Graph Neural Networks with Multi-Scale Attention for Accelerated Drug–Target Interaction Prediction

Sarah Chen[1,*] Marcus J. Rivera[2] Priya K. Patel[1] David Okafor[1] Elena Voss[2]
[1]Nexa Pharmaceuticals, Computational Discovery Division [2]Synapse AI Research, Cambridge, MA
sarah.chen@nexapharma.com www.nexapharma.com/research

Abstract

Predicting drug–target interactions (DTIs) remains a central bottleneck in early-stage pharmaceutical discovery. We present **MultiGNN-DTI**, a graph neural network architecture that integrates multi-scale attention mechanisms across molecular and protein graphs to predict binding affinities with state-of-the-art accuracy. Trained on a curated dataset of 1.2M experimentally validated interactions spanning 4,800 human protein targets and 2.3M drug-like compounds, MultiGNN-DTI achieves a mean AUROC of 0.941 and a Pearson correlation of 0.873 on held-out binding assays, outperforming existing methods by 12–18%. We further demonstrate successful prospective validation against 12 kinase targets, identifying 3 novel inhibitors subsequently confirmed in biochemical assays.

Introduction & Motivation

The average cost to bring a new drug to market exceeds $2.6 billion, with target identification and lead optimization consuming 4–6 years of the pipeline. High-throughput screening remains expensive and low-yield; computational approaches promise to accelerate this phase dramatically.

Key challenges in learning-based DTI prediction include:

- Capturing hierarchical molecular representations beyond 1D SMILES strings or fixed fingerprints.
- Modeling protein binding pockets with sufficient structural granularity.
- Generalizing to unseen targets and chemotypes outside the training distribution.
- Providing interpretable predictions that guide medicinal chemistry decisions.

MultiGNN-DTI addresses each challenge through a modular, attention-driven architecture that jointly encodes ligand graphs, protein surface graphs, and pairwise interaction features. The model is trained end-to-end with contrastive and regression objectives, enabling both classification (bind / no-bind) and affinity prediction (pK_d, pIC_{50}, K_i) from a unified representation.

Data Curation & Preprocessing

We assembled **DrugTargetDB-v2**, a consolidated dataset merging BindingDB, ChEMBL v33, and proprietary Nexa internal assays, totaling 1,204,831 unique compound–target pairs after deduplication and quality filtering.

- **Ligand featurization:** Each compound is represented as a molecular graph $G_m = (V_m, E_m)$ with nodes encoding atom type, hybridization, formal charge, chirality, and aromaticity (54 features) and edges encoding bond type, conjugation, ring membership, and stereo (12 features).
- **Protein featurization:** Target structures from the PDB and AlphaFold DB are converted to residue-level contact graphs $G_p = (V_p, E_p)$. Node features include residue type (one-hot, 20), secondary structure (8-class DSSP), solvent accessibility, and evolutionary conservation scores (PSSM).
- **Negative sampling:** Hard negatives are generated using Tanimoto similarity thresholds ($T_c > 0.6$) to compounds known to bind different targets, preventing trivial discrimination.
- **Splits:** Time-based split (80/10/10) ensures no temporal leakage; an additional scaffold split evaluates generalization to novel chemotypes.

Model Architecture

MultiGNN-DTI consists of three coupled modules connected through a cross-attention fusion layer.

1. Ligand Encoder. A 6-layer Graph Isomorphism Network (GIN) with edge-feature conditioning processes each molecular graph. Jumping-knowledge concatenation aggregates representations across all layers, followed by a Set2Set readout producing a 256-dimensional ligand embedding $\mathbf{h}_m \in \mathbb{R}^{256}$.

2. Protein Encoder. A 4-layer Graph Attention Network (GATv2) with multi-head attention (8 heads, 64 dimensions each) operates over the residue contact graph. A structure-aware positional encoding based on 3D coordinates (when available) or AlphaFold-predicted distances augments node features before message passing. The protein graph is pooled via attention-weighted sum to yield $\mathbf{h}_p \in \mathbb{R}^{256}$.

3. Cross-Attention Fusion. Ligand and protein embeddings interact through a bilinear attention mechanism:

$$\alpha_{ij} = \frac{\exp(\mathbf{h}_{m,i}^\top \mathbf{W} \mathbf{h}_{p,j})}{\sum_{k} \exp(\mathbf{h}_{m,i}^\top \mathbf{W} \mathbf{h}_{p,k})}$$

The attended ligand features are concatenated with the protein embedding and passed through a 3-layer MLP with residual connections, LayerNorm, and GELU activation to produce the final prediction $\hat{y}$.

4. Multi-Task Training. The model jointly optimizes:

- Binary cross-entropy for binding classification.
- Huber loss for affinity regression.
- InfoNCE contrastive loss to align ligand–target pairs in latent space.

Training & Optimization

Hyperparameter	Value
Optimizer	AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Learning rate	5×10^{-4} with cosine annealing
Batch size	2,048 compound–target pairs
Epochs	200 (early stopping, patience = 15)
Weight decay	1×10^{-5}
Dropout	0.2 (all MLP layers)
Label smoothing	0.1 (classification head)
Gradient clipping	1.0 (global norm)
Mixed precision	bfloat16 (A100 GPUs)

Training on $8\times$ Lumina A100-80GB GPUs required approximately 18 hours for full convergence. We employed distributed data-parallel training with gradient accumulation (effective batch size = 8,192) and FlashAttention-2 for the multi-head attention layers in the protein encoder.

Ablation Analysis

To quantify the contribution of each architectural component, we performed systematic ablations on the full DrugTargetDB-v2 benchmark, measuring AUROC and RMSE on the held-out test set.

Model Variant	AUROC	RMSE
Full MultiGNN-DTI	**0.941**	**0.487**
− Cross-attention fusion	0.907	0.568
− Protein GAT (replace w/ CNN)	0.921	0.524
− Jumping knowledge (ligand)	0.928	0.513
− Contrastive loss	0.934	0.501
− PSSM features (protein)	0.929	0.508
− Set2Set readout (mean pool)	0.918	0.532

The cross-attention fusion and Set2Set readout provide the largest individual gains, confirming the importance of fine-grained ligand–protein interaction modeling and expressive graph pooling.

Benchmark Results

MultiGNN-DTI was evaluated against six competitive baselines on the DrugTargetDB-v2 test set (both time-split and scaffold-split). Results are averaged over 5 independent runs with different seeds.

Method	AUROC	Pearson r
DeepDTA (CNN + 1D sequences)	0.782	0.641
GraphDTA (GCN + CNN)	0.813	0.692
MolTrans (Transformer)	0.848	0.741
DrugBAN (Bilinear Attention)	0.876	0.776
ConPLex (Contrastive)	0.894	0.803
HyperPCM (Hypergraph)	0.903	0.822
MultiGNN-DTI (ours)	**0.941**	**0.873**

On the scaffold-split benchmark (testing generalization to unseen chemotypes), MultiGNN-DTI achieves AUROC = 0.887 (vs. second-best 0.812), demonstrating robustness to distribution shift.

Prospective Validation: Kinase Case Study

We applied MultiGNN-DTI to screen the Nexa compound library (4.8M compounds) against 12 clinically relevant kinase targets: EGFR, BRAF, CDK4/6, PI3Kα, BTK, JAK2, ALK, MEK1, AKT1, mTOR, VEGFR2, and FLT3.

Screening pipeline:

1. **Virtual screening:** MultiGNN-DTI scored all 4.8M compounds against each target (∼57.6M predictions in ∼3 hours on 8 GPUs).
2. **Top-k selection:** Top 500 compounds per target were filtered by drug-likeness (QED > 0.4, MW < 550 Da, logP < 5), PAINS alerts, and synthetic accessibility (SA score < 4).
3. **Docking:** Selected compounds were docked using AutoDock Vina; the top 50 consensus hits (agreement between GNN prediction and docking score) per target advanced to experimental validation.
4. **Biochemical assays:** ADP-Glo kinase assays at 10 μM and 1 μM, followed by dose-response curves (IC_{50}) for confirmed hits.

Results: From 600 tested compounds, we identified:

- 47 compounds with $IC_{50} < 10$ μM (hit rate = 7.8%).
- 12 compounds with $IC_{50} < 1$ μM (lead-like).
- 3 novel sub-micromolar inhibitors of CDK4/6, EGFR(T790M), and BTK, respectively, with selectivity indices $> 50\times$ over related kinases.

The CDK4/6 inhibitor (**NX-2731**) shows an IC_{50} of 84 nM against CDK4 and 127 nM against CDK6, representing a new chemotype distinct from palbociclib. Lead optimization is ongoing.

Conclusions & Future Work

Summary:

- MultiGNN-DTI establishes a new state-of-the-art for drug–target interaction prediction, leveraging graph neural networks with multi-scale cross-attention.
- The architecture performs robustly under both temporal and scaffold-based splits, demonstrating genuine generalization capability.
- Prospective validation against 12 kinase targets yielded 3 confirmed novel inhibitors, validating the pipeline for real-world drug discovery.

Ongoing and future directions:

- Extending to protein–protein interaction interfaces and PROTAC ternary complex prediction.
- Incorporating 3D structural dynamics via molecular dynamics simulation embeddings.
- Active learning loop integrating experimental feedback for iterative model refinement.
- Deployment within the Nexa discovery platform for ongoing screening campaigns against GPCR and ion channel target families.

References & Acknowledgments

1. Öztürk H. et al. (2018). DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17), i821–i829.
2. Nguyen T. et al. (2021). GraphDTA: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8), 1140–1147.
3. Huang K. et al. (2021). MolTrans: molecular interaction transformer for drug–target interaction prediction. *Bioinformatics*, 37(6), 830–836.
4. Bai P. et al. (2023). DrugBAN: bilinear attention networks for drug–target binding affinity prediction. *arXiv*:2301.12732.
5. Singh R. et al. (2023). ConPLex: contrastive learning for protein–ligand interaction prediction. *ICML 2023*.

This work was supported by the Nexa Pharmaceuticals Discovery Accelerator Program and the Meridian Foundation (grant MRF-2024-DD-1187). We thank the Apex High-Performance Computing Center for GPU resources.