Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DeepRelCDR: A Deep Relation-Aware Molecular Graph Attention Framework for Cancer Drug Response Prediction

Sarvin Chitsaz
Arak University
Majid Sepahvand
Arak University
Maytham N. Meqdad
Al-Mustaqbal University
Masoud Daneshtalab
masoud.daneshtalab@mdu.se
Mälardalen University

Article

Keywords:

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10988328/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

DeepRelCDR: A Deep Relation-Aware Molecular Graph Attention Framework for Cancer Drug Response Prediction

Sarvin Chitsaz[a], Majid Sepahvand[a], Maytham N. Meqdad[b], Masoud Daneshtalab[c,*]

[a]*Department of Computer Engineering, Arak University, Markazi, Iran*
[b]*Intelligent Medical Systems Department, Al-Mustaqbal University, Hillah 51001, Babil, Iraq*
[c]*DeepHERO Lab, Mälardalen University, Västerås, Sweden*

Abstract

Predicting cancer cell-line response to therapeutic compounds is central to precision oncology, yet graph neural network models typically represent each drug only through covalent-bond connectivity, leaving ring-system and functional-group structure implicit; the relation set used here is fixed by chemical domain knowledge and evaluated on a single pharmacogenomic resource rather than learned or validated across cohorts. We introduce DeepRelCDR, a deep relation-aware molecular graph attention framework that represents each drug as three explicit relations, covalent bonds, ring co-membership, and RDKit-derived functional-group co-membership, each aggregated independently in a depthwise-separable fashion by a transform-free attention head on raw atom features and fused by a single shared pointwise layer; the resulting drug embedding is combined with a cell-line gene-expression embedding to predict a standard drug-sensitivity readout, LN_IC50. Trained on 235 drugs and 199,903 drug–cell-line pairs from GDSC2, DeepRelCDR reaches a test R^2 of 0.762 (95% CI [0.756, 0.769]), outperforming a single-relation attention baseline by 6.0 points, a Ridge-regression baseline by 13.6 points, and a GraphDRP-style reimplementation by 6.7 points of R^2. Gradient-based gene attribution recovers Schlafen-11, an independently established chemosensitivity biomarker gene, as the top camptothecin-sensitivity gene across three cancer types, and a personalized drug-ranking analysis reaches an 80.3% top-1 hit rate versus 53.5% for a non-personalized baseline.

Keywords: Graph neural networks, Drug response prediction, Multi-relation graph attention, Depthwise-separable convolution, SLFN11, Precision oncology

1. Introduction

Cancer cell lines that share a tissue of origin, and even tumors within a single patient, can respond very differently to the same therapeutic compound: in the pharmacogenomic screen analyzed here, selecting a single best-on-average drug rather than a patient-specific

*Corresponding author. Email: masoud.daneshtalab@mdu.se
Email addresses: s.chitsaz.02@bsc.araku.ac.ir (Sarvin Chitsaz), m-sepahvand@araku.ac.ir (Majid Sepahvand), maytham.meqdad@uomus.edu.iq (Maytham N. Meqdad), masoud.daneshtalab@mdu.se (Masoud Daneshtalab)

one costs 26.8 percentage points of top-1 sensitivity accuracy (Section 4.11). Large pharmacogenomic screens such as the Genomics of Drug Sensitivity in Cancer (GDSC) project pair hundreds of drugs against hundreds to thousands of molecularly profiled cell lines, reporting a dose-response statistic, LN_IC_{50}, for each drug–cell-line pair [1, 2]. Learning to predict LN_IC_{50} from a drug's chemical structure and a cell line's gene expression is an attractive computational proxy for precision oncology.

The timing reflects two developments: GDSC-scale consortia now provide enough pairs to train graph-based encoders rather than hand-crafted descriptors [1, 2], and toolkits such as RDKit [3] make it practical to extract ring systems and functional groups from a SMILES string at scale, together making it newly feasible to move beyond a bond-only graph.

Graph neural networks (GNNs) are the dominant approach here, parsing SMILES into a graph of atoms and bonds processed by neighbor-aggregating architectures [4, 5], outperforming fingerprint- and string-based representations [6] across contrastive, transformer-based, and multi-task variants [7, 8, 9, 10, 11, 12]. The graph, however, is almost universally restricted to a *single* relation, covalent-bond adjacency, leaving ring co-membership (cyclic, aromatic, conformational dependencies not distinguished from an arbitrary bond chain) and functional-group membership (a shared local chemical context) implicit, forcing the model to infer them from connectivity rather than receiving explicit priors.

Addressing this requires a chemical inductive bias giving ring and functional-group membership as explicit edges, and a parameter-efficiency solution, since naive multi-relation extensions trade topology for a larger, overfitting-prone budget given modest drug counts. Section 2 reviews this trade-off and the principle of decoupling spatial aggregation from feature mixing, borrowed from *depthwise-separable convolution*.

We propose DeepRelCDR, representing each drug with three explicit relations: covalent bonds, ring co-membership, and RDKit-derived functional-group co-membership. A *depthwise* stage aggregates each relation via its own attention head on raw features; a *pointwise* stage fuses the three outputs through a shared linear layer, determining *where* each relation aggregates from and *how* signals mix across features. The drug embedding combines with a cell-line gene-expression embedding to predict LN_IC_{50}.

The key contributions of this paper are as follows:

- DeepRelCDR, a multi-relation graph attention architecture decoupling covalent-bond, ring, and functional-group semantics into three parallel relations rather than a single bond-only graph.
- A depthwise-separable relation-wise attention layer replacing per-relation transformation matrices with a shared pointwise fusion layer; a dedicated ablation shows attention weighting alone contributes 4.8 points of R^2 (Section 4.5.1).
- On 235 drugs and 199,903 GDSC2 pairs, the full model exceeds single- and two-relation ablations beyond seed variance, and outperforms Ridge, GraphDRP-style, and DeepAEG-style baselines by 13.6, 6.7, and 2.7 points of R^2 (Section 4.6).
- Accuracy is not narrowly tuned: sensitivity analysis changes R^2 by at most 1.5 points (Section 4.8); the model generalizes to CCLE, CTRPv2 (Section 4.12), and an unrelated task, ESOL, reaching $R^2 = 0.7391$ vs. 0.3842 for Ridge (Section 4.13).

- Two independent diagnostics, attention-weighted output magnitude and pointwise fusion weight norm, both show non-negligible contribution from all three relations.
- The model is interpretable at gene and atom level: attribution recovers SLFN11, a known camptothecin-sensitivity biomarker, across three cancer types; personalized drug ranking reaches an 80.3% top-1 hit rate versus 53.5% for a non-personalized baseline.

Section 2 reviews related work; Section 3 presents the architecture; Section 4 reports experiments; Section 5 discusses results, limitations, and future work; Section 6 concludes.

2. Related work

2.1. Graph neural networks

Early drug-response models represented drugs as fixed fingerprints or SMILES strings via convolutional/recurrent encoders paired with a separate cell-line encoder [6], discarding atom-and-bond topology. This persists: DeepTraCDR [13] combines expression with a fixed ECFP6 (Extended-Connectivity Fingerprint, diameter 6) fingerprint via transformer and GCN over a bipartite drug–cell-line network, but still flattens the drug's own structure. GraphDRP [7] replaced fingerprints with a molecular graph via GCN/GAT, and DeepCDR [8] paired a similar drug encoder with a multi-omics cell-line encoder; both, with Nguyen et al. [14], confirmed graphs outperform flattened representations.

This work, however, adopted the simplest graph: single covalent-bond relation. Subsequent work refined the backbone without revisiting the relation: DualGCN [9] and GraphCDR [10] apply dual-branch/contrastive convolution; TGSA [15] adds protein-association similarity; DeepTTA [11], Chu et al. [16], GPDRP [17], DeepCoVDR [18] use transformers; SubCDR [19] and DeepAEG [20] target interpretability; ASGCL [21] and GNNDRP [12] add contrastive/multi-task objectives. Each changes how the bond graph is processed, not what it represents: ring and functional-group co-occurrence stay implicit across the GCN [4], GAT [5], MPNN [22], GraphSAGE [23], GIN [24] formulations these build on, motivating DeepRelCDR (Section 3); Section 4.6 compares against a bond-only reimplementation.

2.2. Multi-relational graph neural networks

Multi-relational GNNs outside drug response add relations for heterogeneous edges, notably knowledge graphs. R-GCN [25] gives each relation its own weight matrix, summing transformed messages, resolving single-relation limits but growing size multiplicatively with relation count.

Within drug response, CETExplainer [26] applies R-GCN to a heterogeneous drug–cell-line graph, leaving each molecular graph single-relation. Chain-aware GNN [27] represents atoms via shortest-path chains; MMGNN [28] decomposes into overlapping atom-type-pair subgraphs via a shared backbone; neither gives an explicit chemical relation nor separates relation-specific aggregation from cross-relation mixing.

None closes the gap alone: R-GCN reintroduces the parameter cost; interaction graphs leave molecules single-relation; atom-type decompositions forgo chemical semantics. A parallel trade-off is resolved in image models: *depthwise-separable convolution* [29] decouples

spatial filtering from channel mixing. We adapt this to the relation dimension: three chemically named relations (bond, ring, functional group) are aggregated independently by a transform-free attention head before a pointwise layer mixes across relations (Section 3.4), removing R-GCN's size-relation coupling and keeping every relation interpretable.

Table 1 positions DeepRelCDR against representative prior architectures along relation count, per-relation transformation, and empirical evaluation under the shared protocol used here (Section 4).

Table 1: Qualitative comparison of DeepRelCDR with representative prior GNN-based drug-response architectures.

Model	Relations	Transform	Evaluated
GraphDRP [7]	1	n/a	Yes‡
DeepAEG [20]	1	n/a	Yes‡
DeepCDR [8]	1	n/a	No
R-GCN-style [25]	Any	Full $C_{in}\times C_{out}$	No
Single-rel. GAT (ablation)	1	n/a	Yes
DeepRelCDR (ours)	3	None†	Yes

†Attention-only per relation; a single shared pointwise layer mixes across relations. "Transform" denotes each model's per-relation feature transformation. ‡Reimplemented from the published architecture description (Section 4.6, Table 6), not the original authors' codebase or reported numbers.

3. Method

3.1. Problem formulation

Given a drug represented by its SMILES (Simplified Molecular Input Line Entry System) string [30] and a cell line's gene expression profile, the task is to predict LN_IC_{50} (lower values indicate higher sensitivity). Let $G_d = (\mathcal{V}, \mathcal{E}_{bond}, \mathcal{E}_{ring}, \mathcal{E}_{fg}, X)$ be the multi-relation graph of drug d: $\mathcal{V}$ is the set of N heavy atoms, $X \in \mathbb{R}^{N\times C_{in}}$ the atom feature matrix, and $\mathcal{E}_{bond}, \mathcal{E}_{ring}, \mathcal{E}_{fg}$ three relation-specific edge sets (Section 3.3). Let $g_c \in \mathbb{R}^{1000}$ be cell line c's expression vector, restricted to the top 1000 variable training-split genes (Section 4.2). Given $\mathcal{D}_{train}$ with labels $y_{d,c} \in \mathbb{R}$, the model learns $f_\theta : (G_d, g_c) \mapsto \hat{y}_{d,c}$ by solving

$$\theta^\star = \arg\min_\theta \frac{1}{|\mathcal{D}_{train}|} \sum_{(G_d, g_c, y_{d,c}) \in \mathcal{D}_{train}} \Big(f_\theta(G_d, g_c) - y_{d,c} \Big)^2, \tag{1}$$

with θ shared across all relation-specific attention heads. Validation uses $\mathcal{D}_{val}$; final accuracy on disjoint $\mathcal{D}_{test}$, split at the cell-line level.

3.2. Framework overview

Fig. 1 summarizes the pipeline. RDKit parses a SMILES string into a graph with three edge types, bond, ring, and functional-group co-membership (Section 3.3). Two stacked DeepRelCDR layers aggregate each relation via depthwise attention, then fuse via a shared pointwise layer (Section 3.4). Global mean pooling yields drug embedding $d \in \mathbb{R}^{128}$; a two-layer MLP yields cell-line embedding $c \in \mathbb{R}^{128}$ (Section 3.5). The two are concatenated and fused to predict LN_IC_{50} (Section 3.6).

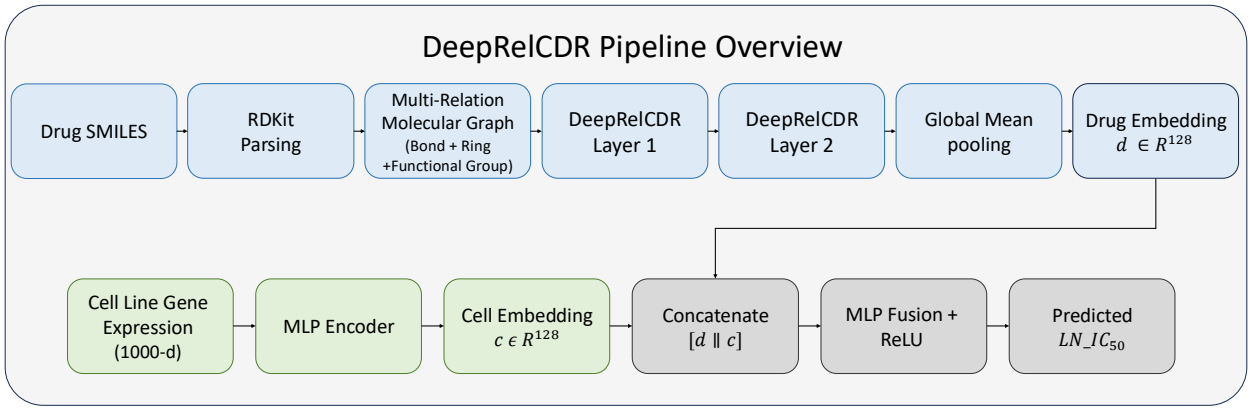


Figure 1: Overall DeepRelCDR pipeline: SMILES parsing, two DeepRelCDR layers, and pooling into a drug embedding $d \in \mathbb{R}^{128}$, fused with a cell-line embedding $c \in \mathbb{R}^{128}$ to predict LN_IC_{50}.

3.3. Multi-relation molecular graph construction

We represent every drug as a graph with three explicit relations sharing the same node set, illustrated in Fig. 2 for Nutlin-3a.

3.3.1. Node features

For each atom v_i, RDKit extracts a 20-dimensional feature vector

$$x_i = \big[\,\text{onehot}(\text{symbol}_i) \,\|\, \deg_i \,\|\, q_i \,\|\, a_i \,\|\, h_i \,\|\, r_i \,\|\, \text{onehot}(\text{hyb}_i)\,\big] \in \mathbb{R}^{20} \tag{2}$$

where $\text{onehot}(\text{symbol}_i) \in \{0,1\}^{10}$ encodes atom type over $\{\text{C}, \text{N}, \text{O}, \text{S}, \text{F}, \text{Cl}, \text{Br}, \text{I}, \text{P}, \text{H}\}$, $\deg_i$ is heavy-atom degree, q_i formal charge, $a_i \in \{0,1\}$ aromaticity, h_i attached-hydrogen count, $r_i \in \{0,1\}$ ring membership, and $\text{onehot}(\text{hyb}_i) \in \{0,1\}^5$ hybridization over $\{sp, sp^2, sp^3, sp^3d, sp^3d^2\}$ (Table 2).

3.3.2. Bond relation

$\mathcal{E}_{bond}$ connects every pair of atoms joined by a covalent bond, in both directions: $\mathcal{E}_{bond} = \{(i,j),(j,i) : (i,j) \in \text{Bonds}(mol)\}$, capturing local, first-order connectivity.

3.3.3. Ring relation

$\mathcal{E}_{ring}$ is constructed from RDKit's symmetrized smallest set of smallest rings (SSSR); for every ring ρ returned by `GetRingInfo().AtomRings()`, all atom pairs within ρ are fully connected:

$$\mathcal{E}_{ring} = \bigcup_{\rho \in \text{SSSR}(mol)} \{(i,j) : i,j \in \rho,\ i \neq j\}. \tag{3}$$

This places every ring atom within one hop of every other, letting a fused or aromatic scaffold act as a coherent unit rather than a bond chain.

Fig. 3 shows the combined graph for four diverse drugs: (a) LMP744, (b) staurosporine, (c) tretinoin, (d) Telomerase Inhibitor IX, spanning scaffold sizes without tuning.

Table 2: Composition of the 20-dimensional atom feature vector x_i (Eq. (2)).

Feature group	Dimensions	Encoding
Atom symbol	10	one-hot over C, N, O, S, F, Cl, Br, I, P, H
Heavy-atom degree	1	integer count
Formal charge	1	integer
Aromaticity	1	binary flag
Total attached hydrogens	1	integer count
Ring membership	1	binary flag
Hybridization	5	one-hot over $sp, sp^2, sp^3, sp^3d, sp^3d^2$
Total	20	

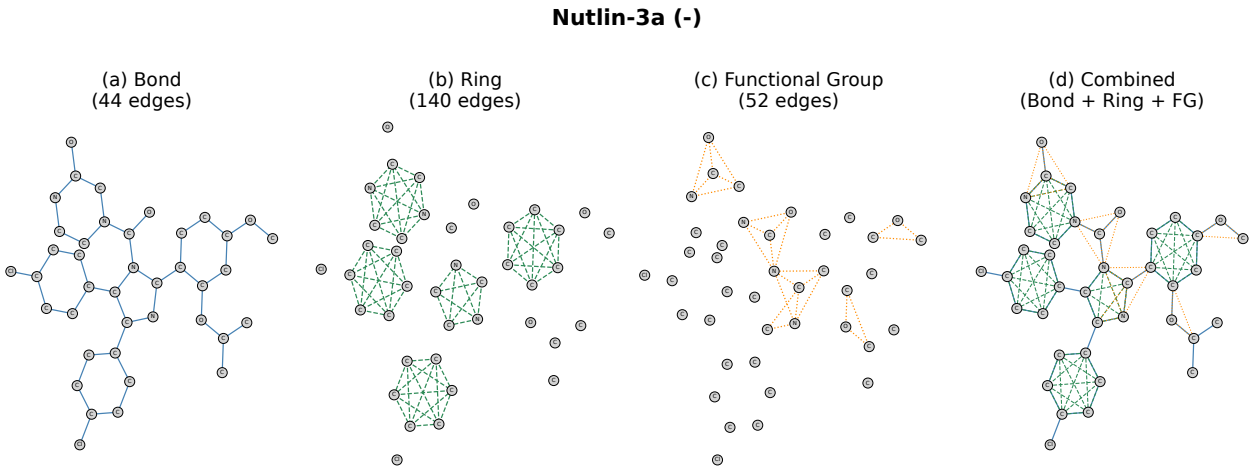


Figure 2: Decomposition of Nutlin-3a into the three relation graphs used by DeepRelCDR and their union: (a) Bond (44 edges), (b) Ring (140 edges), (c) Functional Group (52 edges), (d) Combined.

3.3.4. Functional-group relation

$\mathcal{E}_{fg}$ is built from 65 SMARTS patterns: 64 of RDKit's built-in `Fragments` patterns, after excluding 17 ring-scaffold or metabolic-site patterns (e.g. `fr_benzene`) that overlap with the ring relation, plus one custom boronic-acid pattern. The default `fr_ether` and `fr_amide` patterns were replaced with stricter SMARTS excluding ester/lactone oxygens and carbamate/urea nitrogens (Fig. 4). For every substructure match μ, all atom pairs in μ are fully connected, analogous to Eq. (3).

Self-loops are not stored in any $\mathcal{E}_r$; they are added dynamically inside the attention layer, so each relation independently weights self versus neighbor features.

All three relations derive from the 2D molecular graph, with no 3D conformer or stereochemistry: GDSC2 provides only a SMILES string, so 3D structure would be computationally

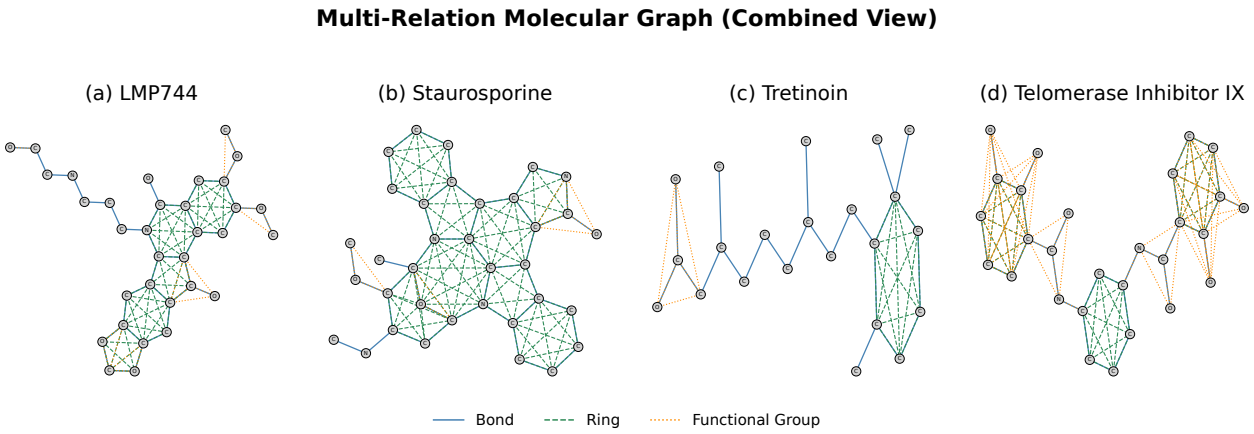


Figure 3: Combined (bond + ring + functional-group) multi-relation graph for four structurally diverse drugs: (a) LMP744, (b) staurosporine, (c) tretinoin, (d) Telomerase Inhibitor IX.

generated rather than measured; prior 2D encoders, GraphDRP [7] and DeepCDR [8] among them, already outperform fingerprint-based representations (Section 2). 3D/stereochemical relations are left as future work (Section 5).

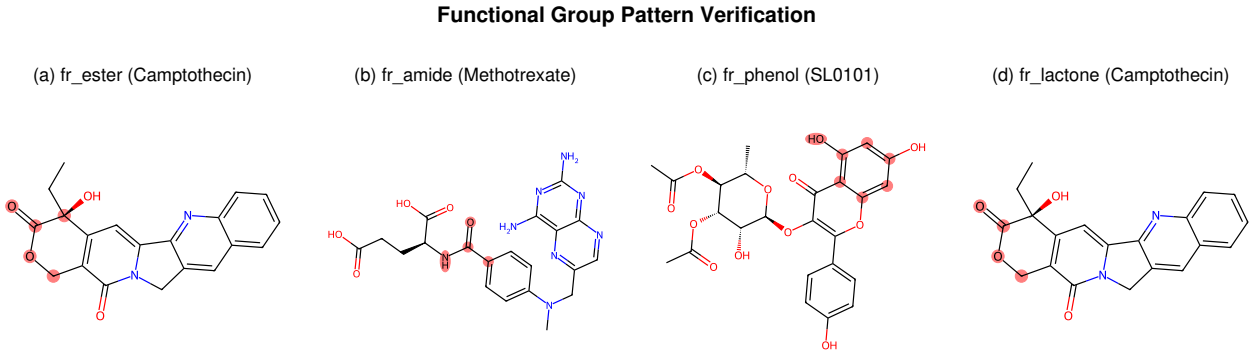


Figure 4: Verification of four functional-group SMARTS patterns: (a) ester and (d) lactone on camptothecin, (b) amide on methotrexate, (c) phenol on SL0101.

3.4. Depthwise-separable multi-relation graph attention

This section details the internals of the DeepRelCDR layer introduced in Fig. 1.

A naive multi-relation extension equips each relation r with its own full transformation matrix $W_r \in \mathbb{R}^{C_{in} \times C_{out}}$ over a static, unweighted aggregation, outputs summed (Fig. 5), coupling topology cost to representational capacity. We instead decouple the two: each relation performs *only* attention-weighted aggregation on raw features, and a shared pointwise layer performs all feature mixing jointly (Fig. 6), mirroring depthwise-separable convolution [29].

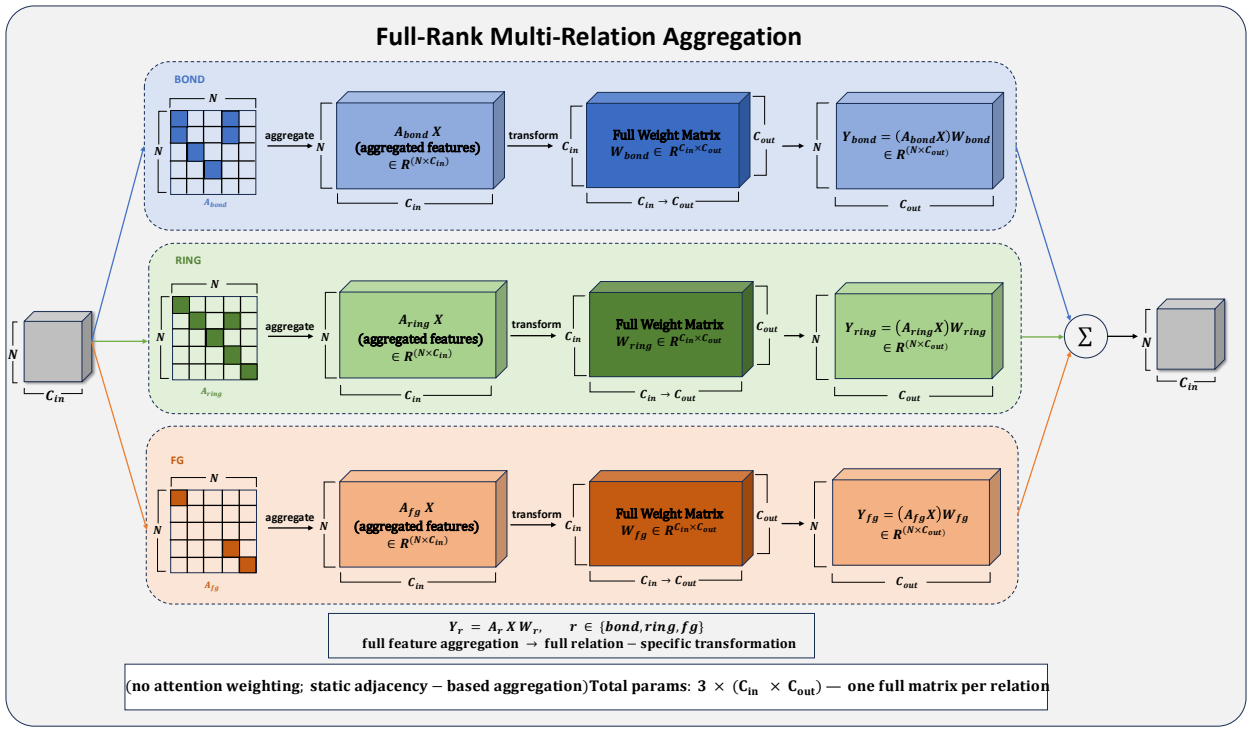


Figure 5: Full-rank multi-relation aggregation: each relation aggregates $A_r X$ and applies its own transformation matrix $W_r \in \mathbb{R}^{C_{in} \times C_{out}}$, with outputs summed.

3.4.1. Relation-wise depthwise attention

Let $X \in \mathbb{R}^{N \times C_{in}}$ be the node feature matrix at a given layer. For each relation $r \in \mathcal{R} = \{bond, ring, fg\}$, self-loops are first added,

$$\hat{\mathcal{E}}_r = \mathcal{E}_r \cup \{(i, i) : i = 1, \ldots, N\}. \tag{4}$$

For every edge $(i, j) \in \hat{\mathcal{E}}_r$, an attention logit is computed by a scoring vector $a_r \in \mathbb{R}^{2C_{in}}$ private to relation r:

$$e_{ij}^r = \mathrm{LeakyReLU}_{0.2}\big(a_r^\top \left[x_i \,\|\, x_j\right]\big)\,. \tag{5}$$

This uses no transforming matrix, so relation r contributes only $2C_{in}$ parameters. Logits are normalized by softmax over node i's in-neighborhood,

$$\alpha_{ij}^r = \frac{\exp\big(e_{ij}^r\big)}{\sum_{k\,:\,(i,k)\in\hat{\mathcal{E}}_r} \exp(e_{ik}^r)}, \tag{6}$$

followed by dropout ($p = 0.2$). The relation-specific output is the attention-weighted, *untransformed* aggregation of neighbor features,

$$z_i^r = \sum_{j\,:\,(i,j)\in\hat{\mathcal{E}}_r} \alpha_{ij}^r\, x_j \;\in\; \mathbb{R}^{C_{in}}. \tag{7}$$

Since z_i^r retains dimensionality C_{in} with no cross-channel mixing, Eqs. (5)–(7) constitute the depthwise stage: three independently parameterized attention heads, one per relation, each writing back only a reweighted version of the input.

3.4.2. Pointwise relation fusion

The three relation-specific outputs are concatenated,

$$z_i = \left[\, z_i^{bond} \,\|\, z_i^{ring} \,\|\, z_i^{fg} \,\right] \;\in\; \mathbb{R}^{3C_{in}}, \tag{8}$$

and passed through a single shared pointwise layer producing the layer output,

$$h_i = \Theta\, z_i + b, \qquad \Theta \in \mathbb{R}^{C_{out} \times 3C_{in}},\; b \in \mathbb{R}^{C_{out}}. \tag{9}$$

This learns a single joint linear combination over all $3C_{in}$ channels, letting the model weight bond-, ring-, and functional-group-derived signals arbitrarily rather than combining them additively. A DeepRelCDR layer thus contains

$$|\theta_{layer}| = \underbrace{3 \cdot 2C_{in}}_{\text{depthwise attention}} + \underbrace{3C_{in}C_{out} + C_{out}}_{\text{pointwise fusion}} \tag{10}$$

learnable parameters. Writing the full layer as

$$H = \mathrm{DeepRelCDRLayer}(X; \{\mathcal{E}_{bond}, \mathcal{E}_{ring}, \mathcal{E}_{fg}\}), \tag{11}$$

each layer is followed by batch normalization and LeakyReLU (slope 0.01),

$$H' = \mathrm{LeakyReLU}_{0.01}\big(\mathrm{BN}(H)\big), \tag{12}$$

and we stack two such blocks (Fig. 1), with dropout ($p = 0.2$) on the first block's output:

$$H^{(1)} = \mathrm{Dropout}_{0.2}\Big(\mathrm{LeakyReLU}_{0.01}\big(\mathrm{BN}(\mathrm{DeepRelCDRLayer}_1(X))\big)\Big), \tag{13}$$

$$H^{(2)} = \mathrm{LeakyReLU}_{0.01}\big(\mathrm{BN}(\mathrm{DeepRelCDRLayer}_2(H^{(1)}))\big), \tag{14}$$

where $\mathrm{DeepRelCDRLayer}_1$ maps $C_{in} = 20 \to C_{out} = 128$ and $\mathrm{DeepRelCDRLayer}_2$ maps $128 \to 128$. $H^{(2)} \in \mathbb{R}^{N \times 128}$ is reduced to a drug embedding by global mean pooling (Eq. (15)).

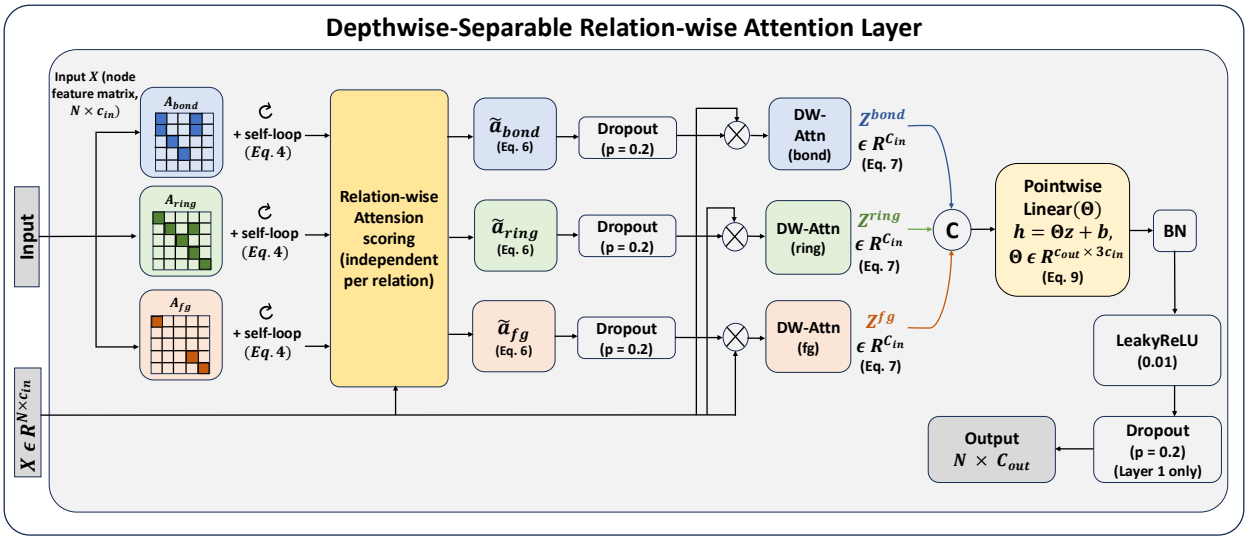


Figure 6: Depthwise-separable relation-wise attention (this work): each relation aggregates independently on raw features, then a shared pointwise layer fuses them and applies BN and LeakyReLU.

3.5. Drug and cell-line encoders

3.5.1. Drug encoder

The drug encoder applies two stacked DeepRelCDR layers to X (Eqs. (13)–(14)), producing $H^{(2)} \in \mathbb{R}^{N\times 128}$, reduced by global mean pooling to the drug embedding

$$d = \frac{1}{N}\sum_{i=1}^{N} H^{(2)}_{i,:} \;\in\; \mathbb{R}^{128}. \tag{15}$$

3.5.2. Cell-line encoder

The gene expression vector $g_c \in \mathbb{R}^{1000}$ passes through a two-layer MLP with batch normalization and LeakyReLU (slope 0.01), with dropout ($p = 0.2$) after the first block:

$$u = \mathrm{LeakyReLU}\big(\mathrm{BN}(W_1 g_c + b_1)\big), \tag{16}$$

$$c = \mathrm{LeakyReLU}\big(\mathrm{BN}(W_2\,\mathrm{Dropout}(u) + b_2)\big), \tag{17}$$

with $W_1 \in \mathbb{R}^{128\times 1000}$, $W_2 \in \mathbb{R}^{128\times 128}$, yielding $c \in \mathbb{R}^{128}$.

3.6. Multimodal fusion and prediction

The drug and cell-line embeddings are concatenated and passed through a fully connected layer with ReLU,

$$f = \mathrm{ReLU}\big(W_f\,[\,d \,\|\, c\,] + b_f\big), \tag{18}$$

$$W_f \in \mathbb{R}^{128\times 256}, \tag{19}$$

and a linear output layer produces the scalar prediction,

$$\hat{y}_{d,c} = w_o^{\top} f + b_o, \qquad w_o \in \mathbb{R}^{128}. \tag{20}$$

The full model is trained end-to-end by minimizing the mean squared error between $\hat{y}_{d,c}$ and the experimentally measured LN_IC_{50},

$$\mathcal{L} = \frac{1}{|\mathcal{B}|} \sum_{(d,c) \in \mathcal{B}} \big(\hat{y}_{d,c} - y_{d,c}\big)^2. \tag{21}$$

3.7. End-to-end training procedure

Algorithm 1 summarizes the full training procedure, composing graph construction, drug/cell-line encoders, and fusion into a single training loop with early stopping.

Algorithm 1 DeepRelCDR training procedure

Require: Training set $\mathcal{D}_{train} = \{(G_d, g_c, y_{d,c})\}$; validation set $\mathcal{D}_{val}$; max epochs T; patience P; learning rate η

Ensure: Trained parameters $\theta^\star$

Initialize θ (drug encoder, cell-line encoder, fusion head) randomly; $\theta^\star \leftarrow \theta$; $b_{val} \leftarrow \infty$; $p \leftarrow 0$
for $t = 1$ **to** T **do**
  for each minibatch $\mathcal{B} \subset \mathcal{D}_{train}$ **do**
    for each $(G_d, g_c, y_{d,c}) \in \mathcal{B}$ **do**
      $H^{(1)} \leftarrow \mathrm{Dropout}\big(\mathrm{LeakyReLU}(\mathrm{BN}(\mathrm{DeepRelCDRLayer}_1(X_d)))\big)$ ▷ Eq. (13)
      $H^{(2)} \leftarrow \mathrm{LeakyReLU}\big(\mathrm{BN}(\mathrm{DeepRelCDRLayer}_2(H^{(1)}))\big)$ ▷ Eq. (14)
      $d \leftarrow \frac{1}{N}\sum_{i=1}^{N} H^{(2)}_{i,:}$ ▷ global mean pooling, Eq. (15)
      $c \leftarrow \mathrm{MLP}_{cell}(g_c)$ ▷ cell-line encoder, Section 3.5
      $\hat{y}_{d,c} \leftarrow w_o^\top \mathrm{ReLU}(W_f[d \| c] + b_f) + b_o$
    end for
    $\mathcal{L}_{\mathcal{B}} \leftarrow \frac{1}{|\mathcal{B}|}\sum_{(d,c)\in\mathcal{B}}(\hat{y}_{d,c} - y_{d,c})^2$
    $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{\mathcal{B}}$ ▷ Adam update
  end for
  $\mathcal{L}_{val} \leftarrow$ mean squared error of current θ on $\mathcal{D}_{val}$
  if $\mathcal{L}_{val} < b_{val}$ **then**
    $b_{val} \leftarrow \mathcal{L}_{val}$; $\theta^\star \leftarrow \theta$; $p \leftarrow 0$
  else
    $p \leftarrow p + 1$
    if $p \geq P$ **then break** ▷ early stopping
    end if
  end if
end for
return $\theta^\star$

3.8. Complexity analysis

Let N be the number of heavy atoms and $E_r = |\hat{\mathcal{E}}_r|$ the edge count (including self-loops) in relation r. The depthwise stage scores and aggregates every edge once in $O(C_{in})$ time; the pointwise stage applies a dense $C_{out} \times 3C_{in}$ transform at every node in $O(N\, C_{in}\, C_{out})$ time.

Per-layer time complexity is therefore

$$O\big((E_{bond} + E_{ring} + E_{fg})\, C_{in} + N\, C_{in}\, C_{out}\big). \tag{22}$$

Unlike E_{bond}, which is linear in N, the ring and functional-group relations fully connect every atom pair within each matched substructure, so E_{ring} and E_{fg} scale quadratically with the largest single ring or SMARTS match, $O(N)$ in the worst case but in practice bounded (for the 25-atom Nutlin-3a, $E_{ring} = 140$, $E_{fg} = 52$, well below $N^2 = 625$). Space complexity per layer is $O(C_{in}C_{out})$ for Θ plus $O(C_{in})$ per attention vector, independent of N; activation memory scales as $O(N\, C_{in})$ per relation. The cell-line MLP contributes a further $O(B \times 1000 \times 128)$ time per batch, independent of drug graph size.

4. Experiments

4.1. Dataset

Drug–cell-line response pairs are drawn from GDSC2, the methodologically updated of the two GDSC screening releases [1, 2] (242,036 raw dose–response records; Fig. 7, Table 3). Records merged with RNA-seq expression retain 235,800 records from 944 cell lines. Of 286 unique drug names, SMILES were resolved via PubChem for 235 (82.2%); the remaining 51 (38 unresolved, 13 unverifiable CIDs) were excluded. All 235 parsed successfully via RDKit [3], leaving 199,903 records. RNA-seq (RSEM TPM) provided 41,144 genes; the top 1000 variable genes were selected *using the training split only* to avoid leakage, z-scored using training-set statistics. Cell lines were split 70/15/15 (660/142/142; `random_state`=42), so none appears in more than one split.

Table 3: Dataset statistics for GDSC2 drug–cell-line response pairs, from raw records through the train/validation/test split.

Quantity	Count
Raw GDSC2 dose–response records	242,036
Response records with paired expression data	235,800
Unique drug names with paired response data	286
Drug names resolved to SMILES via PubChem	235 (82.2%)
Drugs successfully parsed into multi-relation graphs	235 (100%)
Response records with valid SMILES	199,903
Unique genes in RNA-seq expression matrix	41,144
Genes retained for modeling (training-set top-variance)	1,000
Cell lines with paired response and expression data	944
Training cell lines	660
Validation cell lines	142
Test cell lines	142
Training drug–cell-line pairs	140,372
Validation drug–cell-line pairs	29,639
Test drug–cell-line pairs	29,892

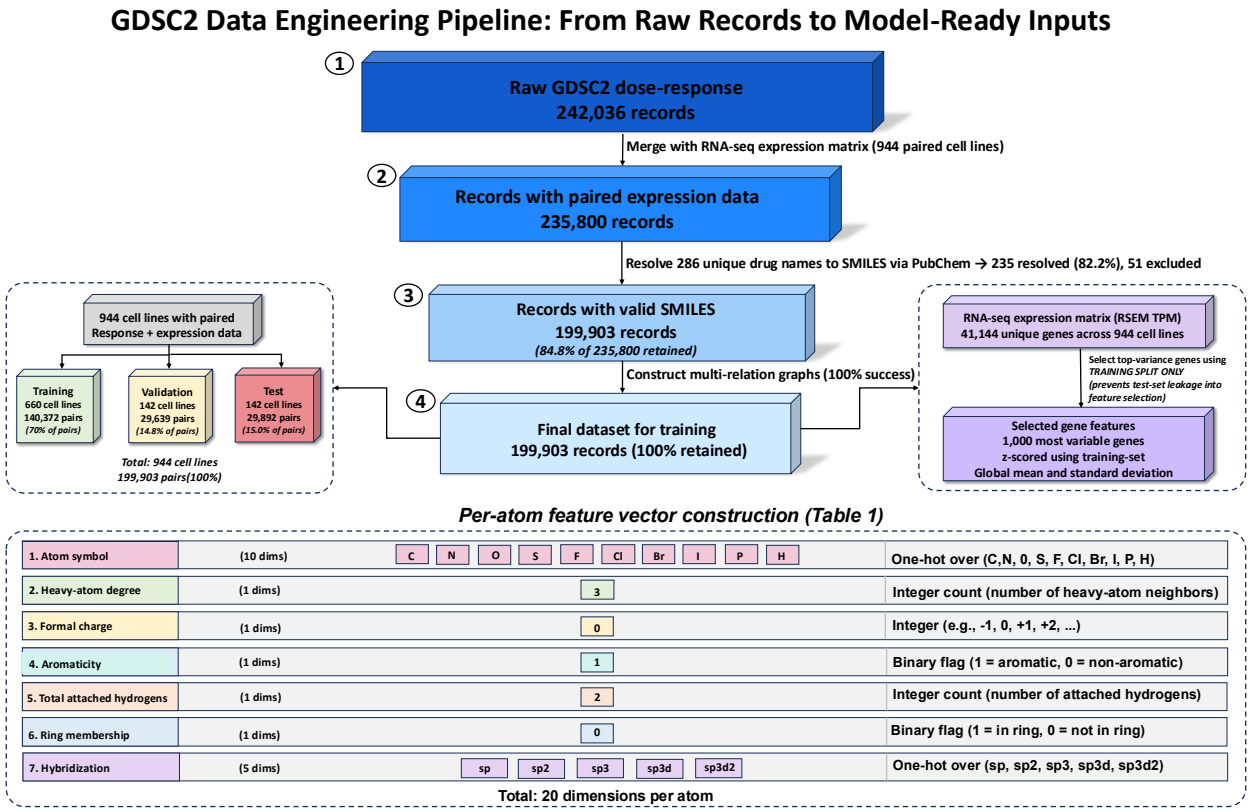


Figure 7: GDSC2 data engineering pipeline, from raw dose–response records to model-ready inputs: filtering by expression availability and SMILES resolution, then a cell-line-level train/validation/test split.

4.2. Implementation details

Table 4 summarizes the training configuration. The full model has 236,665 trainable parameters (58,488 drug encoder, 145,152 cell-line encoder, 33,025 fusion/output). Training uses early stopping (patience 5), restoring best weights before evaluation. All models are trained on CPU under Python 3.11.4, PyTorch 2.7.1, RDKit 2026.03.4, a full run completing in under an hour.

Table 4: Implementation and training configuration.

Parameter	Value
Optimizer	Adam
Learning rate	1×10^{-3}
Weight decay	1×10^{-5}
Loss function	Mean squared error
Batch size	32
Max epochs	20
Early stopping patience	5 (on validation loss)
Hidden dimension (C_{out})	128
Dropout rate	0.2
Attention LeakyReLU slope	0.2
Encoder LeakyReLU slope	0.01
Atom feature dimension	20
Gene feature dimension	1000
Functional-group SMARTS patterns	65
Trainable parameters	236,665
Random seed (primary run)	42
Compute device	CPU (no GPU)
Python version	3.11.4
PyTorch version	2.7.1
RDKit version	2026.03.4
Approx. training time per model	<1 hour (20 epochs)

4.3. Main results

On the held-out test set of 29,892 pairs, DeepRelCDR reaches $R^2 = 0.7617$. A bootstrap (1000 resamples) yields a 95% CI of $[0.7555, 0.7686]$ (Fig. 8), indicating stability rather than a small-sample artifact.

This quantifies sampling uncertainty, not training-run variance. Retraining with two additional seeds (7, 123) gives $R^2 = 0.7619 \pm 0.0030$ (range $[0.7590, 0.7650]$), consistent with seed-42 and an order of magnitude narrower than most ablation gaps (Section 4.5). The same protocol applies to every ablation configuration (Table 5).

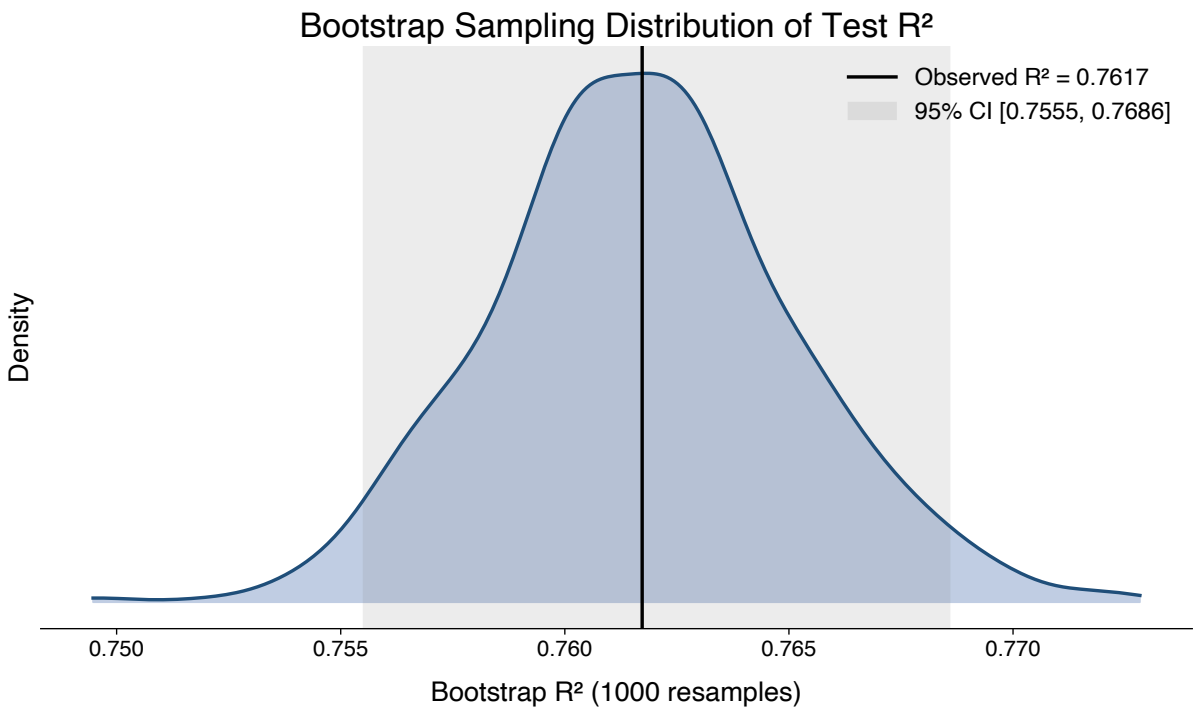


Figure 8: Bootstrap sampling distribution of test R^2 (1000 resamples). The observed $R^2 = 0.7617$ falls at the center, with a 95% confidence interval of $[0.7555, 0.7686]$.

Fig. 9 shows predicted versus actual LN_IC_{50} as a density plot, concentrated around the identity line ($y = \hat{y}$) where most pairs lie; a small fraction, mostly at range extremes, deviate more.

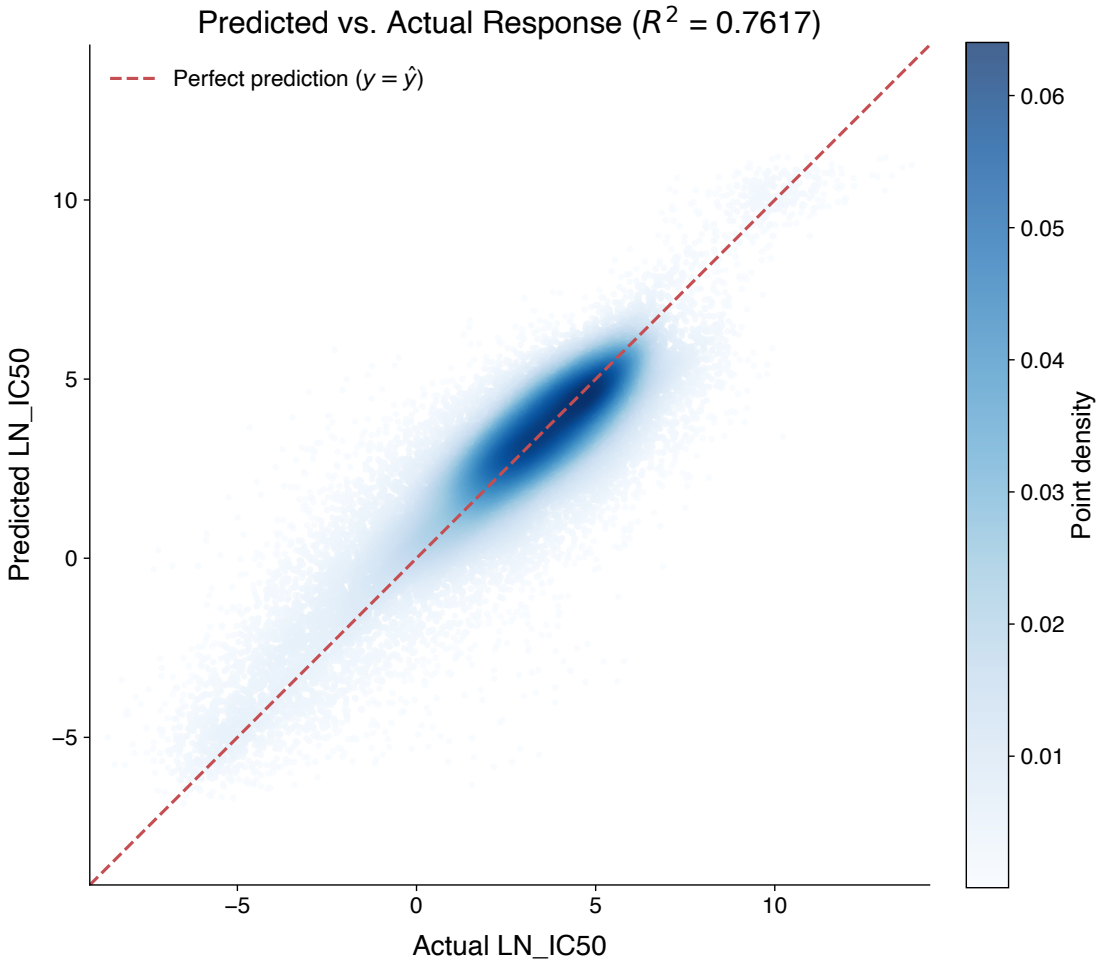


Figure 9: Predicted versus actual LN_IC_{50} for all 29,892 test-set pairs. The dashed line marks perfect prediction ($y = \hat{y}$); color indicates local point density.

4.4. Error analysis by cancer type and drug

Fig. 10 breaks down mean absolute error (MAE) by cancer type and drug. Error is not uniform: pancreatic/colorectal carcinoma with docetaxel and B-cell non-Hodgkin's lymphoma with nutlin-3a show higher MAE, while glioblastoma with acetalax is predicted with very low error, plausibly reflecting uneven GDSC2 representation and intrinsically variable dose-response for some drugs.

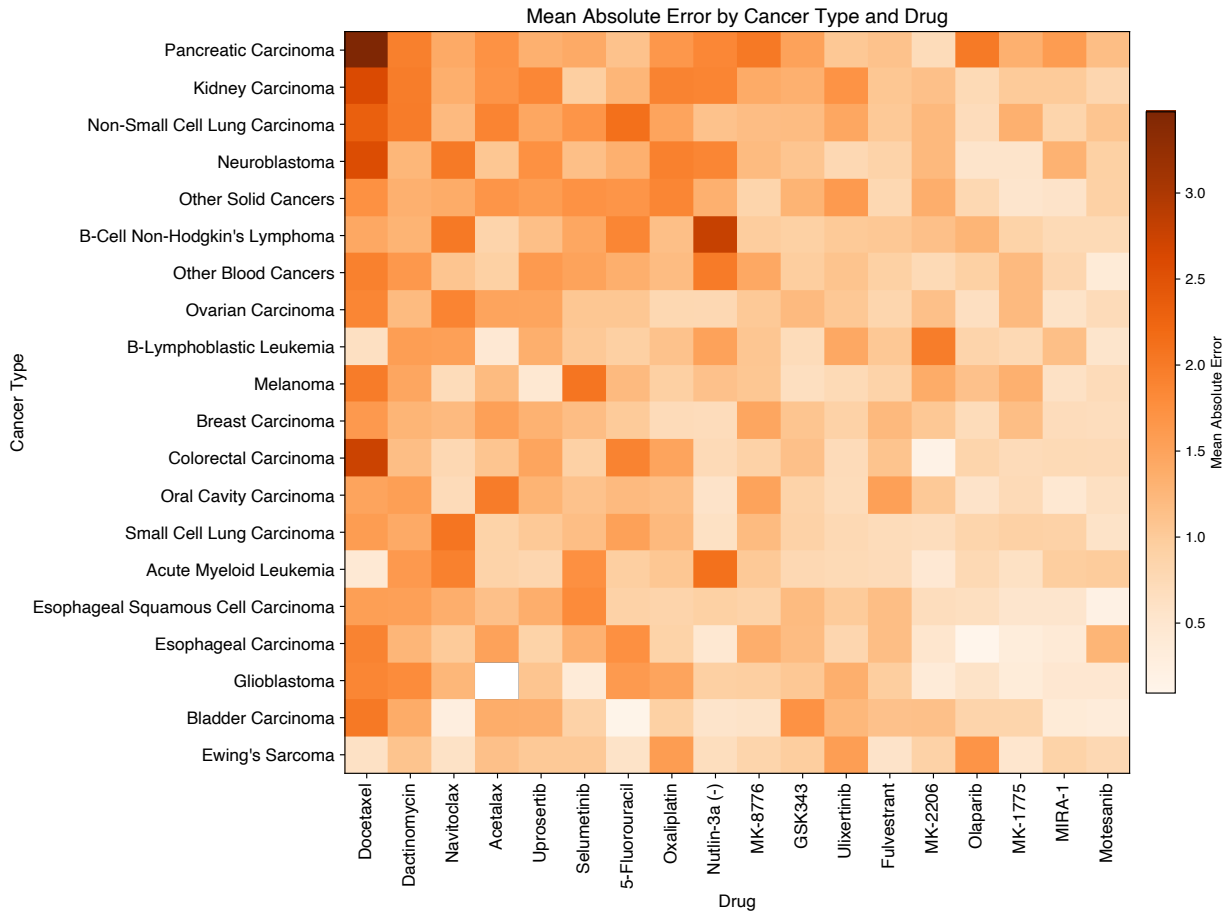


Figure 10: Mean absolute error on the test set, broken down by cancer type and drug. Darker cells indicate higher error.

4.5. Relation-level ablation study

To isolate each relation's contribution, we train three ablated variants restricted to a relation subset, plus a single-relation attention baseline (bond only, no decomposition), replacing every DeepRelCDR layer with a standard graph attention layer [5] scoring *transformed* features:

$$e_{ij} = \mathrm{LeakyReLU}_{0.2}\big(a^\top \big[Wx_i \,\|\, Wx_j\big]\big), \qquad W \in \mathbb{R}^{C_{in}\times C_{out}},\ a \in \mathbb{R}^{2C_{out}}, \tag{23}$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k\,:\,(i,k)\in\hat{\mathcal{E}}_{bond}} \exp(e_{ik})}, \tag{24}$$

aggregating transformed features directly,

$$h'_i = \sum_{j\,:\,(i,j)\in\hat{\mathcal{E}}_{bond}} \alpha_{ij}\, Wx_j \ \in\ \mathbb{R}^{C_{out}}, \tag{25}$$

coupling what decomposition removes. Each baseline layer contains

$$|\theta_{baseline\ layer}| = \underbrace{C_{in}C_{out}}_{\text{transform } W} + \underbrace{2C_{out}}_{\text{attention vector } a} \tag{26}$$

parameters versus $3 \cdot 2C_{in} + 3C_{in}C_{out} + C_{out}$ for DeepRelCDR: a 19,968-parameter baseline encoder versus 58,488, smaller from one relation and no fusion stage, not weaker approximation. Table 5 reports test R^2 per configuration.

Table 5: Relation-level ablation on the GDSC2 test set. The seed-42 column is the primary run reported elsewhere; the mean ± SD column summarizes three seeds (42, 7, 123).

Configuration	Test R^2 (seed 42)	Mean ± SD (3 seeds)
Single-relation GAT (bond only)	0.7020	0.7017 ± 0.0035
Bond + Functional Group	0.7193	0.7193 ± 0.0033
Bond + Ring	0.7529	0.7526 ± 0.0030
Bond + Ring + Functional Group (full DeepRelCDR)	0.7617	0.7619 ± 0.0030

Every relation added to bond-only improves accuracy, so both carry information not recoverable from bonds alone. Ring contributes more: Bond+Ring reaches 0.7529 versus 0.7193 for Bond+FG; removing ring costs $\Delta = -0.0424$, over four times removing functional-group ($\Delta = -0.0088$). The full model outperforms every ablation, so relations are complementary, not redundant.

These gaps exceed training noise: two additional seeds give SD of 0.0030–0.0035, an order of magnitude smaller. The full model exceeds the single-relation baseline by 22.6 standard errors and Bond+Ring by 3.8 (smallest margin); every other gap exceeds 13. With SD from only three runs, these indicate a large, consistently-signed gap rather than a precise significance level; ring's contribution is directionally unambiguous, functional-group's consistently positive across seeds.

4.5.1. Is attention necessary, or would unweighted aggregation suffice?

Every configuration above uses learned attention. A separate question is whether accuracy mainly reflects having three relations regardless of aggregation. We train a No-Attention variant, keeping the architecture unchanged but replacing α_{ij}^r with a fixed, unweighted average,

$$\alpha_{ij}^r = \frac{1}{\left|\{k : (i,k) \in \hat{\mathcal{E}}_r\}\right|}, \qquad \forall (i,j) \in \hat{\mathcal{E}}_r, \tag{27}$$

so Eq. (7) reduces to a degree-normalized mean with no learnable a_r. This reaches $R^2 = 0.7135$, below the full model (0.7617) and even Bond+FG (0.7193) despite all three relations: exposing more relations alone is not sufficient; learned attention recovers 4.8 points of R^2 over unweighted averaging at fixed relation count.

For Nutlin-3a, ring contributes 140 edges versus 52 for functional-group from the same 25 atoms, suggesting a larger receptive field explanation for ring's cost. Tested across 228 training-set drugs, correlating E_{ring}/E_{fg} with each drug's cost of removing ring, the quantities are not positively correlated (Pearson $r = -0.076$, $p = 0.25$; Spearman $r = -0.114$, $p = 0.086$), indistinguishable from zero. We therefore do not treat edge count as explaining ring's larger cost; the source remains open beyond ring co-membership marking cyclic, aromatic, conformationally constrained classes not distinguished by bond connectivity alone.

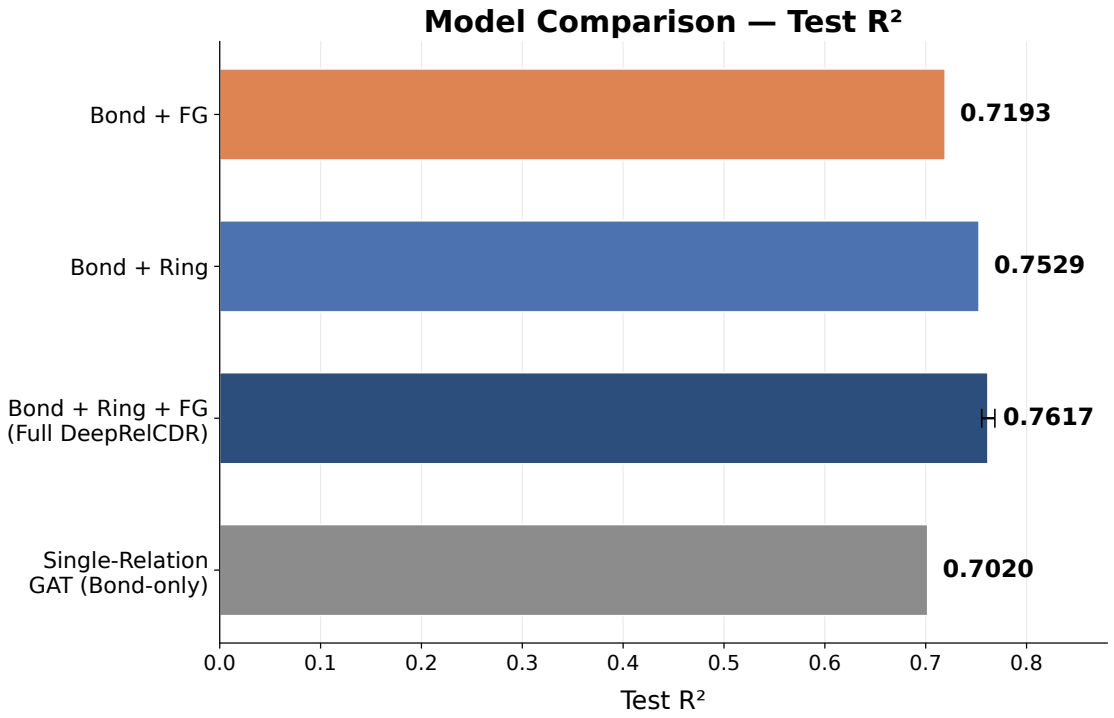


Figure 11: Test R^2 for the four relation-level ablation configurations of Table 5, with the error bar on the full model showing the 95% bootstrap confidence interval from Fig. 8.

4.6. Comparison against external baselines

The ablation isolates the relation set within the DeepRelCDR family. To situate this against outside approaches, we train three baselines under conditions identical to DeepRelCDR (same split, gene selection, optimizer, protocol; Section 4.2).

The first is Ridge regression on a 1024-bit Morgan fingerprint (radius 2) plus the 1000-gene vector. The second reimplements GraphDRP [7] from its description: a three-layer GCN over the bond graph with max pooling, a 1D-CNN gene encoder, fused via MLP. The third reconstructs DeepAEG [20]: SMILES recombination augmentation, a hybrid GCN with joint atom–bond updates, and a Transformer SMILES encoder. All are trained end-to-end with the same MSE loss.

Table 6: Comparison against external baselines trained under identical conditions (same GDSC2 split, gene selection, and training protocol as DeepRelCDR) on the held-out test set.

Configuration	Test R^2
Ridge regression (Morgan fingerprint + expression)	0.6259
GraphDRP-style (3-layer GCN + 1D-CNN) [7]	0.6948
DeepAEG-style (augmented graph + Transformer) [20]	0.7350
Full DeepRelCDR (this work)	0.7617

DeepRelCDR outperforms GraphDRP-style by 6.7 points of R^2, DeepAEG-style by 2.7, and Ridge by 13.6 (Table 6). DeepAEG-style closing roughly two-thirds of the gap shows re-

cent refinements (augmentation, joint updates, Transformer encoding) recover accuracy without multi-relation structure, but not entirely; the remaining margin is consistent with ring and functional-group structure contributing beyond encoder capacity alone. With pipeline, split, and protocol fixed across Table 6, these gaps reflect architecture rather than dataset differences. These are good-faith reimplementations, not expected to exactly reproduce original accuracy under a different population; they indicate relative architecture design under one shared protocol.

4.7. Relation contribution diagnostics

The ablation measures contribution by removing and retraining; here we probe the trained model directly via two diagnostics: mean attention-weighted output magnitude $\|z_i^r\|$ per relation, averaged over all atoms, and the ℓ_2 norm of pointwise fusion matrix Θ's columns per relation. Table 7 and Fig. 12 report both as percentages.

Table 7: Relation contribution diagnostics on the trained full model, by two independent metrics (percentage of total across the three relations).

Metric	Bond	Ring	Functional Group
Mean attn-weighted output magnitude	34.2%	32.3%	33.5%
Pointwise fusion weight norm	36.2%	32.4%	31.4%

Both diagnostics place all three relations within a narrow band (31–36%), bond contributing somewhat more, none dominating or near-zero, consistent with all three being used, though measuring internal signal flow rather than ablation-deletion cost.

4.8. Sensitivity to hyperparameters

Main results use 1000 top-variance genes and $C_{out} = 128$ (Table 4). We retrain, holding all else fixed, varying (i) gene count and (ii) C_{out}.

Table 8: Sensitivity of test R^2 to the number of selected genes and to the hidden dimension, holding all other hyperparameters and the data split fixed.

Configuration	Test R^2
Genes = 500	0.7541
Genes = 1000 (main configuration)	0.7617
Genes = 2000	0.7649
Hidden dim. = 64	0.7486
Hidden dim. = 128 (main configuration)	0.7617
Hidden dim. = 256	0.7638

Table 8 shows test R^2 changes by at most 1.5 points across a 4-fold gene-count range and 1.3 across hidden dimension, with diminishing returns. The main configuration was not a narrow optimum; nearby settings stay near $R^2 = 0.7617$, so ablation conclusions are unlikely to be an artifact of this choice.

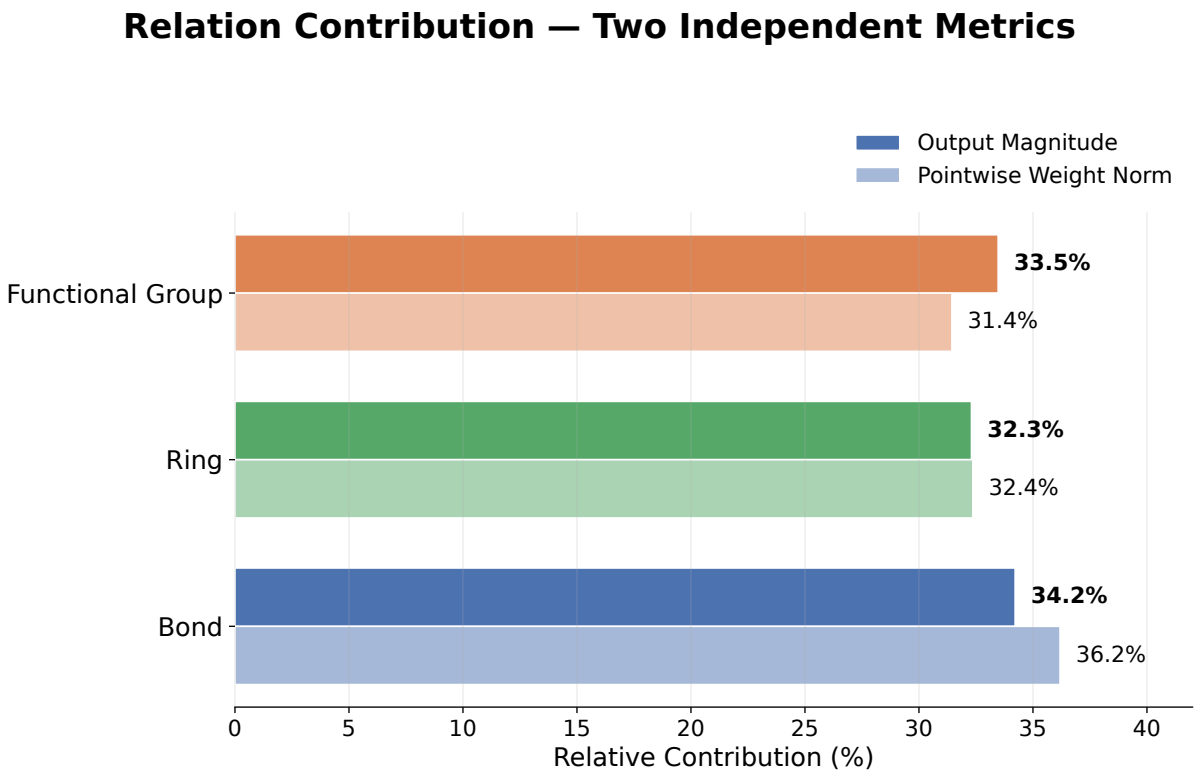


Figure 12: Relative contribution of each relation to the trained model, by mean attention-weighted output magnitude and by pointwise fusion weight norm (Table 7).

4.9. Biological validation: SLFN11 gene attribution

Schlafen-11 (SLFN11) is a putative DNA/RNA helicase whose expression is a consistently reported determinant of sensitivity to DNA-damaging chemotherapy, including camptothecin [31]. As an independent check, we compute gradient-based gene attribution for camptothecin predictions across three cancer types: NSCLC ($n = 15$), SCLC ($n = 11$), breast carcinoma ($n = 8$), reporting the top five genes (Table 9).

Fig. 13 presents these rankings as bar charts.

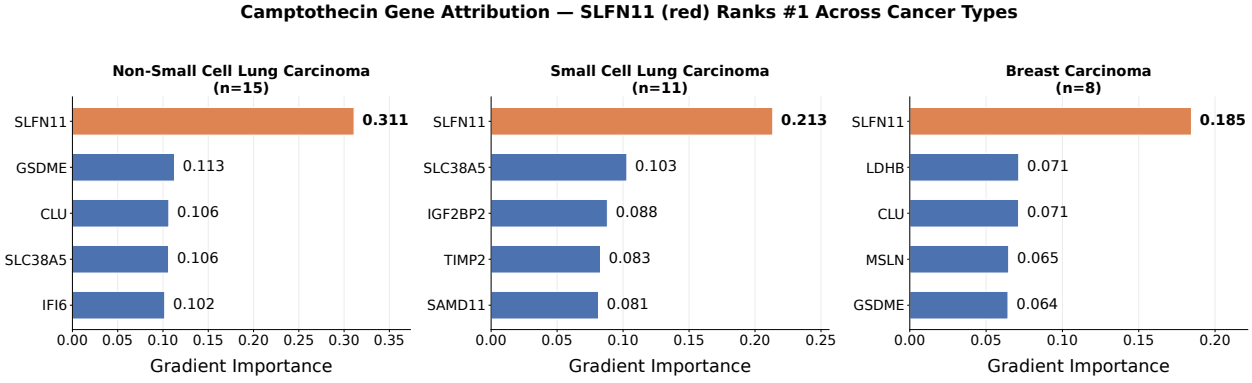


Figure 13: Gradient-based gene attribution for camptothecin sensitivity, by cancer type. SLFN11 (highlighted) ranks first by a wide margin in all three (Table 9).

Table 9: Top-5 gene attributions for camptothecin sensitivity, by cancer type. SLFN11 ranks first in every cancer type tested.

Cancer type (n)	Gene	Score
NSCLC (15)	SLFN11	0.311
	GSDME	0.113
	CLU	0.106
	SLC38A5	0.106
	IFI6	0.102
SCLC (11)	SLFN11	0.213
	SLC38A5	0.103
	IGF2BP2	0.088
	TIMP2	0.083
	SAMD11	0.081
Breast carcinoma (8)	SLFN11	0.185
	LDHB	0.071
	CLU	0.071
	MSLN	0.065
	GSDME	0.064

SLFN11 ranks first by a wide margin in all three cancer types, with attribution 2.1–2.8 times the second-ranked gene. With no SLFN11-specific supervision, this was not targeted; we interpret its recovery as evidence the learned gene-level signal is biologically grounded.

4.10. Atom-level interpretability: ring-relation attention

To probe the drug side, Fig. 14 visualizes, for four drugs, the per-atom output magnitude of the ring-relation attention branch overlaid on structure. Darker markers indicate atoms weighted more heavily; self-loops give every atom a nonzero magnitude, so this reflects self- versus neighbor-contribution balance, not a strict on/off indicator. Magnitude varies atom to atom: (a) LMP744 and (c) tretinoin each have one dominant cyclic region, while (b) staurosporine and (d) Telomerase Inhibitor IX vary internally across several ring systems, a qualitative diagnostic rather than a universal rule.

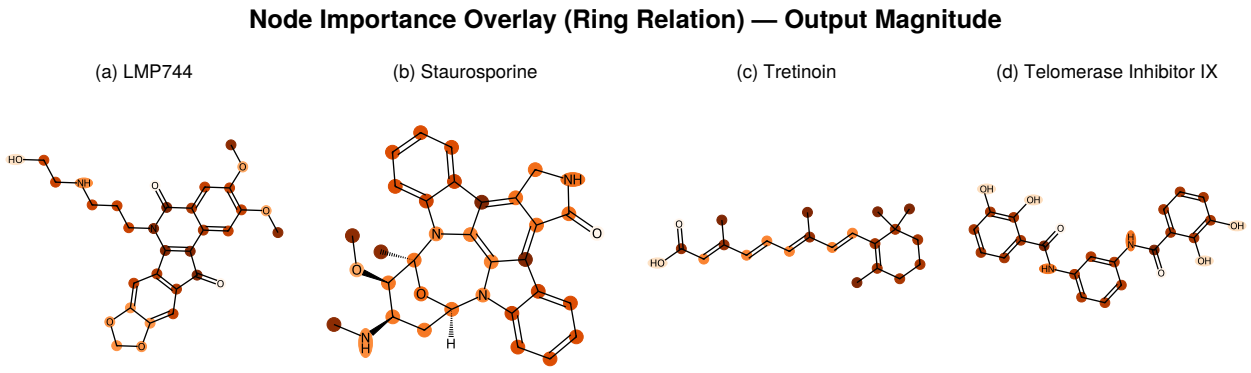


Figure 14: Per-atom output magnitude of the ring-relation attention branch for four drugs: (a) LMP744, (b) staurosporine, (c) tretinoin, (d) Telomerase Inhibitor IX. Darker markers indicate higher magnitude.

4.11. Clinical decision support: personalized drug ranking

To assess practical utility, we simulate personalized drug selection on the 142 test-set cell lines screened against at least two drugs: the personalized ranking selects the drug with highest predicted sensitivity per cell line; the non-personalized baseline always selects the best average training-set drug (5-Fluorouracil). A recommendation is a hit if actual response falls below the drug's training-derived median LN_IC_{50}.

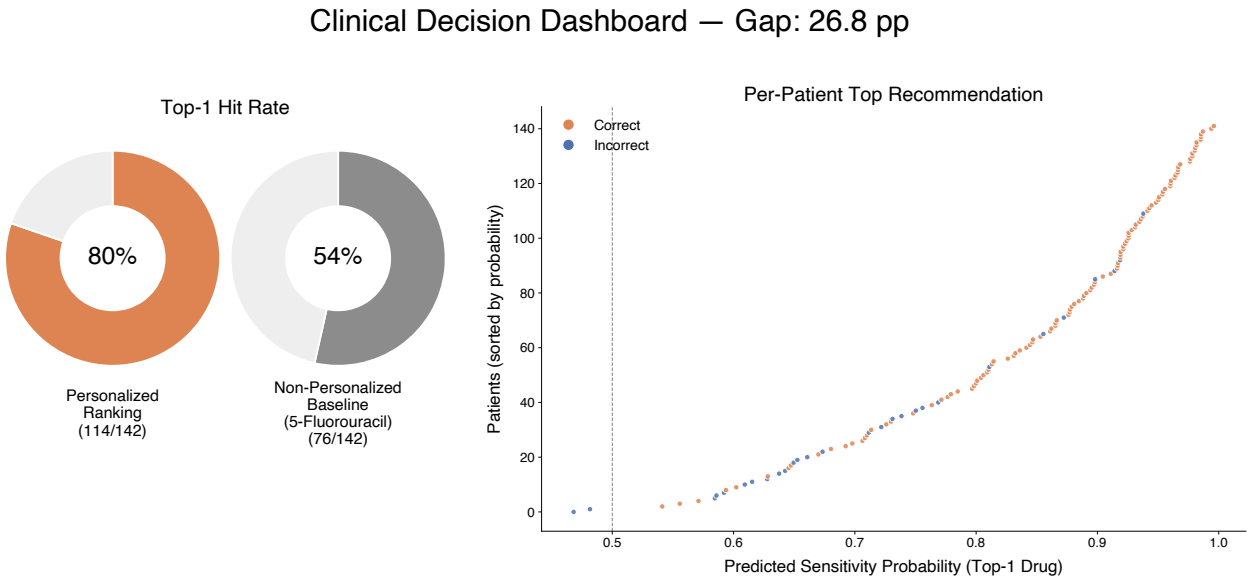


Figure 15: Personalized versus non-personalized drug recommendation on 142 test-set cell lines. Left: top-1 hit rate (80% vs. 54%). Right: per-patient sensitivity probability, colored by correctness.

Personalized ranking achieves 80.3% (114/142) versus 53.5% (76/142) for baseline, a 26.8-point gap. Among 52 disagreeing cell lines, personalized is correct and baseline wrong on 45, reverse on 7 (McNemar's, $p < 10^{-6}$); bootstrap gives a 95% CI of [17.6, 35.9] points. Correct recommendations concentrate at higher probabilities, incorrect near 0.5, so confidence is informative. This is an illustrative clinical-utility estimate, not a validated tool (Section 5).

4.12. Generalization to other pharmacogenomic resources

All results so far use GDSC2. To assess generalization, we apply the identical pipeline and protocol (Table 4) to two independent screens: CCLE [32] and CTRPv2 [33], retraining from scratch on each since they profile different drugs and cell lines. Table 10 summarizes each after the same filtering pipeline and the resulting test R^2.

Table 10: DeepRelCDR retrained from scratch, under an identical pipeline and protocol, on two additional pharmacogenomic resources and on GDSC2-internal controls.

Resource	Drugs	Cell lines	Final pairs	Test R^2
GDSC2 (main; Section 4.1)	235	944	199,903	0.7617
GDSC2 (24-drug subset, seed 42)	24	944	22,785	0.8378
GDSC2 (24-drug subset, 5 seeds)	24	944	—	0.785 ± 0.060
GDSC2 (sparsity-matched to CTRPv2)	235	944	122,307	0.7476
CCLE [32]	24	504	10,184	0.6812
CTRPv2 [33]	546	886	266,709	0.6687

DeepRelCDR reaches positive, comparable-order R^2 on both resources, though 8–9 points lower than GDSC2. To situate CCLE's accuracy, we ran a drug-count-matched control: five random draws of 24 GDSC2 drugs, matching the pipeline (including gene re-selection) and retraining. Test R^2 ranged 0.6692–0.8378 (mean 0.785, SD 0.060), an order of magnitude more variable than three-seed training noise; the lowest draw is close to CCLE's own accuracy (0.6812).

For CTRPv2, we tested whether its sparser matrix (55.1% fill versus GDSC2's 90.1%) explains its lower accuracy: subsampling GDSC2 rows to match this fill rate reaches $R^2 = 0.7476$, only 1.4 points below the full run, versus CTRPv2's 9.3-point gap.

4.13. Generalization beyond drug-response prediction

CCLE/CTRPv2 hold the *task* fixed while varying resource. A separate question is whether the relation-decomposition principle transfers to an unrelated molecular-property task with no cell-line input. We attach the unchanged drug encoder to a linear output head, applying it to ESOL [34], 1128 compounds with measured aqueous solubility ($\log S$), using identical features and protocol, 70/15/15 split (789/169/170; `random_state`=42), against a Ridge baseline on Morgan fingerprints.

Table 11: DeepRelCDR's drug encoder, with no changes beyond a linear output head, transferred to an unrelated molecular-property prediction task (ESOL aqueous solubility).

Configuration	Test R^2
Ridge regression (Morgan fingerprint)	0.3842
DeepRelCDR drug encoder (this work)	0.7391

The drug encoder reaches $R^2 = 0.7391$ versus 0.3842 for Ridge (Table 11), with no architectural change and no cancer-specific supervision. Since ESOL is physicochemical with no gene-expression input, this addresses a different question than CCLE/CTRPv2: evidence the design is general-purpose, not pharmacogenomics-specific. This is a single split on one benchmark, not MoleculeNet-wide, indicating transfer rather than characterizing where best; a broader suite is future work.

5. Discussion

The results support two hypotheses. The ablation (Section 4.5) shows explicit ring and functional-group relations substantially improve accuracy (0.7020 to 0.7617 R^2), not achievable through either alone (Bond+Ring: 0.7529; Bond+FG: 0.7193) or without learned attention (No-Attention: 0.7135, Section 4.5.1): the relations are complementary, and aggregation matters as much as relation count. SLFN11 attribution and personalized ranking provide independent, non-accuracy signals of usable representations. DeepRelCDR exceeds GraphDRP-style by 6.7 points and DeepAEG-style by 2.7 under identical protocol (Section 4.6), so accuracy reflects architecture, not merely a neural network, or a recent one. Table 1 positions DeepRelCDR against prior work; these results fill in its "Evaluated" column.

Two further properties matter. Because ring and functional-group membership are explicit, named relations, attention weights are interpretable atom by atom (Section 4.10), not via post-hoc filter interpretation. The depthwise/pointwise decomposition keeps the representation parameter-efficient: the full model matches the single-relation baseline's encoder size, so gains reflect structure, not size. The relation set was fixed a priori; 3D distance or stereochemistry are not modeled. Learned attention is load-bearing: removing it while keeping all three relations costs 4.8 points of R^2.

A third property concerns testing rather than proposing mechanisms: Nutlin-3a's edge counts (140 ring vs. 52 functional-group) suggested an explanation for ring's ablation cost, but across 228 drugs no correlation supports this (Pearson $r = -0.076$, $p = 0.25$; Spearman $r = -0.114$, $p = 0.086$); we report this null result rather than the anecdote. A fourth illustrates single-draw hazards: a drug-count control initially appeared to rule out drug count as explaining CCLE's lower accuracy ($R^2 = 0.8378$ on one draw), but five seeds gave 0.6692–0.8378 (mean 0.785, SD 0.060), the lowest close to CCLE's own accuracy (0.6812), retracting the initial reading; a sparsity-matched CTRPv2 control explains only 1.4 of its 9.3-point gap. A fifth concerns cross-task generalization: the unmodified encoder reaches $R^2 = 0.7391$ on ESOL versus 0.3842 for Ridge, evidence the design is general-purpose, not pharmacogenomics-specific.

Limitations: the 1000-gene set is selected by training-set variance alone ($\leq$1.5-point sensitivity, Section 4.8, though clinically relevant genes beyond SLFN11 may still be excluded); main results use a single split, though bootstrap and three-seed variance are quantified, and the architecture retrains at reduced accuracy on CCLE/CTRPv2 without hyperparameter changes, establishing robustness to resource choice, not zero-shot transfer; the personalized-ranking analysis uses a training-derived threshold rather than patient outcomes; and the external comparison omits DeepCDR [8] and a generic R-GCN-style baseline [25], and self-reported numbers, since protocols are not always comparable across the literature [35, 36].

Future work: learning or expanding the relation set; extending baselines to DeepCDR and R-GCN-style models under a shared protocol; k-fold cross-validation; characterizing favorable drug subsets given the 0.669–0.838 variance; identifying CTRPv2's remaining gap beyond sparsity; broadening cross-task evaluation beyond ESOL; and validating personalized ranking against patient outcomes.

6. Conclusion

We introduced DeepRelCDR, a deep relation-aware molecular graph attention framework for cancer drug response prediction that represents each drug as three explicit relations, covalent bonds, ring co-membership, and functional-group co-membership, rather than the single bond-only graph used by most prior GNN-based methods, with each relation aggregated independently and fused by a shared pointwise layer. Trained on 235 drugs and 199,903 drug–cell-line pairs from GDSC2, DeepRelCDR reaches a test R^2 of 0.762 (95% CI [0.756, 0.769]), exceeding a single-relation ablation by 6.0 points, and, under identical conditions, a Ridge baseline by 13.6, a GraphDRP-style reimplementation by 6.7, and a 2024-era DeepAEG-style reimplementation by 2.7 points of R^2. Learned attention weighting alone accounts for 4.8 of these points. The architecture retrains successfully on CCLE and CTRPv2, and its unmodified drug encoder reaches $R^2 = 0.7391$ on an unrelated task, ESOL solubility prediction, versus 0.3842 for Ridge. Gradient-based gene attribution independently recovers SLFN11 as the top camptothecin-sensitivity gene across three cancer types, and a personalized drug-ranking analysis reaches an 80.3% top-1 hit rate versus 53.5% for a non-personalized baseline. Future work includes learning the relation set rather than fixing it a priori and extending baseline comparisons to DeepCDR and R-GCN-style models under a shared protocol.

CRediT authorship contribution statement

Sarvin Chitsaz: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. Majid Sepahvand: Supervision, Validation, Writing – review & editing. Maytham N. Meqdad: Validation, Writing – review & editing. Masoud Daneshtalab: Supervision, Funding acquisition, Writing – review & editing.

Data availability

The GDSC2 drug sensitivity data are publicly available at https://www.cancerrxgene.org. Processed data and code are available at https://github.com/SarvinChitsaz/DeepRel-CDR/tree/main/assets/results/model_comparison.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI use

The authors declare that no generative artificial intelligence (AI) tools were used in the preparation, writing, editing, or revision of this manuscript.

References

[1] W. Yang, et al., Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, Nucleic Acids Research 41 (D1) (2013) D955–D961.

[2] F. Iorio, et al., A landscape of pharmacogenomic interactions in cancer, Cell 166 (3) (2016) 740–754.

[3] G. Landrum, RDKit: Open-source cheminformatics, https://www.rdkit.org (accessed 2026).

[4] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: International Conference on Learning Representations (ICLR), 2017.

[5] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, in: International Conference on Learning Representations (ICLR), 2018.

[6] D. Baptista, P.G. Ferreira, M. Rocha, Deep learning for drug response prediction in cancer, Briefings in Bioinformatics 22 (1) (2021) 360–379.

[7] T. Nguyen, G.T.T. Nguyen, T. Nguyen, D.-H. Le, Graph convolutional networks for drug response prediction, IEEE/ACM Transactions on Computational Biology and Bioinformatics 19 (1) (2022) 146–154.

[8] Q. Liu, Z. Hu, R. Jiang, M. Zhou, DeepCDR: a hybrid graph convolutional network for predicting cancer drug response, Bioinformatics 36 (Supplement_2) (2020) i911–i918.

[9] T. Ma, Q. Liu, H. Li, M. Zhou, R. Jiang, X. Zhang, DualGCN: a dual graph convolutional network model to predict cancer drug response, BMC Bioinformatics 23 (2022) 129.

[10] X. Liu, C. Song, F. Huang, H. Fu, W. Xiao, W. Zhang, GraphCDR: a graph neural network method with contrastive learning for cancer drug response prediction, Briefings in Bioinformatics 23 (1) (2022) bbab457.

[11] L. Jiang, C. Jiang, X. Yu, R. Fu, S. Jin, X. Liu, DeepTTA: a transformer-based model for predicting cancer drug response, Briefings in Bioinformatics 23 (3) (2022) bbac100.

[12] C. Song, X. Liu, Z. Xiong, L. Liu, W. Zhang, GNNDRP: graph neural network with multi-task learning for drug response prediction, IEEE Transactions on Computational Biology and Bioinformatics 22 (3) (2025) 1052–1059.

[13] Z. Asgari, M.H. Eslami, A. Kianfar, K. Abbasi, M. Barshan Tashnizi, H.R. Rabiee, DeepTraCDR: prediction of cancer drug response using multimodal deep learning with transformers, Knowledge-Based Systems 352 (Part A) (2026) 116933.

[14] G.T. Nguyen, H.D. Vu, D.-H. Le, Integrating molecular graph data of drugs and multiple-omic data of cell lines for drug response prediction, IEEE/ACM Transactions on Computational Biology and Bioinformatics 19 (2) (2022) 710–717.

[15] Y. Zhu, et al., TGSA: protein–protein association-based twin graph neural networks for drug response prediction with similarity augmentation, Bioinformatics 38 (2) (2022) 461–468.

[16] T. Chu, T.T. Nguyen, B.D. Hai, Q.H. Nguyen, T. Nguyen, Graph transformer for drug response prediction, IEEE/ACM Transactions on Computational Biology and Bioinformatics (2022) 1–9.

[17] Y. Yang, P. Li, GPDRP: a multimodal framework for drug response prediction with graph transformer, BMC Bioinformatics 24 (2023) 484.

[18] Z. Huang, P. Zhang, L. Deng, DeepCoVDR: deep transfer learning with graph transformer and cross-attention for predicting COVID-19 drug response, Bioinformatics 39 (Suppl. 1) (2023) i475–i483.

[19] X. Liu, W. Zhang, A subcomponent-guided deep learning method for interpretable cancer drug response prediction, PLOS Computational Biology 19 (3) (2023) e1011382.

[20] C. Lao, P. Zheng, H. Chen, Q. Liu, F. An, Z. Li, DeepAEG: a model for predicting cancer drug response based on data enhancement and edge-collaborative update strategies, BMC Bioinformatics 25 (2024) 105.

[21] Y. Dong, Y. Zhang, Y. Qian, Y. Zhao, Z. Yang, X. Feng, ASGCL: adaptive sparse mapping-based graph contrastive learning network for cancer drug response prediction, PLOS Computational Biology 21 (1) (2025) e1012748.

[22] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, G.E. Dahl, Neural message passing for quantum chemistry, in: Proceedings of the 34th International Conference on Machine Learning (ICML), PMLR 70, 2017, pp. 1263–1272.

[23] W.L. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, in: Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 1024–1034.

[24] K. Xu, W. Hu, J. Leskovec, S. Jegelka, How powerful are graph neural networks?, in: International Conference on Learning Representations (ICLR), 2019.

[25] M. Schlichtkrull, T.N. Kipf, P. Bloem, R. van den Berg, I. Titov, M. Welling, Modeling relational data with graph convolutional networks, in: European Semantic Web Conference (ESWC), 2018, pp. 593–607.

[26] X. Li, J. Gui, Q. Gao, H. Shi, Z. Yue, Controllable edge-type-specific interpretation in multi-relational graph neural networks for drug response prediction, arXiv preprint arXiv:2408.17129 (2024).

[27] H. Wang, A. Zhang, Y. Zhong, J. Tang, K. Zhang, P. Li, Chain-aware graph neural networks for molecular property prediction, Bioinformatics 40 (10) (2024) btae574.

[28] T. Nguyen, D.D. Nguyen, MMGNN: multi-level, multi-color graph neural networks for molecular property prediction, arXiv preprint arXiv:2606.20906 (2026).

[29] A.G. Howard, et al., MobileNets: efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017).

[30] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, Journal of Chemical Information and Computer Sciences 28 (1) (1988) 31–36.

[31] G. Zoppoli, et al., Putative DNA/RNA helicase Schlafen-11 (SLFN11) sensitizes cancer cells to DNA-damaging agents, Proceedings of the National Academy of Sciences 109 (37) (2012) 15030–15035.

[32] J. Barretina, et al., The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity, Nature 483 (7391) (2012) 603–607.

[33] A. Basu, et al., An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules, Cell 154 (5) (2013) 1151–1161.

[34] J.S. Delaney, ESOL: estimating aqueous solubility directly from molecular structure, Journal of Chemical Information and Computer Sciences 44 (3) (2004) 1000–1005.

[35] A. Partin, et al., Benchmarking community drug response prediction models: datasets, models, tools, and metrics for cross-dataset generalization analysis, Briefings in Bioinformatics 27 (1) (2026) bbaf667.

[36] J. Bernett, P. Iversen, M. Picciani, M. Wilhelm, K. Baum, M. List, From hype to health check: critical evaluation of drug response prediction models with DrEval, bioRxiv preprint 2025.05.26.655288 (2025).