RECOVERY OVER AGREEMENT: CROSS-VIEW MASKED GRAPH LEARNING FOR DRUG REPOSITIONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Computational drug repositioning seeks new indications for existing drugs, yet drug–disease association (DDA) prediction is hampered by sparse labels and by evidence scattered across heterogeneous entity attributes and auxiliary relations. Existing multi-view methods encode attribute- and relation-derived graphs independently and align them only through agreement objectives, which reward redundancy rather than complementarity; they typically leave supervised edges visible to relation message passing, and multi-task variants force DDA, drug–protein, and disease–protein prediction onto a single fused representation. We propose **Cross-view Attribute–Relation Exchange (CARE)**, which replaces cross-view agreement with cross-view reconstruction. Under a single node- and edge-masking process applied only to the relation encoder, the relation view reconstructs masked entity representations. In contrast, the attribute view ranks masked associations above type-compatible unknown pairs, so each view must supply what the other withholds and target edges never enter relation message passing. A type-aware contrastive objective keeps entity semantics consistent across views, and task-conditioned gates let drug–protein and disease–protein supervision strengthen the shared encoders without imposing a common fusion on the target task. On three benchmarks, CARE consistently outperforms existing methods, with the largest gains in recall, indicating that our approach helps recover associations that sparse relational evidence alone misses and showing the benefit of coupling multi-task learning with reciprocal cross-view reconstruction for sparse DDA prediction.

1 INTRODUCTION

Developing a new drug through conventional pipelines demands substantial time, cost, and clinical risk (Li et al., 2016). Drug repositioning offers a complementary route by finding new indications for compounds whose pharmacology and safety are already partly known (Pushpakom et al., 2019; Hua et al., 2022), centering on drug–disease association (DDA) prediction, which ranks candidate links for experimental follow-up. The task is hard because known DDAs cover only a small fraction of possible pairs, and an unrecorded pair usually reflects missing knowledge rather than a true negative. Computational methods increasingly cast drugs, diseases, and proteins as nodes in a biomedical graph, combining chemical fingerprints, disease phenotypes, protein sequences, and observed associations to propagate evidence beyond individual pairs (Zhao et al., 2022; Yu et al., 2021; Cai et al., 2021). Multi-view models build such graphs and use contrastive objectives to preserve shared information (Gao et al., 2023; Jia et al., 2025). At the same time, multi-task learning couples DDA with drug–protein interaction (DPI) and disease–protein association (DPA) prediction so protein-centered relations can inform repositioning (Sun et al., 2026; Huang et al., 2020; Cui et al., 2026).

These advances suggest the problem is less the number of sources than how they interact. Attributes and relations offer complementary but asymmetric evidence. Attribute similarity can support a candidate with sparse connections, while relation neighborhoods situate entities in shared biological context, yet neither suffices, as similarities can be noisy or task-dependent and relations remain sparse and biased toward well-studied entities. Existing multi-view pipelines typically build fixed graphs, encode each view separately, and fuse features only near the prediction head (Figure 1a), rewarding redundancy rather than requiring either view to recover what the other lacks. Worse,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

training labels often remain in the relation graph itself, letting the relation view partly read rather than infer its targets. Multi-task prediction adds a second constraint. DDA, DPI, and DPA share entities and mechanisms, yet need not draw on attribute and relation evidence equally; a single fused representation conflates what should pass between views with how much of each view a task should use. Together, these observations motivate this work's central question, whether attribute and relation views can recover one another without exposing supervised targets to relation message passing, while adapting their contributions to each task.

We answer this question with CARE, a cross-view masked graph framework for multi-task drug repositioning (Figure 1b). CARE maintains two views: a learnable attribute graph, induced by type-specific encoders over drug, disease, and protein attributes, and a typed relation graph over observed DDA, DPI, and DPA edges. A reciprocal masked autoencoder couples the views through a single node- and edge-masking process applied only to the relation-encoder input: the relation view must reconstruct masked entities' attribute representations from their relational context, while the clean attribute view must rank each masked association above type-compatible unknown pairs. Since the same masked edges serve as supervised positives for all three tasks, target edges never enter relation message passing. A type-restricted alignment loss further ties the two views together on unmasked entities, and a task-conditioned gate fuses attribute and relation features per relation type and feeds relation-specific decoders, with DPI and DPA providing auxiliary supervision for the primary DDA objective. We evaluate CARE on three benchmarks via ten-fold cross-validation, targeting DDA prediction, and observe consistently better performance across metrics, with CARE outperforming the strongest baseline in both AUC and AUPR. Ablations demonstrate the effectiveness of reciprocal masked reconstruction, type-restricted alignment, and auxiliary multi-task supervision.

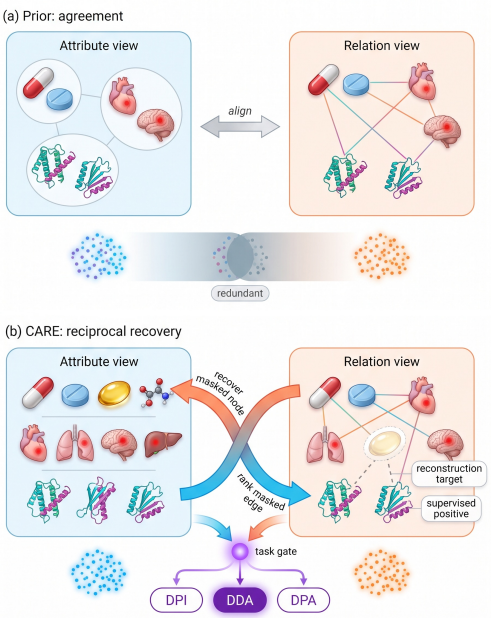

Figure 1: **Agreement versus recovery.** (a) Existing multi-view models encode fixed attribute and relation graphs separately and pull their outputs together, an objective satisfied when both views carry the *same* information, while target edges stay visible to message passing. (b) CARE masks the relation-view input once and asks each view to *recover* what the other withholds, then gates the two views per task.

- We introduce a two-view graph-learning framework for DDA prediction that keeps adaptive within-type attribute structure separate from observed cross-type biomedical relations, so heterogeneous evidence is not collapsed into a single representation too early.
- We replace cross-view agreement with cross-view *recovery*: under one shared corruption process, each view reconstructs what the other withholds, and target edges are never exposed to the relation encoder. This makes the two views more complementary.
- We develop a DDA-centered multi-task strategy in which relation-specific gates decide how much attribute and relational evidence each task uses. Thus, DPI and DPA supervision strengthens the shared entity representations without forcing all tasks to use a single fusion ratio.

2 Related Work

Graph Learning for Drug Repositioning. Drug–disease association (DDA) prediction is increasingly formulated as link prediction over heterogeneous biomedical graphs that link drugs and diseases through similarities, proteins, and other entities. HINGRL combines attribute and topological representations over a heterogeneous network (Zhao et al., 2022), while DRHGCN and REDDA use graph convolution and relation modeling to reconstruct drug–disease links (Cai et al., 2021; Gu et al., 2022). Recent methods improve encoder expressiveness from complementary directions: DAHNGC

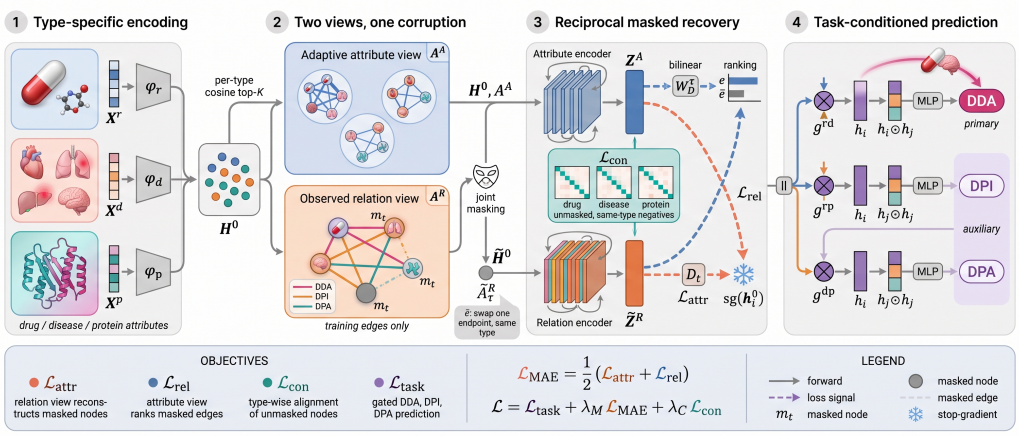

Figure 2: Overview of CARE. Type-specific encoders map drug, disease, and protein attributes into a shared latent space, inducing an adaptive within-type attribute view. At the same time, training-set DDA, DPI, and DPA links form a typed relation view. Node and edge masking corrupt only the relation-encoder input: the relation view reconstructs masked entity representations, and the clean attribute view ranks masked associations above type-compatible unknown pairs. Type-aware contrastive learning aligns unmasked entities across views, and task-conditioned prediction enhances DDA prediction.

jointly models homogeneous and heterogeneous structures (Zhong et al., 2023), GCNGAT combines convolution with attention for pair-specific representations (Yang et al., 2024), and Transformer-based models capture broader dependencies (He et al., 2024). Beyond DDA-specific graphs, KG-Predict integrates diverse biomedical entities in a knowledge graph for repurposing (Gao et al., 2022), while SFRLDDA enriches representations with semantic and functional similarities (Zhao et al., 2023), together confirming the value of heterogeneous context.

Multi-View Representation Learning. Multi-view learning exploits complementary structures from molecular similarity, biological attributes, and observed associations. SMGCL builds multiple similarity- and association-driven views and aligns them via graph contrastive learning (Gao et al., 2023), while DRGCL contrasts semantic and topological graph representations (Jia et al., 2025). Other work emphasizes view construction and supervision quality: hierarchical negative sampling distinguishes structurally different negatives for contrastive DDA learning (Wang et al., 2024); DRMAHGC combines heterogeneous graph augmentation, high-order representations, and positive-fusion negative sampling (Liu et al., 2024b); and MDGCN learns across multiple dependency graphs with contrastive objectives (Gan et al., 2025). Though these methods improve robustness under sparse associations, their auxiliary views are typically optimized for a single target relation, leaving other tasks' supervision only partially exploited.

Multi-Task Learning & Generalization. Multi-task learning directly transfers information across related biomedical relations. Collective factorization jointly models multiple association matrices (Huang et al., 2020), while MTLDR adds auxiliary drug-, disease-, protein-, and miRNA-related tasks to a shared graph representation (Ye et al., 2023). Most closely related, MTP-DDA jointly predicts drug–disease, drug–protein, and disease–protein associations, combining multi-view graph convolution with contrastive learning (Cui et al., 2026). TxGNN studies a complementary challenge: knowledge-graph-based repurposing under zero-shot and sparsely observed diseases (Huang et al., 2024). Despite these advances, multi-task DDA prediction remains nontrivial: auxiliary relations differ in density, semantics, and relevance to the target task, so naive parameter sharing or fixed weighting can introduce negative transfer. This motivates models that selectively transfer relation-specific knowledge while preserving complementary views for sparse DDA prediction.

3 The Proposed Method

We model drug repositioning using two complementary graph views over drugs, diseases, and proteins. As shown in Figure 2, CARE exchanges information through reciprocal masked reconstruction and aligns the two views with type-aware contrastive learning. Let $\mathcal{S} = \{r, d, p\}$ denote the

three entity types, where r, d, and p refer to drugs, diseases, and proteins, respectively. Their node sets are $\mathcal{V}^r = \{v_i^r\}_{i=1}^{N_r}$, $\mathcal{V}^d = \{v_i^d\}_{i=1}^{N_d}$, and $\mathcal{V}^p = \{v_i^p\}_{i=1}^{N_p}$. The complete node set is $\mathcal{V} = \mathcal{V}^r \cup \mathcal{V}^d \cup \mathcal{V}^p$, $N = N_r + N_d + N_p$. Each entity type provides a distinct attribute matrix $\mathbf{X}^t \in \mathbb{R}^{N_t \times F_t}$ for $t \in \mathcal{S}$. The feature dimensions F_t may differ because the three matrices describe different biological modalities. We represent observed drug–disease, drug–protein, and disease–protein relations by $\mathbf{Y}^{rd} \in \{0,1\}^{N_r \times N_d}$, $\mathbf{Y}^{rp} \in \{0,1\}^{N_r \times N_p}$, $\mathbf{Y}^{dp} \in \{0,1\}^{N_d \times N_p}$. An entry equal to one denotes an observed association. An entry equal to zero denotes a currently unobserved pair rather than a confirmed biological negative. Let $\mathcal{T} = \{rd, rp, dp\}$ denote the prediction tasks, and $\mathcal{P}^{rd} = \mathcal{V}^r \times \mathcal{V}^d$, $\mathcal{P}^{rp} = \mathcal{V}^r \times \mathcal{V}^p$, $\mathcal{P}^{dp} = \mathcal{V}^d \times \mathcal{V}^p$ be their type-compatible pair spaces. For each $\tau \in \mathcal{T}$, the training edge set $\mathcal{E}_{\mathrm{tr}}^{\tau} \subset \mathcal{P}^{\tau}$ contains only observed training associations. Given the entity attributes and these training edges, it learns a score $s_{ij}^{\tau} = f_{\tau}(v_i, v_j)$ for every candidate pair $(v_i, v_j) \in \mathcal{P}^{\tau}$. Drug–disease association prediction is the primary task, while drug–protein and disease–protein prediction provide auxiliary supervision.

3.1 Initial Multi-View Graph Construction

Drug, disease, and protein data provide complementary evidence, but attributes and observed associations characterize different aspects of the same biomedical system. We preserve this distinction by learning an attribute view from entity representations while retaining observed cross-type associations in a separate relation view. These two graph structures provide the inputs to the subsequent reconstruction and alignment objectives.

Initial Entity Representations. Type-specific encoders first map the heterogeneous attributes into a shared latent space. For node v_i^t of type $t \in \mathcal{S}$, we define $\mathbf{h}_i^{t,0} = \phi_t(\mathbf{x}_i^t) \in \mathbb{R}^d$, where ϕ_r, ϕ_d, and ϕ_p retain separate parameters. We stack the resulting representations in a fixed drug–disease–protein order as $\mathbf{H}^0 = \left[\mathbf{H}^{r,0}; \mathbf{H}^{d,0}; \mathbf{H}^{p,0}\right] \in \mathbb{R}^{N \times d}$.

Adaptive Attribute View. The attribute view connects biologically similar entities of the same type. For every $t \in \mathcal{S}$, a trainable projection produces $\mathbf{u}_i^t = \mathbf{W}_G^t \mathbf{h}_i^{t,0} \in \mathbb{R}^{d_g}$, $\mathbf{W}_G^t \in \mathbb{R}^{d_g \times d}$. The cosine affinity between two distinct entities of type t is $e_{ij}^t = (\mathbf{u}_i^t)^\top \mathbf{u}_j^t \,/\, (\|\mathbf{u}_i^t\|_2 \|\mathbf{u}_j^t\|_2)$, $i \neq j$. Let $\mathcal{N}_i^t$ contain the K largest affinities for node v_i^t, with $K < N_t$. For $j \in \mathcal{N}_i^t$, the directed attribute weight is $\alpha_{ij}^t = \exp(e_{ij}^t/\tau_A) / \sum_{k \in \mathcal{N}_i^t} \exp(e_{ik}^t/\tau_A)$, and $\alpha_{ij}^t = 0$ otherwise, where $\tau_A > 0$ controls the sharpness of the neighborhood distribution. An undirected weighted adjacency matrix is constructed with $a_{ij}^t = \frac{1}{2}(\alpha_{ij}^t + \alpha_{ji}^t)$, $a_{ii}^t = 0$. The full attribute adjacency matrix takes the block-diagonal form $\mathbf{A}^A = \mathrm{blkdiag}\left(\mathbf{A}_A^r, \mathbf{A}_A^d, \mathbf{A}_A^p\right)$.

Observed Relation View. The relation view preserves direct biomedical evidence without converting association profiles into additional within-type similarities. Let $\mathbf{Y}_{\mathrm{tr}}^{rd}$, $\mathbf{Y}_{\mathrm{tr}}^{rp}$, and $\mathbf{Y}_{\mathrm{tr}}^{dp}$ denote the three association matrices after the train–test split. We define

$$\mathbf{A}^R = \begin{pmatrix} \mathbf{0} & \mathbf{Y}_{\mathrm{tr}}^{rd} & \mathbf{Y}_{\mathrm{tr}}^{rp} \\ (\mathbf{Y}_{\mathrm{tr}}^{rd})^\top & \mathbf{0} & \mathbf{Y}_{\mathrm{tr}}^{dp} \\ (\mathbf{Y}_{\mathrm{tr}}^{rp})^\top & (\mathbf{Y}_{\mathrm{tr}}^{dp})^\top & \mathbf{0} \end{pmatrix} \in \{0,1\}^{N \times N}. \tag{1}$$

For later message passing, $\mathbf{A}_{\tau}^R$ retains the two symmetric off-diagonal blocks associated with τ. Consequently, $\mathbf{A}^R = \sum_{\tau \in \mathcal{T}} \mathbf{A}_{\tau}^R$. The attribute view models within-type proximity, while the relation view records observed cross-type interactions.

3.2 Cross-View Masked Graph Autoencoding

The two graph views expose complementary information over the same entities. We encode them with separate graph networks and corrupt only the information presented to the relation encoder during training. This construction supports relation-to-attribute reconstruction and attribute-to-relation reconstruction without allowing a target edge to appear in the relation encoder input. We first add self-loops to the attribute graph and apply symmetric normalization as $\widehat{\mathbf{A}}^A = \mathbf{A}^A + \mathbf{I}_N$, $\overline{\mathbf{A}}^A = (\mathbf{D}^A)^{-1/2} \widehat{\mathbf{A}}^A (\mathbf{D}^A)^{-1/2}$, where $D_{ii}^A = \sum_j \widehat{A}_{ij}^A$. The attribute encoder stacks a fixed number of residual message-passing layers using the recurrence $\mathbf{H}_A^{(l+1)} = \mathbf{H}_A^{(l)} + \mathrm{ReLU}(\overline{\mathbf{A}}^A \mathbf{H}_A^{(l)} \mathbf{W}_A^{(l)})$, $\mathbf{H}_A^{(0)} = \mathbf{H}^0$, where $\mathbf{W}_A^{(l)} \in \mathbb{R}^{d \times d}$, and all hidden layers have width d. Its output $\mathbf{Z}^A$ is taken after the final layer of this recurrence. The relation encoder accepts a

collection of relation-specific adjacency matrices $\{\mathbf{B}_\tau\}_{\tau\in\mathcal{T}}$ and an input feature matrix $\mathbf{X}$. For each relation, it computes $\overline{\mathbf{B}}_\tau = \mathbf{D}_\tau^{-1/2}\mathbf{B}_\tau\mathbf{D}_\tau^{-1/2}$, $(D_\tau)_{ii} = \sum_j (B_\tau)_{ij}$. We set $(D_\tau)_{ii}^{-1/2} = 0$ for an isolated node. Relation-specific transformations preserve the semantics of the three interaction types through

$$\mathbf{H}_R^{(l+1)} = \mathbf{H}_R^{(l)} + \mathrm{ReLU}\Big(\sum_{\tau\in\mathcal{T}} \overline{\mathbf{B}}_\tau \mathbf{H}_R^{(l)} \mathbf{W}_{R,\tau}^{(l)}\Big),\ \mathbf{H}_R^{(0)} = \mathbf{X}. \tag{2}$$

Here $\mathbf{W}_{R,\tau}^{(l)} \in \mathbb{R}^{d\times d}$ for every layer and relation type, and the relation encoder is stacked to the same depth as the attribute encoder so that both views are compared at a matched receptive field. We denote the resulting mapping, taken after its final layer, by $E_R\left(\{\mathbf{B}_\tau\}_{\tau\in\mathcal{T}}, \mathbf{X}\right)$.

Joint View Corruption. Every training iteration applies node masking and relation masking together. For each $t \in \mathcal{S}$, we sample a node set $\mathcal{M}_V^t \subset \mathcal{V}^t$ by drawing each $v_i \in \mathcal{V}^t$ independently with masking probability $p_V \in [0,1]$, so that $\mathbb{E}[|\mathcal{M}_V^t|] = p_V N_t$, and replace its input representation with a type-specific token $\mathbf{m}_t \in \mathbb{R}^d$, such that $\widetilde{\mathbf{h}}_i^0 = \mathbf{m}_t$ if $v_i \in \mathcal{M}_V^t$, and $\widetilde{\mathbf{h}}_i^0 = \mathbf{h}_i^0$ otherwise. Here $\mathbf{h}_i^0$ denotes the row of the initial representations corresponding to v_i. We define $\mathcal{M}_V = \bigcup_{t\in\mathcal{S}} \mathcal{M}_V^t$ and collect the corrupted rows in $\widetilde{\mathbf{H}}^0$. For each $\tau \in \mathcal{T}$, we likewise sample $\mathcal{M}_E^\tau \subset \mathcal{E}_{\mathrm{tr}}^\tau$ by drawing each training edge independently with a masking probability $p_E \in [0,1]$, so that $\mathbb{E}[|\mathcal{M}_E^\tau|] = p_E|\mathcal{E}_{\mathrm{tr}}^\tau|$, and remove both directions of every selected edge from $\mathbf{A}_\tau^R$. The remaining adjacency matrix is $\widetilde{\mathbf{A}}_\tau^R$. Node and edge masks are sampled independently; thus, masking a node does not imply masking its incident edges, and vice versa. Node and edge masks are sampled independently and redrawn at every training iteration with fixed rates p_V and p_E; no corruption is applied at inference. The corruption yields one relation representation $\widetilde{\mathbf{Z}}^R = E_R\big(\{\widetilde{\mathbf{A}}_\tau^R\}_{\tau\in\mathcal{T}}, \widetilde{\mathbf{H}}^0\big)$. For every masked positive edge $e \in \mathcal{M}_E^\tau$, we sample one type-compatible unobserved pair $\bar{e} \in \mathcal{P}^\tau \setminus \mathcal{E}_{\mathrm{tr}}^\tau$ by replacing one endpoint with an entity of the same type.

Relation-to-Attribute Reconstruction. The corrupted relation view must recover the hidden representation of every masked entity. A type-specific decoder gives $\widehat{\mathbf{h}}_i^0 = D_t\big(\widetilde{\mathbf{z}}_i^R\big),\ v_i \in \mathcal{M}_V^t$. Here $D_t\colon \mathbb{R}^d \to \mathbb{R}^d$ maps relation-view embeddings back to the shared initial feature space. We prevent the learnable reconstruction target from moving toward the decoder output by applying a stop-gradient operator. The reconstruction loss is

$$\mathcal{L}_{\mathrm{attr}} = (|\mathcal{M}_V|)^{-1} \sum_{v_i\in\mathcal{M}_V} \Big(1 - (\widehat{\mathbf{h}}_i^0)^\top \mathrm{sg}(\mathbf{h}_i^0) / (\|\widehat{\mathbf{h}}_i^0\|_2 \|\,\mathrm{sg}(\mathbf{h}_i^0)\|_2)\Big)^2. \tag{3}$$

Only the target branch is detached in the loss. The modality encoders and the attribute graph encoder remain trainable through other objectives.

Attribute-to-Relation Reconstruction. Edge masking withholds $\mathcal{M}_E^\tau \subset \mathcal{E}_{\mathrm{tr}}^\tau$ from relation-view message passing, but retains these observed associations as cross-view reconstruction targets. We require the attribute representation to recover their relational evidence without accessing the relation graph. This view-specific objective is important because the downstream predictor could otherwise fit a masked association primarily through the relation representation and the fusion module, without requiring $\mathbf{Z}^A$ to preserve information relevant to cross-type interactions. Let $\tau = (t_s, t_o)$ denote a relation whose ordered endpoints belong to $\mathcal{V}^{t_s}$ and $\mathcal{V}^{t_o}$, respectively. For each masked edge $e = (v_i^{t_s}, v_j^{t_o}) \in \mathcal{M}_E^\tau$, we construct the type-compatible corruption set $\mathcal{C}_e^\tau = \big[(\mathcal{V}^{t_s} \setminus \{v_i\}) \times \{v_j\}\big] \cup \big[\{v_i\} \times (\mathcal{V}^{t_o} \setminus \{v_j\})\big] \setminus \mathcal{E}_{\mathrm{tr}}^\tau$. We then sample one counterpart $\bar{e} \sim \mathrm{Unif}(\mathcal{C}_e^\tau)$. Thus, $\bar{e}$ is obtained by replacing exactly one endpoint of e with another entity of the same type, while excluding every association observed in the training split. The procedure preserves the endpoint types of task τ and does not consult validation or test labels; consequently, $\bar{e}$ remains an unknown pair rather than a confirmed biological negative. A relation-specific bilinear decoder scores an ordered candidate pair as $q_{ij}^{A,\tau} = (\mathbf{z}_i^A)^\top \mathbf{W}_D^\tau \mathbf{z}_j^A,\ \mathbf{W}_D^\tau \in \mathbb{R}^{d\times d}$. Because this decoder operates exclusively on $\mathbf{Z}^A$, its prediction cannot use the corrupted relation representation $\widetilde{\mathbf{Z}}^R$. For each masked association, we optimize the pairwise logistic ranking loss

$$\mathcal{L}_{\mathrm{rel}} = (M_E)^{-1} \sum_{\tau\in\mathcal{T}} \sum_{e\in\mathcal{M}_E^\tau} \mathrm{Softplus}\big(q_{\bar{e}}^{A,\tau} - q_e^{A,\tau}\big),\ M_E = \sum_{\tau\in\mathcal{T}} |\mathcal{M}_E^\tau|, \tag{4}$$

3.3 Cross-View Alignment & Multi-Task Learning

Masked reconstruction transfers information across views, but it neither enforces entity-level consistency between them nor specifies how each prediction task should use them. We therefore align the two views of each entity contrastively, and let drug–disease, drug–protein, and disease–protein prediction fuse the views differently while sharing the two graph encoders.

Cross-View Contrastive Alignment. For each entity type $t \in \mathcal{S}$, we align the clean attribute representation $\mathbf{z}_i^A$ with the corrupted relation representation $\widetilde{\mathbf{z}}_i^R$ of the same entity over the unmasked set $\mathcal{U}^t = \mathcal{V}^t \setminus \mathcal{M}_V^t$. Masked nodes are excluded so that their clean attributes cannot act as contrastive targets, and node masks are sampled such that $|\mathcal{U}^t| \geq 2$. Negatives are restricted to the same entity type, which prevents the model from solving the objective through drug, disease, and protein type differences alone. Given a query view $\mathbf{p}$ and a candidate view $\mathbf{q}$, the type-wise InfoNCE loss is

$$\ell^t(\mathbf{p}, \mathbf{q}) = -(|\mathcal{U}^t|)^{-1} \sum_{v_i \in \mathcal{U}^t} \log \frac{\exp(\mathrm{sim}(\mathbf{p}_i, \mathbf{q}_i)/\tau_C)}{\sum_{v_j \in \mathcal{U}^t} \exp(\mathrm{sim}(\mathbf{p}_i, \mathbf{q}_j)/\tau_C)}, \tag{5}$$

where $\mathrm{sim}(\cdot,\cdot)$ denotes cosine similarity and $\tau_C > 0$ is the contrastive temperature. Symmetrizing the loss function gives $\mathcal{L}_{\mathrm{con}} = (2|\mathcal{S}|)^{-1} \sum_{t \in \mathcal{S}} \big(\ell^t\big(\mathbf{z}^A, \widetilde{\mathbf{z}}^R\big) + \ell^t\big(\widetilde{\mathbf{z}}^R, \mathbf{z}^A\big)\big)$.

Task-Conditioned Prediction. Known DDAs are too sparse to supervise the shared encoders alone, whereas DPI and DPA involve the same drugs and diseases and encode protein-mediated mechanisms that often underlie therapeutic associations. Task-conditioned prediction exploits this by jointly training DDA, DPI, and DPA on top of the shared type-specific encoders, adaptive attribute graph, and graph encoders, so auxiliary gradients enrich the drug and disease representations used by the target task. Because the three tasks need not rely on attribute and relational evidence in the same proportions, a single fused representation would let auxiliary tasks impose their preferences on DDA and risk negative transfer. Each task $\tau \in \mathcal{T}$ therefore conditions the fusion of the two views on its own vector-valued gate $\mathbf{g}_i^\tau = \mathrm{sigmoid}\big(\mathbf{W}_g^\tau[\mathbf{z}_i^A \| \widetilde{\mathbf{z}}_i^R] + \mathbf{b}_g^\tau\big) \in (0,1)^d$, with $\mathbf{W}_g^\tau \in \mathbb{R}^{d \times 2d}$ and $\mathbf{b}_g^\tau \in \mathbb{R}^d$, yielding the task-specific representation $\mathbf{h}_i^\tau = \mathbf{g}_i^\tau \odot \mathbf{z}_i^A + (\mathbf{1} - \mathbf{g}_i^\tau) \odot \widetilde{\mathbf{z}}_i^R$, where $\|$ denotes concatenation and $\odot$ element-wise multiplication. Knowledge is thus shared through the encoders, while each gate decides, dimension by dimension, how much of each view its task consumes. For an ordered, type-compatible pair $e = (v_i, v_j) \in \mathcal{P}^\tau$, a task-specific decoder computes $\widehat{y}_e^\tau = \mathrm{sigmoid}\big(\mathrm{MLP}_\tau([\mathbf{h}_i^\tau \| \mathbf{h}_j^\tau \| \mathbf{h}_i^\tau \odot \mathbf{h}_j^\tau])\big)$, where $\mathrm{MLP}_\tau \colon \mathbb{R}^{3d} \to \mathbb{R}$ outputs one logit. The positive targets are the masked edges $\mathcal{M}_E^\tau$, which are absent from the relation-encoder input, and each is paired with its sampled counterpart $\bar{e}$: $\mathcal{L}_{\mathrm{task}}^\tau = -(2|\mathcal{M}_E^\tau|)^{-1} \sum_{e \in \mathcal{M}_E^\tau} \big[\log \widehat{y}_e^\tau + \log(1 - \widehat{y}_{\bar{e}}^\tau)\big]$. The unobserved pair receives an operational rather than biological negative label, and normalizing by $|\mathcal{M}_E^\tau|$ prevents a relation type with more observations from dominating. Since DDA is the target and DPI (rp) and DPA (dp) serve only as auxiliary supervision, we weight the losses by a single coefficient $\eta_{\mathrm{aux}} \geq 0$ as $\mathcal{L}_{\mathrm{task}} = \mathcal{L}_{\mathrm{task}}^{rd} + \frac{\eta_{\mathrm{aux}}}{2}\big(\mathcal{L}_{\mathrm{task}}^{rp} + \mathcal{L}_{\mathrm{task}}^{dp}\big)$. Setting $\eta_{\mathrm{aux}} = 0$ disables explicit DPI/DPA prediction while retaining the auxiliary relations for message passing and cross-view reconstruction, corresponding to the auxiliary-relations-only variant, whereas an overly large η_{aux} may bias the shared representations toward the auxiliary tasks. With the two reconstruction directions combined into $\mathcal{L}_{\mathrm{MAE}} = \frac{1}{2}(\mathcal{L}_{\mathrm{attr}} + \mathcal{L}_{\mathrm{rel}})$, The complete objective is $\mathcal{L} = \mathcal{L}_{\mathrm{task}} + \lambda_{\mathrm{M}}\mathcal{L}_{\mathrm{MAE}} + \lambda_{\mathrm{C}}\mathcal{L}_{\mathrm{con}}$, where $\lambda_{\mathrm{M}} \geq 0$ and $\lambda_{\mathrm{C}} \geq 0$ control masked reconstruction and cross-view alignment.

4 Experiments

4.1 Experimental Setup

Table 1: Statistics of the three benchmark datasets.

Dataset	Drugs	Diseases	Proteins	Drug–disease associations	Drug–protein associations	Disease–protein associations	Sparsity
Cdataset	663	409	993	2,532	3,773	10,734	0.0093
Bdataset	269	598	1,021	18,416	3,110	5,898	0.1145
Fdataset	593	313	2,741	1,933	3,243	54,265	0.0104

Datasets & Baselines. We evaluate CARE on three benchmarks for computational drug repositioning: Cdataset (Luo et al., 2016), Bdataset (Zhang et al., 2018), and Fdataset (Gottlieb et al., 2011). Table 1 summarizes their statistics. All three contain drugs, diseases, and proteins, together

with drug–disease associations (DDAs), drug–protein interactions (DPIs), and disease–protein associations (DPAs), providing a common setting for DDA-centered multi-task learning. Cdataset and Fdataset are derived from DrugBank (Knox et al., 2024) and Online Mendelian Inheritance in Man (OMIM) (Hamosh et al., 2005), whereas Bdataset is constructed from the Comparative Toxicogenomics Database (Davis et al., 2023). We compare CARE with 12 methods, including DRHGCN (Cai et al., 2021), DRWBNCF (Meng et al., 2022), DRGBCN (Tang et al., 2025), SMGCL (Gao et al., 2023), DRAGNN (Meng et al., 2024), DRGCL (Jia et al., 2025), AMDGT (Liu et al., 2024a), MTP-DDA (Cui et al., 2026), MilGNet (Gu et al., 2025), DFDRNN (Zhu et al., 2025), SIGDR (Cui et al., 2025), and DIAM (Xu et al., 2026). These baselines span graph convolution and attention encoders, neural collaborative filtering, (multi-view) graph contrastive learning, multimodal and cross-domain dual-feature modeling. Further details are provided in Appendix B.

Implementation Details. CARE is trained in a DDA-centered multi-task setting, where DDA prediction is the primary objective and DPI/DPA prediction provides auxiliary supervision through shared entity encoders. Graph construction and training use only training-split associations, and the relation encoder never sees held-out DDA links. Baselines use hyperparameters from the original papers when available and, otherwise, public implementations, under identical 10-fold partitions, and we report the mean over all folds. AUC, AUPR, accuracy, precision, recall, and F1 are reported. Further details are in Appendix C.

4.2 Performance Comparison

CARE ranks first on all six metrics on all three datasets (Table 2), for 18 first-place results. Gains over the strongest baseline, DIAM, are consistent across datasets: AUC/AUPR improve by 0.0014/0.0012 on Cdataset and 0.0020/0.0015 on Fdataset, widening to 0.0041/0.0061 on Bdataset, where baselines are weakest. Threshold metrics follow the same pattern: recall rises by 0.0146–0.0192 and F1 by 0.0089–0.0127 over DIAM across the three datasets, while precision still improves rather than trading off against recall. Against the purely contrastive baseline DRGCL, AUC/AUPR gains remain substantial. The gains are thus not reducible to extra relation types or a generic cross-view agreement loss. These results match how CARE controls information flow: a learnable within-type attribute graph avoids fixed similarity networks; the typed relation graph keeps DDA, DPI, and DPA semantics distinct; reciprocal masked reconstruction forces each view to recover information withheld from the other; and type-aware alignment regularizes drug, disease, and protein embeddings without collapsing roles. DDA-conditioned fusion then selects the attribute and relation dimensions useful for the primary task, so DPI and DPA act as auxiliary signals rather than co-equal targets.

4.3 Ablation Study

We test whether reciprocal masked reconstruction and type-aware alignment give complementary signals, and whether DPI/DPA help DDA via graph context, auxiliary supervision, or both. All variants use the same splits, sampled unknown pairs, and protocol as Full CARE. Table 3 reports Recall and F1, which track positive-association recovery more clearly.

Masked Reconstruction & Cross-View Alignment. Both reconstruction directions improve positive recovery. Dropping relation-to-attribute reconstruction lowers Recall by 0.0098, 0.0122, and 0.0097 on C/B/F, with mean drops of 0.0106 Recall and 0.0072 F1. Dropping attribute-to-relation reconstruction is stronger: Recall falls by 0.0122, 0.0152, and 0.0114; F1 by 0.0081, 0.0114, and 0.0084. Reconstructing masked typed associations from the attribute view is the stronger signal, but both directions count. Removing both directions degrades more: Recall by 0.0206–0.0252 and F1 by 0.0145–0.0198. The extra loss versus one-way ablations supports reciprocal reconstruction: the relation view recovers entity attributes, while the attribute view must keep evidence of hidden typed links. Removing type-aware contrastive alignment yields smaller mean drops of 0.0083 Recall and 0.0057 F1. Alignment aids fusion compatibility rather than providing the main supervision.

Auxiliary Relations & Multi-Task Supervision. Pure DDA (no DPI/DPA edges or objectives) is the weakest variant. Recall falls by 0.0230, 0.0314, and 0.0261; F1 by 0.0163, 0.0248, and 0.0202 (means 0.0268 Recall, 0.0204 F1). The larger Recall gap shows that auxiliary evidence mainly recovers hard positives, which ranking metrics hide. Either auxiliary relation recovers much of the loss. DDA+DPI keeps mean deficits of 0.0137 Recall and 0.0103 F1; DDA+DPA reduces them to 0.0102 and 0.0074. DPA is more useful here, not as a general biological claim. Keeping both DPI and DPA as graph context without their prediction losses leaves mean deficits of 0.0067 Recall

Table 2: Performance comparison on the three datasets under 10-fold cross-validation. For each dataset–metric pair, the top five mean results are color-ranked: light green, blue, purple, orange, and yellow indicate ranks 1–5, respectively. The best mean is shown in bold dark red, and the second-best mean in underlined dark blue.

Cdataset						
Method	AUC↑	AUPR↑	Acc.↑	Prec.↑	Rec.↑	F1↑
DRHGCN	$0.9608_{\pm 0.0033}$	$0.9675_{\pm 0.0036}$	$0.9102_{\pm 0.0069}$	$0.9193_{\pm 0.0075}$	$0.9001_{\pm 0.0067}$	$0.9090_{\pm 0.0047}$
DRWBNCF	$0.8984_{\pm 0.0046}$	$0.9132_{\pm 0.0030}$	$0.8272_{\pm 0.0047}$	$0.8354_{\pm 0.0139}$	$0.8199_{\pm 0.0103}$	$0.8256_{\pm 0.0092}$
DRGBCN	$0.8832_{\pm 0.0043}$	$0.8743_{\pm 0.0029}$	$0.8100_{\pm 0.0096}$	$0.7781_{\pm 0.0147}$	$0.8704_{\pm 0.0107}$	$0.8204_{\pm 0.0067}$
SMGCL	$0.9426_{\pm 0.0045}$	$0.9542_{\pm 0.0020}$	$0.8934_{\pm 0.0045}$	$0.9168_{\pm 0.0050}$	$0.8661_{\pm 0.0085}$	$0.8901_{\pm 0.0093}$
DRAGNN	$0.9325_{\pm 0.0042}$	$0.9594_{\pm 0.0050}$	$0.8829_{\pm 0.0063}$	$0.9161_{\pm 0.0085}$	$0.8780_{\pm 0.0121}$	$0.8956_{\pm 0.0088}$
DRGCL	$0.9629_{\pm 0.0036}$	$0.9691_{\pm 0.0041}$	$0.9070_{\pm 0.0075}$	$0.9380_{\pm 0.0095}$	$0.8720_{\pm 0.0111}$	$0.9035_{\pm 0.0076}$
AMDGT	$0.9671_{\pm 0.0029}$	$0.9693_{\pm 0.0054}$	$0.9010_{\pm 0.0034}$	$0.8837_{\pm 0.0094}$	$0.9241_{\pm 0.0116}$	$0.9033_{\pm 0.0070}$
MTP-DDA	$0.9734_{\pm 0.0042}$	$0.9730_{\pm 0.0044}$	$0.9185_{\pm 0.0044}$	$0.9233_{\pm 0.0074}$	$0.9127_{\pm 0.0070}$	$0.9182_{\pm 0.0050}$
MilGNet	$0.9148_{\pm 0.0038}$	$0.9216_{\pm 0.0042}$	$0.8427_{\pm 0.0053}$	$0.8489_{\pm 0.0055}$	$0.8320_{\pm 0.0117}$	$0.8404_{\pm 0.0072}$
DFDRNN	$0.9578_{\pm 0.0038}$	$0.9635_{\pm 0.0051}$	$0.9048_{\pm 0.0036}$	$0.9291_{\pm 0.0049}$	$0.8742_{\pm 0.0140}$	$0.9008_{\pm 0.0079}$
SIGDR	$0.9635_{\pm 0.0046}$	$0.9658_{\pm 0.0034}$	$0.9177_{\pm 0.0030}$	$0.9085_{\pm 0.0050}$	$0.9293_{\pm 0.0086}$	$0.9186_{\pm 0.0068}$
DIAM	$0.9948_{\pm 0.0016}$	$0.9954_{\pm 0.0020}$	$0.9667_{\pm 0.0026}$	$0.9909_{\pm 0.0049}$	$0.9420_{\pm 0.0094}$	$0.9657_{\pm 0.0052}$
CARE	$\mathbf{0.9962}_{\pm 0.0011}$	$\mathbf{0.9966}_{\pm 0.0013}$	$\mathbf{0.9750}_{\pm 0.0021}$	$\mathbf{0.9934}_{\pm 0.0024}$	$\mathbf{0.9566}_{\pm 0.0058}$	$\mathbf{0.9746}_{\pm 0.0032}$
Bdataset						
Method	AUC↑	AUPR↑	Acc.↑	Prec.↑	Rec.↑	F1↑
DRHGCN	$0.8950_{\pm 0.0038}$	$0.9048_{\pm 0.0033}$	$0.8213_{\pm 0.0107}$	$0.8306_{\pm 0.0092}$	$0.8047_{\pm 0.0198}$	$0.8174_{\pm 0.0111}$
DRWBNCF	$0.8819_{\pm 0.0046}$	$0.8915_{\pm 0.0048}$	$0.8048_{\pm 0.0073}$	$0.8136_{\pm 0.0058}$	$0.7894_{\pm 0.0062}$	$0.8013_{\pm 0.0075}$
DRGBCN	$0.8517_{\pm 0.0072}$	$0.8426_{\pm 0.0064}$	$0.7751_{\pm 0.0089}$	$0.7448_{\pm 0.0142}$	$0.8357_{\pm 0.0187}$	$0.7876_{\pm 0.0108}$
SMGCL	$0.8794_{\pm 0.0054}$	$0.8922_{\pm 0.0038}$	$0.8119_{\pm 0.0081}$	$0.8294_{\pm 0.0110}$	$0.7860_{\pm 0.0169}$	$0.8071_{\pm 0.0117}$
DRAGNN	$0.8708_{\pm 0.0024}$	$0.9037_{\pm 0.0057}$	$0.8032_{\pm 0.0061}$	$0.8427_{\pm 0.0138}$	$0.7745_{\pm 0.0111}$	$0.8072_{\pm 0.0118}$
DRGCL	$0.9038_{\pm 0.0050}$	$0.9109_{\pm 0.0036}$	$0.8297_{\pm 0.0062}$	$0.8551_{\pm 0.0156}$	$0.7954_{\pm 0.0156}$	$0.8242_{\pm 0.0058}$
AMDGT	$0.9276_{\pm 0.0049}$	$0.9247_{\pm 0.0029}$	$0.8549_{\pm 0.0055}$	$0.8543_{\pm 0.0093}$	$0.8557_{\pm 0.0076}$	$0.8549_{\pm 0.0111}$
MTP-DDA	$0.9389_{\pm 0.0041}$	$0.9367_{\pm 0.0050}$	$0.8726_{\pm 0.0044}$	$0.8808_{\pm 0.0106}$	$0.8612_{\pm 0.0176}$	$0.8709_{\pm 0.0098}$
MilGNet	$0.8596_{\pm 0.0064}$	$0.8714_{\pm 0.0057}$	$0.7865_{\pm 0.0079}$	$0.7921_{\pm 0.0144}$	$0.7748_{\pm 0.0106}$	$0.7834_{\pm 0.0126}$
DFDRNN	$0.8947_{\pm 0.0046}$	$0.9056_{\pm 0.0061}$	$0.8198_{\pm 0.0090}$	$0.8374_{\pm 0.0140}$	$0.7932_{\pm 0.0102}$	$0.8147_{\pm 0.0093}$
SIGDR	$0.9212_{\pm 0.0063}$	$0.9186_{\pm 0.0043}$	$0.8522_{\pm 0.0089}$	$0.8452_{\pm 0.0143}$	$0.8626_{\pm 0.0173}$	$0.8537_{\pm 0.0044}$
DIAM	$0.9861_{\pm 0.0019}$	$0.9809_{\pm 0.0030}$	$0.9514_{\pm 0.0032}$	$0.9796_{\pm 0.0057}$	$0.9220_{\pm 0.0080}$	$0.9499_{\pm 0.0060}$
CARE	$\mathbf{0.9902}_{\pm 0.0015}$	$\mathbf{0.9870}_{\pm 0.0018}$	$\mathbf{0.9635}_{\pm 0.0028}$	$\mathbf{0.9850}_{\pm 0.0035}$	$\mathbf{0.9412}_{\pm 0.0067}$	$\mathbf{0.9626}_{\pm 0.0039}$
Fdataset						
Method	AUC↑	AUPR↑	Acc.↑	Prec.↑	Rec.↑	F1↑
DRHGCN	$0.9473_{\pm 0.0048}$	$0.9574_{\pm 0.0032}$	$0.8960_{\pm 0.0072}$	$0.9141_{\pm 0.0132}$	$0.8748_{\pm 0.0139}$	$0.8937_{\pm 0.0084}$
DRWBNCF	$0.8897_{\pm 0.0025}$	$0.9068_{\pm 0.0044}$	$0.8254_{\pm 0.0080}$	$0.8260_{\pm 0.0100}$	$0.8300_{\pm 0.0101}$	$0.8263_{\pm 0.0056}$
DRGBCN	$0.8670_{\pm 0.0035}$	$0.8627_{\pm 0.0083}$	$0.7941_{\pm 0.0063}$	$0.7601_{\pm 0.0095}$	$0.8636_{\pm 0.0158}$	$0.8078_{\pm 0.0040}$
SMGCL	$0.9264_{\pm 0.0049}$	$0.9402_{\pm 0.0044}$	$0.8678_{\pm 0.0081}$	$0.8733_{\pm 0.0142}$	$0.8616_{\pm 0.0109}$	$0.8667_{\pm 0.0050}$
DRAGNN	$0.9076_{\pm 0.0029}$	$0.9462_{\pm 0.0050}$	$0.8540_{\pm 0.0055}$	$0.9039_{\pm 0.0082}$	$0.8355_{\pm 0.0126}$	$0.8673_{\pm 0.0082}$
DRGCL	$0.9512_{\pm 0.0051}$	$0.9588_{\pm 0.0038}$	$0.8939_{\pm 0.0063}$	$0.9269_{\pm 0.0121}$	$0.8567_{\pm 0.0117}$	$0.8897_{\pm 0.0070}$
AMDGT	$0.9565_{\pm 0.0021}$	$0.9600_{\pm 0.0033}$	$0.8843_{\pm 0.0065}$	$0.8714_{\pm 0.0141}$	$0.9037_{\pm 0.0096}$	$0.8865_{\pm 0.0081}$
MTP-DDA	$0.9663_{\pm 0.0028}$	$0.9681_{\pm 0.0068}$	$0.9097_{\pm 0.0068}$	$0.9119_{\pm 0.0134}$	$0.9072_{\pm 0.0145}$	$0.9095_{\pm 0.0044}$
MilGNet	$0.8934_{\pm 0.0065}$	$0.9067_{\pm 0.0024}$	$0.8321_{\pm 0.0082}$	$0.8408_{\pm 0.0152}$	$0.8197_{\pm 0.0157}$	$0.8301_{\pm 0.0064}$
DFDRNN	$0.9548_{\pm 0.0022}$	$0.9616_{\pm 0.0043}$	$0.8995_{\pm 0.0041}$	$0.9227_{\pm 0.0083}$	$0.8708_{\pm 0.0130}$	$0.8960_{\pm 0.0072}$
SIGDR	$0.9594_{\pm 0.0027}$	$0.9630_{\pm 0.0022}$	$0.9043_{\pm 0.0058}$	$0.8964_{\pm 0.0101}$	$0.9157_{\pm 0.0153}$	$0.9054_{\pm 0.0070}$
DIAM	$0.9935_{\pm 0.0015}$	$0.9945_{\pm 0.0019}$	$0.9614_{\pm 0.0041}$	$0.9902_{\pm 0.0045}$	$0.9322_{\pm 0.0088}$	$0.9602_{\pm 0.0051}$
CARE	$\mathbf{0.9955}_{\pm 0.0012}$	$\mathbf{0.9960}_{\pm 0.0014}$	$\mathbf{0.9718}_{\pm 0.0024}$	$\mathbf{0.9928}_{\pm 0.0027}$	$\mathbf{0.9504}_{\pm 0.0061}$	$\mathbf{0.9711}_{\pm 0.0035}$

and 0.0048 F1. Auxiliary edges supply most of the gain, while explicit DPI/DPA prediction adds a smaller, consistent signal to our model.

4.4 Case Study

To assess CARE's practical value for drug repositioning, we conduct a case study on breast cancer (OMIM 114480), where first-line chemotherapy and endocrine regimens are well established, but many patients develop resistant or refractory disease, motivating continued interest in repositioning existing drugs. Excluding the 23 drugs already annotated as historical positives, we rank the remaining candidates by predicted score and examine the top 10. As shown in Table 4, all 10 top-ranked candidates are supported by at least preclinical or mechanistic evidence, and 9 of 10 have direct human-trial or clinical-use evidence specific to breast cancer. Everolimus ranks first despite being absent from the historical Fdataset label matrix, making it a case of temporal validation rather than label recovery. The mTOR inhibitor was not linked to breast cancer in the dataset's

Table 3: Ablation on DDA prediction. Recall and F1 are emphasized. Full results are shown in Appendix D.1.

Variant	Cdataset		Bdataset		Fdataset	
	Recall↑	F1↑	Recall↑	F1↑	Recall↑	F1↑
Full CARE	$0.9566_{\pm 0.0058}$	$0.9746_{\pm 0.0032}$	$0.9412_{\pm 0.0067}$	$0.9626_{\pm 0.0039}$	$0.9504_{\pm 0.0061}$	$0.9711_{\pm 0.0035}$
Cross-view training objectives						
w/o relation-to-attribute reconstruction	$0.9468_{\pm 0.0069}$	$0.9682_{\pm 0.0038}$	$0.9290_{\pm 0.0086}$	$0.9540_{\pm 0.0048}$	$0.9407_{\pm 0.0075}$	$0.9644_{\pm 0.0041}$
w/o attribute-to-relation reconstruction	$0.9444_{\pm 0.0074}$	$0.9665_{\pm 0.0040}$	$0.9260_{\pm 0.0090}$	$0.9512_{\pm 0.0051}$	$0.9390_{\pm 0.0081}$	$0.9627_{\pm 0.0044}$
w/o bidirectional reconstruction	$0.9360_{\pm 0.0088}$	$0.9601_{\pm 0.0050}$	$0.9160_{\pm 0.0107}$	$0.9428_{\pm 0.0061}$	$0.9282_{\pm 0.0095}$	$0.9543_{\pm 0.0052}$
w/o type-aware contrastive alignment	$0.9490_{\pm 0.0065}$	$0.9696_{\pm 0.0036}$	$0.9320_{\pm 0.0081}$	$0.9560_{\pm 0.0045}$	$0.9424_{\pm 0.0072}$	$0.9656_{\pm 0.0039}$
Auxiliary relations and auxiliary supervision						
Pure DDA	$0.9336_{\pm 0.0090}$	$0.9583_{\pm 0.0051}$	$0.9098_{\pm 0.0115}$	$0.9378_{\pm 0.0064}$	$0.9243_{\pm 0.0100}$	$0.9509_{\pm 0.0056}$
DDA + DPI	$0.9456_{\pm 0.0075}$	$0.9669_{\pm 0.0042}$	$0.9243_{\pm 0.0094}$	$0.9497_{\pm 0.0053}$	$0.9372_{\pm 0.0083}$	$0.9609_{\pm 0.0046}$
DDA + DPA	$0.9482_{\pm 0.0071}$	$0.9690_{\pm 0.0040}$	$0.9282_{\pm 0.0089}$	$0.9529_{\pm 0.0050}$	$0.9411_{\pm 0.0079}$	$0.9643_{\pm 0.0043}$
Auxiliary relations only	$0.9514_{\pm 0.0065}$	$0.9709_{\pm 0.0036}$	$0.9327_{\pm 0.0081}$	$0.9564_{\pm 0.0045}$	$0.9440_{\pm 0.0072}$	$0.9665_{\pm 0.0039}$

Table 4: Top-10 candidate drugs for breast cancer (OMIM 114480) ranked by CARE. Detailed analysis is provided in Appendix E.

Rank	Drug	DrugBank ID	Representative Evidence
1	Everolimus	DB01590	BOLERO-2 phase III trial (Baselga et al., 2012)
2	Carboplatin	DB00958	Guideline chemotherapy agent, HR+/TNBC regimens
3	Vinorelbine	DB00361	Guideline cytotoxic agent, metastatic setting
4	Cisplatin	DB00515	Guideline agent, triple-negative regimens
5	Dasatinib	DB01254	Phase II trial, HER2+/HR+ disease (Mayer et al., 2011)
6	Metformin	DB00331	Observational / adjuvant trial evidence
7	Simvastatin	DB00641	Observational cohort evidence
8	Propranolol	DB00571	Perioperative / observational trial evidence
9	Sorafenib	DB00398	Phase II combination trial evidence
10	Disulfiram	DB00822	Preclinical repositioning evidence

2011-era annotations, yet the subsequent phase III BOLERO-2 trial showed that adding everolimus to exemestane more than doubled progression-free survival in hormone-receptor-positive, HER2-negative advanced breast cancer, leading to FDA approval in July 2012 (Baselga et al., 2012). Since this association postdates the training labels, CARE's top prediction corresponds to a real, independently confirmed repositioning success rather than memorization. Dasatinib, ranked fifth, illustrates supportive but heterogeneous evidence. A phase II trial in HER2-positive and/or hormone-receptor-positive advanced breast cancer reported clinical activity with modest, variable response (Mayer et al., 2011), while trials in triple-negative disease found low response rates. CARE thus surfaces a mechanistically plausible, clinically tested candidate whose benefit likely depends on patient selection or combination therapy. The model also demotes literature-prominent but disproven candidates: celecoxib and sunitinib fall to ranks 12 and 13. Adding sunitinib to capecitabine in pretreated metastatic breast cancer failed to improve survival and increased toxicity (Crown et al., 2013); celecoxib likewise showed no consistent breast-cancer-specific benefit in randomized trials. That CARE ranks these agents low despite frequent literature co-occurrence indicates that high literature visibility alone is insufficient to explain CARE's ranking. Overall, CARE's top candidates for breast cancer are disproportionately supported by clinical or trial-level evidence.

5 Conclusion

CARE treats attribute and relation views as sources that must recover each other, not merely agree. In sparse settings, performance depends less on the volume of available evidence than on whether training forces each view to carry what the other lacks. Complementarity, on this account, is something a training signal has to impose; it does not follow from holding several views. The same logic governs auxiliary relations, which help chiefly by reshaping the context in which entities are represented rather than by adding another target. Keeping supervised edges out of the graph the model reasons over, and letting views exchange rather than duplicate information, are principles that should transfer to other sparse biomedical link-prediction problems.

REFERENCES

José Baselga, Mario Campone, Martine Piccart, Howard A. Burris, Hope S. Rugo, Tarek Sahmoud, Shinzaburo Noguchi, Michael Gnant, Kathleen I. Pritchard, Fabienne Lebrun, J. Thaddeus Beck, Yoshinori Ito, Denise Yardley, Ines Deleu, Alejandra Perez, Thomas Bachelot, Luc Vittori, Zhiying Xu, Pabak Mukhopadhyay, David Lebwohl, and Gabriel N. Hortobagyi. Everolimus in postmenopausal hormone-receptor–positive advanced breast cancer. *New England Journal of Medicine*, 366(6):520–529, 2012. doi: 10.1056/NEJMoa1109653.

Lijun Cai, Changcheng Lu, Junlin Xu, Yajie Meng, Peng Wang, Xiangzheng Fu, Xiangxiang Zeng, and Yansen Su. Drug repositioning based on the heterogeneous information fusion graph convolutional network. *Briefings in Bioinformatics*, 22(6):bbab319, 2021. doi: 10.1093/bib/bbab319.

John P. Crown, Véronique Diéras, Elzbieta Staroslawska, Denise A. Yardley, Thomas Bachelot, Neville Davidson, Hans Wildiers, Peter A. Fasching, Olivier Capitain, Manuel Ramos, Richard Greil, Francesco Cognetti, George Fountzilas, Maria Blasinska-Morawiec, Cornelia Liedtke, Rolf Kreienberg, Wilson H. Miller, Vanessa Tassell, Xin Huang, Jolanda Paolini, Kenneth A. Kern, and Gilles Romieu. Phase iii trial of sunitinib in combination with capecitabine versus capecitabine monotherapy for the treatment of patients with pretreated metastatic breast cancer. *Journal of Clinical Oncology*, 31(23):2870–2878, 2013. doi: 10.1200/JCO.2012.43.3391.

Hai Cui, Meiyu Duan, Jianyuan Yuan, Ziwei Ma, and Yijia Zhang. Sign-aware graph contrastive learning for drug repositioning. *IEEE Journal of Biomedical and Health Informatics*, 29(10): 7704–7716, 2025. doi: 10.1109/JBHI.2025.3571801.

Ming-Li Cui, Cui-Na Jiao, Dao-Hui Ge, Chun-Hou Zheng, Hui Yang, Ying-Lian Gao, Yan-Li Wang, and Jin-Xing Liu. MTP-DDA: Enhanced drug-disease associations prediction via multi-task learning with multi-view graph convolutional networks and contrastive learning. In *Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26), Special Track on AI and Health*, 2026.

Allan Peter Davis, Thomas C. Wiegers, Robin J. Johnson, Daniela Sciaky, Jolene Wiegers, and Carolyn J. Mattingly. Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Research*, 51(D1):D1257–D1262, 2023. doi: 10.1093/nar/gkac833.

Yanglan Gan, Shengnan Li, Guangwei Xu, Cairong Yan, and Guobing Zou. Multidependency graph convolutional networks and contrastive learning for drug repositioning. *Journal of Chemical Information and Modeling*, 65(6):3090–3103, 2025. doi: 10.1021/acs.jcim.4c02424.

Zhenxiang Gao, Pingjian Ding, and Rong Xu. KG-Predict: A knowledge graph computational framework for drug repurposing. *Journal of Biomedical Informatics*, 132:104133, 2022. doi: 10.1016/j.jbi.2022.104133.

Zihao Gao, Huifang Ma, Xiaohui Zhang, Yike Wang, and Zheyu Wu. Similarity measures-based graph co-contrastive learning for drug–disease association prediction. *Bioinformatics*, 39(6): btad357, 2023. doi: 10.1093/bioinformatics/btad357.

Assaf Gottlieb, Gideon Y. Stein, Eytan Ruppin, and Roded Sharan. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Molecular Systems Biology*, 7(1):496, 2011. doi: 10.1038/msb.2011.26.

Yaowen Gu, Si Zheng, Qijin Yin, Rui Jiang, and Jiao Li. REDDA: Integrating multiple biological relations to heterogeneous graph neural network for drug-disease association prediction. *Computers in Biology and Medicine*, 150:106127, 2022. doi: 10.1016/j.compbiomed.2022.106127.

Yaowen Gu, Si Zheng, Bowen Zhang, Hongyu Kang, Rui Jiang, and Jiao Li. Deep multiple instance learning on heterogeneous graph for drug–disease association prediction. *Computers in Biology and Medicine*, 184:109403, 2025. doi: 10.1016/j.compbiomed.2024.109403.

Ada Hamosh, Alan F. Scott, Joanna S. Amberger, Carol A. Bocchini, and Victor A. McKusick. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research*, 33(suppl_1):D514–D517, 2005. doi: 10.1093/nar/gki033.

Shihui He, Lijun Yun, and Haicheng Yi. Fusing graph transformer with multi-aggregate GCN for enhanced drug–disease associations prediction. *BMC Bioinformatics*, 25:79, 2024. doi: 10.1186/s12859-024-05705-w.

Yi Hua, Xiaowen Dai, Yuan Xu, Guomeng Xing, Haichun Liu, Tao Lu, Yadong Chen, and Yanmin Zhang. Drug repositioning: Progress and challenges in drug discovery for various diseases. *European Journal of Medicinal Chemistry*, 234:114239, 2022.

Feng Huang, Yang Qiu, Qiaojun Li, Shichao Liu, and Fuchuan Ni. Predicting drug-disease associations via multi-task learning based on collective matrix factorization. *Front Bioeng Biotechnol*, 8:218, 2020.

Kexin Huang, Payal Chandak, Qianwen Wang, Shreyas Havaldar, Akhil Vaid, Jure Leskovec, Girish N. Nadkarni, Benjamin S. Glicksberg, Nils Gehlenborg, and Marinka Zitnik. A foundation model for clinician-centered drug repurposing. *Nature Medicine*, 30:3601–3613, 2024. doi: 10.1038/s41591-024-03233-x.

Xiao Jia, Xinliang Sun, Kaili Wang, and Min Li. Drgcl: Drug repositioning via semantic-enriched graph contrastive learning. *IEEE Journal of Biomedical and Health Informatics*, 29(3):1656–1667, 2025.

Craig Knox, Mike Wilson, Christen M. Klinger, Mark Franklin, Eponine Oler, Alex Wilson, Allison Pon, Jordan Cox, Na Eun Chin, Seth A. Strawbridge, et al. DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Research*, 52(D1):D1265–D1275, 2024. doi: 10.1093/nar/gkad976.

Jiao Li, Si Zheng, Bin Chen, Atul J. Butte, S. Joshua Swamidass, and Zhiyong Lu. A survey of current trends in computational drug repositioning. *Briefings in Bioinformatics*, 17(1):2–12, 2016.

Junkai Liu, Shixuan Guan, Quan Zou, Hongjie Wu, Prayag Tiwari, and Yijie Ding. AMDGT: Attention aware multi-modal fusion using a dual graph transformer for drug–disease associations prediction. *Knowledge-Based Systems*, 284:111329, 2024a. doi: 10.1016/j.knosys.2023.111329.

Junkai Liu, Fuyuan Hu, Quan Zou, Prayag Tiwari, Hongjie Wu, and Yijie Ding. Drug repositioning by multi-aspect heterogeneous graph contrastive learning and positive-fusion negative sampling strategy. *Information Fusion*, 112:102563, 2024b. doi: 10.1016/j.inffus.2024.102563.

Huimin Luo, Jianxin Wang, Min Li, Junwei Luo, Xiaoqing Peng, Fang-Xiang Wu, and Yi Pan. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics*, 32(17):2664–2671, 2016. doi: 10.1093/bioinformatics/btw228.

Erica L. Mayer, Jean-François Baurain, Joseph Sparano, Lewis Strauss, Mario Campone, Pierre Fumoleau, Hope Rugo, Ahmad Awada, Oumar Sy, and Antonio Llombart-Cussac. A phase 2 trial of dasatinib in patients with advanced her2-positive and/or hormone receptor-positive breast cancer. *Clinical Cancer Research*, 17(21):6897–6904, 2011. doi: 10.1158/1078-0432.CCR-11-1500.

Yajie Meng, Changcheng Lu, Min Jin, Junlin Xu, Xiangxiang Zeng, and Jialiang Yang. A weighted bilinear neural collaborative filtering approach for drug repositioning. *Briefings in Bioinformatics*, 23(2):bbab581, 2022. doi: 10.1093/bib/bbab581.

Yajie Meng, Yi Wang, Junlin Xu, Changcheng Lu, Xianfang Tang, Tao Peng, Bengong Zhang, Geng Tian, and Jialiang Yang. Drug repositioning based on weighted local information augmented graph neural network. *Briefings in Bioinformatics*, 25(1):bbad431, 2024.

Sudeep Pushpakom, Francesco Iorio, Patrick A. Eyers, K. Jane Escott, Shirley Hopper, Andrew Wells, Andrew Doig, Tim Guilliams, Joanna Latimer, Christine McNamee, Alan Norris, Philippe Sanseau, David Cavalla, and Munir Pirmohamed. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery*, 18(1):41–58, 2019.

Chang Sun, Zichen Qin, Minglei Li, Yanfei Li, Rong Tang, Shengquan Chen, Yuxiang Wang, Yanqiang Liu, Jinmao Wei, and Jian Liu. A unified framework combining linear and 3d molecular features for robust drug-protein interaction prediction. *Cell Reports Methods*, 6:101304, 2026.

Xianfang Tang, Chang Zhou, Changcheng Lu, Yajie Meng, Junlin Xu, Xinrong Hu, Geng Tian, and Jialiang Yang. Enhancing drug repositioning through local interactive learning with bilinear attention networks. *IEEE Journal of Biomedical and Health Informatics*, 29(3):1644–1655, 2025.

Yuanxu Wang, Jinmiao Song, Qiguo Dai, and Xiaodong Duan. Hierarchical negative sampling based graph contrastive learning approach for drug-disease association prediction. *IEEE Journal of Biomedical and Health Informatics*, 28(5):3146–3157, 2024. doi: 10.1109/JBHI.2024.3360437.

Kerui Xu, Keyuan Xu, Shuheng Yin, and Hang Qiu. DIAM: Adaptive drug repositioning with decoupled biological mechanism and instance-aware modulation. In *Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-ECAI 2026), Special Track on AI and Health*, pp. 6956–6964, 2026.

Runtao Yang, Yao Fu, Qian Zhang, and Lina Zhang. GCNGAT: Drug–disease association prediction based on graph convolution neural network and graph attention network. *Artificial Intelligence in Medicine*, 150:102805, 2024. doi: 10.1016/j.artmed.2024.102805.

Shengwei Ye, Weizhong Zhao, Xianjun Shen, Xingpeng Jiang, and Tingting He. An effective multi-task learning framework for drug repurposing based on graph representation learning. *Methods*, 218:48–56, 2023. doi: 10.1016/j.ymeth.2023.07.008.

Zhouxin Yu, Feng Huang, Xiaohan Zhao, Wenjie Xiao, and Wen Zhang. Predicting drug–disease associations through layer attention graph convolutional network. *Briefings in Bioinformatics*, 22 (4):bbaa243, 2021. doi: 10.1093/bib/bbaa243.

Wen Zhang, Xiang Yue, Weiran Lin, Wenjian Wu, Ruoqi Liu, Feng Huang, and Feng Liu. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinformatics*, 19(1):233, 2018. doi: 10.1186/s12859-018-2220-4.

Bo-Wei Zhao, Lun Hu, Zhu-Hong You, Lei Wang, and Xiao-Rui Su. Hingrl: predicting drug disease associations with graph representation learning on heterogeneous information networks. *Briefings in Bioinformatics*, 23(1):bbab515, 2022.

Bo-Wei Zhao, Xiao-Rui Su, Yue Yang, Dong-Xu Li, Guo-Dong Li, Peng-Wei Hu, Yong-Gang Zhao, and Lun Hu. Drug-disease association prediction using semantic graph and function similarity representation learning over heterogeneous information networks. *Methods*, 220:106–114, 2023. doi: 10.1016/j.ymeth.2023.10.014.

Jiancheng Zhong, Pan Cui, Yihong Zhu, Qiu Xiao, and Zuohang Qu. DAHNGC: A graph convolution model for drug–disease association prediction by using heterogeneous network. *Journal of Computational Biology*, 30(9):1019–1033, 2023. doi: 10.1089/cmb.2023.0135.

Enqiang Zhu, Xiang Li, Chanjuan Liu, and Nikhil R. Pal. Boosting drug-disease association prediction for drug repositioning via dual-feature extraction and cross-dual-domain decoding. *Journal of Chemical Information and Modeling*, 65(11):5745–5757, 2025. doi: 10.1021/acs.jcim.5c00070.

APPENDIX

A Author Declarations **14**

A.1 Use of LLMs 14

A.2 Ethics Statement 14

A.3 Reproducibility 14

A.4 Limitations & Future Work 14

B Details of Baselines **14**

C Details of Experimental Setup **16**

D Details of Experimental Results **17**

D.1 Full Results of Ablation Study 17

D.2 Full Results of Hyperparameter Sensitivity Analysis 19

E Details of Case Study **24**

A Author Declarations

A.1 Use of LLMs

Large language models were used as assistive tools for language editing, literature organization, methodological and experimental feedback, refinement of ablation and sensitivity analyses, and assistance in drafting and revising portions of the manuscript. All model suggestions, references, mathematical formulations, experimental designs, numerical results, and interpretations were independently checked by the authors. The authors conducted and remain fully responsible for the research, implementation, experiments, analysis, and conclusions.

A.2 Ethics Statement

This work uses only publicly available, de-identified biomedical datasets (Cdataset, Bdataset, Fdataset) derived from DrugBank, OMIM, and CTD; no new patient data, clinical records, or human-subjects data were collected. CARE is intended to prioritize candidate drug–disease associations for downstream experimental and clinical validation, not to make direct treatment recommendations; predicted associations, including those reported in the case study, reflect statistical evidence from historical databases and have not been experimentally confirmed by the authors. We caution against using model outputs as a substitute for expert pharmacological or clinical judgment, particularly for candidates without independent supporting evidence. We are not aware of other ethical concerns specific to this work.

A.3 Reproducibility

Code is available in Supplementary Materials. All datasets used in the paper are publicly available. Experimental setups, training details, and hyperparameters are delineated in the appendices.

A.4 Limitations & Future Work

CARE has several limitations. First, the relation graph is restricted to drug–disease, drug–protein, and disease–protein associations; richer biomedical evidence such as pathway membership, gene co-expression, side-effect profiles, or textual literature is not modeled, and incorporating it would require extending the typed relation view and its masking scheme to additional edge and node types. Second, the corruption process masks nodes and edges independently at random within each type; more structured corruption strategies, such as masking by biological pathway or by drug class, might yield harder and more informative reconstruction targets. Third, CARE is evaluated under a transductive, cross-validation setting in which every entity is observed during training; its behavior on strictly cold-start drugs or diseases with no training-time associations, and its scalability to substantially larger knowledge graphs, remain to be characterized.

B Details of Baselines

DRHGCN (Cai et al., 2021). DRHGCN argues that simply merging drug–drug similarity, disease–disease similarity, and drug–disease association networks into a single heterogeneous graph discards the distinct topological information carried by each domain. It therefore designs an intra-domain feature extraction module that applies graph convolution within the drug and disease similarity networks, and an inter-domain module that applies graph convolution over the drug–disease association network. The intra- and inter-domain embeddings are fused in parallel to obtain more representative drug and disease embeddings. A layer-attention mechanism then combines the outputs of multiple graph convolution layers, and a decoder reconstructs the drug–disease association matrix to score candidate pairs.

DRWBNCF (Meng et al., 2022). DRWBNCF starts from the observation that latent factor models, which directly factorize the association matrix, implicitly treat neighbors as independent and therefore capture localized information poorly. It constructs three networks: the known drug–disease association network and drug–drug and disease–disease similarity networks. The similarity networks keep only each entity's nearest neighbors to filter out noisy, weakly similar entities. An integration component based on weighted bilinear graph convolution then encodes the known associations, along with each drug's and disease's neighborhood and neighborhood interactions, into a unified

representation. Finally, a prediction component uses an MLP optimized with an α-balanced focal loss and graph regularization to model the complex drug–disease associations.

DRGBCN (Tang et al., 2025). DRGBCN integrates heterogeneous information through a deep bilinear attention network to infer candidate drugs for specific diseases. It first builds a comprehensive drug–disease network that incorporates multiple drug and disease similarity networks. It then learns node embeddings with graph convolution, using a layer-attention mechanism to weight the embeddings produced at different convolution layers. Rather than scoring pairs from independently computed embeddings, it constructs a bilinear attention network that explicitly captures pairwise local interactions between the drug and the disease in each candidate pair. An MLP maps the resulting interaction representation to an association score.

SMGCL (Gao et al., 2023). SMGCL applies graph contrastive learning to mitigate sparse DDA supervision. It notes that conventional graph contrastive methods generate views by corrupting the raw graph, which is ill-suited to association data, and that they model node interactions poorly. Its multi-source view construction module builds a known-DDA view together with drug- and disease-similarity views based on nearest neighbors. Its context-aware neighborhood aggregation module uses a bilinear GNN to capture complex local features in the DDA view, and a global-aware attention mechanism to compensate for the limited receptive field of bilinear aggregation. A graph co-contrastive objective then co-trains the node representations to maximize cross-view agreement. The prediction and contrastive tasks are unified in a primary-and-auxiliary learning paradigm.

DRAGNN (Meng et al., 2024). DRAGNN is motivated by the observation that prior models largely overlook the relevance between different node embeddings when modeling drug–disease associations. It uses a graph attention mechanism to assign adaptive coefficients to heterogeneous drug and disease neighbors, which makes information collection for each target node more effective. To avoid packing too much information into a limited embedding space, it deliberately omits self-node aggregation, which emphasizes informative heterogeneous and homogeneous neighborhood information. It also applies average pooling during neighbor aggregation to strengthen local information while keeping the architecture simple. An MLP then produces the final association predictions.

DRGCL (Jia et al., 2025). DRGCL argues that existing GNN-based repositioning models focus on the drug–disease association graph but ignore semantic information, so their learned features lack consistency between global topology and local semantics. It treats the known drug–disease associations as a topology graph. It then builds a separate semantic graph by connecting each drug and disease to its top-K most similar neighbors based on similarity information, rather than generating views through generic graph augmentation, which preserves the data's semantic content. Encoders learn embeddings on both graphs, and graph contrastive learning pulls together the embeddings of the same entity across the two embedding spaces. This alignment strengthens both the topological and semantic features used for prediction.

AMDGT (Liu et al., 2024a). AMDGT addresses two shortcomings of earlier methods: reliance on similarity data alone and simple concatenation for modality integration. It integrates multimodal similarities from multiple perspectives. It also introduces pretrained language-model embeddings to construct a heterogeneous drug–disease network enriched with biochemical information, including protein associations. Two graph transformer modules learn representations from the homogeneous similarity networks and from the heterogeneous network, respectively, and their attention mechanisms capture both neighbor-level and global context. A modality-interaction module based on a transformer encoder then fuses the attention-aware representations from the different modalities before prediction.

MTP-DDA (Cui et al., 2026). MTP-DDA casts DDA prediction as one task in a multi-task framework that also predicts drug–protein and disease–protein associations. It constructs three graphs: an attribute collaborative graph from KNN-sparsified fingerprint, phenotype, and sequence similarities; a GIP collaborative graph from KNN-sparsified Gaussian interaction profile kernels; and an interaction graph from the three association matrices. We combine these into two sub-views and one main view. Each sub-view is encoded by a residual GCN on both the original graph and a corrupted version with row-shuffled node features. A discriminator then contrasts local node features against a pooled global summary, following a Deep Graph Infomax-style objective. In the main view, a GCN extracts low-order features and Node2Vec extracts high-order features, and an attention module fuses the two. The features from the three views are concatenated and linearly projected, a

dot-product decoder scores each pair, and task-specific loss weights balance the unequal sample sizes of the three tasks.

MilGNet (Gu et al., 2025). MilGNet introduces deep multiple-instance learning to drug repositioning. It observes that earlier meta-path-based methods rarely learn representations at the level of individual path instances end-to-end. A heterogeneous GNN encoder first learns drug and disease node embeddings. We then treat each drug–disease pair as a bag, and a pseudo meta-path generator produces multiple instances of a dedicated quadruplet meta-path from the network topology. A bidirectional instance encoder enhances the representation of each meta-path instance. A multi-scale interpretable predictor aggregates the instances with an attention mechanism, yielding predictions at both the bag and instance levels, and the instance-level attention can be used to inspect plausible treatment mechanisms.

DFDRNN (Zhu et al., 2025). DFDRNN addresses a limitation of methods that mine information only from adjacent nodes: they neglect relationships between drug and disease feature spaces, which can bias the encoding. It represents every drug and disease with two features, a similarity feature and an association feature. A single-domain dual-feature extraction module extracts these features within the drug or disease domain. A cross-domain dual-feature extraction module operates on the mixed drug–disease domain and converts similarity features into association features and vice versa. Both modules use self-attention to mine neighbor information, and a layer-attention mechanism combines the fused features into final representations. A cross-dual-domain decoder then predicts drug–disease associations in both the drug and disease domains.

SIGDR (Cui et al., 2025). SIGDR observes that GNN-based repositioning methods are generally designed for unsigned graphs and therefore cannot exploit the complementary information carried by negative links. It also notes that sign-aware GNNs from general domains cannot be transferred directly to drug repositioning. First, they do not specify how to explicitly connect nodes within the same set (drug–drug and disease–disease). Second, they do not define a suitable contrastive objective for signed graphs. SIGDR resolves both issues. It accounts for both positive and negative links from signed biological networks, explicitly models same-type connections, and designs a sign-aware graph contrastive learning objective to infer potential drug–disease associations.

DIAM (Xu et al., 2026). DIAM targets two limitations of prevailing DDA predictors. Random negative sampling treats all unobserved pairs uniformly, yielding loose decision boundaries, and static, monolithic fusion cannot adapt to which biological mechanism dominates a given pair. Its dual-encoder backbone combines a heterogeneous graph transformer over the drug–disease–protein graph with a hypergraph convolutional network over KNN-based drug and disease hypergraphs. A dual-stream mechanism-decoupling module computes two signals for each pair. The first is a global structural-propagation signal that aggregates drug-side and disease-side similarity to known neighbors. The second is a local molecular-interaction signal that sums the degree-based importance of protein targets shared by the drug and the disease. A prior-confidence projection maps the two signals to a per-pair scalar. This scalar weights a biological-plausibility score used to mine hard negatives above a threshold, and it drives an adaptive residual gate that amplifies the pair's interaction embedding. Cross-view and intra-view InfoNCE objectives align the two encoders, and they are jointly optimized with the classification loss.

C Details of Experimental Setup

Environments. We run all experiments on Linux workstations with NVIDIA A5000 GPUs. The codebase is built upon PyTorch 2.4.0, PyTorch Geometric 2.6.1, and PyTorch Cluster 1.6.3, running under CUDA 12.1 with Python 3.9.

Hyperparameters Setup. Table 5 lists the hyperparameters of CARE together with the default configuration used to produce the main results in Table 2. K controls the neighborhood size used to induce the adaptive attribute graph; p_V and p_E are the independent node- and edge-masking probabilities applied during joint view corruption; λ_M and λ_C weight the masked-reconstruction and contrastive-alignment losses in the complete objective; and η_{aux} weights the combined DPI/DPA auxiliary task loss relative to the primary DDA task loss. All hyperparameters are shared across Cdataset, Bdataset, and Fdataset; no dataset-specific tuning is performed.

Table 6: Cross-view objective ablations. Entries are cross-dataset averages absolute decreases from Full CARE on the $[0, 1]$ metric scale. Larger values indicate a greater loss after removing the corresponding objective.

Variant	AUC	AUPR	Accuracy	Precision	Recall	F1
w/o relation-to-attribute reconstruction	0.0027	0.0031	0.0069	0.0037	0.0106	0.0072
w/o attribute-to-relation reconstruction	0.0035	0.0039	0.0088	0.0053	0.0129	0.0093
w/o both reconstruction directions	0.0066	0.0072	0.0158	0.0108	0.0227	0.0170
w/o contrastive alignment	0.0021	0.0024	0.0053	0.0029	0.0083	0.0057

Table 5: Hyperparameters of CARE and their default values, shared across all three datasets. Explored ranges and sensitivity results for the first six rows are reported in Table 8.

Symbol	Description	Default
K	Attribute-graph neighborhood size	20
p_V	Node-masking probability	0.3
p_E	Edge-masking probability	0.3
λ_M	Masked-reconstruction loss weight	0.1
λ_C	Contrastive-alignment loss weight	0.1
η_{aux}	DPI/DPA auxiliary-task weight	1
τ_A	Attribute-affinity temperature	0.5
τ_C	Contrastive temperature	0.1
d	Shared hidden dimension	128
—	Optimizer	AdamW
—	Weight decay	applied
—	Early stopping	100 epochs
—	Gradient clipping norm	1.0
—	Learning-rate schedule	exponential warm-up, 200 steps

Training. We adopt the AdamW optimizer with weight decay and employ early stopping after 100 epochs. We cap gradient clipping at 1.0 to ensure training stability. The learning rate follows an exponential warm-up schedule with 200 warm-up steps by default.

D Details of Experimental Results

D.1 Full Results of Ablation Study

This section reports the complete results of the ablation experiments presented in the main text. Whereas the main discussion emphasizes Recall and F1, we include AUC, AUPR, Accuracy, Precision, Recall, and F1 here to determine whether the conclusions remain consistent across ranking-based and threshold-dependent metrics. Each value in the tables is averaged over Cdataset, Bdataset, and Fdataset. Because the vertical range is adapted separately in each panel to expose small differences, the numerical axis labels, rather than the apparent bar heights across panels, should be used for cross-metric comparisons.

Cross-View Learning Objectives. We first isolate the contributions of reciprocal masked reconstruction and type-aware contrastive alignment. The relation-to-attribute objective predicts masked entity attributes from the relation view, whereas the attribute-to-relation objective predicts masked typed associations from the attribute view. Table 6 reports the average absolute decrease from Full CARE after removing either direction separately, removing both directions jointly, or removing contrastive alignment. The results are visualized in Figure 3. Removing either reconstruction direction degrades all six metrics on average, but the attribute-to-relation direction has the larger effect throughout. Its removal decreases AUC, AUPR, Accuracy, Precision, Recall, and F1 by 0.0035, 0.0039, 0.0088, 0.0053, 0.0129, and 0.0093, respectively; the corresponding decreases after removing relation-to-attribute reconstruction are 0.0027, 0.0031, 0.0069, 0.0037, 0.0106, and 0.0072. Thus, reconstructing typed relations from the attribute view provides a stronger individual signal, but relation-to-attribute reconstruction contributes consistently rather than redundantly. The joint ablation makes this complementarity clearer. Removing both directions produces the largest degradation in this group, with average absolute decreases of 0.0066 in AUC, 0.0072 in AUPR, 0.0158 in Accuracy, 0.0108 in Precision, 0.0227 in Recall, and 0.0170 in F1. These losses are substantially

Table 7: Auxiliary-information ablations. Entries are cross-dataset averages absolute decreases from Full CARE. Larger values indicate a greater loss relative to the complete model.

Variant	AUC	AUPR	Accuracy	Precision	Recall	F1
Pure DDA	0.0081	0.0088	0.0192	0.0131	0.0268	0.0204
DDA+DPI	0.0038	0.0043	0.0095	0.0064	0.0137	0.0103
DDA+DPA	0.0028	0.0031	0.0067	0.0042	0.0102	0.0074
Auxiliary relations only	0.0019	0.0021	0.0045	0.0027	0.0067	0.0048

larger than those caused by removing either direction alone. The two objectives therefore constrain different parts of the shared representation: one transfers typed relational structure into attribute recovery, while the other requires attribute-view embeddings to preserve information sufficient for typed association prediction. The effect is consistently stronger on Recall than on Precision. Under the joint reconstruction ablation, the Recall decrease ranges from 0.0206 to 0.0252 across the three datasets, whereas the Precision decrease ranges from 0.0080 to 0.0137. The cross-dataset averages, 0.0227 and 0.0108, respectively, show the same asymmetry. This pattern indicates that reciprocal reconstruction is especially useful for recovering valid associations that are more difficult to separate, rather than only increasing confidence for already well-identified pairs. The accompanying F1 decrease confirms that the Recall change is not obtained through a compensating deterioration that leaves the overall classification balance unchanged. Removing type-aware contrastive alignment causes a smaller but uniform decline: 0.0021 in AUC, 0.0024 in AUPR, 0.0053 in Accuracy, 0.0029 in Precision, 0.0083 in Recall, and 0.0057 in F1. Contrastive alignment is therefore not the dominant learning signal, but its effect extends across both ranking and classification metrics. Together with the reconstruction ablations, this result supports the intended division of labor: reciprocal reconstruction preserves cross-view predictive content, while type-aware alignment encourages the two encoders to organize that content consistently.

Auxiliary Relations & Auxiliary Tasks. We next separate the value of auxiliary biomedical structure from that of auxiliary prediction supervision. Pure DDA removes both DPI/DPA edges and their prediction objectives. DDA+DPI and DDA+DPA retain one auxiliary relation type and optimize its corresponding auxiliary task. Auxiliary relations only retain both DPI and DPA edges in the heterogeneous graph, including their contribution to representation learning and reconstruction, but optimizes no auxiliary prediction head. Full CARE retains both relation types and both auxiliary tasks. We evaluate all variants only on the primary DDA prediction task. The results are visualized in Figure 4. Removing all auxiliary information produces the largest losses in this group. Relative to Full CARE, Pure DDA decreases AUC by 0.0081, AUPR by 0.0088, Accuracy by 0.0192, Precision by 0.0131, Recall by 0.0268, and F1 by 0.0204 on average. The effect is most pronounced on Bdataset, where Recall and F1 decrease by 0.0314 and 0.0248, respectively. As in the cross-view objective ablations, Recall is approximately twice as sensitive as Precision on average. Auxiliary biomedical evidence therefore contributes most strongly to the recovery of positive DDA pairs that cannot be identified reliably from the observed DDA graph alone. Both single-relation variants recover a substantial part of the Pure-DDA loss, although DPA is more beneficial than DPI under the present results. DDA+DPI retains average deficits of 0.0038, 0.0043, 0.0095, 0.0064, 0.0137, and 0.0103 across AUC, AUPR, Accuracy, Precision, Recall, and F1, respectively. The corresponding deficits for DDA+DPA are smaller at 0.0028, 0.0031, 0.0067, 0.0042, 0.0102, and 0.0074. This comparison shows that each auxiliary relation is informative, while their relative utility depends on the relation semantics and the predictive evidence available in the associated graph. Auxiliary relations only is the closest ablated variant to Full CARE. Its remaining average decreases are 0.0019 in AUC, 0.0021 in AUPR, 0.0045 in Accuracy, 0.0027 in Precision, 0.0067 in Recall, and 0.0048 in F1. Comparing this variant with Pure DDA shows that introducing both auxiliary relation types recovers 0.0201 of the average Recall gap and 0.0156 of the average F1 gap even without auxiliary prediction losses. The richer heterogeneous context and its reconstruction targets therefore account for most of the improvement. The remaining difference between Auxiliary relations only and Full CARE demonstrates a smaller but consistent additional benefit from explicitly predicting DPI and DPA during multi-task training. Taken together, the two ablation groups support three conclusions. First, both directions of reciprocal reconstruction are useful and complementary. Second, type-aware contrastive alignment provides an additional, broadly consistent gain beyond reconstruction. Third, auxiliary relations improve DDA prediction through both heterogeneous structural context and supervised auxiliary tasks, with the structural contribution accounting for the larger share under

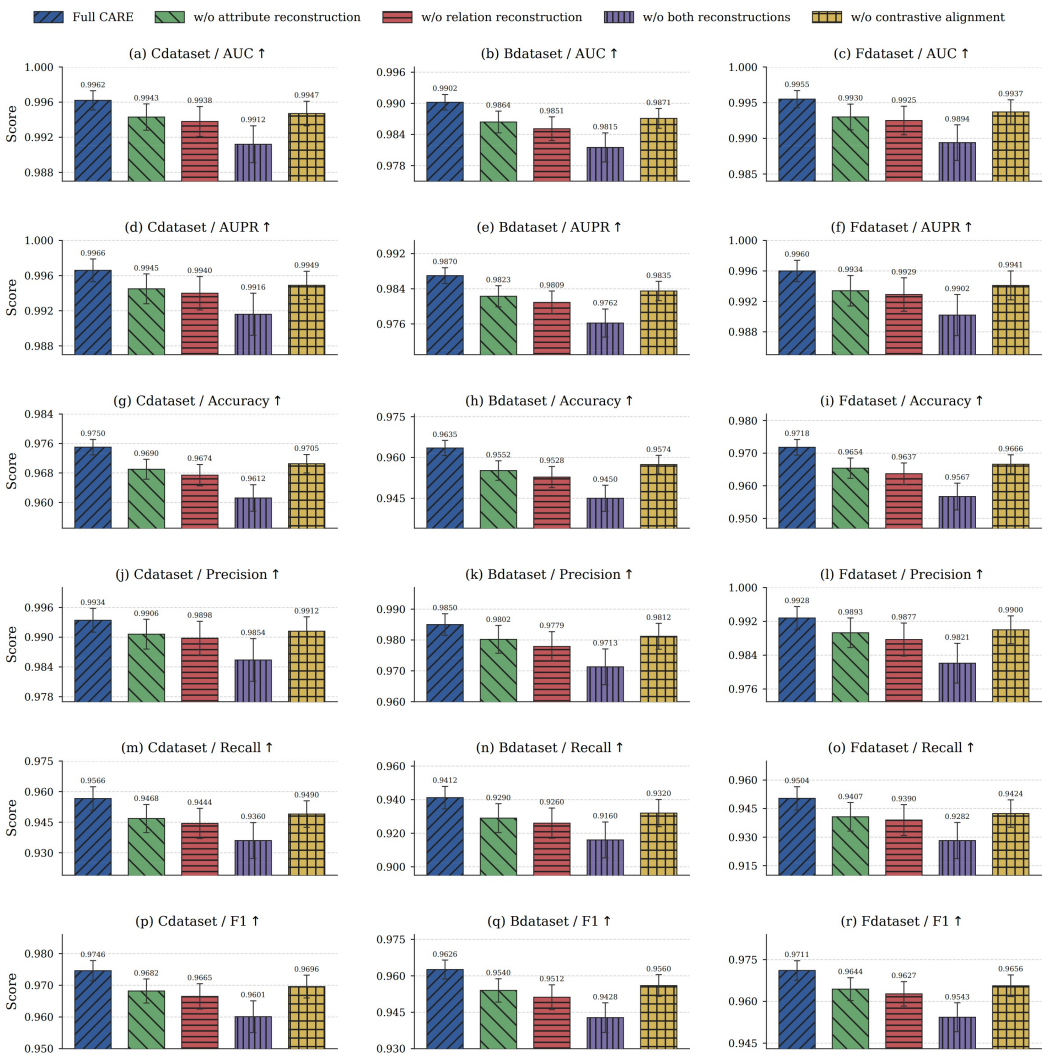

Figure 3: Complete ablation results for reciprocal masked reconstruction and type-aware contrastive alignment under 10-fold cross-validation. Rows report AUC, AUPR, Accuracy, Precision, Recall, and F1; columns report Cdataset, Bdataset, and Fdataset. Bars and error bars denote the mean and one standard deviation, respectively. Panel-specific vertical scales are shown explicitly.

these experiments. The larger changes in Recall and F1 than in the nearly saturated AUC and AUPR further justify emphasizing positive-pair recovery in the main analysis while retaining the complete six-metric results here.

D.2 Full Results of Hyperparameter Sensitivity Analysis

We examine the sensitivity of CARE to nine hyperparameters covering graph construction, self-supervised learning, multi-task optimization, and model capacity: the neighborhood size K, node- and edge-masking rates p_V and p_E, reconstruction and contrastive weights λ_M and λ_C, auxiliary-task weight η_{aux}, attribute-affinity temperature τ_A, contrastive temperature τ_C, and shared hidden dimension d. Each hyperparameter is varied independently while the remaining ones are fixed to the default configuration $(K, p_V, p_E, \lambda_M, \lambda_C, \eta_{\text{aux}}, \tau_A, \tau_C, d) = (20, 0.3, 0.3, 0.1, 0.1, 1, 0.5, 0.1, 128)$. Because AUC is already close to saturation and may obscure practically meaningful changes in positive-pair recovery, we emphasize Recall and F1. Figures 5 and 7 report the results for the first six hyperparameters, while Figure 8 reports the results for τ_A, τ_C, and d. All reported changes are absolute differences

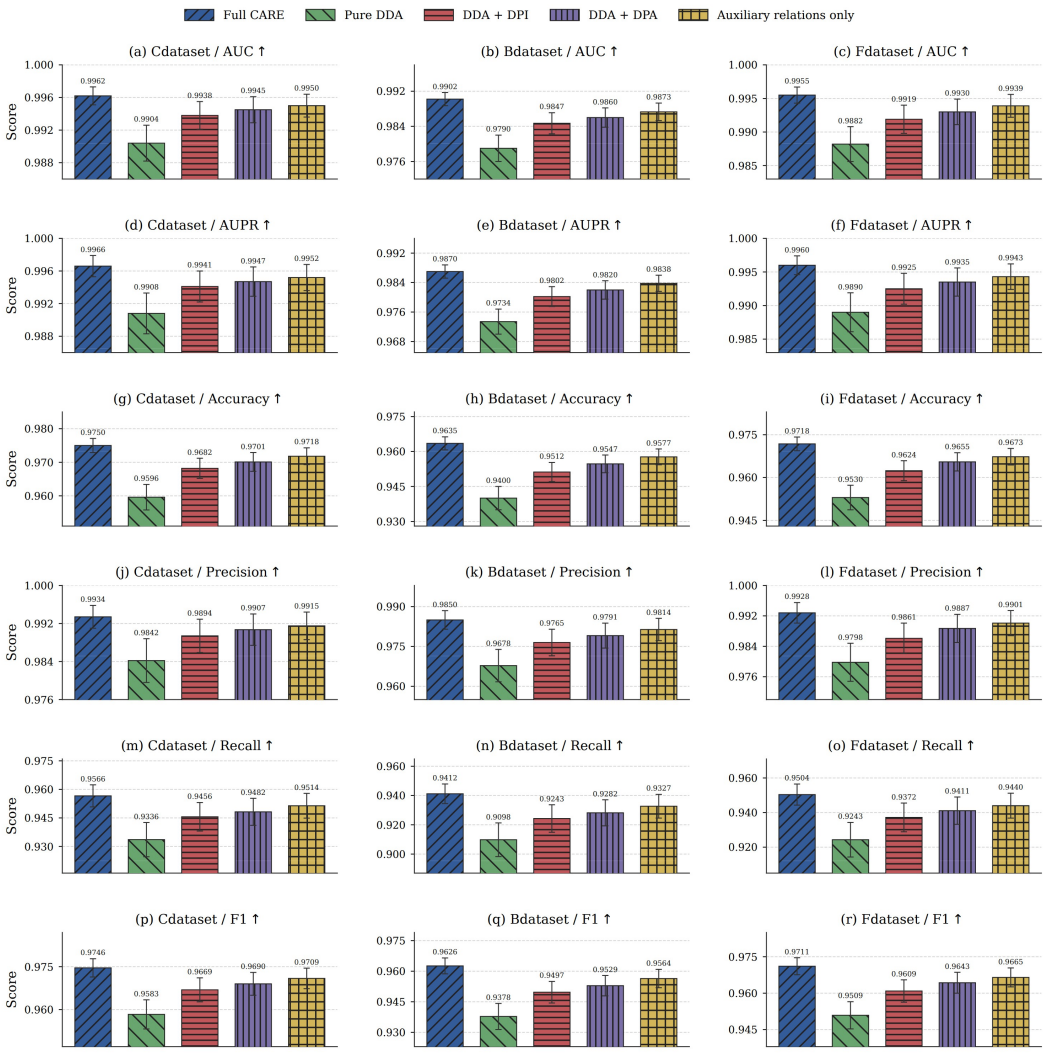

Figure 4: Complete auxiliary-information ablations under 10-fold cross-validation. Rows report AUC, AUPR, Accuracy, Precision, Recall, and F1; columns report Cdataset, Bdataset, and Fdataset. Pure DDA excludes DPI/DPA edges and objectives; DDA+DPI and DDA+DPA retain one auxiliary relation and its prediction objective; Auxiliary relations only retain both auxiliary edge types but optimize only the DDA prediction head. Bars and error bars denote the mean and one standard deviation, respectively. Panel-specific vertical scales are shown explicitly.

on the $[0, 1]$ metric scale. For example, a change from 0.9500 to 0.9474 corresponds to an absolute decrease of 0.0026.

Graph Construction & Representation Capacity. CARE remains stable across broad ranges of the graph-construction hyperparameters. Within $K \in [10, 30]$, the largest absolute deviation in cross-dataset average Recall from the result at $K = 20$ is only 0.0026, and the largest deviation on any individual dataset is 0.0038. The dataset-specific Recall maxima occur at $K = 30$, 15, and 30 on Cdataset, Bdataset, and Fdataset, respectively, but their improvements over the shared default do not exceed 0.0004. Thus, $K = 20$ provides a robust common setting rather than a dataset-specific optimum. Performance deteriorates more clearly outside this range: $K = 5$ decreases average Recall by 0.0102, indicating that a very small neighborhood omits useful local evidence, whereas $K = 40$ decreases it by 0.0052, suggesting that an overly large neighborhood introduces less relevant context. The two masking rates exhibit similarly broad stable regions. Within $p_V \in [0.2, 0.4]$,

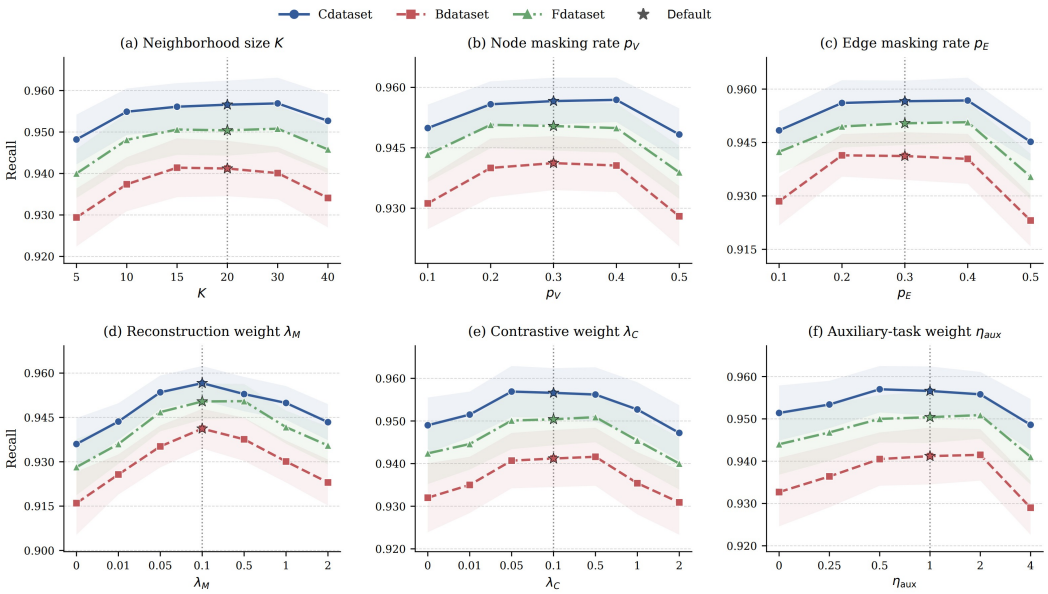

Figure 5: Hyperparameter sensitivity of CARE in terms of Recall under 10-fold cross-validation. Curves show the mean across the ten folds, shaded regions indicate one standard deviation, and stars mark the default settings.

the largest absolute deviations in cross-dataset average Recall and F1 are only 0.0006 and 0.0005, respectively. Within $p_E \in [0.2, 0.4]$, the corresponding deviations are 0.0004 and 0.0003. Moreover, the best node-masking rates are 0.4, 0.3, and 0.2, while the best edge-masking rates are 0.4, 0.2, and 0.4 on Cdataset, Bdataset, and Fdataset, respectively. These dispersed local maxima support the existence of stable performance plateaus. Performance declines at both extremes: a masking rate of 0.1 produces an insufficiently challenging reconstruction task, whereas masking half of the graph removes too much task-relevant information. The latter effect is particularly evident for edge masking, for which $p_E = 0.5$ decreases average Recall and F1 by 0.0148 and 0.0102, respectively. The attribute-affinity temperature and hidden dimension further demonstrate that CARE does not depend on a narrowly specified graph representation. Within $\tau_A \in [0.25, 1]$, the maximum absolute deviations in cross-dataset average Recall and F1 are only 0.00033 and 0.00027. The corresponding Recall maxima occur at $\tau_A = 0.25$, 1, and 1 on the three datasets and exceed the shared default by no more than 0.0003. By contrast, $\tau_A = 0.1$ decreases average Recall and F1 by 0.00523 and 0.00393, whereas $\tau_A = 2$ produces decreases of 0.00383 and 0.00287. These results are consistent with a broad useful temperature range: an excessively small τ_A makes the affinity distribution overly concentrated, while an excessively large value weakens the distinction between informative and uninformative neighbors. The hidden dimension exhibits a capacity-saturation pattern. Reducing d to 32 decreases average Recall and F1 by 0.01337 and 0.00990, indicating that a small representation cannot adequately encode the two graph views and their task-specific information. At $d = 64$, the remaining deficits decrease to 0.00283 and 0.00207. Increasing the dimension from 128 to 256 changes average Recall and F1 by only 0.00033 and 0.00023, respectively, whereas $d = 512$ produces modest decreases of 0.00330 and 0.00247. Therefore, $d = 128$ provides a favorable trade-off between predictive performance and representation size, while the near-equivalent results at 128 and 256 indicate that CARE is not tied to one exact hidden dimension.

Training Objectives & Task Balancing. The reconstruction weight produces the clearest trend among the objective coefficients. Setting $\lambda_M = 0$ exactly recovers the variant without bidirectional reconstruction and decreases average Recall and F1 by 0.0227 and 0.0170, respectively. Nevertheless, CARE remains stable once a moderate reconstruction contribution is introduced: within $\lambda_M \in [0.05, 0.5]$, the maximum absolute deviations from the default are 0.0042 in average Recall and 0.0031 in average F1. Bidirectional reconstruction is therefore important, but its coefficient does not require precise tuning within this range. The contrastive loss is less sensitive to both its weight and temperature. For $\lambda_C \in [0.05, 0.5]$, the maximum absolute deviations in cross-dataset average

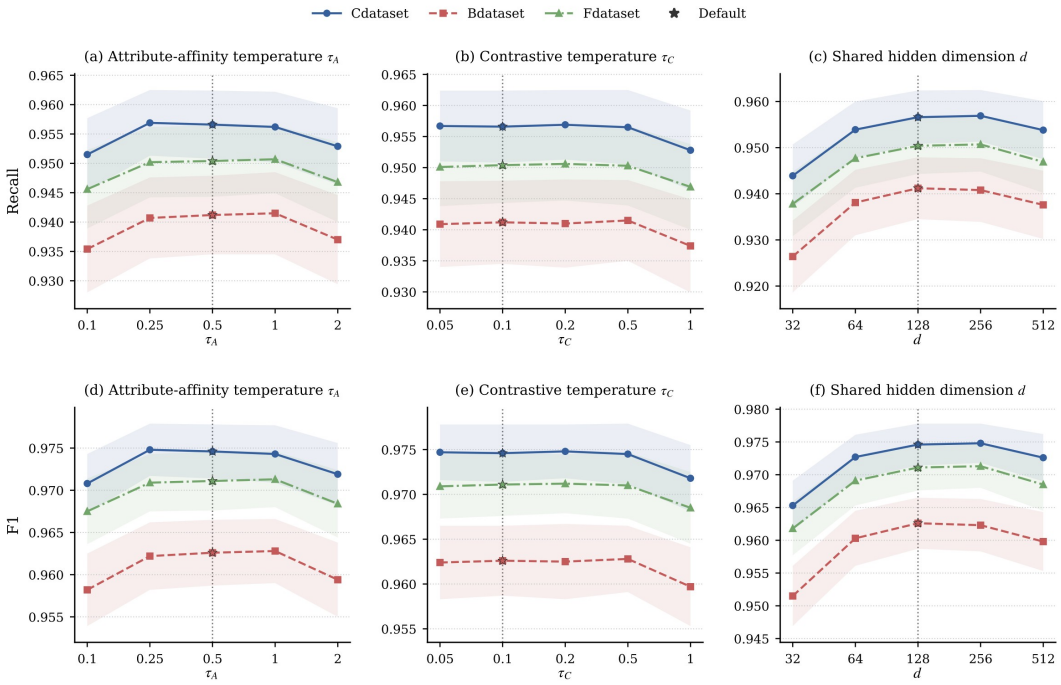

Figure 6: Sensitivity of CARE to the attribute-affinity temperature τ_A, contrastive temperature τ_C, and shared hidden dimension d under 10-fold cross-validation. The upper and lower rows report Recall and F1, respectively. Curves show the mean across the ten folds, shaded regions indicate one standard deviation, and stars mark the default settings.

Recall and F1 are only 0.0002 and 0.0001. The dataset-specific Recall maxima occur at $\lambda_C = 0.05$, 0.5, and 0.5, showing that the default $\lambda_C = 0.1$ lies on a broad plateau rather than at a common dataset-specific optimum. Similarly, varying τ_C from 0.05 to 0.5 produces maximum average deviations of only 0.00023 in Recall and 0.00017 in F1. The best Recall values occur at $\tau_C = 0.2$, 0.5, and 0.2 on Cdataset, Bdataset, and Fdataset, respectively. Only the comparatively large value $\tau_C = 1$ causes a clearer degradation, decreasing average Recall and F1 by 0.00370 and 0.00277. These results indicate that type-aware contrastive alignment does not rely on a narrowly tuned loss weight or softmax temperature. The auxiliary-task weight also admits a broad stable region. Within $\eta_{\text{aux}} \in [0.5, 2]$, the maximum absolute deviations in average Recall and F1 are only 0.0003 and 0.0002, with dataset-specific Recall maxima at 0.5, 2, and 2. Removing auxiliary-task supervision by setting $\eta_{\text{aux}} = 0$ recovers the "auxiliary relations only" endpoint and decreases average Recall and F1 by 0.0067 and 0.0048, respectively. Conversely, assigning excessive weight to the auxiliary tasks at $\eta_{\text{aux}} = 4$ produces larger decreases of 0.0099 and 0.0078. Auxiliary supervision is therefore beneficial, but excessive emphasis on DPI and DPA prediction can divert the shared representations from the primary DDA objective.

Figure 7 provides the complementary F1 results for the first six hyperparameters. Within their respective robust windows, the maximum absolute deviations in cross-dataset average F1 are 0.0018 for K, 0.0005 for p_V, 0.0003 for p_E, 0.0031 for λ_M, 0.0001 for λ_C, and 0.0002 for η_{aux}. The lower row of Figure 8 shows the corresponding F1 results for τ_A, τ_C, and d, whose maximum stable deviations are 0.00027, 0.00017, and 0.00023, respectively. The agreement between Recall and F1 indicates that the observed robustness is not merely caused by a shift in the decision operating point; the balance between precision and recall is largely preserved under moderate hyperparameter changes. Several neighboring settings outperform the shared default on an individual dataset by only 0.0002–0.0005, which is small relative to the variation across folds. Moreover, the locations of the local maxima differ across the three datasets. CARE's robustness is therefore evidenced by small, non-monotonic fluctuations across broad parameter ranges rather than by every dataset attaining its numerical optimum at the shared default configuration. Overall, the results show that CARE is robust to moderate changes in graph construction, temperature scaling, loss weighting,

Table 8: Summary of the hyperparameter sensitivity results. "Maximum stable deviation" denotes the largest absolute difference from the default result within the listed robust window. All differences are measured on the $[0, 1]$ metric scale and averaged across the three datasets.

Hyperparameter	Explored values	Default	Robust window	Maximum stable deviation	
				Recall	F1
K	$\{5, 10, 15, 20, 30, 40\}$	20	$[10, 30]$	0.0026	0.0018
p_V	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$	0.3	$[0.2, 0.4]$	0.0006	0.0005
p_E	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$	0.3	$[0.2, 0.4]$	0.0004	0.0003
λ_M	$\{0, 0.01, 0.05, 0.1, 0.5, 1, 2\}$	0.1	$[0.05, 0.5]$	0.0042	0.0031
λ_C	$\{0, 0.01, 0.05, 0.1, 0.5, 1, 2\}$	0.1	$[0.05, 0.5]$	0.0002	0.0001
η_{aux}	$\{0, 0.25, 0.5, 1, 2, 4\}$	1	$[0.5, 2]$	0.0003	0.0002
τ_A	$\{0.1, 0.25, 0.5, 1, 2\}$	0.5	$[0.25, 1]$	0.00033	0.00027
τ_C	$\{0.05, 0.1, 0.2, 0.5, 1\}$	0.1	$[0.05, 0.5]$	0.00023	0.00017
d	$\{32, 64, 128, 256, 512\}$	128	$[128, 256]$	0.00033	0.00023

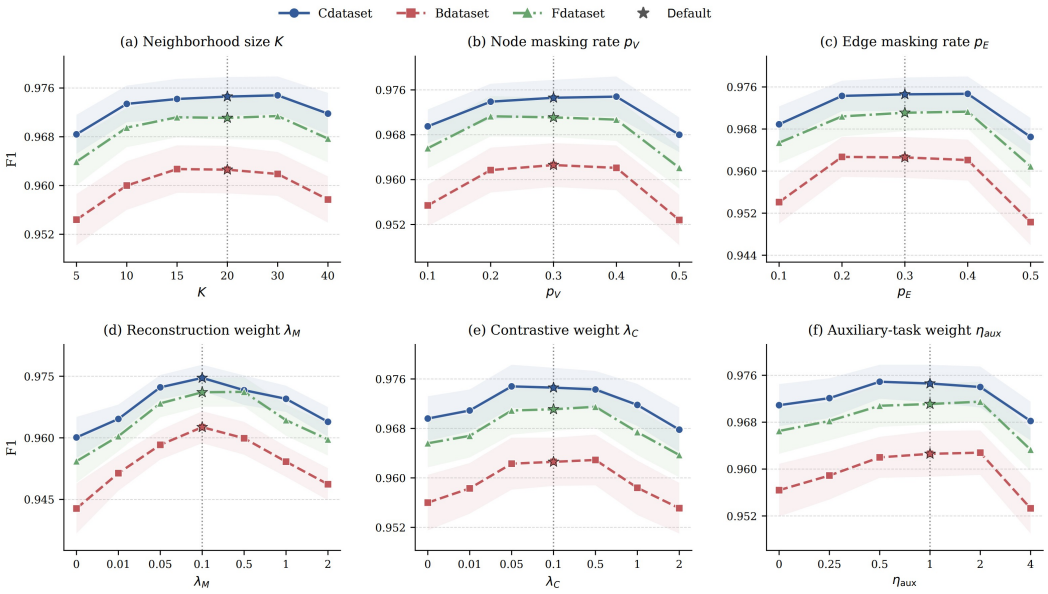

Figure 7: Hyperparameter sensitivity of CARE in terms of F1 under 10-fold cross-validation. Curves show the mean across the ten folds, shaded regions indicate one standard deviation, and stars mark the default settings.

task balancing, and representation capacity, while extreme masking, insufficient reconstruction, or inadequate model capacity produces clearer degradation.

Affinity Temperature, Contrastive Temperature, and Hidden Dimension. We further evaluate the attribute-affinity temperature τ_A, contrastive temperature τ_C, and shared hidden dimension d. Their default values are 0.5, 0.1, and 128, respectively. As in the preceding analysis, one hyperparameter is varied at a time, while all remaining settings are fixed to their default values.

The attribute-affinity temperature controls the concentration of the learned neighborhood distribution. CARE remains stable for $\tau_A \in [0.25, 1]$: within this interval, the largest cross-dataset average absolute deviations from the default are 0.00033 in Recall and 0.00027 in F1. The dataset-specific Recall maxima occur at 0.25, 1, and 1 on Cdataset, Bdataset, and Fdataset, respectively, and exceed the shared default by at most 0.0003. In contrast, setting $\tau_A = 0.1$ decreases average Recall and F1 by 0.00523 and 0.00393, while $\tau_A = 2$ yields decreases of 0.00383 and 0.00287. These results are consistent with a broad useful temperature range: an excessively small value produces an overly concentrated affinity distribution, whereas an excessively large value weakens the distinction between informative and uninformative neighbors. The contrastive temperature is even less sensitive. Across $\tau_C \in [0.05, 0.5]$, the maximum cross-dataset average absolute deviations are only 0.00023

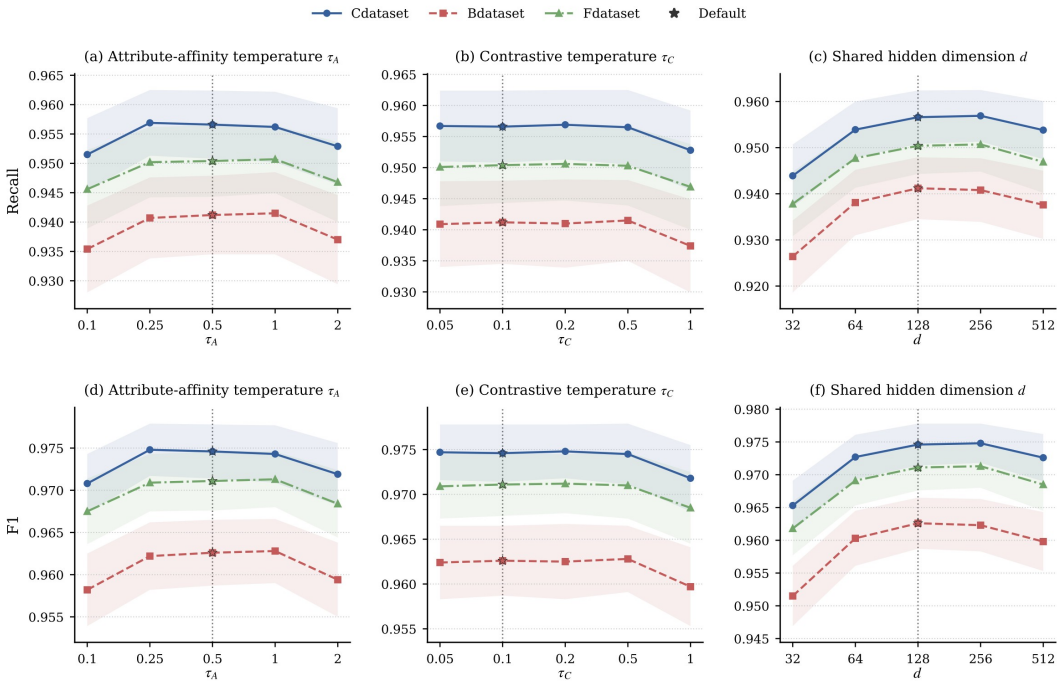

Figure 8: Sensitivity of CARE to the attribute-affinity temperature τ_A, contrastive temperature τ_C, and shared hidden dimension d under balanced 10-fold cross-validation. The upper and lower rows report Recall and F1, respectively. Curves show the mean across the ten folds, shaded regions indicate one standard deviation, and stars mark the shared default settings.

in Recall and 0.00017 in F1. The best Recall values occur at $\tau_C = 0.2$, 0.5, and 0.2 on Cdataset, Bdataset, and Fdataset, respectively. Thus, $\tau_C = 0.1$ is a stable shared choice rather than a dataset-specific numerical optimum. Performance declines only when the temperature becomes excessively large: $\tau_C = 1$ reduces average Recall and F1 by 0.00370 and 0.00277, respectively. This behavior indicates that the type-aware contrastive objective does not depend on a narrowly tuned softmax temperature, although an overly diffuse contrastive distribution weakens its discriminative signal. The hidden dimension exhibits a capacity-saturation pattern rather than a sharp optimum. Reducing d to 32 decreases average Recall and F1 by 0.01337 and 0.00990, indicating that a very small representation cannot adequately encode the two graph views and their task-specific information. Increasing the dimension from 64 to 128 provides a smaller improvement: at $d = 64$, the remaining average deficits are 0.00283 in Recall and 0.00207 in F1. Results at $d = 256$ differ from those at $d = 128$ by only 0.00033 in average Recall and 0.00023 in average F1, with Cdataset and Fdataset attaining slightly higher values at 256 and Bdataset remaining slightly higher at 128. Expanding the representation to $d = 512$ does not produce a further benefit and instead decreases average Recall and F1 by 0.00330 and 0.00247. Therefore, $d = 128$ provides a favorable trade-off between predictive performance and representation size, while the near-equal results at 128 and 256 show that CARE is not tied to one exact hidden dimension.

E Details of Case Study

We first audited the original PREDICT supplementary files against the commonly used Fdataset matrix. The indication spreadsheet contains 1,933 unique drug–disease records, covering 593 drug names and 313 OMIM disease names. The released Fdataset matrix has dimensions 313×593 and contains exactly 1,933 positive entries. Its disease identifiers are OMIM-based and its drug identifiers are DrugBank-based. The rows for breast cancer (OMIM 114480), lung cancer (OMIM 211980), and Alzheimer disease (OMIM 104300) reproduce the supplementary indications exactly. Across the complete matrix, only five cells differ after automatic identifier normalization; all five

Table 9: Evidence tiers used for post-hoc candidate validation. A literature mention alone is not treated as proof of efficacy.

Tier	Meaning	Interpretation
A	Established clinical use	Regulatory/guideline or current clinical-use evidence for the disease.
B	Direct human evidence	Disease-specific trial or human study; direction may be supportive or mixed.
C	Preclinical/mechanistic	Disease-specific animal, cellular, or mechanistic evidence only.
I	Indirect	Supportive-care, symptom-only, class-level, or non-disease-modifying evidence.
N	Negative	Direct evidence is predominantly negative or contradictory.
U	Unverified	No qualifying direct evidence was located under the stated protocol.

are explained by legacy DrugBank migrations involving valproic acid/divalproex and bismuth subsalicylate. None of them affect the diseases studied here.

Table 10: Top-15 candidates for Breast Cancer (OMIM 114480) on Fdataset. DrugBank membership and exclusion from the historical positive row are verified against DrugBank. Evidence tiers are defined in Table 9.

Rank	**CARE**	DIAM	MTP-DDA
1	**Everolimus** (DB01590; 0.9681; A)	Cisplatin (DB00515; 0.9518; B)	Dasatinib (DB01254; 0.9342; B)
2	**Carboplatin** (DB00958; 0.9613; B)	Everolimus (DB01590; 0.9427; A)	Norethindrone (DB00717; 0.9230; U)
3	**Vinorelbine** (DB00361; 0.9529; B)	Carboplatin (DB00958; 0.9367; B)	Prednisone (DB00635; 0.9149; I)
4	**Cisplatin** (DB00515; 0.9480; B)	Vinorelbine (DB00361; 0.9243; B)	Cisplatin (DB00515; 0.8999; B)
5	**Dasatinib** (DB01254; 0.9373; B)	Dasatinib (DB01254; 0.9161; B)	Levonorgestrel (DB00367; 0.8899; U)
6	**Metformin** (DB00331; 0.9302; B)	Metformin (DB00331; 0.9020; B)	Daunorubicin (DB00694; 0.8727; C)
7	**Simvastatin** (DB00641; 0.9180; B)	Sorafenib (DB00398; 0.8950; B)	Vincristine (DB00541; 0.8653; B)
8	**Propranolol** (DB00571; 0.9124; B)	Sunitinib (DB01268; 0.8841; N)	Cytarabine (DB00987; 0.8515; C)
9	**Sorafenib** (DB00398; 0.8984; B)	Celecoxib (DB00482; 0.8689; N)	Etoposide (DB00773; 0.8324; B)
10	**Disulfiram** (DB00822; 0.8919; C)	Simvastatin (DB00641; 0.8621; B)	Norgestimate (DB00957; 0.8232; U)
11	**Chloroquine** (DB00608; 0.8821; C)	Prednisone (DB00635; 0.8450; I)	Diethylstilbestrol (DB00255; 0.8068; B)
12	**Celecoxib** (DB00482; 0.8670; N)	Dexamethasone (DB01234; 0.8357; I)	Dexamethasone (DB01234; 0.7947; I)
13	**Sunitinib** (DB01268; 0.8592; N)	Etoposide (DB00773; 0.8222; B)	Teniposide (DB00444; 0.7745; C)
14	**Itraconazole** (DB01167; 0.8464; C)	Norgestimate (DB00957; 0.8040; U)	Ethinyl Estradiol (DB00977; 0.7631; B)
15	**Norgestimate** (DB00957; 0.8294; U)	Diethylstilbestrol (DB00255; 0.7930; B)	Carboplatin (DB00958; 0.7443; B)

Breast Cancer. The breast-cancer node already contains 23 historical positives, including tamoxifen, anastrozole, paclitaxel, docetaxel, lapatinib, and doxorubicin; these drugs are therefore excluded from candidate ranking. The CARE list instead prioritizes later or externally supported associations that remain zeros in the historical matrix. Everolimus provides the clearest temporal validation because the NCI now lists it as an approved breast-cancer agent, yet it remains absent from the 2011 positive row. Carboplatin, vinorelbine, and cisplatin add direct clinical support, whereas dasatinib, metformin, simvastatin, propranolol, and sorafenib illustrate candidates with human but heterogeneous evidence. The list also retains mechanistically grounded repositioning candidates, including disulfiram and chloroquine. Importantly, celecoxib and sunitinib are not counted as confirmations because subsequent human evidence is negative or insufficient, and Norgestimate remains explicitly unverified. This directional treatment of evidence avoids the common but misleading practice of counting any co-mention as successful validation.

Lung Cancer. The generic lung-cancer node (OMIM 211980) contains only doxorubicin as a historical positive, even though the separate small-cell-lung-cancer node contains five additional cytotoxic drugs. This ontology split makes lung cancer a useful temporal-recovery test: many drugs that are clinically established today remain unlabeled for the generic 211980 node. The CARE ranking places erlotinib and pemetrexed first, followed by platinum agents, taxanes, gemcitabine, vinorelbine, etoposide, and topotecan. Twelve of its top fifteen candidates now have established clinical-use evidence. DIAM retrieves a similar core but weights exploratory agents more heavily, whereas the MTP-DDA list includes several hematologic or endocrine drugs with weak direct lung-cancer evidence. The result should be interpreted as recovering missing historical labels, not discovering fifteen genuinely novel therapies.

Table 11: Top-15 candidates for Lung Cancer (OMIM 211980) on Fdataset. DrugBank membership and exclusion from the historical positive row are verified against the data. Evidence tiers are defined in Table 9.

Rank	CARE	DIAM	MTP-DDA
1	**Erlotinib** (DB00530; 0.9744; A)	Erlotinib (DB00530; 0.9572; A)	Tamoxifen (DB00675; 0.9287; C)
2	**Pemetrexed** (DB00642; 0.9676; A)	Pemetrexed (DB00642; 0.9481; A)	Cisplatin (DB00515; 0.9175; A)
3	**Carboplatin** (DB00958; 0.9592; A)	Cisplatin (DB00515; 0.9421; A)	Vincristine (DB00541; 0.9094; B)
4	**Cisplatin** (DB00515; 0.9543; A)	Carboplatin (DB00958; 0.9297; A)	Sorafenib (DB00398; 0.8944; B)
5	**Paclitaxel** (DB01229; 0.9436; A)	Paclitaxel (DB01229; 0.9215; A)	Tretinoin (DB00755; 0.8844; C)
6	**Docetaxel** (DB01248; 0.9365; A)	Docetaxel (DB01248; 0.9074; A)	Sunitinib (DB01268; 0.8672; B)
7	**Gemcitabine** (DB00441; 0.9243; A)	Gemcitabine (DB00441; 0.9004; A)	Fludarabine (DB01073; 0.8598; U)
8	**Vinorelbine** (DB00361; 0.9187; A)	Vinorelbine (DB00361; 0.8895; A)	Methotrexate (DB00563; 0.8460; A)
9	**Etoposide** (DB00773; 0.9047; A)	Etoposide (DB00773; 0.8743; A)	Cladribine (DB00242; 0.8269; U)
10	**Topotecan** (DB01030; 0.8982; A)	Topotecan (DB01030; 0.8675; A)	Prednisone (DB00635; 0.8177; I)
11	**Everolimus** (DB01590; 0.8884; A)	Sorafenib (DB00398; 0.8504; B)	Docetaxel (DB01248; 0.8013; A)
12	**Methotrexate** (DB00563; 0.8733; A)	Sunitinib (DB01268; 0.8411; B)	Methylprednisolone (DB00959; 0.7892; I)
13	**Sorafenib** (DB00398; 0.8655; B)	Methotrexate (DB00563; 0.8276; A)	Mercaptopurine (DB01033; 0.7690; U)
14	**Sunitinib** (DB01268; 0.8527; B)	Tretinoin (DB00755; 0.8094; C)	Dacarbazine (DB00851; 0.7576; C)
15	**Celecoxib** (DB00482; 0.8357; B)	Celecoxib (DB00482; 0.7984; B)	Flutamide (DB00499; 0.7388; U)

Table 12: Top-15 candidates for Alzheimer Disease (OMIM 104300) on Fdataset. DrugBank membership and exclusion from the historical positive row are verified against the data. Evidence tiers are defined in Table 9.

Rank	CARE	DIAM	MTP-DDA
1	**Lithium** (DB01356; 0.9296; B)	Tacrolimus (DB00864; 0.9113; C)	Prednisone (DB00635; 0.8875; U)
2	**Tacrolimus** (DB00864; 0.9228; C)	Progesterone (DB00396; 0.9022; C)	Epoprostenol (DB01240; 0.8763; U)
3	**Progesterone** (DB00396; 0.9144; C)	Sertraline (DB01104; 0.8962; B)	Methamphetamine (DB01577; 0.8682; U)
4	**Zileuton** (DB00744; 0.9095; C)	Methamphetamine (DB01577; 0.8838; U)	Vigabatrin (DB01080; 0.8532; U)
5	**Sertraline** (DB01104; 0.8988; B)	Prednisone (DB00635; 0.8756; U)	Sertraline (DB01104; 0.8432; B)
6	**Fluoxetine** (DB00472; 0.8917; B)	Vigabatrin (DB01080; 0.8615; U)	Progesterone (DB00396; 0.8260; C)
7	**Pioglitazone** (DB01132; 0.8795; N)	Epoprostenol (DB01240; 0.8545; U)	Cyclosporine (DB00091; 0.8186; C)
8	**Cyclosporine** (DB00091; 0.8739; C)	Zileuton (DB00744; 0.8436; C)	Tacrolimus (DB00864; 0.8048; C)
9	**Simvastatin** (DB00641; 0.8599; N)	Cyclosporine (DB00091; 0.8284; C)	Zileuton (DB00744; 0.7857; C)
10	**Minocycline** (DB01017; 0.8534; N)	Dextroamphetamine (DB01576; 0.8216; I)	Dextroamphetamine (DB01576; 0.7765; I)
11	**Ibuprofen** (DB01050; 0.8436; N)	Lithium (DB01356; 0.8045; B)	Fluoxetine (DB00472; 0.7601; B)
12	**Naproxen** (DB00788; 0.8285; N)	Pioglitazone (DB01132; 0.7952; N)	Simvastatin (DB00641; 0.7480; N)
13	**Rosiglitazone** (DB00412; 0.8207; N)	Minocycline (DB01017; 0.7817; N)	Pioglitazone (DB01132; 0.7278; N)
14	**Dextroamphetamine** (DB01576; 0.8079; I)	Ibuprofen (DB01050; 0.7635; N)	Minocycline (DB01017; 0.7164; N)
15	**Vigabatrin** (DB01080; 0.7909; U)	Naproxen (DB00788; 0.7525; N)	Naproxen (DB00788; 0.6976; N)

Alzheimer Disease. The Alzheimer's disease node contains seven historical positives: donepezil, rivastigmine, memantine, galantamine, valproic acid, vitamin E, and tacrine. The remaining candidate space is substantially harder than the cancer cases because most repositioning signals are preclinical, symptomatic, or contradicted by later trials. The CARE list ranks lithium first because a meta-analysis of placebo-controlled studies reported a modest cognitive signal. Tacrolimus is supported by recent mouse-model evidence, and zileuton has mechanistic support through reduced tau phosphorylation. Progesterone and sertraline are retained but labeled mixed because their evidence is either mechanistic or symptom-directed. Pioglitazone, minocycline, statins, and NSAIDs are not presented as successful validations: later human studies are neutral or negative. This case therefore tests whether a model ranks the *quality and direction* of evidence, rather than merely retrieving drugs frequently discussed with dementia.

Robustness. The evidence-discounted comparison is more informative than a raw confirmation ratio. In breast and lung cancer, CARE concentrates established or human-tested candidates near the top. In Alzheimer disease, it promotes the limited supportive signals and demotes candidates contradicted by randomized trials. Thus, the intended robustness claim is not that every predicted association is clinically effective, but that the ranking is less sensitive to ontology incompleteness and less likely to overvalue indirect literature co-occurrence.

Table 13: Evidence-oriented comparison of the rankings. Evidence@K counts candidates with tier A–C evidence, whereas Support@15 requires directionally supportive evidence. E-DCG@15 discounts weaker and lower-ranked evidence.

Disease	Model	Evidence@5	Evidence@10	Support@15	Negative	Unverified	E-DCG@15
Breast Cancer	CARE	5/5	10/10	6/15	2	1	3.387
Breast Cancer	DIAM	5/5	8/10	6/15	2	1	3.255
Breast Cancer	MTP-DDA	2/5	6/10	5/15	0	3	2.151
Lung Cancer	CARE	5/5	10/10	12/15	0	0	5.439
Lung Cancer	DIAM	5/5	10/10	11/15	0	0	5.230
Lung Cancer	MTP-DDA	5/5	7/10	4/15	0	4	2.446
Alzheimer Disease	CARE	5/5	7/10	3/15	6	1	1.988
Alzheimer Disease	DIAM	3/5	5/10	3/15	4	4	1.480
Alzheimer Disease	MTP-DDA	1/5	5/10	2/15	4	4	0.874