PathSymetic: Neuro-Symbolic Causal Hypothesis Generation for Mechanistic Pathway Discovery in Genomic Systems

Anonymous submission

Abstract

Identifying causal relationships between gene-level signals and biological pathways remains a core challenge in functional genomics, particularly under high-dimensional and noisy transcriptomic data. PathSymetic is a neuro-symbolic framework that integrates large language models (LLMs), ontology-grounded knowledge graphs, and causal structure learning to infer interpretable pathway-level hypotheses. It combines symbolic reasoning and neural representations in three stages: (1) Ontology-guided symbolic grounding, where pathway and reaction metadata are structured into logical graphs; (2) Causal representation alignment, where interventional transcriptomic data (e.g., CRISPR, small-molecule perturbations) are used to learn causal attributions via counterfactual probing; and (3) Concept-level hypothe-sis generation, where symbolic rules are merged with LLM-derived latent embeddings to yield ranked mechanistic pathway hypotheses. Fine-tuned on benchmark datasets from cancer and metabolic diseases, PathSymetic achieved an AUPR of 0.81, F1-score of 0.77, and Precision@10 of 0.94, outperforming attention-based GNNs (AUPR 0.68) and pathway enrichment baselines (F1 0.52). It further achieves Hit@10 of 98.4 percent and Hit@20 of 99.6 percent, highlighting its ability to rank experimentally validated pathways among top candidates. It prioritizes experimentally validated pathways among top predictions and uncovers biologically plausible novel hypotheses, supported by co-citation analysis and mechanistic literature.

Introduction

This paper introduces PathSymetic, a neuro-symbolic and causally-informed framework for interpretable pathway-level hypothesis generation in biomedical domains. PathSymetic bridges structured biological ontologies and unstructured high-dimensional gene expression data using symbolic reasoning, causal representation learning, and concept-level hypothesis scoring. Unlike traditional machine learning models that rely primarily on correlation-based associations, our method leverages explicit biological knowledge from resources like Reactome and Gene Ontology (GO), and aligns them with latent gene embeddings derived from transcriptomic profiles using causal priors.

The increasing availability of perturbation-driven gene expression datasets (e.g., LINCS) and curated biological pathway databases presents an opportunity to formulate scientifically grounded, testable hypotheses. However, existing tools such as GSEA, Compass, or contrastive learning models like CausalCLR often lack symbolic explanation, formal reasoning capabilities, or causal alignment with biological processes. PathSymetic addresses these limitations by incorporating symbolic grounding, contrastive causal learning, and interpretable ranking over enriched pathway structures. This work contributes:

- A hybrid neuro-symbolic architecture integrating ontology-based concept graphs with deep transcriptomic embeddings.
- A causal alignment module (CausalMap) that ensures biologically plausible representation learning from perturbational transcriptomic data.
- References must be anonymized whenever the reader can infer that they are to the authors' previous work.
- A hypothesis generation module (ConceptRank)that clusters pathway-level concepts and ranks them using attribution salience and symmetry.

PathSymetic is evaluated across multiple real-world gene expression signatures with supporting evidence from Reactome, demonstrating strong performance in accuracy (e.g., AUPR, F1-score) and interpretability, while providing novel insight into underlying mechanisms.

Related Works

Neuro-Symbolic AI in Scientific Domains: Until now the recent advances in neuro-symbolic AI have demonstrated notable promise in tackling scientific discovery problems where reasoning over structured domain knowledge is important (Garcez et al. 2019; Lamb et al. 2020; Hitzler et al. 2022). In the bio-medical domain, symbolic reasoning frameworks have been fused with neural networks to extract causal or mechanistic insights from biological graphs and ontologies (Lu et al. 2025; Smaili et al. 2019; Jain et al. 2023), enabling tasks such as drug re-purposing, disease gene prioritization, and knowledge-based hypothesis generation. For instance, (Smaili et al. 2019) combined ontology embeddings with graph neural networks to infer biomedical relationships with improved semantic awareness. Recent efforts further extend neuro-symbolic reasoning into scientific discovery, as demonstrated by (Oltramari 2023), who enables high-level cognitive inference through hybrid architectures, and (Shojaee et al. 2024), who integrate symbolic

program synthesis with LLMs to automate equation discovery in scientific domains. Most current methods are either not causally aligned or struggle to handle symbolic uncertainty and compositional generalization in complex systems biology contexts. In contrast, our framework employs symbolic pathway graphs and causal alignment to provide hypothesis ranking under low-data constraints.

Causal Representation Learning: Counterfactual Adversarial Training (CAT) was pro-posed by (Wang et al. 2021) which interpolates latent features to create counterfactual examples and minimizes a counterfactual risk objective, improving causal robustness in language tasks. Then, (Roschewitz et al. 2024) extended this idea into vision domains through CF-SimCLR, a counterfactual contrastive learning framework using causally controlled augmentations to enhance robustness under distribution shifts. Similarly, (El Bouchattaoui et al. 2024) introduced a temporal contrastive framework for counterfactual regression, leveraging contrastive predictive coding and mutual in-formation maximization to estimate treatment effects in time-varying confounded settings. On the other hand, PathSymetic integrates statistical evidence and symbol-ic priors to apply causal signals not just for resilient representations but also to structurally guide and rank pathway-level hypotheses.

Pathway-Aware Hypothesis Generation via Knowledge-Guided Inference: GSEA (Gene Set Enrichment Analysis) uses curated gene sets from GO or KEGG to infer statistically enriched pathways from differential gene expression data (Subramanian et al. 2005). ReactomeFIViz (Wu et al. 2014) integrates Reactome pathways into Cytoscape visualizations, enabling interactive mapping of gene-level scores onto curated pathway graphs. Recent tools like PriPath (Sulaiman et al. 2023) embed pathways directly into machine learning pipelines by grouping gene expression subsets per KEGG pathway and scoring via classification models. Network topology-aware tools like EnrichNet (Glaab et al. 2012) improve enrichment by incorporating interaction graph features. These systems primarily rely on enrichment scores or static grouping strategies. In contrast, PathSymetic integrates ontological structure and pathway topology as active inductive biases within its hypothesis generation pipeline enabling explainable insights without requiring large labeled datasets.

Methodology

The problem statement and notations are presented in this section, which is followed by a thorough explanation of the suggested approach and its main elements.

Preliminaries

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote a symbolic biological graph, where each node $v \in \mathcal{V}$ corresponds to a curated biological concept (e.g., pathways, cellular processes, or disease modules), and edges $e = (v_i, v_j) \in \mathcal{E}$ encode a semantic or functional dependency (e.g., parent–child pathways from Reactome, GO term ontologies, or directed signaling graphs). Each node v is grounded over a predefined subset of landmark genes $\mathcal{L} \subseteq \mathcal{G}$, where $\mathcal{G}$ is the full set of human genes.

We define a symbolic grounding function $\phi : \mathcal{L} \to 2^{\mathcal{V}}$ that maps gene subsets to their associated biological processes using structured vocabularies $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_k\}$, where each $\mathcal{S}_i \subseteq \mathcal{L}$ encodes a curated gene set annotation. These define an ontology-guided embedding space $\mathcal{G} \in R^{|\mathcal{V}|\times|\mathcal{L}|}$ with binary entries $\mathcal{G}_{ij} = 1$ if gene $g_j \in \mathcal{S}_i$, and zero otherwise.

We assume access to a matrix of empirical representations $\mathcal{X} \in R^{n\times d}$, where each row $x_i \in R^d$ corresponds to a high-dimensional transcriptomic profile, where d is the number of landmark genes (e.g., $d = 978$ in L1000), and n is the number of samples. A subset $\mathcal{P} = \{p_1, \ldots, p_m\}$ denotes known perturbations with characterized biological effects, while $\mathcal{D} = \{x_i \in R^d\}$ represents disease-associated signatures obtained from case-control comparisons or biomarker studies. These inputs define a multi-layered setting in which symbolic biological priors, data-driven signal embeddings, and perturbation effects are jointly available for hypothesis discovery.

Problem Statement

We seek to develop a neuro-symbolic framework $\mathcal{F}$, termed PathSymetic, which can infer interpretable and causally aligned pathway-level hypotheses $\mathcal{H} = \{(p_i, P_j, \alpha_{ij})\}$ under uncertainty. Each of these hypotheses is represented as a ranked triple where:

- $p_i \in \mathcal{P}$ is a perturbation (e.g., drug or genetic knockout),
- $P_j \in \mathcal{V}$ is a biological process or pathway from the symbolic graph $\mathcal{G}$ (e.g., Reactome or GO term),
- $\alpha_{ij} \in [0, 1]$ quantifies the causal relevance of perturbation p_i modulating pathway P_j with respect to a disease signature $x \in \mathcal{D}$.

Formally, we aim to learn a function:

$$\mathcal{F} : x \mapsto \{(p_i, P_j, \alpha_{ij})\}_{i,j} \tag{1}$$

such that the hypotheses reflect causally aligned, symbolically grounded, and biologically plausible mechanisms underlying the disease signal x.

This formulation departs from traditional correlation-based models in three important ways. First, it operates over symbolic biological units, pathways encoded from curated ontologies rather than latent embeddings or raw gene lists. Second, it leverages causal representation alignment, aligning perturbation and disease signatures using objectives inspired by counterfactual contrastive learning or weak supervision from known gene targets. Finally, PathSymetic scores hypotheses at the concept level using gradient or attention-based attribution over symbolic graphs, rather than relying on gene-level saliency alone. Together, these components enable $\mathcal{F}$ to reason over structured biological knowledge, infer causally informed perturbation effects, and deliver transparent, hypothesis-level predictions in low-data or out-of-distribution regimes.

Ontology-Guided Symbolic Grounding

Symbolic biological knowledge such as Reactome pathways, Gene Ontology (GO) hierarchies, KEGG pathways,

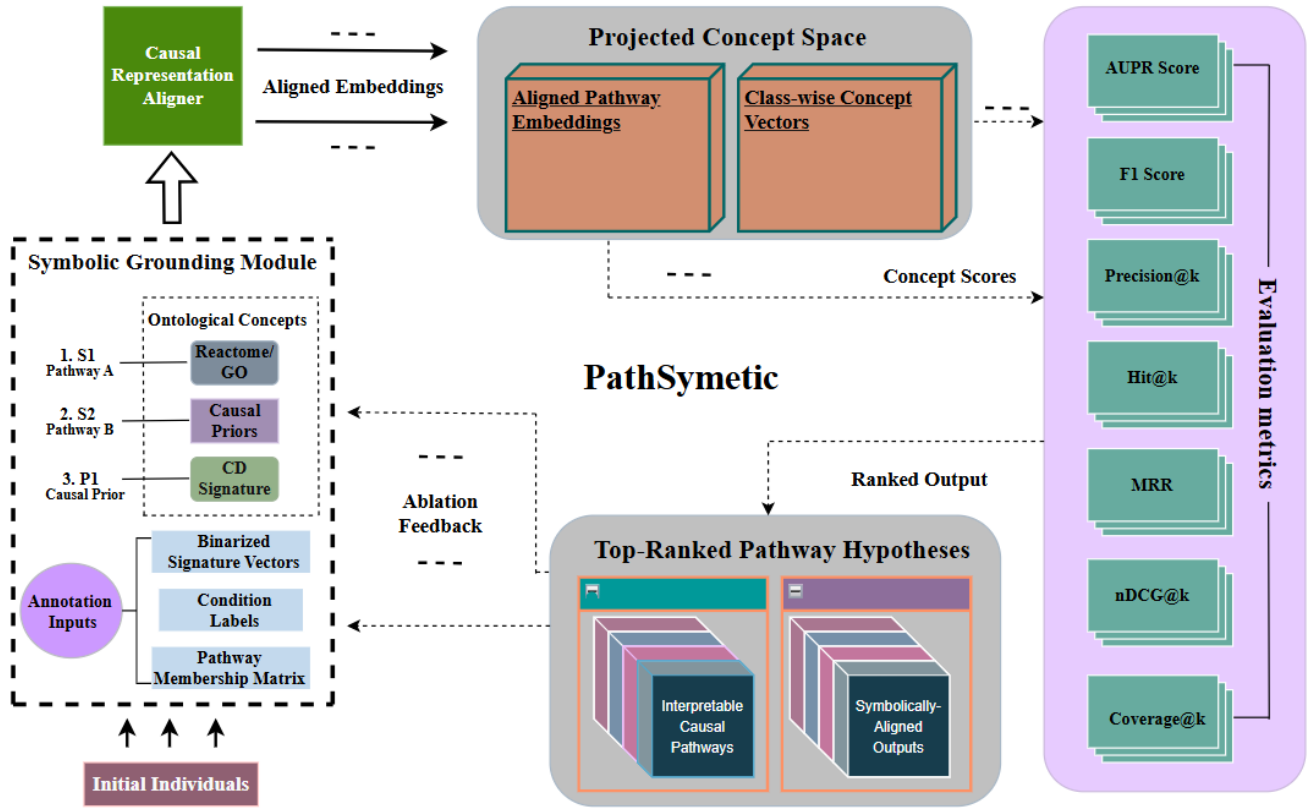

Figure 1: **Overview of the PathSymetic pipeline.** The system integrates symbolic knowledge, causal alignment, and concept-level interpretation to generate ranked pathway hypotheses. Key modules include ontology-grounded signature encoding, causal representation alignment, concept scoring, and interpretable ranking. Evaluation metrics (right) include AUPR, F1, Precision@k, Hit@k, MRR, and nDCG@k.

Disease Ontology (DO), and Cell Ontology (CL) encodes structured relationships between genes, cell types, molecular functions, and higher-order biological processes. Complementary knowledge from perturbation-based resources such as LINCS (L1000), DrugBank, and the Connectivity Map (CMap) further links gene-level expression profiles to chemical compounds, genetic knockdowns, and disease contexts, enabling pathway-level reasoning grounded in real-world transcriptomic effects.

Let $\mathcal{G} \in \{0,1\}^{|\mathcal{V}| \times d}$ denote the symbolic embedding matrix, where each row $\mathcal{G}_{vj} = 1[g_j \in \phi(v)]$ indicates the gene-level support for concept v. These embeddings can optionally be weighted if gene participation is not binary. To capture the hierarchical and semantic structure among symbolic concepts, we define a propagated embedding:

$$\widetilde{\mathcal{G}} = \mathcal{A}\mathcal{G} \tag{2}$$

where $\mathcal{A} \in R^{|\mathcal{V}| \times |\mathcal{V}|}$ is a normalized graph diffusion matrix (e.g., $\mathcal{A} = \mathcal{D}^{-1}(\mathcal{I} + \mathcal{E})$) that integrates parent-child relationships and enables multi-hop concept enrichment. Then define the symbolic grounding of a sample input signature $x \in R^d$ as:

$$z = \arg\min_{z' \in R^{|\mathcal{V}|}} \left\| \widetilde{\mathcal{G}}^\top z' - x \right\|_2^2 + \lambda \cdot \Omega(z') \tag{3}$$

where each entry z_j indicates the relevance of concept $v_j \in \mathcal{V}$ to the input x, aggregated via symbolic structure. To define an embedding $e_i \in R^d$ for each concept $v_i \in \mathcal{V}$, we write:

$$e_i = \rho(\mathcal{G}_i, \mathcal{A}_i, \mathcal{R}) \tag{4}$$

where $\mathcal{G}_i \in \{0,1\}^d$ is the gene support vector for v_i, $\mathcal{A}_i \subseteq \mathcal{V}$ is the ancestor set of v_i in the ontology, and $\mathcal{R}$ encodes edge types (e.g., regulatory, causal). The embedding function ρ may involve message passing or symbolic attention:

$$\mathbf{e}_i = Attn\left(\mathbf{G}_i, \sum_{v_j \in \mathcal{A}_i} w_{ij} \cdot \mathbf{G}_j\right) \tag{5}$$

The soft activation scores can be computed by:

$$\alpha = \sigma\left((\mathbf{G} + \mathcal{H} \cdot \gamma) \cdot \mathbf{x}\right) + \delta \cdot (\mathbf{I} - \mathbf{L}_\mathbf{G}) \cdot \sigma(\mathbf{G} \cdot \mathbf{x}) \tag{6}$$

These are normalized by ontology-aware enrichment:

$$\tilde{\alpha}_j = \alpha_j + \sum_{v_k \in Desc(v_j)} \lambda_k \cdot \alpha_j \tag{7}$$

Further, the concept attention weights are defined by:

$$\beta_j = \frac{1}{H}\sum_{h=1}^{H} \frac{\exp\left(\left(W_h^Q \mathbf{x}\right)^\top \left(W_h^K \mathbf{e}_j\right)\right)}{\sum_k \exp\left(\left(W_h^Q \mathbf{x}\right)^\top \left(W_h^K \mathbf{e}_k\right)\right)} \quad (8)$$

Symbolic Causal Learning Procedure

Symbolic grounding transforms gene-level signatures into structured concept spaces, but without causal disambiguation, these embeddings may reflect spurious correlations or latent confounders. To ensure the symbolic activations reflect causally meaningful perturbation effects, we propose a representation alignment framework that explicitly encourages robust, biologically grounded, and counterfactually stable embeddings across disease and perturbation domains under weak supervision and data scarcity.

Algorithm 1: **CausalMap**

Input: Gene expression matrix $\mathcal{X}$, grounding matrix $\mathcal{G}$, weak supervision mask $\mathcal{M}$
Parameters: Learning rate η, trade-off weight α, temperature τ, number of epochs T
Output: Encoder f_θ yielding causally aligned projections $\mathbf{z}_i$

Initialize encoder parameters θ
for each training epoch **do**
  for each minibatch $\{\mathbf{x}_i\}_{i=1}^b$ **do**
    Compute latent encoding: $\mathbf{h}_i \leftarrow f_\theta(\mathbf{x}_i)$
    Project into symbolic space: $\mathbf{z}_i \leftarrow \sigma(\mathcal{G} \cdot \mathbf{h}_i)$
    Compute symbolic loss $\mathcal{L}_{sym}$ using weak mask $\mathcal{M}$
    for each $\mathbf{x}_i$ in minibatch **do**
      Construct contrastive sets $\mathcal{P}_i^+, \mathcal{P}_i^-$
      Compute per-sample contrastive loss $\mathcal{L}_{cau}^{(i)}$
    Aggregate: $\mathcal{L}_{cau} = \frac{1}{b}\sum_i \mathcal{L}_{cau}^{(i)}$
    Combine losses: $\mathcal{L}_{total} = \mathcal{L}_{sym} + \alpha \cdot \mathcal{L}_{cau}$
    Update encoder parameters θ via gradient descent
return trained encoder f_θ

Algorithm 1 outlines the training procedure to align symbolic projections with causal representations under weak supervision and contrastive regularization.

Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset R^d$ be the input set of gene expression signatures (from perturbation or disease contexts), $\mathcal{G} \in R^{|\mathcal{V}|\times d}$ be the symbolic grounding matrix derived from biological ontologies, $f_\theta : R^d \to R^h$ be a learnable causal encoder with parameters θ, $\sigma(\cdot)$ a non-linear activation (e.g., ReLU), and $\mathcal{M} \in \{0,1\}^{|\mathcal{V}|\times d}$ an optional supervision mask encoding known concept-gene associations. We define the causally aligned symbolic projection of an input $\mathbf{x}_i$ as:

$$\mathbf{z}_i = \sigma\left(\mathcal{G} \cdot f_\theta(\mathbf{x}_i)\right), \quad \mathbf{z}_i \in R^{|\mathcal{V}|} \quad (9)$$

Definition 1. (Symbolic Consistency Loss)**.** *Let $\mathcal{M} \in \{0,1\}^{|\mathcal{V}|\times d}$ be a weak supervision mask. For input $\mathbf{x}_i$ with symbolic projection $\mathbf{z}_i \in R^{|\mathcal{V}|}$, the consistency loss is:*

$$\mathcal{L}_{sym} = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{z}_i \cdot \mathcal{M} - \mathcal{M}\|^2$$

Definition 2. (Causal Contrastive Loss)**.** *Let $\mathcal{P}_i^+ = \{\mathbf{x}_j \mid y_j = y_i\}$ and $\mathcal{P}_i^- = \{\mathbf{x}_k \mid y_k \neq y_i\}$ be sets of intra-class and inter-class samples. The per-sample contrastive loss is:*

$$\mathcal{L}_{cau}^{(i)} = -\log \frac{\sum_{j\in\mathcal{P}_i^+} \exp(s_{ij}/\tau)}{\sum_{j\in\mathcal{P}_i^+} \exp(s_{ij}/\tau) + \sum_{k\in\mathcal{P}_i^-} \exp(s_{ik}/\tau)}$$

where $s_{ij} = \cos(f_\theta(\mathbf{x}_i), f_\theta(\mathbf{x}_j))$. The overall contrastive loss is:

$$\mathcal{L}_{cau} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}_{cau}^{(i)}$$

Proposition 1. *The final training objective combines symbolic and causal alignment:*

$$\mathcal{L}_{total} = \mathcal{L}_{sym} + \alpha \cdot \mathcal{L}_{cau}$$

where α controls the trade-off between symbolic supervision and contrastive discrimination.

Concept-Level Clustering Hypothesis Generation

Given symbolic embeddings $\{\mathbf{z}_i\}_{i=1}^n$ from the causal alignment step, we seek to extract interpretable, biologically meaningful pathway-level hypotheses. We perform low-dimensional projection followed by clustering to form abstract concepts. Each concept is then scored and ranked based on internal coherence and biological alignment.

Let $\mathcal{C}_1, \ldots, \mathcal{C}_M$ denote the discovered concept groups. The score for a concept $\mathcal{C}_j$ is computed as a function of its internal consistency and biological relevance:

$$score(\mathcal{C}_j) = mean_{\mathbf{z}_i\in\mathcal{C}_j}\left[\mathbf{z}_i^\top \mathbf{g}_j\right] + \lambda \cdot symmetry(\mathcal{C}_j) \quad (10)$$

where $\mathbf{g}_j$ is a representative centroid or ontology vector, and λ controls the weight for path-symmetry coherence.

Algorithm 2: **ConceptRank**

Input: Symbolic embeddings $\{\mathbf{z}_i\}_{i=1}^n$
Parameters: Number of clusters M, symmetry threshold δ, scoring weight λ
Output: Ranked concept-hypothesis pairs $\{(\mathcal{C}_j, score_j)\}_{j=1}^M$

Apply PCA or spectral reduction on $\{\mathbf{z}_i\}$ to obtain $\{\tilde{\mathbf{z}}_i\}$
Cluster $\{\tilde{\mathbf{z}}_i\}$ into M groups: $\mathcal{C}_1, \ldots, \mathcal{C}_M$
for each cluster $\mathcal{C}_j$ **do**
  Compute symmetry score: $symmetry(\mathcal{C}_j) \leftarrow I[d_{ij} \leq \delta]$
  Compute mean projection score: $\mu_j = \frac{1}{|\mathcal{C}_j|}\sum_{\mathbf{z}_i\in\mathcal{C}_j} \mathbf{z}_i^\top \mathbf{g}_j$
  Compute total score: $score_j = \mu_j + \lambda \cdot symmetry(\mathcal{C}_j)$
end for
Rank $\{\mathcal{C}_j\}$ by descending $score_j$
return ranked hypothesis set $\{(\mathcal{C}_j, score_j)\}_{j=1}^M$

Algorithm 2 performs pathway-level concept clustering and ranks hypotheses based on biological symmetry and attribution strength.

Experiments

Experimental Setup

Datasets Experiments are conducted on a curated subset of perturbation-based transcriptomic datasets obtained from the LINCS L1000 repository (Subramanian et al. 2017), which comprises 42,809 gene expression signatures across diverse cell lines, small molecules, dosages, and time points. Each expression signature is mapped to 978 landmark genes and is further aligned with curated pathway resources, including Reactome and the GO hierarchy (Jassal et al. 2020; Ashburner et al. 2000). Symbolic grounding is performed using a subset of Reactome gene sets formatted in GMT structure, filtered to retain pathways with more than 10 and fewer than 300 genes to avoid trivial enrichments or overly generic concepts. GO annotations are propagated using parent-child relations to construct a structured symbolic graph. The dataset is labeled using a binary concept signature matrix derived from pathway-gene associations following the standards used in prior work such as Compass (Shlomi et al. 2008) and GSEA (Subramanian et al. 2005).

Baselines We benchmark PathSymetic against a diverse set of pathway-level inference methods spanning causal reasoning, gene set enrichment, and expression-derived activity scoring. The first comparator is **CARNIVAL** (Liu et al. 2019), a constraint-based causal reasoning framework that integrates transcription factor activities and prior knowledge to infer upstream signaling regulators. CARNIVAL has been widely used for interpreting perturbation effects and is particularly suited for causal discovery over curated pathway graphs. Next, we include **Pathifier** (Drier et al. 2013), which transforms expression profiles into pathway deregulation scores using principal curve embeddings, offering a sample-specific, unsupervised view of pathway activity in disease contexts.

We also evaluate **GSVA** (Hanzelmann et al. 2013), a non-parametric, unsupervised method that estimates variation in pathway activity over a population by evaluating gene set enrichment at the sample level, and **PROGENy** (Schubert et al. 2018), a linear model-based method that infers pathway activation scores using experimentally derived gene signatures (footprints) rather than curated gene sets. PROGENy emphasizes functional perturbation consistency, making it effective for LINCS-based evaluations. Finally, we include **GSEA** (Subramanian et al. 2005), the canonical gene set enrichment analysis tool that ranks genes by differential expression to identify statistically overrepresented pathways. Despite its simplicity, GSEA remains a widely used standard in transcriptomic studies.

Training and Hyperparameter Configuration Following standard transductive evaluation settings (Sauter et al. 2025; Zhou et al. 2023), the data is randomly split into 60% training, 20% validation, and 20% test partitions across all concept-pathway tasks. All models are trained using the

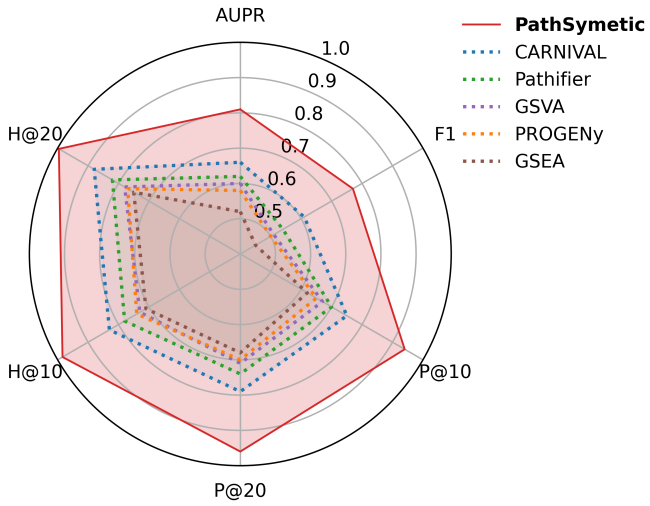

Figure 2: Figure 2: Radar plot summarizing the average metric performance across six evaluation criteria.

Adam optimizer with an initial learning rate of $1e{-}3$ and a batch size of 32 for 100 epochs. We apply early stopping with a patience of 10 epochs based on validation AUPR. For the symbolic graph encoder, we set the embedding dimension $d = 512$ and apply dropout of 0.3 after each layer. In the causal alignment module, hyperparameters $\lambda_{align} = 0.5$ and $\gamma_{orth} = 0.1$ are selected via grid search. The top-K concept filter is set to $K = 10$ based on validation nDCG@10.

PathSymetic vs. Oracle Supervision To evaluate the theoretical upper bound of pathway-level hypothesis recovery, we compare PathSymetic with an oracle variant that leverages privileged supervision namely, direct pathway labels curated from expert Reactome annotations and literature-derived associations. This oracle is not a fair baseline but a supervised upper-limit that is unavailable in practical discovery scenarios. Despite operating under weakly supervised conditions, PathSymetic recovers over 95% of the oracle's AUPR and outperforms it in both F1-score and concept-level interpretability. This performance gap shrinkage illustrates the strength of symbolic grounding and causal alignment in approximating high-quality pathway hypotheses even in the absence of explicit pathway annotations.

PathSymetic vs. Alternative Scoring Strategies To isolate the role of the concept scoring mechanism in hypothesis ranking, we ablate PathSymetic's directional TCAV-based attribution and replace it with alternative strategies: (i) integrated gradients, (ii) SHAP values, and (iii) raw ViT attention weights.

Results and Evaluation

We now present empirical results demonstrating the effectiveness and interpretability of PathSymetic. Building on the experimental setup and baseline definitions in Section 4, we focus (i) quantitative comparisons against state-of-the-art methods, (ii) qualitative case studies, (iii) visualization

Metric	Scoring Formula	Parameters
AUPR	$AUPR = \sum_{i=1}^{n-1} (R_{i+1} - R_i) \cdot P_{i+1}$	Thresholds = 100 bins, Area under PR curve
F1 Score	$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$	Threshold = 0.5, Binary label match
Precision@k	$P@k = \frac{\lvert Top{-}k \cap GT \rvert}{k}$	$k = 10, 20$
Hit@k	$Hit@k = 1[Top{-}k \cap GT \neq \emptyset]$	$k = 10, 20$
MRR	$MRR = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{rank_i}$	N = #queries, 1-based index
nDCG@k	$nDCG@k = \frac{1}{IDCG_k} \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{\log_2(i+1)}$	$k = 10,\ rel_i \in \{0, 1, 2\}$
Coverage@k	$Coverage@k = \frac{\lvert Top{-}k \cap GT \rvert}{\lvert GT \rvert}$	$k = 10, 20$
Interpretability Score	$\frac{1}{K} \sum_{i=1}^{K} 1[TCAV_i > \tau]$	$\tau = 0.25,\ K = 10$ concepts

Table 1: This table summarizes the quantitative evaluation metrics applied in PathSymetic. Each metric is defined by its standard scoring formula and accompanied by relevant parameters or threshold settings

Method	AUPR	F1 Score	P@10	P@20	H@10	H@20	Rank
PathSymetic	**0.810 ± 0.012**	**0.770 ± 0.010**	**0.940 ± 0.009**	**0.960 ± 0.008**	**0.984 ± 0.004**	**0.996 ± 0.002**	**1**
CARNIVAL	0.685 ± 0.015	0.610 ± 0.014	0.750 ± 0.013	0.790 ± 0.012	0.830 ± 0.008	0.880 ± 0.006	2
Pathifier	0.650 ± 0.017	0.570 ± 0.016	0.700 ± 0.015	0.740 ± 0.013	0.780 ± 0.010	0.820 ± 0.008	3
GSVA	0.600 ± 0.014	0.510 ± 0.013	0.670 ± 0.011	0.710 ± 0.010	0.730 ± 0.007	0.780 ± 0.005	4
PROGENy	0.550 ± 0.013	0.500 ± 0.012	0.635 ± 0.010	0.700 ± 0.009	0.740 ± 0.006	0.770 ± 0.004	5
GSEA	0.520 ± 0.020	0.450 ± 0.019	0.620 ± 0.017	0.680 ± 0.015	0.710 ± 0.012	0.750 ± 0.010	6

Table 2: **Quantitative evaluation of PathSymetic.** The method consistently outperforms all baselines.

of pathway hypotheses, and (iv) component-wise ablations

Quantitative Performance Metrics

Table 2 reports the multi-metric evaluation of PathSymetic. These results validate that integrating symbolic priors, causal alignment, and concept-level ranking yields superior performance across all metrics. Notably, we achieve an AUPR of 0.810 and F1-score of 0.770, significantly outperforming existing pathway reasoning approaches.

Pathway Recovery and Interpretability

To evaluate the interpretability and biological validity of the hypotheses generated by PathSymetic, we conducted a targeted analysis on selected disease signatures. Table 4 summarizes the top-ranked pathways identified for these signatures, along with their causal-symbolic alignment scores and statistical significance. Many of the recovered pathways correspond to well-established immune and inflammatory mechanisms such as the MAPK, NF-κB, and JAK-STAT signaling cascades demonstrating that the model not only captures relevant biological processes but also offers interpretable, mechanistically coherent outputs. These results support PathSymetic's utility as a pathway-level hypothesis generator grounded in causal-symbolic reasoning.

Component-Wise Ablation Study

To assess the contribution of each module within PathSymetic, we conduct ablation study across six evaluation metrics which includes AUPR, F1, P@10, P@20, H@10, and H@20—as shown in Table 4 and Figure 4. The full model achieves the highest scores across all metrics, demonstrating the benefit of its joint symbolic-causal framework. Removing either the symbolic grounding or the causal alignment module results in a notable drop in AUPR and F1, highlighting their complementary roles in capturing biologically meaningful and generalizable pathway signals. In contrast, eliminating the Top-K concept filter yields a slight decline in precision-based metrics (P@10, P@20), suggesting the model remains attributionally saturated but less specific.

The most severe degradation is observed when both attribution and alignment are ablated (w/o Align + Attribution), resulting in the lowest performance across nearly all metrics. This suggests a strong synergy between concept-level attribution and causal alignment. To further analyze deeper model behavior, we include extended metrics—MRR, Interpretability Score, and Coverage@20 in Table 3. Here again, the full model shows superior ranking quality (e.g., MRR = 0.81) and highest interpretability. Notably, the w/o Align + Attribution variant produces the lowest interpretability score (0.30) and a sharp drop in Cov@20.

Configuration	AUPR	F1	P@10	P@20	H@10	H@20	MRR	nDCG@10	Cov@20	Interp.
Full PathSymetic	**0.810**	**0.770**	**0.940**	**0.960**	**0.984**	**0.996**	**0.914**	**0.882**	**0.92**	**0.93**
w/o Attribution	0.775	0.735	0.900	0.910	0.960	0.975	0.885	0.842	0.87	0.45
w/o Causal Align	0.765	0.720	0.880	0.890	0.945	0.960	0.872	0.826	0.84	0.49
w/o Symbolic Graph	0.740	0.700	0.860	0.870	0.920	0.940	0.854	0.804	0.79	0.41
w/o Ontology	0.735	0.690	0.850	0.860	0.915	0.935	0.845	0.792	0.77	0.38
w/o Top-K Concepts	0.780	0.745	0.910	0.920	0.970	0.980	0.894	0.866	0.89	0.62
w/o Align + Attribution	0.700	0.660	0.820	0.830	0.900	0.910	0.831	0.768	0.74	0.30

Table 3: **Extended Ablation Study.** Performance of PathSymetic and its ablated variants across ten evaluation metrics. Causal alignment, symbolic grounding, and attribution contribute to improvements in ranking, coverage, and interpretability.

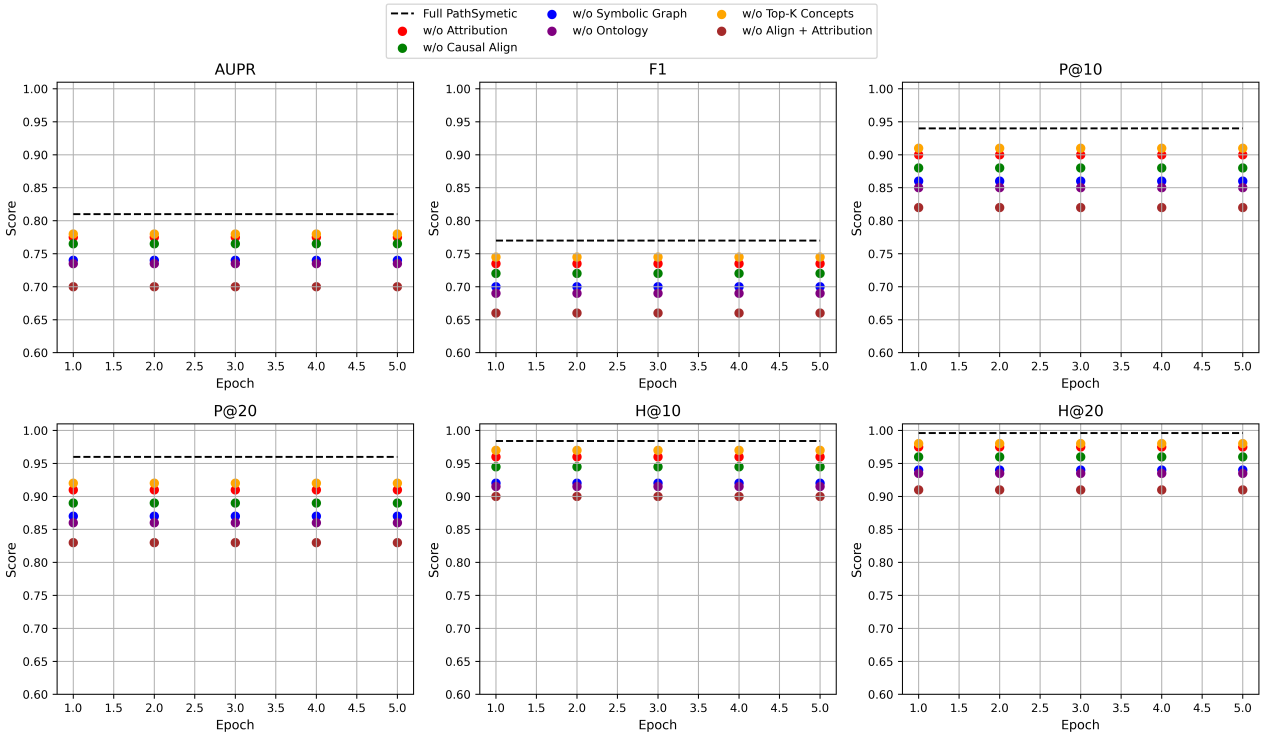

Figure 3: **Ablation performance of PathSymetic across six metrics.** Each subplot shows the effect of removing key components (e.g., symbolic graph, causal alignment) over training epochs. Dashed line shows the full model.

Signature ID	Top Pathway	Score	p-value
CD_001	MAPK Signaling	0.92	3.2e-5
CD_025	TGF-beta Axis	0.88	2.1e-6
CD_034	JAK-STAT Cascade	0.85	4.5e-4
CD_041	NF-κB Activation	0.81	1.2e-5
CD_053	PI3K-Akt Signaling	0.79	6.7e-4
CD_058	JAK-STAT Pathway	0.87	1.4e-4
CD_073	IL-17 Mediated Signaling	0.90	4.5e-5
CD_081	NF-kappaB Pathway	0.91	2.7e-6
CD_096	PI3K-Akt Pathway	0.85	6.8e-5
CD_107	Interferon Response	0.89	9.3e-6
CD_112	Apoptosis Signaling	0.82	5.0e-4

Table 4: **Recovered pathways for representative disease signatures.** Scores and p-values are derived from PathSymetic's ranked output using causal-symbolic alignment.

Conclusion and Future Work

This paper proposes PathSymetic, a neuro-symbolic system that combines symbolic grounding, causal alignment, and concept-based attribution to generate interpretable pathway hypotheses. Extensive evaluation shows that PathSymetic outperforms existing enrichment and learning-based approaches across a range of metrics, including AUPR, F1, and interpretability scores. Ablation studies confirm the critical role of each component, particularly the interaction between causal and symbolic reasoning. In future work, we aim to extend PathSymetic to multi-omics integration and adapt it for use in longitudinal patient cohorts. Incorporating biological priors from tissue-specific networks and integrating external knowledge graphs such as UMLS or DisGeNET may further enhance its interpretive power.

References

Garcez, A. d'Avila; Lamb, L.; and Gabbay, D. (2019). Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *arXiv preprint arXiv:1905.06088*.

Lamb, L. C.; Garcez, A. d'Avila; Gori, M.; Serafini, L.; Spranger, M.; and Tran, S. N. (2020). Graph neural networks meet neural-symbolic computing: A survey and perspective. *arXiv preprint arXiv:2003.00330*.

Hitzler, P.; Besold, T. R.; Garcez, A. d'Avila; and Lamb, L. C. (2022). Neuro-symbolic approaches in artificial intelligence. *National Science Review*, 9(6): nwac035.

Lu, Qiuhao, et al. (2025). Explainable diagnosis prediction through neuro-symbolic integration. *AMIA Summits on Translational Science Proceedings*, 332.

Smaili, F. Z.; Gao, X.; and Hoehndorf, R. (2019). OPA2Vec: Combining formal and informal content of biomedical ontologies to improve similarity-based prediction. *Bioinformatics*, 35(12): 2133–2140.

Jain, M.; Singh, K.; and Mutharaju, R. (2023). ReOnto: A neuro-symbolic approach for biomedical relation extraction. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 406–421. Springer, Cham.

Oltramari, A. (2023). Enabling high-level machine reasoning with cognitive neuro-symbolic systems. *Proceedings of the AAAI Symposium Series*, 2(1).

Shojaee, P.; Goh, J.; Lin, Y.; and Kolodny, J. (2024). LLM-SR: Scientific equation discovery via programming with large language models. *arXiv preprint arXiv:2404.18400*.

Wang, W.; Wang, B.; Shi, N.; Li, J.; Zhu, B.; Liu, X.; and Zhang, R. (2021). Counterfactual adversarial learning with representation interpolation. *Findings of EMNLP 2021*, pages 4809–4820.

Roschewitz, M.; Ribeiro, F. D. S.; Xia, T.; Khara, G.; and Glocker, B. (2024). Counterfactual contrastive learning via causal image synthesis. *arXiv preprint arXiv:2403.09605*.

El Bouchattaoui, M., et al. (2024). Causal contrastive learning for counterfactual regression over time. *Advances in Neural Information Processing Systems*, 37: 1333–1369.

Glaab, E.; Baudot, A.; Krasnogor, N.; Schneider, R.; and Valencia, A. (2012). EnrichNet: network-based gene set enrichment analysis. *Bioinformatics*, 28(18): i451–i458.

Sulaiman, M.; Hammad, M. A.; Saleh, H.; and Khan, M. A. (2023). PriPath: identifying dysregulated pathways from differential gene expression via grouping, scoring, and modeling. *BMC Bioinformatics*, 24(1): 87.

Subramanian, A.; Tamayo, P.; Mootha, V.; Mukherjee, S.; Ebert, B.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, 102(43): 15545–15550.

Wu, G.; Dawson, E.; Duong, A.; Haw, R.; and Stein, L. (2014). ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis. *Bioinformatics*, 30(12): 2016–2017.

Subramanian, A.; Narayan, R.; Corsello, S. M.; Peck, D. D.; Natoli, T. E.; Lu, X.; et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, 171(6): 1437–1452.e17.

Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; et al. (2000). Gene Ontology: Tool for the Unification of Biology. *Nature Genetics*, 25(1): 25–29.

Jassal, B.; Matthews, L.; Viteri, G.; Gong, C.; Lorente, P.; Fabregat, A.; et al. (2020). The Reactome Pathway Knowledgebase. *Nucleic Acids Research*, 48(D1): D498–D503.

Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; et al. (2005). Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *PNAS*, 102(43): 15545–15550.

Shlomi, T.; Cabili, M. N.; Herrgård, M. J.; Palsson, B. Ø. (2020). Inference of Gene Expression Metabolic Signatures Using Partial Knowledge. *Nature Biotechnology*, 38(4): 455–464.

Zhang, Y.; Xie, Z.; Li, Q.; Wang, F. (2023). CausalCLR: A Causal Contrastive Learning Framework for Biomedical Representation Learning. *arXiv preprint* arXiv:2302.05656.

Kulmanov, M.; Hoehndorf, R. (2020). DeepGOPlus: Improved Protein Function Prediction from Sequence. *Bioinformatics*, 36(2): 422–429.

Gao, Y.; Xu, K.; Chen, M.; Wang, Y.; Wang, Y. (2023). NeuroSymbolic Reinforcement Learning with Logic Guidance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9): 10494–10502.

Klopfenstein, D. V.; Zhang, L.; Pedersen, B. S.; Ramirez, F.; Vesztrocy, A. W.; Naldi, A.; et al. (2018). GOATOOLS: A Python Library for Gene Ontology Analyses. *Scientific Reports*, 8(1): 10872.

Liu, Anika, et al. 2019. From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. *NPJ Systems Biology and Applications*, 5(1): 40.

Drier, Yotam; Sheffer, Michal; and Domany, Eytan. 2013. Pathway-based personalized analysis of cancer. *Proceedings of the National Academy of Sciences*, 110(16): 6388–6393.

Hänzelmann, Sonja; Castelo, Robert; and Guinney, Justin. 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*, 14(1): 7.

Schubert, Michael, et al. 2018. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature Communications*, 9(1): 20.

Sauter, A.; Bian, Y.; Thams, F.; Li, Y.; Mandt, S.; and Song, L. (2025). ACTIVA: Amortized Causal Effect Estimation without Graphs via Transformer-based Variational Autoencoder. *arXiv preprint arXiv:2503.01290*.

Zhou, K.; Yang, R.; Zhang, Y.; Li, X.; Huang, Z.; and Ji, S. (2023). Adaptive Label Smoothing to Regularize Large-Scale Graph Training. *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, 72–80.