Journal of Innovations in Pharmaceutical and Biological Sciences (JIPBS)
ISSN: 2349-2759
Available online at https://jipbs.com/index.php/journal

Research article

DrugRepur AI: An Explainable Knowledge Graph Embedding Platform for Drug Repurposing with Multi-Level Biological Validation

Momin Mohammad Fuzail, Barrawaz Aateka Yahya*, Shaikh Shoaib Iftekhar, Sarfaraz Khan

Maulana Azad Educational Trust's Y. B. Chavan College of Pharmacy, Aurangabad, Maharashtra, India.

Received on: 08/05/2026, Revised on: 24/06/2026, Accepted on: 27/06/2026, Published on: 30/06/2026.

***Corresponding Author: Dr. Barrawaz Aateka Yahya**, Maulana Azad Educational Trust's Y. B. Chavan College of Pharmacy, Aurangabad, Maharashtra, India.
Phone No: 9923350939.
Email id: barrawazqa@gmail.com

Keywords: Drug repurposing, knowledge graph embeddings, RotatE, explainable AI, biological validation, pharmacovigilance, pathway analysis.

Vol. 13 (1): 67-75, Jan-June, 2026.

DOI: http://doi.org/10.56511/JIPBS.2026.13108

Abstract

Background: Most computational drug repurposing systems generate predictions without providing mechanistic justification, biological context, or clinical evidence. This disconnect between prediction and understanding limits practical adoption by pharmaceutical researchers. **Objective:** To design and validate an integrated platform combining knowledge graph embedding-based prediction with multi-level biological validation for drug repurposing hypothesis generation. **Methods:** A RotatE knowledge graph embedding model was trained on a DrugBank-derived knowledge graph containing 16,698 entities and 2.94 million relational triples using PyKEEN. The prediction engine was wrapped in a validation architecture comprising dual-tiered explainability (path-based and embedding-based reasoning), disease pathway analysis, drug target identification with druggability scoring, biomarker discovery via transcriptomic reversal analysis, chemical similarity search, ClinicalTrials.gov integration, literature mining, pharmacovigilance-based safety profiling, and novelty assessment. **Results:** Internal validation against RepoDB yielded an MRR of 0.422, Hits@10 of 65.4%, and AUC-ROC of 0.847. External validation against the independent PREDICT dataset produced Hits@10 of 40%, confirming cross-dataset generalization. Case studies on Metformin and Hydroxychloroquine demonstrated the system across known associations, novel embedding-based predictions, and path-supported hypotheses with convergent biological evidence from pathway, transcriptomic, and chemical similarity analyses. **Conclusion:** Knowledge graph embeddings, when integrated within structured biological validation, can produce drug repurposing hypotheses that are scientifically defensible and clinically contextualized. The multi-level evidence architecture transforms numerical predictions into testable scientific hypotheses suitable for guiding early-stage experimental investigation.

Introduction

Drug repurposing identifies new therapeutic applications for existing compounds and reduces both the cost and timeline of bringing treatments to patients. Traditional drug development requires twelve to fifteen years and costs between one and two billion dollars per approved compound, with failure rates exceeding ninety percent during clinical trials [1, 2]. Repurposed drugs bypass much of the preclinical pipeline because their safety profiles, pharmacokinetic properties, and manufacturing processes are already established, reducing time-to-clinic to as few as three years [3].

Computational approaches have expanded the scope of repurposing beyond serendipity. Structure-based methods use molecular fingerprints and similarity coefficients to identify compounds with shared chemical features. Network-based methods construct heterogeneous graphs

linking drugs, diseases, targets, and pathways, then apply guilt-by-association reasoning across multi-hop relational paths [4]. Machine learning classifiers train on known associations to score unobserved drug-disease pairs. Each approach captures one dimension of the problem, and each carries limitations: structural methods miss pharmacological complexity, network methods depend on explicit graph edges, and classifiers require hand-engineered features and cannot handle cold-start entities absent from training data [5].

Knowledge graph embedding (KGE) models address several of these limitations. By learning continuous vector representations of entities and relations through geometric optimization, KGE models encode the full relational structure of biomedical knowledge in a dense, low-dimensional space that supports prediction without feature engineering and generalizes to cold-start scenarios through compositional inference [6]. RotatE [7] models relations as rotations in complex vector space, naturally handling symmetric, antisymmetric, compositional, and inverse relational patterns found in biomedical knowledge graphs.

However, most published KGE-based repurposing systems share a critical weakness: predictions are delivered as ranked numerical scores without mechanistic reasoning, biological context, or clinical evidence. A confidence score of 0.89 for a drug-disease pair tells the researcher nothing about why the prediction was made, whether pathway biology supports the association, whether clinical trials already exist, or whether pharmacovigilance data raises safety concerns. This gap between prediction and understanding is the problem addressed in this work.

This paper presents DrugRepur AI, a platform that integrates RotatE-based drug repurposing prediction with a multi-level biological validation architecture spanning eight evidence layers: path-based and embedding-based explainability, disease pathway analysis, drug target identification, biomarker discovery through gene expression reversal, chemical similarity analysis, clinical trial integration, literature mining, and pharmacovigilance-based safety profiling. The system is demonstrated through case studies on Metformin and Hydroxychloroquine covering known associations, novel predictions, and the distinction between evidence-backed and speculative hypotheses.

The COVID-19 pandemic demonstrated both the potential and the limitations of drug repurposing at scale. Hydroxychloroquine, Remdesivir, Dexamethasone, and Baricitinib were all investigated as repurposed candidates with varying clinical success. The urgency of the pandemic response revealed that computational prediction alone was insufficient; researchers needed integrated biological reasoning, clinical trial awareness, and safety context to prioritize candidates efficiently. The experience underscored the need for systems that deliver not just ranked predictions but structured, multi-level scientific evidence supporting each hypothesis [8].

The present work addresses this need through an architecture that assembles independent evidence from graph-level, pathway-level, target-level, transcriptomic-level, chemical-level, clinical-level, and safety-level sources for every predicted drug-disease association. Each evidence layer operates on different data, applies different analytical methods, and captures different aspects of biological plausibility. The convergent weight of multiple independent evidence streams is what transforms a numerical prediction score into a scientific hypothesis worthy of experimental investigation.

2. Materials and Methods

2.1 Knowledge Graph Construction

The knowledge graph was constructed from the DrugBank database [9], accessed through an approved academic download account. Three entity types were extracted: drugs, diseases, and biological targets. Relational triples in (head, relation, tail) format encoded multiple relationship types including treats, targets, interacts with, and failed for. The final graph contained approximately 16,698 entities and 2,939,813 triples. RepoDB [10] provided internal validation data, and the PREDICT dataset [11] served as an independent external benchmark.

2.2 RotatE Model Training

RotatE represents each relation as an element-wise rotation in complex vector space. For a triple (h, r, t), the scoring function computes $d(h, r, t) = \|h \circ r - t\|$, where $\circ$ denotes element-wise complex multiplication and each element of r is constrained to unit modulus. The model was trained using PyKEEN [12] with 256-dimensional embeddings, 300 epochs, stochastic local closed-world assumption (sLCWA) training, and corrupted-triple negative sampling on CUDA-accelerated hardware. Training completed in approximately 136.6 seconds. (Table 1).

Table 1. Dataset and training configuration.

Parameter	Value
Entities	~16,698
Triples	~2,939,813
Embedding Dimension	256 (128 complex)
Epochs	300
Training Strategy	sLCWA
Negative Sampling	Corrupted triples (head/tail)
Framework	PyKEEN 1.0 (PyTorch/CUDA)

2.3 Evaluation Protocol

Ranking-based metrics were used for evaluation: Mean Reciprocal Rank (MRR), Hits@K (K = 1, 5, 10), Adjusted Mean Rank Index (AMRI), and AUC-ROC. For each test triple, all entities were ranked by predicted score, and the rank of the correct entity was recorded. Internal validation used a RepoDB held-out partition. External validation applied trained embeddings directly to PREDICT data

without retraining. Baseline comparison used Random Forest and Multi-Layer Perceptron classifiers trained on molecular and network features from the same dataset.

2.4 Explainability Framework

A dual-tiered explainability system distinguishes between two categories of predictions. Path-based reasoning uses breadth-first search (maximum three hops, up to five paths) through the knowledge graph to find explicit relational chains connecting a drug to a predicted disease through intermediate biological entities. When no path exists, embedding similarity reasoning identifies the prediction as based on geometric proximity in the 256-dimensional space, flagging the association as a novel hypothesis requiring external validation. This distinction is communicated transparently to the user for every prediction [13].

2.5 Multi-Level Biological Validation

Six validation modules provide independent biological evidence at different scales:

Pathway analysis maps drugs onto curated signaling pathways (AMPK, insulin, longevity) and identifies associated genes (PRKAA1, INSR, SIRT1, etc.) to establish mechanistic consistency between drug action and predicted indication.

Target analysis extracts putative protein targets from web intelligence, computes druggability scores, and visualizes drug-target-response interaction networks with mechanism of action represented as an explicit concept node [14].

Biomarker discovery evaluates whether a drug reverses disease-associated gene expression signatures through differential expression analysis, producing volcano plots and quantitative reversal scores.

Chemical similarity computes Morgan fingerprints with Tanimoto coefficients to identify structurally related compounds with known pharmacological profiles, supporting analogy-based reasoning.

Clinical evidence integrates ClinicalTrials.gov trial data, PubMed and arXiv literature, and patent databases to assess translational maturity and research novelty for each prediction.

Safety profiling combines FAERS pharmacovigilance signals with structural similarity analysis to flag adverse event risks and compute risk scores [15]. (Figure 1).

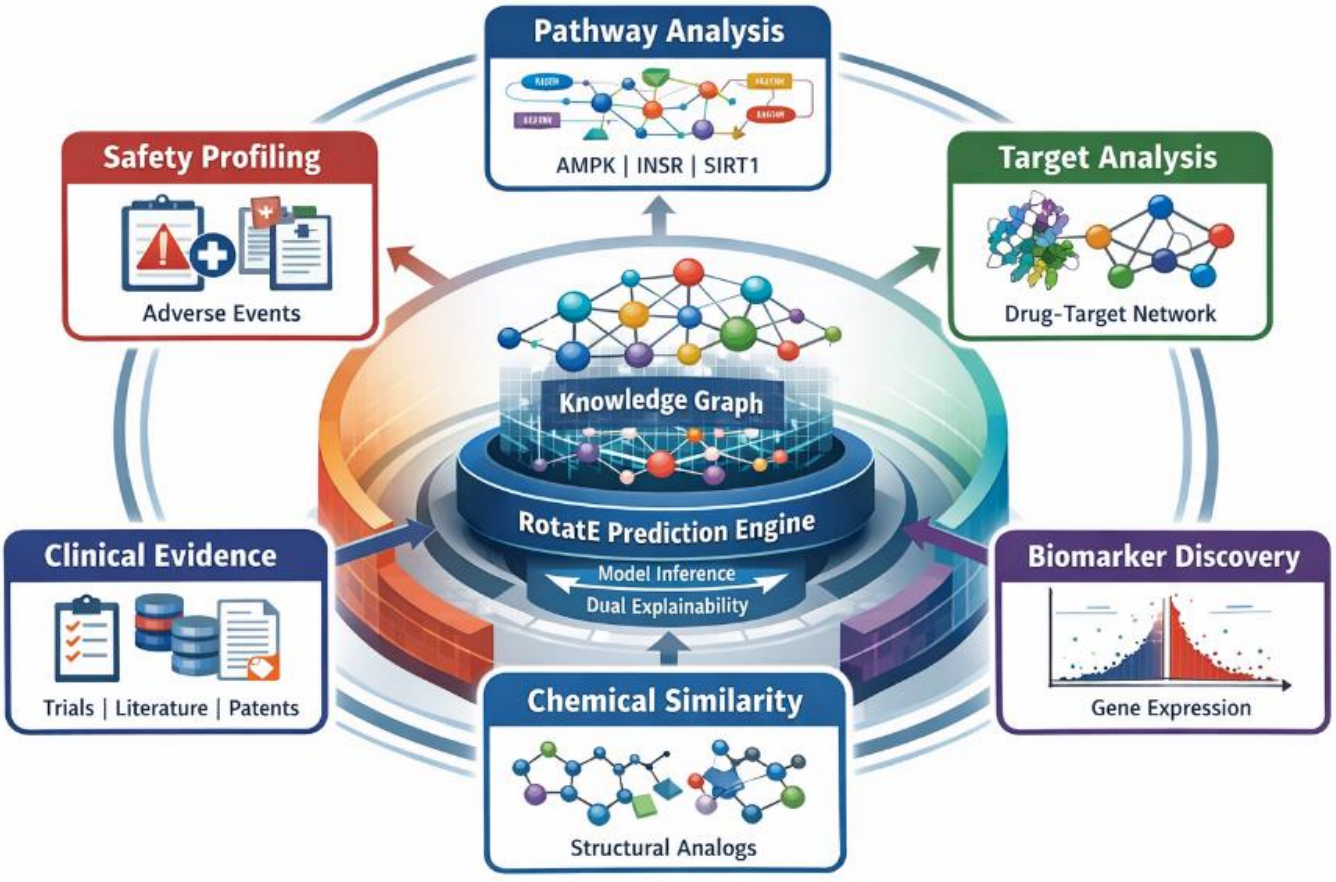


Figure 1. System architecture showing the six-layer validation framework surrounding the RotatE prediction engine. Data flows from knowledge graph construction through model inference, dual explainability, biological validation, clinical evidence integration, and safety profiling.

3. Results

3.1 Model Performance

The training loss curve showed stable convergence across 300 epochs, with rapid initial decrease followed by gradual stabilization at approximately 0.064. No oscillation or divergence was observed. Internal validation against RepoDB yielded an MRR of 0.422 (average correct entity rank ~2.4 out of 16,698), Hits@10 of 65.4%, Hits@1 of 30.5%, AMRI of 0.9153, and AUC-ROC of 0.847. External validation against PREDICT produced MRR of 0.206, Hits@10 of 40%, and Hits@5 of 35%. (Table 2).

Baseline comparison showed that Random Forest (AUC-ROC ~0.968) and neural network (~0.941) classifiers outperformed RotatE on AUC-ROC within feature-complete evaluation. However, both baselines require pre-computed molecular and network features and cannot score entities absent from the training feature matrix. RotatE achieved a cold-start AUC of approximately 0.71 on entities unseen during training, a capability structurally absent from the baseline classifiers. (Figure 2).

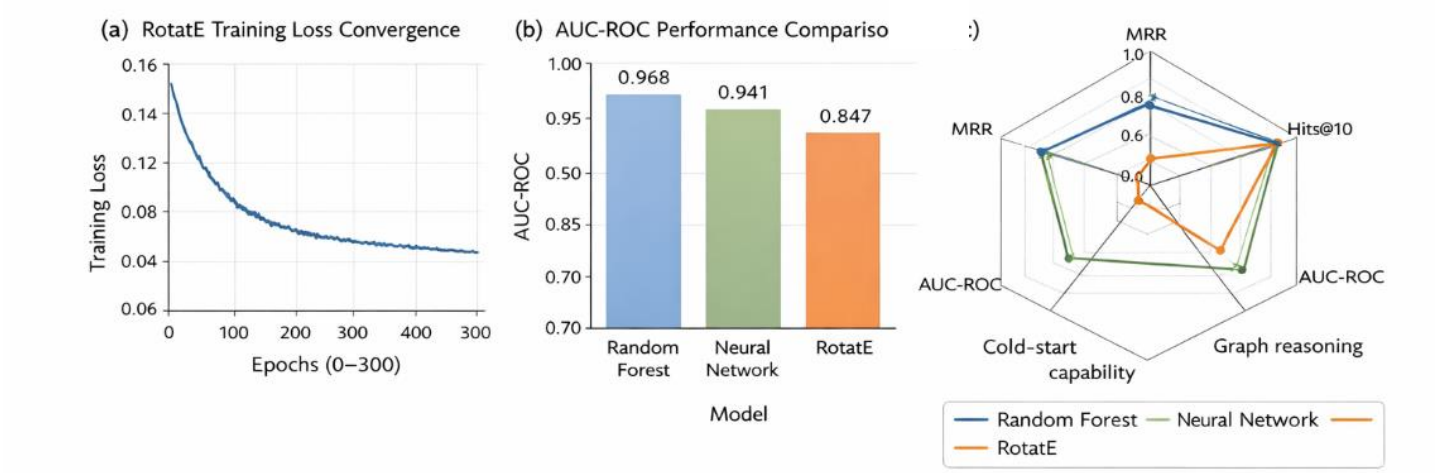
3.2 Case Study: Metformin

Metformin was selected because its established multi-pathway pharmacology provides a rich test case for biological validation. The model predicted Diabetes Mellitus as the top indication at 100% confidence, correctly recovering the primary known association.

Pathway analysis identified three affected signaling cascades consistent with established Metformin pharmacology: the AMPK signaling pathway (PRKAA1, PRKAA2, MTOR), reflecting the primary mechanism of AMPK activation; the insulin signaling pathway (INSR, IRS1, PI3KCA), reflecting insulin sensitivity improvement; and the longevity regulating pathway (SIRT1, FOXO3), connecting Metformin to the anti-aging research that has generated substantial recent interest. ClinicalTrials.gov integration returned active and completed trials across diabetes, cancer, obesity, and aging. The novelty score was 20/100, correctly classifying the association as well-established. (Table 3).

Table 2. Internal and external validation results.

Metric	Internal (RepoDB)	External (PREDICT)	Interpretation
MRR	0.422	0.206	Avg. rank ~2.4
Hits@1	30.5%	—	Top-1 accuracy
Hits@5	—	35%	Top-5 recall
Hits@10	65.4%	40%	Top-10 recall
AMRI	0.9153	—	91.5% above random
AUC-ROC	0.847	—	Discriminative power



(a) (b) (c)

Figure 2. (a) RotatE training loss curve showing three-phase convergence over 300 epochs. (b) AUC-ROC comparison: Random Forest (0.968), Neural Network (0.941), RotatE (0.847). (c) Radar chart comparing models across MRR, Hits@10, AUC-ROC, cold-start capability, and graph reasoning.

Table 3. Metformin pathway analysis results with gene-level resolution.

Pathway	Associated Genes	Pharmacological Relevance
AMPK Signaling	PRKAA1, PRKAA2, MTOR	Primary MoA via AMPK activation
Insulin Signaling	INSR, IRS1, PI3KCA	Peripheral insulin sensitivity
Longevity Regulation	SIRT1, FOXO3	Anti-aging research target

3.3 Case Study: Hydroxychloroquine

Hydroxychloroquine tested the system across therapeutic domains, generating 20 repurposing candidates. The results revealed a critical distinction between two types of high-confidence predictions that the dual explainability framework was specifically designed to communicate.

3.3.1 Arthritis: Novel embedding-based prediction

Arthritis ranked first at 100% confidence. No knowledge graph path connected Hydroxychloroquine to Arthritis. The prediction derived entirely from geometric proximity between the drug and disease embedding vectors in 256-dimensional space. The system flagged this as a novel, speculative hypothesis requiring external validation. Despite the highest possible numerical confidence, the absence of graph-level support means no documented biological relationship within the knowledge graph directly justifies the association.

3.3.2 Malaria: Path-supported mechanistic prediction

Malaria ranked second at 89.4% confidence. The system identified five supporting knowledge graph paths, including multi-hop chains through Chloroquine and shared autoimmune associations via Lupus. The prediction was classified as mechanistically supported and recommended for experimental validation priority. The supporting paths establish that known biological relationships connect the drug to the predicted disease through independently verifiable intermediate entities. (Figure 3, Table 4).

3.3.3 Convergent biological validation

Biomarker discovery analysis for Hydroxychloroquine against the Diabetes gene expression signature produced a drug reversal score of 0.90, with top reversed genes including PGC1A (mitochondrial biogenesis), IRS1 (insulin signaling), AMPK (energy sensing), and RPS6 (mTOR downstream effector). Chemical similarity analysis identified Chloroquine at 86% Tanimoto similarity, consistent with the shared antimalarial and anti-rheumatic profiles of both compounds. Safety profiling returned a risk score of 0.0/5.0 for the Arthritis indication, classified as low risk. The interactive knowledge graph visualization displayed all 20 predicted diseases with confidence-weighted node sizes using Force Atlas 2 force-directed layout. (Figure 3, Figure 4).

UMAP projection of the learned 256-dimensional embeddings onto two dimensions revealed drug clusters corresponding to therapeutic classes and disease clusters corresponding to pathological categories, providing qualitative evidence that the model learned structured biomedical representations rather than memorizing individual training triples.

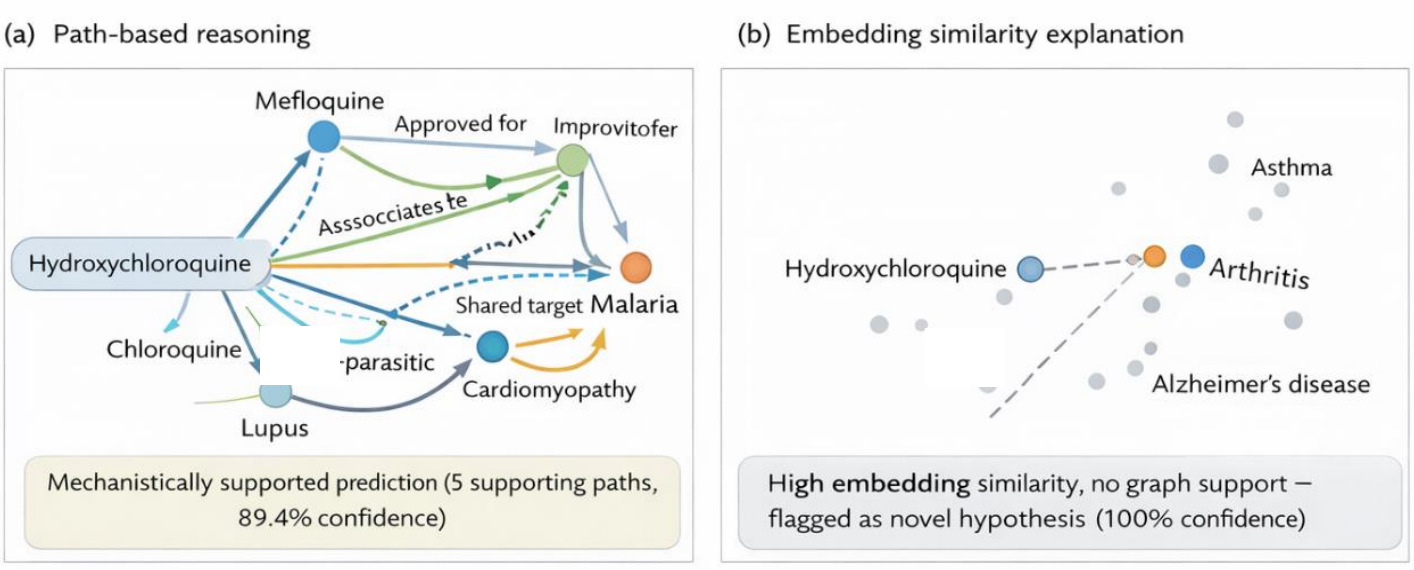


Figure 3. (a) Path-based reasoning for the Hydroxychloroquine-Malaria prediction showing five multi-hop knowledge graph paths through shared compounds and conditions. (b) Embedding similarity explanation for the Hydroxychloroquine-Arthritis prediction: high confidence, no direct graph path, flagged as novel hypothesis.

Table 4. Hydroxychloroquine top predictions with dual explainability analysis.

Rank	Disease	Confidence	Explainability Type	Evidence
1	Arthritis	100%	Embedding similarity	No graph paths (novel)
2	Malaria	89.4%	Path-based reasoning	5 supporting paths

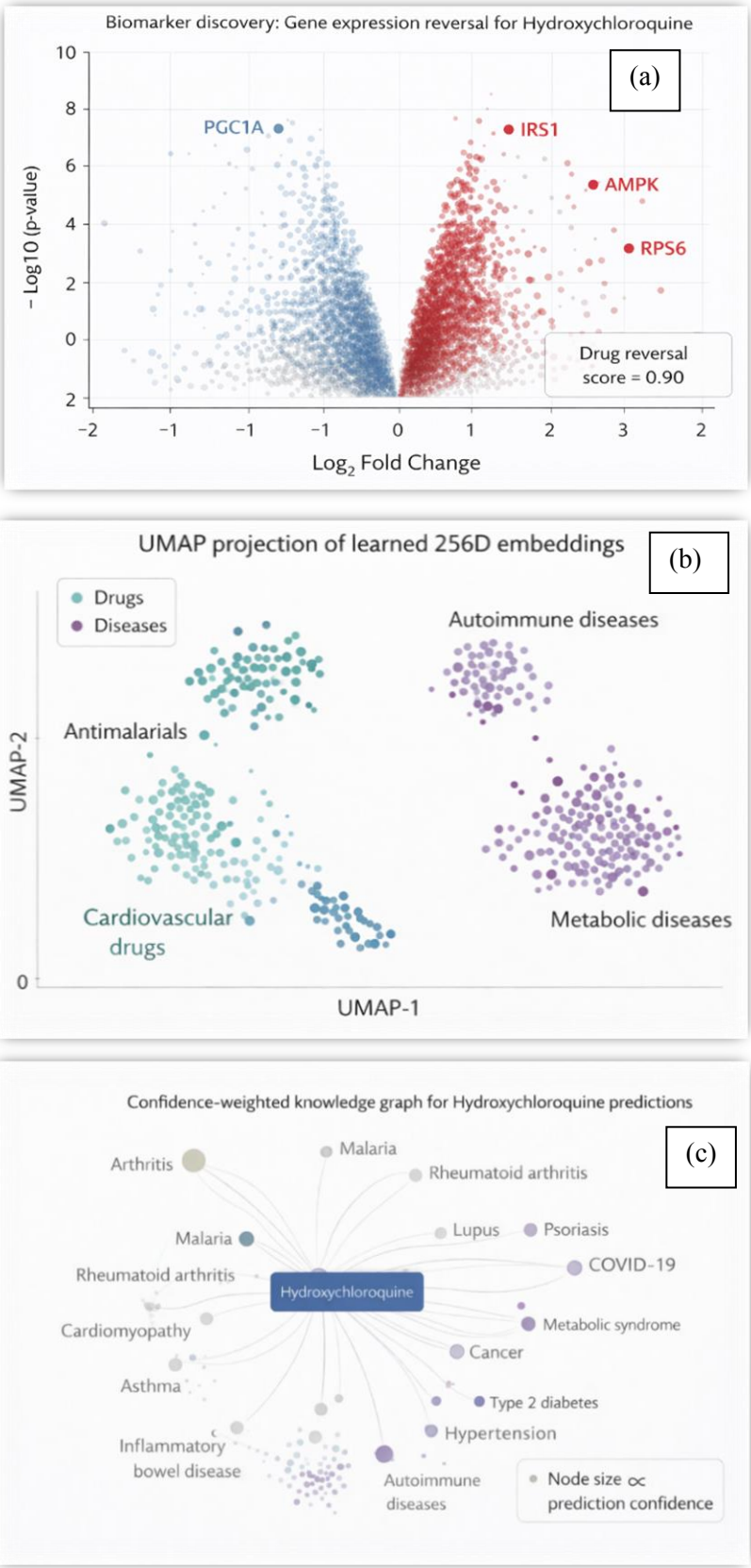


Figure 4. (a) Volcano plot from biomarker discovery showing gene expression reversal for Hydroxychloroquine against Diabetes (reversal score 0.90), (b) UMAP embedding space visualization: drug clusters (teal) and disease clusters (purple) in 2D projection of 256-dimensional learned representations, (c) Interactive knowledge graph for Hydroxychloroquine with confidence-weighted disease nodes.

4. Discussion

4.1 The Integration Architecture as Primary Contribution

The central contribution of this work is not the RotatE model, which applies an established algorithm, but the validation architecture that transforms numerical predictions into structured scientific hypotheses. A confidence score of 0.894 acquires meaning only in the context of five supporting graph paths, pathway analysis showing AMPK and insulin signaling involvement, clinical trials registered on ClinicalTrials.gov, published literature in peer-reviewed journals, 86% structural similarity to Chloroquine, and a low-risk pharmacovigilance profile. Each layer provides independent evidence at a different biological scale, and the convergent weight of these layers is what converts a number into a testable hypothesis.

Most published KGE-based repurposing systems treat prediction as the endpoint. The researcher who receives a ranked list of drug-disease scores must then perform pathway analysis, target identification, literature search, clinical trial review, and safety assessment manually. This manual process negates much of the computational efficiency that motivated the prediction in the first place. By automating multi-level evidence assembly, the platform delivers structured hypotheses rather than raw scores, reducing the gap between computational output and experimental decision-making.

4.2 Dual Explainability and Resource Allocation

The Hydroxychloroquine case study demonstrates why dual-tiered explainability matters for practical research decisions. Both the Arthritis prediction (100% confidence) and the Malaria prediction (89.4% confidence) received high numerical scores, but the biological evidence supporting each differs fundamentally. Arthritis, explained solely by embedding geometry, represents a genuinely novel hypothesis with no documented knowledge graph support. Malaria, supported by five relational paths, represents an evidence-backed inference grounded in known biology. A researcher allocating limited experimental resources would reasonably prioritize the path-supported prediction for initial investigation and reserve the embedding-only prediction for computational follow-up or literature exploration.

Without transparent communication of this distinction, both predictions appear equally supported. The system's explicit flagging of embedding-based predictions as speculative prevents overconfident interpretation of high numerical scores and enables calibrated experimental investment. This transparency is not a convenience feature but a scientific requirement for responsible use of AI-generated hypotheses.

4.3 Convergent Evidence Across Biological Scales

The pathway analysis results for Metformin illustrate how multi-level validation strengthens prediction credibility. The model's top prediction (Diabetes Mellitus) maps onto the insulin signaling pathway (INSR, IRS1, PI3KCA) and the AMPK signaling pathway (PRKAA1, PRKAA2, MTOR), which are precisely the pathways established through decades of experimental pharmacology. This alignment is not circular reasoning; the model was trained on relational triples, not pathway data, and the pathway analysis was performed independently. The convergence of graph-based prediction and pathway-level biology from different data sources provides stronger evidence than either source alone.

The biomarker discovery module operates at yet another biological scale. The 0.90 reversal score for Hydroxychloroquine against the Diabetes gene expression signature shows that the drug's molecular effects oppose disease-associated expression changes, independent of both the knowledge graph structure and the pathway analysis. When graph-level, pathway-level, and transcriptomic-level evidence converge on the same hypothesis from independent analytical perspectives, the collective evidential weight exceeds the sum of its parts.

4.4 Cold-Start Capability

Feature-based classifiers achieved higher AUC-ROC values (0.968 and 0.941 versus 0.847) but cannot score entities absent from the training feature matrix. This limitation is critically important because the most valuable repurposing candidates are often poorly characterized compounds where computational prediction has the greatest marginal utility. RotatE's compositional inference through the knowledge graph enables predictions for cold-start entities based on their relational neighborhood, achieving a cold-start AUC of approximately 0.71. For a repurposing system intended to discover genuinely novel associations, this capability is essential and structurally unavailable from feature-based approaches regardless of their in-distribution performance.

4.5 Limitations and Future Directions

Several limitations should be noted. The knowledge graph was constructed primarily from DrugBank, which is comprehensive but not exhaustive. Associations absent from DrugBank cannot influence model training, and biases in DrugBank curation, such as overrepresentation of well-studied drugs and underrepresentation of rare disease associations, propagate into the learned embeddings. The gene expression reversal analysis uses curated gene-disease associations rather than actual high-throughput transcriptomic datasets from resources such as the Connectivity Map or the Gene Expression Omnibus [16]. Replacing this approximation with experimental transcriptomic data would strengthen the biomarker validation layer substantially and provide more granular molecular evidence for or against each predicted association.

The drug combination synergy estimates are experimental computational approximations rather than validated clinical assessments and should be interpreted with appropriate caution. No systematic expert evaluation with clinical pharmacologists, medicinal chemists, or drug discovery professionals has been conducted to assess the clinical relevance of generated predictions at scale. Such evaluation would provide critical insight into which validation layers are most useful for experimental decision-making and which

predictions carry genuine translational potential versus computational artifacts.

The RotatE model learns exclusively from relational graph structure and does not incorporate molecular descriptors, three-dimensional protein structures, or raw genomic sequences into the embedding process. Multimodal architectures that fuse graph-level relational information with molecular-level chemical descriptors could potentially improve both prediction accuracy and cold-start performance by leveraging complementary information sources. Training on larger, multi-source knowledge graphs combining DrugBank with ChEMBL, PubChem, UniProt, KEGG, and Reactome would expand the relational space and potentially improve cross-domain generalization.

Future work should also explore active learning, where human feedback collected through the annotation interface directly updates model parameters in subsequent training cycles, closing the loop between expert domain knowledge and computational inference. Formal usability studies with end-user pharmaceutical researchers would identify gaps between computational output format and the information needs of experimental scientists, informing interface improvements that increase practical adoption. Systematic benchmarking against other published repurposing platforms on shared evaluation datasets would situate the system within the broader computational repurposing landscape and identify specific strengths and weaknesses relative to alternative approaches.

5. Conclusion

DrugRepur AI demonstrates that knowledge graph embeddings, when embedded within a structured multi-level validation architecture, can produce drug repurposing hypotheses that carry biological meaning, clinical context, and transparent evidential grounding. The RotatE model achieved competitive link prediction performance (MRR 0.422, Hits@10 65.4%) and generalized to an independent external dataset (Hits@10 40%). The validation architecture transforms these numerical outputs into structured hypotheses by assembling pathway-level, target-level, transcriptomic, chemical, clinical, and safety evidence independently for each prediction.

The dual explainability framework communicates the critical distinction between evidence-backed predictions supported by explicit knowledge graph paths and novel embedding-based hypotheses inferred from latent geometric proximity. This transparency enables calibrated experimental investment and prevents overconfident interpretation of high numerical scores that lack biological grounding. The Hydroxychloroquine case study illustrated this distinction concretely: Arthritis (100% confidence, no graph paths) and Malaria (89.4% confidence, five supporting paths) represent fundamentally different categories of prediction requiring different follow-up strategies.

The platform does not replace experimental validation or clinical judgment. What the platform produces are hypotheses. The quality of those hypotheses, measured by the breadth and depth of automatically assembled supporting evidence spanning graph structure, signaling pathways, gene expression, chemical similarity, clinical trial data, published literature, and pharmacovigilance signals, represents the contribution of this work to the computational drug repurposing field. The system demonstrates that meaningful integration of AI prediction with structured biological reasoning is achievable using open-source tools, publicly available databases, and standard computational resources.

Conflict of Interest

The author declares no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability

The DrugBank database is available through academic license application at https://go.drugbank.com. RepoDB is publicly available at https://portal.dbmi.hms.harvard.edu/projects/repodb/. The PREDICT dataset is available as described in Gottlieb et al. [4]. The PyKEEN framework is open-source at https://github.com/pykeen/pykeen. Platform source code and trained model artifacts are available from the corresponding author upon reasonable request.

Acknowledgements

The author acknowledges the DrugBank team at the University of Alberta for approving the academic data access request and the PyKEEN development team at the Ludwig Maximilian University of Munich for maintaining the open-source embedding framework that enabled this work.

Ethics Approval and Consent to Participate

Not applicable.

Author Contribution

All authors are contributed equally in the research.

Declaration of Generative AI

This manuscript has utilized OpenAI's ChatGPT (version 4) to enhance the clarity and coherence of the language. The tool was employed exclusively for language improvement purposes with ethical and academic standards. The authors take full responsibility for the content and integrity of the manuscript.

Funding

None.

References

1. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Econ. 2016;47:20-33.
2. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014;32(1):40-51.
3. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41-58.
4. Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, Hadley D, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. eLife. 2017;6:e26726.
5. Zhang Y, Li X, Guo Y, Shi Y, Liu J, Wang Z. Machine learning and deep learning approaches for drug repurposing. Curr Med Chem. 2021;28(14):2700-18.
6. Wang Q, Mao Z, Wang B, Guo L. Knowledge graph embedding: a survey of approaches and applications. IEEE Trans Knowl Data Eng. 2017;29(12):2724-43.
7. Sun Z, Deng ZH, Nie JY, Tang J. RotatE: Knowledge graph embedding by relational rotation in complex space. In: Proceedings of the International Conference on Learning Representations (ICLR); 2019.
8. Zeng X, Song X, Ma T, Pan X, Zhou Y, Hou Y, et al. Repurpose open data to discover therapeutics for COVID-19 using deep learning. J Proteome Res. 2020;19(11):4624-36.
9. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database. Nucleic Acids Res. 2018;46(D1):D1074-D1082.
10. Brown AS, Patel CJ. A standard database for drug repositioning. Sci Data. 2017;4(1):170029.
11. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications. Mol Syst Biol. 2011;7(1):496.
12. Ali M, Berrendorf M, Hoyt CT, Vermue L, Galkin M, Sharifzadeh S, et al. PyKEEN 1.0: a Python library for training and evaluating knowledge graph embeddings. J Mach Learn Res. 2021;22(82):1-6.
13. Jimenez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-84.
14. Mohamed SK, Novacek V, Nounu A. Discovering protein drug targets using knowledge graph embeddings. Bioinformatics. 2020;36(2):603-10.
15. Sakaeda T, Tamon A, Kadoyama K, Okuno Y. Data mining of the public version of the FDA Adverse Event Reporting System. Int J Med Sci. 2013;10(7):796-803.
16. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929-35.