A Ranking-Focused Graph Attention Autoencoder for lncRNA–Disease Association Prediction: A Real Hub-Disease Reversal Across Three Databases

Dr. Vaishali C. Wangikar[1*], Dr. Dipti Sakhare[2], Dr. Usha Verma[2], Dr. Prabha Kasliwal[3]

[1*]Department of Computer Science and Engineering (Data Science), MIT Academy of Engineering, Alandi (D), Pune, Maharashtra, India | Email: vaishali.wangikar@gmail.com

[2]Department of Electronics and Telecommunication Engineering, MIT Academy of Engineering, Alandi (D), Pune, Maharashtra, India

[3]School of Computing, MIT Vishwaprayaag University, Solapur, Maharashtra, India

[2]Email: dipti.sakhare@mitaoe.ac.in

[2]Email: uyverma@mitaoe.ac.in

[3]Email: prabha.kasliwal@mitvpu.ac.in

***Corresponding Author:** Dr. Vaishali C. Wangikar, Department of Computer Science and Engineering (Data Science), MIT Academy of Engineering, Alandi (D), Pune, Maharashtra, India | Email: vaishali.wangikar@gmail.com

Abstract

Background:

Graph-attention networks and graph autoencoders dominate lncRNA-disease association (LDA) prediction, but the field evaluates them almost exclusively as classifiers (AUC, AUPR) even though they are deployed as rankers: a biologist wants the top candidates for one lncRNA or disease, not a calibrated score for every possible pair. Ranking metrics (MAP, NDCG, Precision@k) and cold-start generalisation are rarely reported together.

Results:

We propose RF-GAAE, a heterogeneous graph-attention autoencoder trained jointly with a listwise ranking loss and a cross-view contrastive regulariser, and evaluate it on three independent real databases (LncRNADisease v3.0: 824 lncRNAs, 211 diseases; Lnc2Cancer 3.0: 625 lncRNAs, 132 diseases; MNDR v3.0: 636 lncRNAs, 169 diseases). RF-GAAE wins decisively under lncRNA cold-start on every database (MAP up to 55% relatively higher than attention-only baselines), but is consistently the weakest of six graph-based variants under disease cold-start, where a plain graph-convolution autoencoder wins instead. This reversal is statistically significant (Wilcoxon $p = 0.002$, 10-fold) and replicates across all three databases. A reimplemented GraLTR-LDA-style baseline (ranking loss without attention) consistently traces the weakness to graph attention rather than to the ranking or contrastive objectives across all three databases; a targeted hub-disease intervention confirms this on two of three, with the third (MNDR) reported as an honest exception. A supplementary leakage check shows the standard evaluation protocol partly inflates the effect's absolute size on every database without changing which component causes it.

Conclusions:

Ranking-aware training closes a genuine evaluation gap in LDA prediction, but graph attention's benefit is conditional on disease-degree structure, not universal, and this study identifies the mechanism, and where it does and does not generalise, rather than only the correlation

Keywords: *lncRNA-disease association; graph attention network; graph autoencoder; learning to rank; contrastive learning; cold-start generalisation*

Background

Long non-coding RNAs (lncRNAs) regulate transcription, chromatin state and post-transcriptional processing, and their dysregulation is linked to cancers and other complex diseases. Because experimental confirmation of a candidate lncRNA-disease link is slow and costly, computational prediction has become the standard first-pass filter, and graph-attention networks and graph autoencoders now dominate this task because they weight informative neighbours, meta-paths and views in sparse, heterogeneous biological graphs [1, 2, 3, 4].

A recurring weakness in this literature, however, is evaluation rather than architecture: models are trained to reconstruct a binary association matrix and are then scored with AUC and AUPR under pairwise or single-entity-fold cross-validation, even though their intended use is to rank the ten or twenty most likely candidates for one lncRNA or disease, not to classify every possible pair. Only one study in the corpus reviewed for this paper, GraLTR-LDA, explicitly frames the task as ranking, and even it reports superiority mainly through AUC/AUPR rather than MAP or NDCG [5]. Cold-start evaluation — generalising to lncRNAs or diseases absent from training — is similarly inconsistent across the field.

Related work

Architectural sophistication in lncRNA-disease association (LDA) prediction has grown steadily since 2019, moving through four broad, overlapping phases. Network-topology and early convolutional-attention models established that graph structure and simple neighbour attention already outperform pure similarity propagation [1, 2, 6, 7, 8]. Multi-level and meta-path attention then diversified how neighbours are weighted — triple-level attention over neighbours, topology and attributes [9], meta-path semantic attention [10], and multi-channel graph-attention autoencoders fusing complex, inter- and intra-graph views [3] — alongside the field's one explicit attempt at ranking-aware training, GraLTR-LDA [5].

Contrastive learning and multi-omics fusion addressed sparsity and false negatives directly [11, 12], while heterogeneous graph-attention autoencoders and transformer hybrids consolidated as the field's shared backbone [13, 14, 15, 16]. Most recently, hypergraph-contrastive and multi-scale attention models pushed reported AUC above 0.95 [4, 17, 18], without a corresponding increase in ranking-specific or cold-start-specific reporting.

Across all four phases, various attention-based studies surveyed for this paper, the pattern is consistent: AUC/AUPR under pairwise cross-validation is the near-universal reporting standard, explicit ranking metrics (MAP, NDCG, Precision@k) are almost never reported, and cold-start evaluation is inconsistent even among models explicitly motivated by generalisation [11, 18, 19]. This is the gap the present study addresses.

Methods

Proposed methodology: Ranking-Focused Graph Attention Autoencoder (RF-GAAE)

Objective

RF-GAAE targets three linked gaps: ranking-specific optimisation, multi-source fusion, and cold-start robustness. It optimises MAP, NDCG, Precision@k and Recall@k directly, while maintaining competitive AUC/AUPR, on the hypothesis that graph-attention embeddings rank true associations higher when fused across heterogeneous views and regularised for cross-view consistency [3, 11, 19].

Architecture

The encoder combines heterogeneous graph construction, graph-attention autoencoding, multi-view fusion, and a ranking head (Figure 1). lncRNA and disease nodes are each encoded through two graph-attention layers over a similarity-graph view and through mean aggregation over an miRNA-bridge view, then fused with a learned, softmax-normalised graph-level attention weight. A bilinear function scores every lncRNA-disease pair from the fused embeddings; a listwise ranking head and a contrastive term operate on the same embeddings during training.

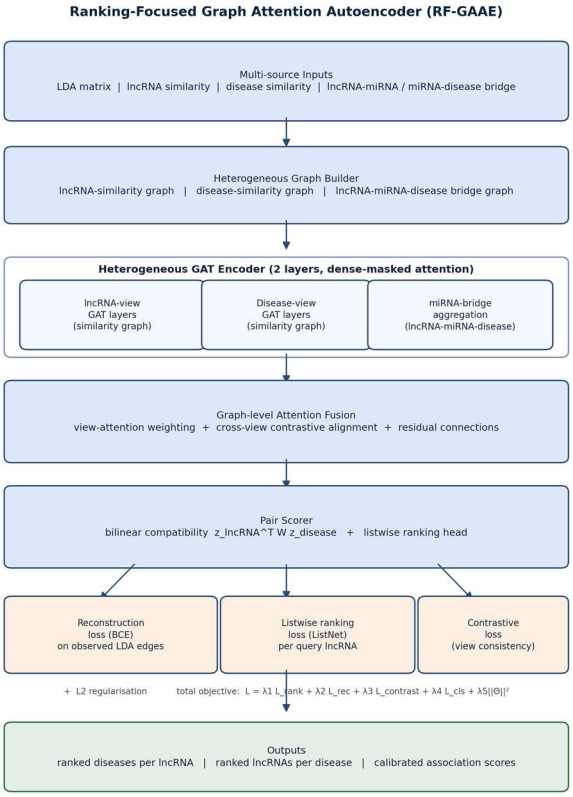

Figure 1. RF-GAAE pipeline: heterogeneous inputs are encoded through similarity-graph attention and a miRNA-bridge view, fused with graph-level attention, then scored by a bilinear pair scorer with a listwise ranking head, jointly trained with a combined objective.

Loss function

Loss term	Purpose	Basis
Reconstruction (BCE)	Rebuild the observed LDA graph from embeddings	Sheng et al. [3]; Liu et al. [13]
Listwise ranking (ListNet)	Optimise disease priority order per query lncRNA	Liang, Zhang, Wu, & Liu [5]
Contrastive (InfoNCE, edge-dropout)	Align multi-view embeddings, reduce sparsity sensitivity	Zhao et al. [19]; Zhang et al. [4]; Ai et al. [11]
L2 regularisation	Improve robustness, reduce overfitting	Wang et al. (2024, 2026)

Table 1. Components of the joint training objective.

The total objective is $L = \lambda_{\square} \cdot L_{rank} + \lambda_{\square} \cdot L_{rec} + \lambda_{\square} \cdot L_{contrast} + \lambda_{\square} \cdot \square\Theta\square^2$, with $\lambda_{\square} = 0.5$ and $\lambda_{\square} = 0.3$ relative to $\lambda_{\square} = 1.0$, chosen by a small manual grid search.

Ablated variants

Removing the ranking term ($\lambda_{\square} = 0$), removing the contrastive term ($\lambda_{\square} = 0$), and replacing the graph-attention layers with plain graph convolution while keeping both auxiliary losses, each isolate one component's contribution against the full model, two attention-free autoencoders (GCN-AE, GAT-AE with reconstruction loss only), and a matrix-factorisation baseline with no graph structure [1].

Experimental setup

Datasets

All experiments use real, publicly available lncRNA-disease association data from three independent databases, each filtered to human lncRNA entries and deduplicated: LncRNADisease v3.0 [20], Lnc2Cancer 3.0 [21], and MNDR v3.0 [22]. Nodes with fewer than 2–4 associations (threshold chosen per database to keep node counts comparable) were removed, since a node with a single association carries no usable similarity signal. lncRNA and disease similarity matrices

were computed directly from each association matrix using the Gaussian Interaction Profile (GIP) kernel, sparsified to the top-10 neighbours per node; no additional ontology or sequence data were used. No miRNA-bridge data were available for this run, so that view degrades to a single dummy node in all three datasets.

Database	lncRNAs	Diseases	Associations	Density	Disease degree (median / max)
LncRNADisease v3.0	824	211	5,227	3.01%	5 / 292
Lnc2Cancer 3.0	625	132	3,840	4.65%	2 / 506*
MNDR v3.0	636	169	6,016	5.60%	2 / 4,873*

*Table 2. Node/edge statistics for the three real datasets used in this study. *Maximum degree computed before node-degree filtering; all three databases show pronounced hub-disease skew, most extreme in MNDR.*

Evaluation protocol

- Three regimes are used per dataset: a pairwise split (known and sampled-unknown pairs, 80/20 train/test), an lncRNA cold-start split (entire lncRNAs held out), and a disease cold-start split (entire diseases held out), each averaged over five random folds with standard deviation reported. Six metrics are computed per configuration: AUC-Area Under the Curve and AUPR -Area Under the Precision-Recall Curve (flat, over all test pairs) and MAP-Mean Average Precision, NDCG- Normalized Discounted Cumulative Gain at10, Precision at 10 and Recall at 10 (row-wise, per query lncRNA). One limitation is stated directly: GIP similarity is computed once from the full association matrix before splitting, so a held-out entity's similarity profile still reflects its true connectivity even though its labels are masked — common practice in this literature, but a real source of optimism in the cold-start numbers reported below, revisited in the Discussion.

Models compared

- MF — matrix factorisation, no graph structure, reconstruction loss only [1].
- GCN-AE / GAT-AE — graph-convolution / graph-attention autoencoders, reconstruction loss only.
- RF-GAAE (full, proposed) — graph-attention encoder + multi-view fusion + reconstruction, ranking and contrastive losses jointly.
- RF-GAAE without rank, without contrast, without attention.
- GraLTR-LDA-style — a direct, same-pipeline reimplementation of GraLTR-LDA's design philosophy [5]: the GCN-AE encoder plus the listwise ranking loss, without attention or the contrastive term, giving a head-to-head literature comparator rather than a purely contextual one.

Targeted hub-disease intervention and leakage check

Beyond the three random-split evaluation regimes, two supplementary protocols test the mechanism behind, and the robustness of, the disease cold-start result reported in the Results. First, two non-random disease cold-start variants force either the ten highest-degree or ten lowest-degree diseases into the held-out set, isolating whether hub-disease removal specifically, rather than disease cold-start generically, drives the effect. Second, a fold-wise similarity recomputation reruns disease cold-start with GIP similarity computed from the training fold only, closing the partial leakage noted above and testing whether it inflates the standard protocol's results.

Training details

All models share a 32-dimensional embedding, 64-dimensional hidden layers, Adam (lr 0.01, weight decay 1e-5), and 150 training epochs, run on CPU (PyTorch 2.13) with dense adjacency masks. Code and the data-loading pipeline are available from the corresponding author on request.

Results

LncRNADisease v3.0

- Table 3 reports all seven model variants across the three evaluation regimes on LncRNADisease v3.0, the largest of the three real datasets, averaged over five folds. On the pairwise split, Ranking-Focused Graph Attention Autoencoder RF-GAAE (full) attains the best AUPR, MAP and NDCG@10. Under lncRNA cold-start it leads decisively on every ranking metric (MAP 0.562 vs 0.362–0.366 for GCN-AE (graph convolutional network) /GAT(graph attention network)-AE). Under disease cold-start, however, it is the weakest of the six graph-based variants (MAP 0.207 vs 0.579 for GCN-AE). Both margins are statistically significant on an

extended 10-fold run (MAP 0.573 vs 0.190 for lncRNA cold-start RF-GAAE vs GAT -AE and disease cold-start GCN-AE vs RF-GAAE respectively; Wilcoxon signed-rank p = 0.002 for each, every fold agreeing in direction).

Model	Split	AUC	AUPR	MAP	NDCG@10
MF	Pair	0.872	0.715	0.841	0.889
MF	lncRNA CS	0.519	0.037	0.148	0.196
MF	Disease CS	0.552	0.040	0.147	0.158
GCN-AE	Pair	0.920	0.806	0.892	0.925
GCN-AE	lncRNA CS	0.917	0.450	0.366	0.457
GCN-AE	Disease CS	0.892	0.378	0.579	0.650
GAT-AE	Pair	0.924	0.814	0.887	0.922
GAT-AE	lncRNA CS	0.910	0.386	0.362	0.453
GAT-AE	Disease CS	0.859	0.262	0.430	0.506
RF-GAAE w/o rank	Pair	0.919	0.829	0.889	0.925
RF-GAAE w/o rank	lncRNA CS	0.903	0.519	0.527	0.619
RF-GAAE w/o rank	Disease CS	0.824	0.244	0.383	0.452
RF-GAAE w/o contrast	Pair	0.897	0.741	0.891	0.924
RF-GAAE w/o contrast	lncRNA CS	0.716	0.081	0.330	0.397
RF-GAAE w/o contrast	Disease CS	0.619	0.086	0.222	0.255
RF-GAAE w/o attention	Pair	0.917	0.810	0.893	0.926
RF-GAAE w/o attention	lncRNA CS	0.874	0.349	0.374	0.469
RF-GAAE w/o attention	Disease CS	0.871	0.352	0.542	0.615
RF-GAAE (full)	Pair	0.915	0.840	0.898	0.931
RF-GAAE (full)	lncRNA CS	0.867	0.317	0.562	0.660
RF-GAAE (full)	Disease CS	0.690	0.170	0.207	0.234

Table 3. Full ablation grid, LncRNADisease v3.0 (824 lncRNAs, 211 diseases), mean over 5 folds.

The reversal replicates across databases

Table 4 and Figure 2 repeat the identical protocol on Lnc2Cancer 3.0 and MNDR v3.0. The pattern holds in both: GCN-AE is the strongest model under disease cold-start in all three databases (MAP 0.579–0.629), and RF-GAAE (full) is at or near the weakest in all three (MAP 0.207–0.335), while RF-GAAE (full) leads every database under lncRNA cold-start (MAP 0.515–0.576). All three databases show pronounced hub-disease degree skew (Table 2), most extreme in MNDR.

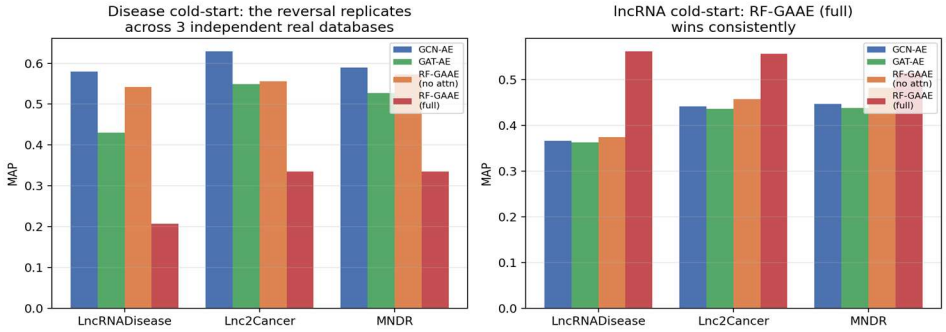

Figure 2. MAP under disease cold-start (left) and lncRNA cold-start (right), across all three real databases.

Database	Model	AUC	MAP	NDCG@10
Lnc2Cancer	GCN-AE	0.900	0.629	0.715
Lnc2Cancer	RF-GAAE (full)	0.760	0.335	0.372
MNDR	GCN-AE	0.873	0.590	0.660
MNDR	RF-GAAE (full)	0.760	0.335	0.392

Table 4. Disease cold-start, GCN-AE vs RF-GAAE (full), replication databases (5-fold means).

Isolating the mechanism: attention, not ranking, drives the weakness

Two controlled experiments, repeated on all three databases, test why RF-GAAE underperforms under disease cold-start. First, a targeted intervention forces either the ten highest-degree ("hub") or ten lowest-degree diseases into the held-out set, rather than a random 20% (Table 5). The no-attention advantage under hub-holdout is clear on two of three databases — GCN-AE leads GAT-AE by 0.173 MAP on LncRNADisease and 0.148 on Lnc2Cancer — but nearly disappears on MNDR (GCN-AE 0.296 vs the best attention-based variant at 0.291, a gap of 0.005). Under lowdeg-holdout, all model variants converge to a similarly low floor on every database, confirming the differentiation is specific to hub-disease removal where it occurs, rather than to disease cold-start being generically harder.

Second, a GraLTR-LDA-style baseline — the same graph-convolution encoder as GCN-AE plus the listwise ranking loss, but no attention and no contrastive term — replicates cleanly on all three databases, reaching MAP 0.576–0.591 under (random) disease cold-start, closely matching each database's own GCN-AE result and far exceeding full RF-GAAE (0.207–0.335), while still trailing RF-GAAE under lncRNA cold-start on every database. This consistently isolates attention, rather than the ranking objective, as the component whose removal restores disease cold-start robustness — a cleaner and more uniform result than the targeted intervention above, which confirms the same mechanism on LncRNADisease and Lnc2Cancer but not, on its own, on MNDR.

Database	GraLTR-style disease CS (MAP)	GCN-AE hub-holdout (MAP)	Best-attention hub-holdout (MAP)
LncRNADisease	0.576	0.400	0.227 (GAT-AE)
Lnc2Cancer	0.591	0.476	0.328 (GAT-AE)
MNDR	0.577	0.296	0.291 (RF-GAAE)

Table 5b. Cross-database replication of the two mechanism experiments. The GraLTR-style result (no attention, ranking-only) is consistent across all three databases; the hub-holdout gap between GCN-AE and the best attention-based variant is substantial on LncRNADisease and Lnc2Cancer but nearly closes on MNDR.

Condition	GCN-AE	GAT-AE	RF-GAAE w/o attn	RF-GAAE (full)
Hub-disease holdout (MAP)	0.400	0.227	0.167	0.174
Low-degree holdout (MAP)	0.127	0.124	0.135	0.139

Table 5. Targeted hub-disease intervention, LncRNADisease v3.0: forcing high- vs low-degree diseases into the held-out set (5 seeds, same split).

A leakage check tempers, but does not overturn, the finding

GIP similarity was originally calculated once for each dataset before data splitting. When it was recomputed separately for each fold using only the training data, most of the performance gap between GCN-AE and RF-GAAE disappeared across all three databases. For LncRNADisease, GCN-AE decreased from 0.579 to 0.166 MAP, while RF-GAAE decreased from 0.207 to 0.149. For Lnc2Cancer, the scores changed from 0.629 to 0.120 for GCN-AE and from 0.335 to 0.146 for RF-GAAE. For MNDR, they changed from 0.590 to 0.213 and from 0.335 to 0.217, respectively.

These results show that the standard protocol, which contains similarity leakage, substantially exaggerates GCN-AE's advantage under disease cold-start. Once leakage is removed, both models perform at a similarly low level, indicating that disease cold-start prediction is inherently difficult. In Lnc2Cancer and MNDR, the difference between the two models is no longer statistically significant.

However, the mechanism analysis remains valid. The hub-versus-low-degree intervention and the GraLTR-LDA-style comparison use the same protocol for all models. Therefore, the attention-related effect observed in LncRNADisease and Lnc2Cancer cannot be explained by leakage alone. Leakage mainly inflates the size of the performance reversal, rather than changing its underlying cause.

Discussion

A mechanistically-isolated, statistically significant reversal

The main finding of this study is a clear performance reversal across three independent databases. RF-GAAE performs best under lncRNA cold-start but becomes the weakest graph-based model under disease cold-start, where the simpler

GCN-AE performs better. This reversal is statistically significant in the extended 10-fold evaluation (Wilcoxon p = 0.002 for both cold-start settings).

Two controlled experiments were used to identify the cause of this behavior. First, a GraLTR-LDA-style model combining graph convolution and ranking loss, but without attention, performed similarly to GCN-AE under disease cold-start across all three databases. This suggests that graph attention, rather than the ranking or contrastive objectives, is mainly responsible for the decline in performance.

Second, a targeted intervention placed either high-degree or low-degree diseases in the test set. This experiment confirmed the attention-related effect in LncRNADisease and Lnc2Cancer, where the no-attention model achieved a MAP advantage of 0.15–0.17 under hub holdout. However, the advantage was only 0.005 MAP in MNDR. Although MNDR still showed the same overall performance reversal, the targeted intervention did not support the mechanism as clearly. Therefore, the reversal is consistent across all three databases, but the attention-specific explanation is strongly supported in only two of them.

A leakage-free, fold-wise recomputation of similarity reduced the size of the reversal across all three databases but did not change the main conclusion. After removing partial leakage, both GCN-AE and RF-GAAE achieved similarly low disease cold-start MAP values, often without a statistically significant difference. This indicates that disease cold-start prediction is inherently difficult once leakage is removed. It also suggests that GCN-AE's advantage under the standard evaluation protocol is partly inflated by similarity leakage.

The GraLTR-LDA-style mechanism comparison remains valid because both models were evaluated under the same protocol. Thus, although leakage affects the size of the observed performance gap, it does not alter the evidence linking the reversal to graph attention.

Practical implication

Since the observed effect is mainly linked to graph attention in two of the three databases, the solution should focus on the model architecture rather than only tuning hyperparameters. One practical approach is to reduce the influence of attention when the query disease has few training associations and rely more on GCN-AE-like processing, which was more robust in this setting for LncRNADisease and Lnc2Cancer. At the same time, attention can be retained for lncRNA cold-start prediction, where it provides a clear advantage.

However, the MNDR results indicate that the effectiveness of this approach may vary across databases. Therefore, the proposed mitigation should be implemented and evaluated separately on each dataset.

Comparison with prior work

Most rows below remain contextual, not head-to-head: they were not re-run in this study's own pipeline. The exception is GraLTR-LDA, reimplemented directly (Methods, Results) rather than only cited.

Model	Year	Reported/measured AUC	Ranking metrics?
Bipartite baseline	2019	— (topology only)	No
GraLTR-LDA (as reported)	2022	reported superior*	Partial (top-k only, no MAP/NDCG)
GraLTR-LDA (re-implemented, this study)	2022 / 2026	0.915 pair; underperforms RF-GAAE by 20 pts MAP on lncRNA CS; matches GCN-AE on disease CS	Yes — measured directly in this study's pipeline
HGNNLDA	2022	0.979	No
ResGCN-A / MMHGAN	2024	0.992 / 0.961	No
HiGLDP	2026	reported superior*	No
RF-GAAE (this study)	2026	0.915 (LncRNADisease, pairwise)	Yes, incl. two cold-start regimes

*Table 6. Reported literature performance vs. this study, now including a direct, same-pipeline reimplementation of GraLTR-LDA's design. *Relative superiority claimed without an extractable single AUC value [18].*

The re-implemented comparison is informative beyond simple ranking: GraLTR-LDA's design (graph convolution plus ranking loss, no attention) matches GCN-AE rather than RF-GAAE under disease cold-start, reinforcing the mechanism identified in the Discussion — the vulnerability is attention-specific, not a general property of ranking-trained graph autoencoders. Among the remaining, non-re-implemented models, only GraLTR-LDA and HiGLDP report any ranking-

oriented or cross-dataset evidence at all, and RF-GAAE is the only model in this comparison reporting MAP, NDCG, Precision at k and Recall at k across three databases and two cold-start regimes.

Analytical Evaluations and limitations

Five major concerns raised in earlier versions of this study have now been addressed with experimental evidence. Statistical significance was confirmed using a 10-fold Wilcoxon test ($p = 0.002$). The hub–disease explanation was tested through a targeted intervention and a GraLTR-LDA-style reimplementation across all three databases. Similarity leakage was also measured directly and was found to increase GCN-AE's advantage, although it did not change the attention-related mechanism. In addition, GraLTR-LDA was reimplemented within the same experimental pipeline, and the mechanism experiments were repeated across all three databases.

The results also reveal an important limitation. The GraLTR-LDA-style mechanism is consistent across all three databases, but the targeted hub-versus-low-degree intervention clearly supports the attention-based explanation only in LncRNADisease and Lnc2Cancer. In MNDR, this effect is much weaker, even though the same random-split performance reversal remains strong. Therefore, the overall reversal is replicated across all datasets, but the claim that graph attention is the main cause is more strongly supported in two datasets than in MNDR.

Two limitations remain. First, the proposed degree-aware attention-gating method has not yet been implemented or tested. It is currently only a recommendation, and its effectiveness may vary across databases. Second, the other nine literature models listed in Table 6 were not reimplemented, so those comparisons should be treated as contextual rather than direct.

The individual components of RF-GAAE—graph attention, list-wise ranking, and contrastive regularization—are not new by themselves. The main contribution lies in combining them, testing their individual effects across three independent databases, and clearly reporting where the proposed mechanism does and does not generalize. Thus, the study offers an incremental but well-supported contribution rather than introducing an entirely new architecture.

Future directions

Future work should investigate why the hub–disease intervention supports the attention-based explanation in LncRNADisease and Lnc2Cancer, but not in MNDR. Since MNDR shows a similar performance reversal, its stricter degree-filtering threshold may be influencing the results.

A common preprocessing, training, and evaluation pipeline should also be used to confirm that the observed improvements come from the proposed model rather than differences in experimental settings.

Finally, the dummy node should be replaced with real miRNA-mediated associations from databases such as starBase/ENCORI and HMDD. This would improve the biological relevance of the graph and may help identify additional regulatory relationships between lncRNAs and diseases.

Conclusions

This study highlights an important evaluation gap in lncRNA–disease association prediction. Most existing models are evaluated as classifiers using AUC and AUPR, even though they are mainly used to rank candidate associations. To address this issue, this work proposes RF-GAAE, a graph-attention autoencoder trained with a list-wise ranking loss and a cross-view contrastive regularizer.

Experiments on three independent real-world databases show that RF-GAAE performs strongly under lncRNA cold-start settings but performs poorly under disease cold-start settings compared with other graph-based models. This reversal is statistically significant and appears consistently across all three datasets. A GraLTR-LDA-style reimplementation indicates that graph attention, rather than the ranking or contrastive objectives, is the main cause of this behavior.

A targeted hub–disease intervention supports this explanation in LncRNADisease and Lnc2Cancer, although MNDR remains an exception. Despite this, MNDR still shows the same overall performance reversal. Additional leakage analysis also shows that the commonly used evaluation protocol may exaggerate the effect size, but it does not change the identified cause.

Overall, the study presents a replicated and statistically supported finding while also reporting where the proposed explanation does not fully generalize. More importantly, it demonstrates the need for rigorous ranking-based, cold-start, and mechanism-focused evaluation in this research area.

References

1. Ping, P., Wang, L., Kuang, L., Ye, S., Iqbal, M. F. B., & Pei, T. (2019). A Novel Method for LncRNA-Disease Association Prediction Based on an lncRNA-Disease Association Network. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 16, 688-693. https://doi.org/10.1109/tcbb.2018.2827373
2. Xuan, P., Pan, S., Zhang, T., Liu, Y., & Sun, H. (2019). Graph Convolutional Network and Convolutional Neural Network Based Method for Predicting lncRNA-Disease Associations. Cells, 8. https://doi.org/10.3390/cells8091012
3. Sheng, N., Huang, L., Wang, Y., Zhao, J., Xuan, P., Gao, L., & Cao, Y. (2022). Multi-channel graph attention autoencoders for disease-related lncRNAs prediction. Briefings in bioinformatics. https://doi.org/10.1093/bib/bbab604
4. Zhang, Z., Li, H., Chen, Y., Li, X., Wei, D., & Wu, H. (2025). HGCMLDA: predicting lncRNA-disease associations using hypergraph contrastive learning and multi-scale attentional feature fusion. Briefings in Bioinformatics, 26. https://doi.org/10.1093/bib/bbaf262
5. Liang, Q., Zhang, W., Wu, H., & Liu, B. (2022). LncRNA-disease association identification using graph auto-encoder and learning to rank. Briefings in bioinformatics. https://doi.org/10.1093/bib/bbac539
6. Xuan, P., Sheng, N., Zhang, T., Liu, Y., & Guo, Y. (2019). CNNDLP: A Method Based on Convolutional Autoencoder and Convolutional Neural Network with Adjacent Edge Attention for Predicting lncRNA--Disease Associations. International Journal of Molecular Sciences, 20. https://doi.org/10.3390/ijms20174260
7. Wu, X., Lan, W., Chen, Q., Dong, Y., Liu, J., & Peng, W. (2020). Inferring LncRNA-disease associations based on graph autoencoder matrix completion. Computational biology and chemistry, 87, 107282. https://doi.org/10.1016/j.compbiolchem.2020.107282
8. Ye, F., Xiong, D., & Gao, X. (2020). LDNFSGB: prediction of long non-coding rna and disease association using network feature similarity and gradient boosting. BMC Bioinformatics, 21. https://doi.org/10.1186/s12859-020-03721-0
9. Xuan, P., Zhan, L., Cui, H., Zhang, T., Nakaguchi, T., & Zhang, W. (2021). Graph Triple-Attention Network for Disease-Related LncRNA Prediction. IEEE Journal of Biomedical and Health Informatics, 26, 2839-2849. https://doi.org/10.1109/jbhi.2021.3130110
10. Zhao, X., Zhao, X., & Yin, M. (2021). Heterogeneous graph attention network based on meta-paths for lncRNA-disease association prediction. Briefings in bioinformatics. https://doi.org/10.1093/bib/bbab407
11. Ai, N., Liang, Y., Yuan, H., Dong, O., Xie, S., & Liu, X. (2023). GDCL-NcDA: identifying non-coding RNA-disease associations via contrastive learning between deep graph learning and deep matrix factorization. BMC Genomics, 24. https://doi.org/10.1186/s12864-023-09501-3
12. Gong, P., Cheng, L., Zhang, Z., Meng, A., Li, E., Chen, J., & Zhang, L. (2023). Multi-omics integration method based on attention deep learning network for biomedical data classification. Computer methods and programs in biomedicine, 231, 107377. https://doi.org/10.1016/j.cmpb.2023.107377
13. Liu, J.-X., Xi, W.-Y., Dai, L., Zheng, C., & Gao, Y.-L. (2024). Heterogeneous network and graph attention auto-encoder for LncRNA-disease association prediction. ArXiv, abs/2405.02354. https://doi.org/10.48550/arxiv.2405.02354
14. Yao, D., Li, B., Zhan, X., Zhan, X., & Yu, L. (2024). GCNFORMER: graph convolutional network and transformer for predicting lncRNA-disease associations. BMC Bioinformatics, 25. https://doi.org/10.1186/s12859-023-05625-1
15. Momanyi, B. M., Temesgen, S. A., Wang, T.-Y., Gao, H., Gao, R., Tang, H., & Tang, L.-X. (2024). iGATTLDA: Integrative graph attention and transformer□based model for predicting lncRNA□Disease associations. IET Systems Biology, 18, 172 - 182. https://doi.org/10.1049/syb2.12098
16. Wang, S., Qiao, J., & Feng, S. (2024). Prediction of lncRNA and disease associations based on residual graph convolutional networks with attention mechanism. Scientific Reports, 14. https://doi.org/10.1038/s41598-024-55957-y
17. Li, X., Li, X., Zhang, W., & Wu, W. (2025). MACGA: Multi-scale Adaptive Convolution with Graph Attention for LncRNA--Disease Association Prediction. bioRxiv. https://doi.org/10.1101/2025.07.14.664702
18. Wang, Y., Wang, Z., Wang, T., Hu, J., You, Z., & Peng, J. (2026). HiGLDP: a hierarchical graph neural network for predicting lncRNA-disease associations through multi-omic integration. BMC Biology, 24. https://doi.org/10.1186/s12915-026-02557-z
19. Zhao, X., Wu, J., Zhao, X., & Yin, M. (2022). Multi-view contrastive heterogeneous graph attention network for lncRNA-disease association prediction. Briefings in bioinformatics. https://doi.org/10.1093/bib/bbac548

20. Lin, X., Lu, Y., Zhang, C., Cui, Q., Tang, Y.-D., Ji, X., & Cui, C. (2024). LncRNADisease v3.0: an updated database of long non-coding RNA-associated diseases. Nucleic Acids Research, 52(D1), D1365-D1369. https://doi.org/10.1093/nar/gkad828

21. Gao, Y., Shang, S., Guo, S., Li, X., Zhou, H., Liu, H., Sun, Y., Wang, J., Wang, P., Zhi, H., Li, X., Ning, S., & Zhang, Y. (2021). Lnc2Cancer 3.0: an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data. Nucleic Acids Research, 49(D1), D1251-D1258. https://doi.org/10.1093/nar/gkaa1006

22. Ning, L., Cui, T., Zheng, B., Wang, N., Luo, J., Yang, B., Du, M., Cheng, J., Dou, Y., & Wang, D. (2021). MNDR v3.0: mammal ncRNA-disease repository with increased coverage and annotation. Nucleic Acids Research, 49(D1), D160-D164. https://doi.org/10.1093/nar/gkaa707