.g., *MAPK10* annotated in KEGG), potentially limiting comprehensive apoptosis modeling.

Pathway databases are managed in distinct ways. KEGG and Reactome rely on expert manual curation from literature and regulatory documents to ensure high accuracy. GO combines manual curation with computational predictions, enabling broader coverage and more frequent updates. MSigDB updates are based on aggregated source contents; however, some contents, such as C2:CP:KEGG_LEGACY subset, cannot be updated due to copyright licensing restrictions, risking outdated pathway information.

Although less commonly used, other notable pathway resources include PathwayCommons [74], NCI-PID [75], and BioCarta [76]. PathwayCommons [74] integrates pathway and interaction data from multiple sources, while NCI-PID [75] and BioCarta [76] provide earlier pathway resources with numerous signaling pathway diagrams. Additionally, specialized pathway datasets, such as LCPathways that was curated from cancer research literature [77], can be leveraged for cancer-related models.

In summary, researchers developing PGI-DLA models must evaluate database structure, update frequency, data quality, and copyright restrictions to select the most suitable database for their specific research tasks.

5. Core PGI-DLA architectural paradigms guided by pathways

The core principle of PGI-DLA models is to enhance predictive performance and model interpretability by leveraging prior biological knowledge, such as the pathway structures in KEGG and Reactome, the functional DAG of GO, and refined gene sets of MSigDB. Their architectures fall into three categories: pathway-informed sparse network architectures, pathway-based graph construction, and Transformer-based pathway modeling (Fig. 4, Tables 1 and S1).

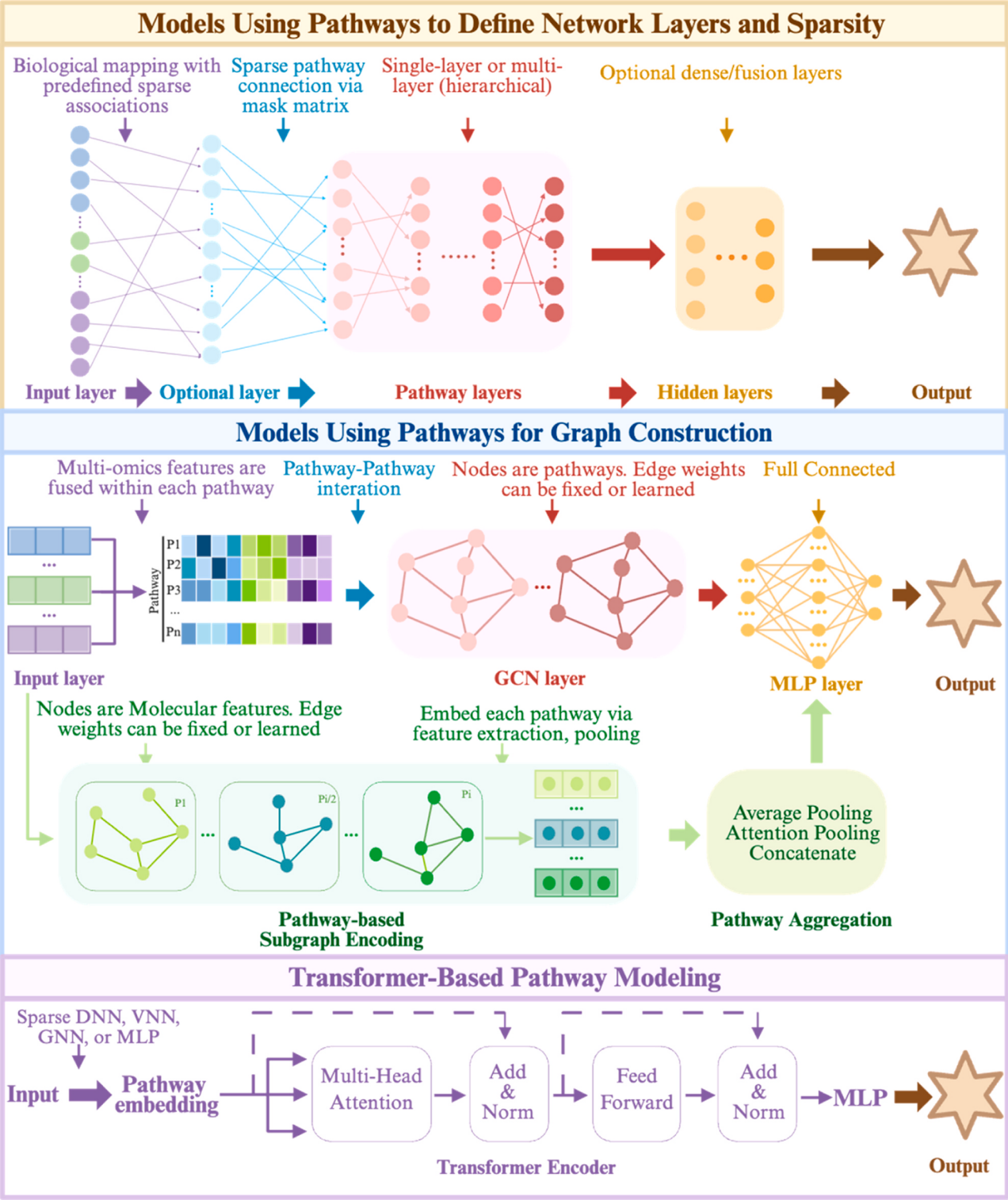


Fig. 4. Three major architectures for PGI-DLA models. Top panel: Models using pathways to define network layers and sparsity, consisting of an input layer, an optional sparse mapping layer, one or more pathway layers, optional hidden or fusion layers, and an output layer. Central panel: Models using pathways for graph construction, either building a pathway-level graph or a subgraph for each pathway. Bottom panel: Transformer-based pathway modeling, capturing self-attention relationships between pathways using a Transformer encoder.

5.1. Pathway-informed Sparse Network Architectures

These models use pathway databases to create sparse neural networks, ensuring information flows along known pathways for improved interpretability. Based on the complexity of the network and the specific

gene-pathway relationships being focused on, they can be further divided into single-layer, hierarchical, and complex embedding models (Fig. 4, Tables 1 and S1).

5.1.1. Single-layer pathway embedding models

This strategy constructs a single-layer sparse pathway hidden layer, with each node corresponding to a specific pathway. Connections between input features (e.g., genes or metabolites) and pathway nodes are typically implemented via a binary mask matrix, denoting whether the feature is included in the pathway. This pathway layer is then fed into one or more fully connected layers to learn interactions among pathways. Tools like SigPrimedNet [78], PiDeeL [79], and GPC-Net [35], PathDNN [80], DeepSigSurvNet [81], consDeepSignaling [82], and Multilevel-GNN [52] (using KEGG) and scGO [28] and Liu et al.'s model [83] (using GO) apply this approach to tasks like cancer grading, survival prediction, disease diagnosis, drug sensitivity, and cell annotation.

In addition to static pathway embedding, dynamic and optimization-adjusted models also enable adaptive network refinement during training. For instance, PathExpSurv [84] employs a two-stage training strategy: first, it initializes gene-pathway connections using KEGG pathways; then, it introduces a fully connected structure, where newly added gene-pathway links are pruned via L1 regularization. Through iterative training, high-confidence supplementary connections are identified to expand pathways for survival risk prediction. In another study, IntNet [16] selects genes from KEGG and PathwayCommons and performs random walk with restart on HumanNet, an orthogonally defined gene relationship database, to compute functional relevance scores. In this way, both curated and data-driven connections contribute to each pathway's sparse mask matrix.

5.1.2. Hierarchical pathway embedding models

These models leverage the hierarchical structure of databases to design multi-layer networks, where genes are sequentially aggregated into lower-level and higher-level pathways. In KEGG, the BRITE framework organizes pathways into a hierarchical structure according to their functional similarity. Based on this reference, XModNN [85] adopts a hierarchical modular network comprising 362 functional pathway modules, 46 system modules, and 6 top-level functional modules. Each module is designed as an independently trainable subnetwork, with a multi-loss hierarchical optimization to extract features. Similarly, the GO's DAG, with hierarchical "is a" and "part of" relationships, is inherently convertible to layer connections and has been widely applied to various prediction tasks like drug response (DrugCell [54], MOViDA [86]), synthetic lethality (e.g., KDDSL [87]), cell growth phenotypes (e.g., DCell [11]), and cancer type prediction (GraphGONet [88]). Additionally, customized architectures can be applied to reduce redundancy in GO terms. For instance, ParsVNN [55] constructs a complete hierarchical network based on GO terms, then applies L0 norm and group Lasso regularization to prune irrelevant connections, retaining only the most predictive gene–neuron and subsystem–neuron links.

5.1.3. Complex embedding models

Complex embedding models integrate multiple subnetworks, each using single-layer or hierarchical pathway embeddings to process specific data modalities independently. For example, SPIN [89] uses KEGG to create separate single-layer pathway embeddings for data from male and female patients, capturing sex-specific differences in survival predictions. Similarly, c-Triadem [46], a PGI-DLA model for Alzheimer's disease, leverages Reactome's hierarchical structure to construct two parallel subnetworks that independently process genotype and gene expression data. Complex embedding can also utilize different GO categories. MULGONET [39] builds independent GO hierarchical networks using BP and MF branches to process multiomics data and then fuses outputs from each branch for prediction. This approach maps feature importance to both molecular function and biological relevance and enables more informative model interpretation.

5.2. Pathway-based graph construction

Beyond encoding pathway knowledge into sequential neural layers,

another type of PGI-DLA model conceptualizes biological pathways as dynamic, computationally tractable graphs to capture intricate interactions within and across pathways. Based on varying modeling approaches, these architectures can be categorized into pathway-level GNNs and gene-level GNNs.

In pathway-level GNNs, each node represents a biological pathway, and edges connect pathways with specific similarities or relationships, such as shared component genes or known interactions between component genes. Since pathway annotations are readily available in major pathway databases, this strategy has been widely adopted by models such as GraphPath (prostate cancer metastasis) [22], IRnet (immunotherapeutic response) [90], Burkhart et al.'s model (disease biochemistry) [91], regX (cell state probabilities) [44], PGLCN (tumor mutation burden) [42], and Cox-Path (survival risk) [51]. During training, graph attention or convolution is applied to update node representations, capturing complex inter-pathway influences. To derive a global representation of the pathway graph, these models use global attention pooling, feature concatenation, or averaging, with the resulting representation fed into a multi-layer perceptron (MLP) for final predictions.

In gene-level GNNs, each pathway is represented as an independent subgraph, with genes as nodes and edges reflecting their biological interactions. For instance, PathGNN [50] creates a graph for each Reactome pathway and uses graph convolutional neural network to embed pathways. These embeddings are concatenated with clinical features and fed into a fully connected neural network for survival prediction. Similarly, DRPreter [92] embeds KEGG pathways, combining them with drug feature embeddings to predict drug response. More complex graph structures can be employed to infer intricate pathway relationships. MCDHGN utilizes diverse gene set annotations in MSigDB to build a heterogeneous GNN for predicting cancer driver genes, with nodes for genes, pathways, and gene sets, and edges representing gene-to-gene relationships and gene-to-gene-set mappings [41]. To learn multi-scale cell representations, Cell Decoder [93] leverages prior knowledge from MSigDB and Reactome to construct subgraphs, each with specific node types (e.g., genes or pathways) and edges (e.g., PPI for gene GNNs or hierarchical relationships for pathway GNNs). During training, Cell Decoder performs intra-layer propagation within the same subgraph and inter-layer aggregation across different subgraphs to enable stratified information integration. In another study, DISHyper [17] adapts a hypergraph neural network (HGNN) to model prior gene sets. Unlike traditional edges that connect pairs of nodes, a hyperedge in an HGNN can link multiple nodes, allowing a single hyperedge to represent an entire gene set. This property enables DISHyper to capture higher-order gene associations and reveal neighborhood information that conventional GNNs may overlook.

5.3. Transformer-based pathway modeling

Originally designed for natural language processing, the Transformer architecture models long-range dependencies through self-attention, which does not assume a fixed input order and directly computes association strengths between any pair of data elements [94]. This architecture has been applied to pathway modeling, where a biological pathway is treated as a "sentence" and its component genes or sub-pathways as "words", thereby enabling the modeling of complex biological pathway interactions. For example, DeePathNet [95] employs a two-layer stacked Transformer encoder to model dependencies among cancer pathways. The first layer applies self-attention to integrate global pathway interactions, capturing direct interaction patterns. The second layer models higher-order regulatory relationships, such as multi-pathway synergy, indirect dependencies, and cross-pathway inhibition. To incorporate additional biological priors, PathMethy [26] introduces a CrossTalk Transformer encoder, which introduces a pathway interaction matrix as an attention bias to the traditional multi-head self-attention mechanism. This matrix captures coordinated

methylation patterns across samples and guides the model to focus on pathway pairs with potential synergy.

Transformer-based architectures are also used for multimodal data integration. For example, MMP [64] and SURVPATH [66] combine histological image and pathway embeddings as Transformer inputs for survival prediction. Large-scale Transformer embedding can impose significant computational burdens that need to be mitigated. SURVPATH reduces complexity by eliminating resource-intensive interactions within image segments and preserving only interactions between pathways and between pathways and image segments. In comparison, MMP performs unsupervised clustering to group image patches into a small number of core patterns before computing attention. Pathformer [43] integrates multiomics data with pathway annotations. In its core input matrix, rows represent biological pathways, and columns correspond to features derived from different omics layers. This structure allows Pathformer to employ a unique “criss-cross attention” mechanism that analyzes pathway crosstalk through column-wise attention and intrinsic regulatory patterns among omics features through row-wise attention.

5.4. Comparative analysis of architectural paradigms

Sparse DNNs/VNNs excel in parameter efficiency (Table 3, Supplementary Table 1). For example, P-NET achieves an AUC of 0.93 for predicting prostate cancer state with only 71,000 parameters, a 3,800-fold reduction compared to 270 million in fully connected networks [23]. GraphGONet matches dense MLP accuracy (AUC ≈ 0.94–0.95) with just 23,900 parameters [88]; while PONET, using only 1.6 % of the parameters of a similar tool, Pathomic Fusion [96], achieves significant improvements in cancer diagnosis [65]. While these models show superior performance in small-sample regimes, few have been applied to large-cohort studies (e.g., n > 1000), raising concerns about their scalability given current model settings. Moreover, although pathway-based input processing and information flow is instructive for biological and

clinical translation, this architectural rigidity might limit complex non-linear learning, compromising their applications on large datasets.

Compared to DNNs/VNNs, pathway-driven GNNs usually have a higher parameter count, thus requiring more samples for training. For instance, the heterogeneous network MCDHGN, which has over two million edges [41], utilizes over 6000 samples for training. Given sufficient training data, GNN-based models can achieve high performance. GraphPath, a model with a similar prediction task to P-NET, achieves an AUC of 0.933, representing a 3–5 % improvement over the latter [22]. IRnet achieves AUCs of 0.6–0.95 for predicting immunotherapeutic response, with 3–25 % improvements over baseline ML models [90]. Although the GNN has been frequently used for modeling biological and clinical questions, the network-based architecture is less straightforward for clinical adoption, which requires independent gene products that can be isolated as biomarkers.

Transformer models demonstrate robust scalability and high performance on large-scale multi-omics datasets (Table 3, Supplementary Table 1). This is due to the nature of Transformers, where the order of input and the network structures (e.g., the number of encoder/decoder blocks) can be flexible. For instance, DeePathNet processes and integrates multiomics data from large-cohort cancer studies (e.g., TCGA and CPTAC) using a pathway encoder to predict cancer drug response; each data modality has over 20,000 features. It significantly outperforms the baseline random forest model in drug response prediction, achieving accuracy > 0.95 across 23 cancer types [95]. In another study, Transformer-based Pathformer [43] outperforms 18 other models for predicting cancer prognosis and drug sensitivity, including ML approaches and sparse DNN/VNN-based models. Despite their superior performance, Transformer models are less straightforward for clinical or mechanistic translation, as the pathway interactions are not intuitive from the architecture alone and must rely on external interpretation tools, such as SHAP (see below).

6. Interpretability analysis of pathway-guided models

For PGI-DLA models, biologically meaningful interpretation is as important as predictive accuracy. This requirement has driven the development of diverse model interpretation techniques, categorized as post-hoc and intrinsic methods (Table 1).

6.1. Post-Hoc interpretability techniques

Post-hoc methods analyze a model’s behavior after training, using external tools to trace how input features or intermediate nodes contribute to predictions. These methods offer high flexibility and are applicable to complex models like GNNs and Transformers. They enable quantitative analyses of feature importance under the context of either local or global interpretations, without modifying the original model architecture.

DeepLIFT [97] is widely used for tracking feature importance. It establishes a baseline activation (e.g., a neutral input) for each neuron, calculates the difference in activation when transitioning to the actual input, and apportions this difference to input features via a backpropagation-like process. Each feature receives a “contribution score” reflecting its positive or negative influence on the output relative to the baseline. While computationally efficient, DeepLIFT’s reliance on a single baseline may limit its flexibility for heterogeneous datasets.

Another popular interpretation algorithm, SHAP [98], establishes the baseline reference as the average prediction across the dataset. It then derives the marginal contribution of each feature by evaluating all possible combinations of feature subsets. SHAP provides a more thorough evaluation of feature importance based on these combinations, albeit at a higher computational cost.

Gradient-based algorithms also enable post-hoc interpretation. Integrated Gradients (IG) [99] calculates feature contributions by accumulating gradients along a path from a baseline (e.g., a mutation-free

Table 3
Performance Benchmarks Across PGI-DLA Architectural Paradigms.

Models	Task Type	Performance Range
Sparse DNNs/VNN	Classification	AUC: 0.662–1 AUPRC: 0.702 – 0.98 Accuracy: 0.65–0.9928 F1: 0.398–0.966 Macro F1-score: 0.9406 – 0.9993 Precision: 0.81 Recall: 0.78–0.83
	Regression	C-index: 0.601–0.99 MAE: 0.685 MSE: 0.0142 R ² : 0.885 Spearman correlation (ρ): 0.80–0.89 Pearson’s r: 0.85–0.97
GNN	Classification	ARI: 0.7909 AUC: 0.6–0.994 AUPR: 0.79–0.894 Accuracy: 0.82–0.88 F1: 0.4505–0.94 Macro F1 score: 0.81
	Regression	MSE: 0.8251 ± 0.0122, MAE: 0.6682 ± 0.0047 Pearson’s r: 0.9467 ± 0.0013 Spearman correlation (ρ): 0.9248 ± 0.0014 C-index: 0.59–0.82
Transformer	Classification	Accuracy > 0.9 F1: 0.93–1.00 Precision: 0.99–1.00 Recall: 0.86–1.00
	Regression	C-index of 0.629–0.665 MSE: 0.8251 ± 0.0122 MAE: 0.6682 ± 0.0047 Pearson’s r: 0.9467 ± 0.0013 Spearman correlation (ρ): 0.9248 ± 0.0014

state) to the actual input (e.g., a patient's gene mutation profile). Grad-CAM [100], originally designed for interpreting CNN models, has been adapted by DRPreter [92] to identify key genes associated with drug response and by Cell Decoder [93] to identify gene signatures and pathways for distinguishing different cell types. To enhance the stability and clarity of gradient-based interpretations, SmoothGrad [101] adds random noise to the input multiple times, computes gradients for each noisy version, and averages the results to yield robust attributions. For instance, DeepSigSurvNet [81] uses SmoothGrad to identify pathways (e.g., mTOR and PI3K-Akt) associated with cancer patient survival, while consDeepSignaling [82] employs it to highlight ErbB and Ras signaling pathways in anticancer drug responses.

Layer-wise Relevance Propagation (LRP) [102] traces prediction scores backward from the output layer to the input layer, distributing the final prediction proportionally across layers, neurons, and input features. This layer-specific interpretation is a key strength for interpreting PGI-DLA models where pathways are organized as single hidden layer or hierarchical layers. For instance, DTox [58] applies LRP to interpret its drug toxicity prediction model and links TP53- and MAPK-pathways to specific toxicity mechanisms. XModNN [85] used LRP to pinpoint key pathways, such as MAPK and PI3K, associated with neuroblastoma drug resistance.

Moreover, specialized feature interpretation tools have been developed for GNN PGI-DLA models. One such tool is GNNExplainer [103], which identifies critical connections and local graph patterns driving the model's output. Its core mechanism learns a mask over the graph structure to maximize mutual information between the original GNN prediction and the explanation derived from the masked subgraph. PGLCN [42] uses GNNExplainer to highlight the graph structure essential for predicting gastric cancer mutation burden.

Recent systematic benchmarks have evaluated these methods across multiple dimensions. SHAP was shown to have the highest theoretical rigor and cross-model applicability, albeit with greater computational cost [104–106]; GNNExplainer provides optimal structural interpretation for graph-based pathway models [104,107]; DeepLIFT and IG offer efficient alternatives balancing performance and implementation simplicity [105,108]; LRP demonstrates strong layer-wise traceability, particularly suited for hierarchical pathway networks [106,108]; and SmoothGrad achieves superior faithfulness in gradient-based visualization tasks [104]. These evaluations indicate that method selection should consider model architecture, computational resources, and specific interpretation objectives.

6.2. Intrinsically interpretable architecture

PGI-DLA models with interpretable architectures embed biological knowledge directly into their design, making the decision-making process transparent without relying on external tools. Specifically, all the nodes and layers, as well as their weights, biases, and activation states, are endowed with biological meanings. For example, GraphGONet [88], PASNet [109], DeepBINN [110], and BioVNN [37] map genes to pathways or GO terms and higher-level biological concepts based on pathway databases or knowledge graphs, enabling the importance of network components to be reflected by parameters (e.g., weights or derivatives) from model training. Since the Transformer framework avoids dimensionality reduction, PGI-DLA models using Transformers can map all input features to attention layers for feature importance interpretation. For instance, in TOSICA [111], a Transformer-based cell type annotation tool, biologically relevant genes, pathways, or regulons are not only identified during model training but also prioritized to contribute to the final prediction. Additionally, abstract embedding models like variational autoencoders (VAEs), which traditionally offer little interpretability, have been improved with pathway-guided architectures. In VEGA (VAE Enhanced by Gene Annotations) [112], decoder connections (latent variables to genes) are guided by gene module membership in pathway databases, where the weights can be used to

rank gene importance within a biological module.

6.3. End-User perspectives and experimental validation

End-users engage with PGI-DLA interpretability outputs in distinct ways. For clinicians, sparse visible neural networks offer intuitive interpretation, as each node corresponds to known biological entities (genes or pathways), enabling direct tracing through processes like 'high-risk driven by PI3K/AKT pathway activation'. Moreover, post-hoc methods, such as SHAP and LRP, provide quantitative feature importance and thus offer valuable gene-specific or patient-specific explanations. For biologists, intrinsic pathway constraints reveal activated subsystems for experimental follow-up, whereas post-hoc methods quantify relative importance and prioritize research targets. These approaches are complementary and valuable for mechanistic hypothesis generation. Multiple studies have demonstrated experimental validation of PGI-DLA hypotheses: P-NET identified *MDM4* as a therapeutic target for prostate cancer, which was validated by CRISPR knockout experiments [23]; IBPGNET predicted the oncogenic *PSMC1/PSMD11* roles in lung cancer, with knockdown experiments confirming significantly reduced cell proliferation and invasion, as well as enhanced sensitivity to the EGFR inhibitor Afatinib [20]; PAUSE identified the involvement of mitochondrial Complex I genes (*NDUFS3*, *NDUFA3*, *NDUFS7*, *NDUFA2*) in Alzheimer's disease, with experiments confirming that inhibiting these genes delays A β -induced neurotoxicity [113].

7. Conclusion and future directions

By integrating authoritative pathway databases like KEGG, Reactome, GO, and MSigDB, PGI-DLA models transform traditionally 'black-box' deep learning models into transparent ones, aligning predictions with known biological processes. These models process diverse data types, such as genomic variations, transcriptomic expression, chemical structures, and pathological images, by converting them into uniform formats like matrices or graphs through preprocessing (e.g., normalization, encoding). Effective preprocessing, integration, and architectures for multimodal datasets are critical to leveraging available data.

Looking ahead, novel PGI-DLA models are expected to incorporate emerging omics data types, such as spatial omics, which provide high-resolution insights into intra-tissue heterogeneity—a key driver of disease progression and treatment response [114,115]. Current PGI-DLA models typically assume spatial homogeneity, overlooking the complex molecular and cellular architectures within tissues. Incorporating spatial omics using spatially aware approaches [116] would facilitate modeling of cell-to-cell and region-to-region interactions, thereby enhancing predictive accuracy and biological relevance. During this integration, the models need to address data sparsity in single-cell and spatial omics that arises from technical dropout and limited sequencing depth. Also, modality heterogeneity can cause distinct statistical properties, measurement scales, and batch effects that confound joint analysis. Pathway-guided strategies, such as pathway-aware batch normalization, pathway-constrained imputation respecting gene co-expression patterns, and hierarchical alignment mapping are plausible solutions by leveraging conserved biological relationships to improve integration.

The choice of pathway database is crucial for implementing a PGI-DLA model, as it lays the foundation for the model architecture. KEGG and Reactome, with extensive annotations of interaction networks, are well-suited for multilayer and GNN-based models. GO's DAG organization is translatable to hierarchical model designs. MSigDB's comprehensive resource of gene sets supports heterogeneous pathway integration without requiring topological data. Optimal database selection should align with specific research objectives, considering factors like pathway numbers (which determine parameter counts) and pathway annotation and organization (which influence model architecture). Notably, although pathway annotations and database

structures vary significantly, no comparative study has yet evaluated how database selection impacts performance and interpretation for specific biological or clinical predictions. Moreover, it remains unclear if combining pathways from multiple databases can further enhance model performance or interpretation by leveraging complementary biological priors, as the increased parameter count might also complicate model training. We envision that a deeper benchmarking of pathway selection and integration is essential to inform PGI-DLA model design and optimization.

PGI-DLA models could benefit from improved pathway annotations. One direction is to use dynamic, real-time biological priors, which can be generated by applying natural language processing tools, such as BioBERT [117], to extract gene-pathway interactions from PubMed literature. Despite regular updates, pathway databases can still lag behind new research, and integrating this adaptive prior knowledge would improve predictive accuracy and provide up-to-date mechanistic insights. Additionally, cross-species knowledge transfer offers a promising solution to address data scarcity in understudied diseases [118]. Most biological pathways are evolutionarily conserved, and model organisms (e.g., mice, *Drosophila*) provide abundant mechanistic data. By developing cross-species pathway databases that align homologous human pathways and genes with those of these organisms and employing domain adaptation or transfer learning techniques [119], PGI-DLA models can harness this additional biological information to enhance predictive performance and generate mechanistic hypotheses.

Author statement

During the preparation of this work, the author(s) did not use generative AI for writing the manuscript. A Generative AI tool (Google Gemini) was used solely for proofreading and correcting grammatical errors.

This manuscript is original and has not been published elsewhere.

CRedit authorship contribution statement

Qi Zhou: Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing. **Naga Sekhar Madala:** Investigation, Resources, Writing – original draft. **Chen Huang:** Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Writing – original draft, Writing – review & editing.

Funding

This work is supported by NIH/NIGMS funding 5R35GM154953 (to C.H.).

Clinical trial number

Not applicable.

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

Declaration of Competing Interest

All authors agree to this manuscript submission and declare no conflicts of interest.

Acknowledgments

We thank Araf Mahmud and Yingnan Song for their useful discussion in developing and finalizing this manuscript. We apologize that we were unable to include and cite all related studies owing to manuscript space limitations.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.10.063.

Data availability

This is a review article, and no new datasets were generated. All data discussed in this manuscript are available from cited references.

References

- [1] Almendro V, Marusyk A, Polyak K. Cellular heterogeneity and molecular evolution in cancer. *Annu Rev Pathol Mech Dis* 2013;8(1):277–302.
- [2] Schneider L, et al. Integration of deep learning-based image analysis and genomic data in cancer pathology: a systematic review. *Eur J Cancer* 2022;160:80–91.
- [3] Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. *Ann Transl Med* 2020;8(11):713.
- [4] Ching T, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 2018;15(141):20170387.
- [5] Sinha T, et al. Artificial intelligence and machine learning in predicting the response to immunotherapy in non-small cell lung carcinoma: a systematic review. *Cureus* 2024;16(5).
- [6] Kanehisa M, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;45(D1):D353–61.
- [7] Consortium GO. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;32(1):D258–D261 (p).
- [8] Milacic M, et al. The reactome pathway knowledgebase 2024. *Nucleic Acids Res* 2024;52(D1):D672–8.
- [9] Liberzon A, et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27(12):1739–40.
- [10] Gao F, et al. DeepCC: a novel deep learning-based framework for cancer molecular subtype classification. *Oncogenesis* 2019;8(9):44.
- [11] Ma J, et al. Using deep learning to model the hierarchical structure and function of a cell. *Nat Methods* 2018;15(4):290–8.
- [12] Wysocka M, et al. A systematic review of biologically-informed deep learning models for cancer: fundamental trends for encoding and interpreting oncology data. *BMC Bioinforma* 2023;24(1):198.
- [13] Thapa K, et al. Strategies to include prior knowledge in omics analysis with deep neural networks. *Patterns (N Y)* 2025;6(3):101203.
- [14] Ortigosa ES, Gonçalves T, Nonato LG. Explainable artificial intelligence (xai)—from theory to methods and applications. *IEEE Access* 2024;12:80799–846.
- [15] Nilsson A, Meimetis N, Lauffenburger DA. Towards an interpretable deep learning model of cancer. *NPJ Precis Oncol* 2025;9(1):46.
- [16] Zhou L, Li J, Tan W. IntNet: A novel framework using reconstructed mutation samples and a biologically informed neural network for pathway analysis. 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2024.
- [17] Deng C, et al. Identifying new cancer genes based on the integration of annotated gene sets via hypergraph neural networks. *Bioinformatics* 2024;40(ement_1): i511–20.
- [18] van Hiltten A, et al. GenNet framework. *Commun Biol* 2021;4(1).
- [19] Hu J, et al. A deep neural network for gastric cancer prognosis prediction based on biological information pathways. *J Oncol* 2022;2022(1):2965166.
- [20] Xu Z, et al. IBPGNET: lung adenocarcinoma recurrence prediction based on neural network interpretability. *Brief Bioinforma* 2024;25(3):bbae080.
- [21] Zhang L, et al. Biologically interpretable deep learning to predict response to immunotherapy in advanced melanoma using mutations and copy number variations. *J Immunother* 2023;46(6):221–31.
- [22] Ma T, Wang J. GraphPath: a graph attention model for molecular stratification with interpretability based on the pathway–pathway interaction network. *Bioinformatics* 2024;40(4):btae165.
- [23] Elmarakeby HA, et al. Biologically informed deep neural network for prostate cancer discovery. *Nature* 2021;598(7880):348–52.
- [24] Voigt J. Biologically-informed shallow classification learning integrating pathway knowledgeBIOTEC1 2024.
- [25] Han F, et al. Optimized Graph neural network-multilayer perceptron fusion classifier for metastatic prostate cancer detection in Western and Asian populations. *Asian J Urol* 2025.
- [26] Xie J, et al. PathMethy: an interpretable AI framework for cancer origin tracing based on DNA methylation. *Brief Bioinforma* 2024;25(6):bbae497.
- [27] Prosz A, et al. Biologically informed deep learning for explainable epigenetic clocks. *Sci Rep* 2024;14(1):1306.

- [28] Wu Y, et al. scGO: interpretable deep neural network for cell status annotation and disease diagnosis. *Brief Bioinforma* 2025;26(1):bbaf018.
- [29] Jin Z, Shi Y, Zhou L. Transparent sparse graph pathway network for analyzing the internal relationship of lung cancer. *Front Genet* 2024;15:1437174.
- [30] Rozanova S, et al. Quantitative Mass Spectrometry-Based Proteomics: An Overview. *Methods Mol Biol* 2021;2228:85–116.
- [31] Ellis MJ, et al. Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discov* 2013;3(10):1108–12.
- [32] Hartman E, et al. Interpreting biologically informed neural networks for enhanced proteomic biomarker discovery and pathway analysis. *Nat Commun* 2023;14(1):5359.
- [33] Webb-Robertson BJ, et al. Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. *J Proteome Res* 2015;14(5):1993–2001.
- [34] Alseekh S, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods* 2021;18(7):747–56.
- [35] Qian X, et al. Exploring Mechanisms and Biomarkers of Breast Cancer Invasion and Migration: an Explainable Gene-Pathway-Compounds Neural Network. *Cancer Med* 2025;14(6):e70769.
- [36] Jemimah S, AlShehhi A. c-Diadem: a constrained dual-input deep learning model to identify novel biomarkers in Alzheimer's disease. *BMC Med Genom* 2023;16(2):244.
- [37] Lin CH, Lichtarge O. Using interpretable deep learning to model cancer dependencies. *Bioinformatics* 2021;37(17):2675–81.
- [38] Ghosh Roy G, et al. MPVNN: Mutated Pathway Visible Neural Network architecture for interpretable prediction of cancer-specific survival risk. *Bioinformatics* 2022;38(22):5026–32.
- [39] Lan W, et al. MULGONET: An interpretable neural network framework to integrate multi-omics data for cancer recurrence prediction and biomarker discovery. *Fundam Res* 2025.
- [40] Zhao L, et al. DeepOmix: a scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis. *Comput Struct Biotechnol J* 2021;19:2719–25.
- [41] Wang L, et al. MCDHGN: heterogeneous network-based cancer driver gene prediction and interpretability analysis. *Bioinformatics* 2024;40(6).
- [42] Liu C, et al. Biological informed graph neural network for tumor mutation burden prediction and immunotherapy-related pathway analysis in gastric cancer. *Comput Struct Biotechnol J* 2023;21:4540–51.
- [43] Liu X, et al. Pathformer: a biological pathway informed transformer for disease diagnosis and prognosis using multi-omics data. *Bioinformatics* 2024;40(5).
- [44] Xi X, et al. A mechanism-informed deep neural network enables prioritization of regulators that drive cell state transitions. *Nat Commun* 2025;16(1):1284.
- [45] Cai Z, et al. Machine learning for multi-omics data integration in cancer. *iScience* 2022;25(2):103798.
- [46] Jemimah S., Abuhantash F., AlShehhi A., c-Triadem: . medRxiv, 2024., and A constrained, explainable deep learning model to identify novel biomarkers in Alzheimer's disease.
- [47] Kañduia MM, et al. ViLoN-a multi-layer network approach to data integration demonstrated for patient stratification. *Nucleic Acids Res* 2023;51(1):e6.
- [48] Tan J, et al. An interpretable survival model for diffuse large B-cell lymphoma patients using a biologically informed visible neural network. *Comput Struct Biotechnol J* 2024;24:523–32.
- [49] Hao J, et al. Interpretable deep neural network for cancer survival analysis by integrating genomic and clinical data. *BMC Med Genom* 2019;12(10):189.
- [50] Liang B, et al. Risk stratification and pathway analysis based on graph neural network and interpretable algorithm. *BMC Bioinforma* 2022;23(1):394.
- [51] Ma T., et al., . 2024. 1-6.Cox-Path: Biological Pathway-Informed Graph Neural Network for Cancer Survival Prediction.
- [52] Yan H, et al. Prior knowledge-guided multilevel graph neural network for tumor risk prediction and interpretation via multi-omics data integration. *Brief Bioinform* 2024;25(3).
- [53] Lee C, et al. Combining clinical embeddings with multi-omic features for improved patient classification and interpretability in Parkinson's Disease. *medRxiv* 2025:2025. 01.17.25320664.
- [54] Kuenzi BM, et al. Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell* 2020;38(5):672–684.e6.
- [55] Huang X, et al. ParsVNN: parsimony visible neural networks for uncovering cancer-specific and drug-sensitive genes and pathways. *NAR Genom Bioinform* 2021;3(4):lqab097.
- [56] Morgan HL. The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *J Chem Doc* 1965; 5(2):107–13.
- [57] Hao Y, Romano JD, Moore JH. Knowledge graph aids comprehensive explanation of drug and chemical toxicity. *CPT Pharmacomet Syst Pharm* 2023;12(8):1072–9.
- [58] Hao Y, Romano JD, Moore JH. Knowledge-guided deep learning models of drug toxicity improve interpretation. *Patterns (N Y)* 2022;3(9):100565.
- [59] Wang Z, et al. XMR: an explainable multimodal neural network for drug response prediction. *Front Bioinform* 2023;3:1164482.
- [60] Zhang F, et al. PathSynergy: a deep learning model for predicting drug synergy in liver cancer. *Brief Bioinform* 2025;26(2).
- [61] Xie J, et al. Interpretable Drug Response Prediction through Molecule Structure-aware and Knowledge-Guided Visible Neural Network. *bioRxiv* 2024. 2024.02.07.579280.
- [62] Dosovitskiy A, et al. arXiv preprint. arXiv:2010.11929, Image Is worth 16x16 words Transform Image Recognit Scale 2020.
- [63] Liu Z, et al. Swin transformer: Hierarchical vision transformer using shifted windows. *Proc IEEE/CVF Int Conf Comput Vis* 2021.
- [64] Song A, et al. Multimodal Prototyp Cancer Surviv Predict 2024.
- [65] Qiu L., Khormali A., Liu K. 2023. arXiv:2301.02383 DOI: 10.48550/arXiv.2301.02383., and Deep Biological Pathway Informed Pathology-Genomic Multimodal Survival Prediction.
- [66] Jaume G., et al., . 2023.Modeling Dense Multimodal Interactions Between Biological Pathways and Histology for Survival Prediction.
- [67] Behan FM, et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* 2019;568(7753):511–6.
- [68] Yin Q, Chen L. Explainable deep learning for identifying cancer driver genes based on the Cancer Dependency Map. *bioRxiv* 2025:2025. 04.28.651122.
- [69] Chowdhury S, Sarkar RR. Comparison of human cell signaling pathway databases—evolution, drawbacks and challenges. *Database* 2015;2015:bau126 (p).
- [70] Tarca AL, et al. A novel signaling pathway impact analysis. *Bioinformatics* 2009; 25(1):75–82.
- [71] Yang Q, et al. Pathway enrichment analysis approach based on topological structure and updated annotation of pathway. *Brief Bioinforma* 2019;20(1): 168–77.
- [72] Massa MS, Chiogna M, Romualdi C. Gene set analysis exploiting the topology of a pathway. *BMC Syst Biol* 2010;4(1):121.
- [73] Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 2006; 22(13):1600–7.
- [74] Cerami EG, et al. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* 2010;39(1):D685–90.
- [75] Schaefer CF, et al. PID: the pathway interaction database. *Nucleic Acids Res* 2009; 37(1):D674–9.
- [76] Breitkreutz B-J, et al. The BioGRID interaction database: 2008 update. *Nucleic Acids Res* 2007;36(1):D637–40.
- [77] Kuenzi BM, Ideker T. Author Correction: A census of pathway maps in cancer systems biology. *Nat Rev Cancer* 2021;21(3):212.
- [78] Gundogdu P, et al. SigPrimedNet: A Signaling-Informed Neural Network for scRNA-seq Annotation of Known and Unknown Cell Types. *Biology* 2023;12(4).
- [79] Kaynar G, et al. PiDeeL: metabolic pathway-informed deep learning model for survival analysis and pathological classification of gliomas. *Bioinformatics* 2023; 39(11).
- [80] Deng L, et al. Pathway-guided deep neural network toward interpretable and predictive modeling of drug sensitivity. *J Chem Inf Model* 2020;60(10): 4497–505.
- [81] Feng J, Zhang H, Li F. Investigating the relevance of major signaling pathways in cancer survival using a biologically meaningful deep learning model. *BMC Bioinforma* 2021;22(1):47.
- [82] Zhang H, Chen Y, Li F. Predicting Anticancer Drug Response With Deep Learning Constrained by Signaling Pathways. *Front Bioinform* 2021;1:639349.
- [83] Liu Y, Zhang Y-z, Imoto S. Microbial gene ontology informed deep neural network for microbe functionality discovery in human diseases. *Plos One* 2023;18(8): e0290307.
- [84] Hou Z, et al. PathExpSurv: pathway expansion for explainable survival analysis and disease gene discovery. *BMC Bioinforma* 2023;24(1):434.
- [85] Oldenburg J, et al. XModNN: explainable modular neural network to identify clinical parameters and disease biomarkers in transcriptomic datasets. *Biomolecules* 2024;14(12):1501.
- [86] Ferraro L, et al. MOVIDA: multiomics visible drug activity prediction with a biologically informed neural network model. *Bioinformatics* 2023;39(7).
- [87] Wang J, et al. An interpretable artificial intelligence framework for designing synthetic lethality-based anti-cancer combination therapies. *J Adv Res* 2024;65: 329–43.
- [88] Bourgeais V, Zehraoui F, Hanczar B. GraphGONet: a self-explaining neural network encapsulating the Gene Ontology graph for phenotype prediction on gene expression. *Bioinformatics* 2022;38(9):2504–11.
- [89] Ko E, et al. SPIN: sex-specific and pathway-based interpretable neural network for sexual dimorphism analysis. *Brief Bioinforma* 2024;25(4):bbae239.
- [90] Jiang Y, et al. IRnet: immunotherapy response prediction using pathway knowledge-informed graph neural network. *J Adv Res* 2024.
- [91] Burkhart JG, et al. Biology-inspired graph neural network encodes reactome and reveals biochemical reactions of disease. *Patterns (N Y)* 2023;4(7):100758.
- [92] Shin J, et al. DRPreter: interpretable anticancer drug response prediction using knowledge-guided graph neural networks and transformer. *Int J Mol Sci* 2022;23(22):13919.
- [93] Zhu J, et al. Decoding cell identity with multi-scale explainable deep learning. *bioRxiv* 2024:2024. 02.05.578922.
- [94] Vaswani, A., et al. *Attention Is All You Need*. 2017. arXiv:1706.03762 DOI: 10.48550/arXiv.1706.03762.
- [95] Cai Z, et al. DeePathNet: a transformer-based deep learning model integrating multiomic data with cancer pathways. *Cancer Res Commun* 2024;4(12):3151–64.
- [96] Chen RJ, et al. Pathomic Fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans Med Imaging* 2022;41(4):757–70.
- [97] Li J, et al. Deep-LIFT: Deep label-specific feature learning for image annotation. *IEEE Trans Cybern* 2021;52(8):7732–41.
- [98] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017:30.
- [99] Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. in *International conference on machine learning*. PMLR; 2017.

- [100] Selvaraju RR, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proc IEEE Int Conf Comput Vis* 2017.
- [101] Smilkov D, Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*, 20172017.
- [102] Bach S, et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One* 2015;10(7):e0130140.
- [103] Ying Z, et al. Gnnexplainer: Generating explanations for graph neural networks. *Adv Neural Inf Process Syst* 2019;32.
- [104] Sharma S, et al. XAI-based data visualization in multimodal medical data. *bioRxiv* 2025. 2025.07. 11.664302.
- [105] Mumuni F, Mumuni A. *arXiv preprint. arXiv:2501.09967, Explain Artif Intell (XAI) inherent Explain Large Lang Models* 2025.
- [106] Lavecchia A. Explainable artificial intelligence in drug discovery: bridging predictive power and mechanistic insight. *Wiley Interdisciplinary Reviews Computational Molecular Science* 2025;15(5):e70049.
- [107] Houssein EH, et al. Explainable artificial intelligence for medical imaging systems using deep learning: a comprehensive review. *Clust Comput* 2025;28(7):469.
- [108] Arreche O, Abdallah M. A comparative analysis of DNN-based white-box explainable AI methods in network security. *EURASIP J Inf Secur* 2025;2025(1): 16.
- [109] Hao J, et al. PASNet: pathway-associated sparse deep neural network for prognosis prediction from high-throughput data. *BMC Bioinforma* 2018;19(1): 510.
- [110] Meirer J., DeepBINN: A tailored biologically-informed neural network for robust biomarker identification. 2024. 246-2492024. 246249.
- [111] Chen J, et al. Transformer for one stop interpretable cell type annotation. *Nat Commun* 2023;14(1):223.
- [112] Seninge L, et al. VEGA is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. *Nat Commun* 2021;12(1):5684.
- [113] Janizek JD, et al. PAUSE: principled feature attribution for unsupervised gene expression analysis. *Genome Biol* 2023;24(1):81.
- [114] Patkulkar PA, et al. Mapping Spatiotemporal Heterogeneity in Tumor Profiles by Integrating High-Throughput Imaging and Omics Analysis. *ACS Omega* 2023;8(7):6126–38.
- [115] Arora R, et al. Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nat Commun* 2023;14(1):5029.
- [116] Liu T, et al. Graph deep learning enabled spatial domains identification for spatial transcriptomics. *Brief Bioinform* 2023;24(3).
- [117] Lee J, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020;36(4):1234–40.
- [118] Nair NU, et al. Cross-species identification of cancer resistance-associated genes that may mediate human cancer risk. *Sci Adv* 2022;8(31):eabj7176.
- [119] Pan S, Yang Q. A survey on transfer learning. *IEEE Transaction on Knowledge Discovery and Data Engineering*, 22. IEEE press; 2010.
- [120] Li R., et al., . *arXiv preprint arXiv:2403.05818, 2024.PR-NET: Leveraging Pathway Refined Network Structures for Prostate Cancer Patient Condition Prediction.*
- [121] Lee S, et al. Cancer subtype classification and modeling by pathway attention and propagation. *Bioinformatics* 2020;36(12):3818–24.
- [122] Holzschek N, et al. Modeling transcriptomic age using knowledge-primed artificial neural networks. *npj Aging Mech Dis* 2021;7(1):15.
- [123] Gut G, et al. PmVAE: Learning interpretable single-cell representations with pathway modules. *bioRxiv* 2021:2021. 01. 28.428664.
- [124] Park S, Huang E, Ahn T. Classification and functional analysis between cancer and normal tissues using explainable pathway deep learning through RNA-sequencing gene expression. *Int J Mol Sci* 2021;22(21):11531.
- [125] Chen H, et al. Comprehensive single-cell RNA-seq analysis using deep interpretable generative modeling guided by biological hierarchy knowledge. *Brief Bioinforma* 2024;25(4).
- [126] Yin Q, Chen L. CellTICS: an explainable neural network for cell-type identification and interpretation based on single-cell RNA-seq data. *Brief Bioinforma* 2023;25(1).
- [127] Sharma D, Xu W. ReGeNNe: genetic pathway-based deep neural network using canonical correlation regularizer for disease prediction. *Bioinformatics* 2023;39(11).
- [128] Snow O, et al. BDKANN - Biological Domain Knowledge-based Artificial Neural Network for drug response prediction. *bioRxiv* 2020:840553.
- [129] Huang T, et al. Explainable drug side effect prediction via biologically informed graph neural network. *medRxiv* 2023.
- [130] van Hilten A, et al. Phenotype prediction using biologically interpretable neural networks on multi-cohort multi-omics data. *NPJ Syst Biol Appl* 2024;10(1):81.
- [131] Sun B, Chen L. Interpretable deep learning for improving cancer patient survival based on personal transcriptomes. *Sci Rep* 2023;13(1):11344.
- [132] Zhu J, et al. Geometric graph neural networks on multi-omics data to predict cancer survival outcomes. *Comput Biol Med* 2023;163:107117.
- [133] Yang J, et al. BioXNet: a biologically inspired neural network for deciphering anti-cancer drug response in precision medicine. *bioRxiv* 2024:2024. 01.29.576766.