DRUG-TARGET-DISEASE EMBEDDINGS IMPROVE THE ACCURACY OF DRUG REPURPOSING MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Drug repurposing offers a promising strategy to accelerate therapeutic development by identifying new indications for approved drugs. Recent work in this area has made extensive use of biomedical knowledge graphs (KGs). However, most computational models built on KGs predict drug-disease associations without explicitly incorporating information about molecular targets of drug molecules. We present a target-aware drug repurposing framework that leverages drug-target-disease (DTD') triplet embeddings to explicitly capture target information. We generated 512-dimensional Node2Vec embeddings for DTD' combinations labeled as either "positives" (fully connected DTD' triplets in ROBOKOP KG) or "negatives" (random DTD' combinations with no known DD', i.e., "treats", connections) and built multiple binary classification machine learning models. Multilayer perceptron (MLP) models using DTD' triplet embeddings outperformed matched drug-disease (DD') doublet models across all classification and ranking metrics such as precision (0.61 vs 0.41, respectively), recall (0.76 vs 0.35), F1-score (0.68 vs 0.38) and enrichment factor (EF1%: 27.54 vs 15.56). These models also outperformed both XGBoost and Random Forest baselines, and Node2Vec embeddings yielded superior performance compared to embeddings from pretrained large language models such as PubMedBERT, and BioBERT. Our results confirm that explicit target information does improve the predictive accuracy of drug repurposing models built with KGs. We employed our models to generate novel repurposing hypotheses for approximately 22,000 human diseases annotated in the Mondo Disease Ontology. We present case studies of top-ranked drug repurposing candidates nominated by our approach for two neurodegenerative diseases (ALS and Huntington's) supported by evidence in PubMed and ClinicalTrials.gov.

1 INTRODUCTION

Modern drug discovery is characterized by high failure rates and substantial costs. Drug repurposing can accelerate search for new medicines by finding new therapeutic uses for approved drugs through integration of evidence from multiple sources, such as clinical trial records, translational studies, off-label use reports, and electronic health records (Pushpakom et al., 2019; Hua et al., 2022; Avram et al., 2021).

In recent years, biomedical knowledge graphs (KGs) have emerged as an effective framework for integrating and analyzing heterogeneous biomedical data including drugs, diseases, molecular targets, biological pathways, phenotypes, and related entities. These graphs represent information in a structured format where nodes denote entities and edges capture functional relationships, interlinked through formal ontologies and semantics (Vasilevsky et al., 2025). Edges encode relationship types such as 'drug treats disease', 'gene associated with disease', etc. (Unni et al., 2022; Glen et al., 2023). Some popular KGs include ROBOKOP (Bizon et al., 2019; Fecho et al., 2026), Hetionet (Himmelstein et al.), RTX-KG2 (Wood et al., 2022), PrimeKG (Chandak et al., 2023),SPOKE (Morris et al., 2023), BioKG (Su et al., 2023), KG-COVID-19 (Reese et al., 2020), PharmKG (Zheng et al., 2021),iBKH (Su et al., 2023), and others. Diverse KG-based inference methodologies have been reported for generating novel drug repurposing hypotheses; examples include TransE (Bordes

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

et al., 2013), TransH (Wang et al., 2014), RotatE (Sun et al., 2019), DistMult (Yang et al., 2015), and ComplEx (Trouillon et al., 2016).

Pre-trained large language models (LLMs) have also emerged as valuable tools for generating node representations suitable for downstream classification and model training tasks. These include domain-specific models like BioBERT (Lee et al., 2020),PubMedBERT(Gu et al., 2021), and BioMamba(Yue et al., 2024) alongside general-purpose embeddings such as Word2Vec(Mikolov et al., 2013) and OpenAI's text embedding models. [1]

Although these models have demonstrated promising results for drug repurposing predictions on KGs, most link prediction approaches focus on modeling direct drug-disease associations without explicitly encoding mechanism-of-action information. Herein, we propose a drug repurposing framework that explicitly incorporates molecular target information, aiming to enhance both predictive accuracy and mechanistic interpretability of the drug repurposing nominations.

Our framework leverages drug-target-disease triplets within ROBOKOP KG. We employed Node2Vec (N2V) embeddings to encode drugs, diseases, and their associated targets and trained a multilayer perceptron (MLP) for binary classification of drug-target-disease triplets (DTD') defined as "positive" or "negative" as described in Methods. This MLP model was internally validated and employed for virtual screening of all theoretically possible DTD' triplets not including those in the training set to identify novel putative repurposing nominations (DD' doublets) within DTD' triplets predicted as "positive". Finally, we employed LLMs to generate human-readable mechanistic rationales by synthesizing evidence from biomedical literature and clinical trial data, thereby bridging computational predictions with clinically actionable insights.

2 Material and Methods

2.1 Problem Definition

We formulate drug repurposing as a binary classification task over DTD' triplets. The objective is to predict whether drug D exerts a therapeutic effect on disease D' through modulation of target T. Formally, we learn a mapping function $f(D, T, D') \rightarrow \{0, 1\}$, where 1 indicates a supported therapeutic relationship and 0 indicates its absence. Each entity is represented by a fixed-length embedding vector derived from the topology of a KG, and the concatenation of these embeddings serves as input to a supervised classifier.

2.2 Ground Truth Extraction

A Cypher (Francis et al., 2018) query was run on ROBOKOP KG to extract "ground truth" Drug-Target-Disease (DTD') therapeutic triplets. The node categories used for the queries were biolink:Drug to biolink:DiseaseOrPhenotypicFeature to capture the drug-disease relationship (DD'), biolink:Drug and biolink:Gene to define the drug-target relationship (DT) and biolink:DiseaseOrPhenotypicFeature and biolink:Gene to establish the disease-target relationship (D'T), thereby making complete triplets. For each edge, multiple biolink predicates implying functional connection between respective entities were permitted (see full list in the Appendix A.1). To ensure biological reliability, the query restricts target entities to human taxa (NCBI Taxonomy ID: 9606). To ensure the triplets include only drugs with a known safety profile, we cross-referenced the triplets against the MeDIC database. MeDIC is an open curated database that provides highly trustworthy drug-disease indications and contraindications (DeLuca et al., 2025).

2.3 Training Set Generation

Triplets with a *contraindicated_in* edge between drug and disease were labeled as negative instances ("negatives"). To achieve a positive-to-negative class ratio of 1:10, additional negative samples were generated by replacing two out of three components (drug, disease, target) in each positive triplet with an identifier of the same type drawn randomly from the pool of positive triplets (i.e., drug substituted with a different drug, a target with a different target, and a disease with a different disease) (Fig. 1.2) . To prevent data leakage, we removed all edges between drug-disease pairs from

[1] `https://platform.openai.com/docs/guides/embeddings`

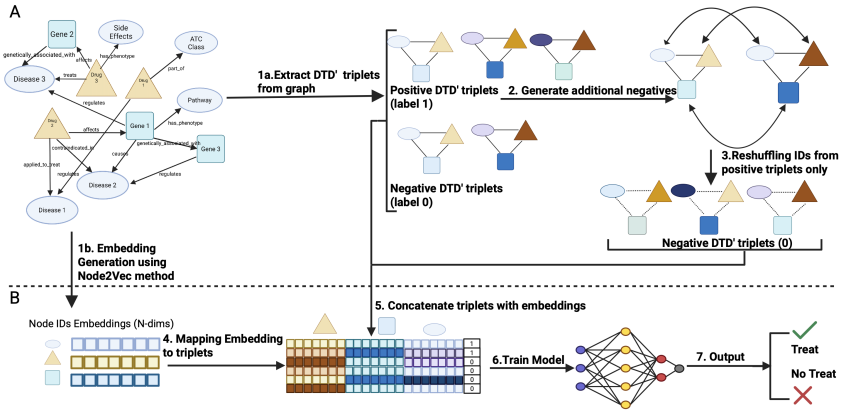


Figure 1: Overview of the KG-based workflow for predicting drug-disease relationships. The workflow consists of two stages: (A) construction and labeling of DTD' triplets from knowledge graph, and (B) embedding-based training of a classifier to predict treatment outcomes. (A) DTD' triplets are extracted from the graph using Cypher queries and labeled as positive (known association, label=1) or negative (label=0). Additional negatives are generated by reshuffling 2 of the 3 IDs in a positive triplet, with replacement IDs drawn only from the pool of positive triplets (2,3). Node embeddings are learned from all graph nodes using N2V (1b.). In second stage (B), each node in DTD' triplet is mapped to its corresponding N2V embedding (4) and the drug, target, and disease embeddings are concatenated into a single feature vector with its associated label (5). The resulting labeled matrix is used to train a neural network classifier (6), which outputs a binary prediction of "treat" or "no-treat" for a given drug-disease pairs.

the graph prior to embedding (Fig. 1b). Duplicate negative triplets were removed. Because the same train/test set was used for both training and evaluation of the DD' model, we also excluded any synthetic triplets that shared the same drug-disease pairs with a positive-label instance, as retaining these would introduce label ambiguity during DD' model training. Data splitting was then performed at the drug-disease pair level, with all associated targets carried along, such that 80% of pairs were allocated to the training set and 20% to the test set. This approach ensures that no drug-disease pair appears in both the training and test sets, while the required ratio is maintained at the triplet level for the DTD' model.

We acknowledge that the drug-disease pairs captured through target-associated triplets may not fully represent the broader landscapes of known drug-disease indications; therefore, for DD' model training, we augmented the training set using a MeDIC-derived indication list. Because the DTD'-based training set consist of a drug-disease with multiple targets, we deduplicated prior to DD' model training. For test set evaluation, we used the same test set for both the DTD' and DD' models. For drug-disease pairs associated with more than one target, we selected maximum predicted scores as score of the association to collapse the DTD' predictions to the DD' level.

2.4 Deep Neural Network Classifier For Indication Hypothesis Generation.

A multi-layer perceptron (MLP) architecture was implemented for the binary classification task of drug-target-disease associations. Each embedding was first passed through two dense sequential layers to learn representations that capture task-relevant features for each drug, target, and disease entity. The DTD' architecture utilizes three distinct feature heads, generating separate learned embeddings for each input type, while the DD' variant only processes drug and disease inputs using two feature heads. The learned embeddings were concatenated and passed through a three-layer fully connected network, with each layer applying L2 regularization to penalize large model weight values, batch normalization to stabilize training, and ReLU (Nair & Hinton, 2010) activation to introduce non-linearity for learning complex patterns. Random dropout regularization was applied between hidden layers to reduce overfitting and enhance generalization of model performance. The network was optimized using Adam optimizer, and a sigmoid activation function was used in the final layer to output binary probabilities.

We adopted focal loss as a loss function for our model as it provides improvement over standard binary cross-entropy (BCE) in class imbalance cases (Lin et al., 2018). Focal loss modifies BCE by introducing a modulating factor that down-weights well-classified examples and emphasizes model to focus on harder-to-classify instances. Bayesian optimization (BO) was used for hyperparameter optimization over batch size, number of training epochs, dropout rate, L2 regularization strength, focal loss parameters (α and and γ), and the latent dimensionality used in the feature-extraction layers for the drug, target, and disease embeddings. Additionally, various baseline models such as XGBoost (Chen & Guestrin, 2016), Random Forest (Breiman, 2001), CatBoost (Prokhorenkova et al., 2019), and Long Short-Term Memory model (Hochreiter & Schmidhuber, 1997) were trained for comparison of metrics. The hyperparameters for these baseline models were optimized using RandomizedSearchCV with appropriate parameter distributions.

2.5 Model Training and Hyperparameter Optimization

We implemented a cross-validation framework with Bayesian optimization (BO) for hyperparameter tuning (Fig. 2a). The initial dataset was split into training (80%) and test (20%) sets. BO explored the hyperparameter space over 33 iterations (3 initialization points + 30 optimization iterations), tuning the following parameters: learning rate (1×10^{-5} to 1×10^{-3}), dropout rate (0.3 to 0.8), L_2 regularization (1×10^{-4} to 1×10^{-2}), batch size (128 to 512), latent dimensions for drug, target, and disease encoders (128 to 256), and focal loss hyperparameters (γ: 2.0 to 5.0, α: 0.35 to 0.75). The DD' model (no target encoder) used the same search space except for dropout (0.3–0.6) and focal α (0.25 to 0.90), and was given a larger optimization budget of 43 configurations (3 initialization + 40 optimization iterations), so that the baseline was not under-tuned relative to DTD'.

2.6 Cross-Validation with Bayesian Optimization and Final Model Training

The mean validation AUC-PR across all five folds served as the objective function for Bayesian optimization. The test set was not used at any point during optimization. The best hyperparameters were then used to train a five-fold ensemble Fig. 2b and each fold was saved independently, with the decision threshold selected from out-of-fold validation scores. All features were standardized using StandardScaler fitted on the training data before model training. The test set was evaluated only after all training procedures were completed.

2.7 Evaluation Metrics

We evaluated the predictive performance of our MLP model using standard classification metrics including precision, recall, Matthew's correlation coefficient (MCC)(Chicco & Jurman, 2023), and F1 scores. We computed the enrichment factor to assess the model's ability to prioritize true associations within the top fraction of ranked candidates. For each run, the decision threshold was chosen by maximizing F1 on the out-of-fold predictions from cross-validation on the training set. All threshold-dependent metrics (F1, precision, recall, MCC, accuracy) were computed on the test set at that per-run threshold and reported as mean ± std across seeds. AUROC and enrichment factor (EF) are threshold-free. Similarly, we implemented ranking-based metrics MRR and Hit@K to demonstrate the model's ability to rank true positives above false candidates (i.e. negative-labelled drug-disease pairs ranked higher than positive-labelled drug-disease candidates.

2.8 Inference Set Generation

To generate novel drug repurposing hypotheses for diseases, we performed inference across all feasible drug-disease-target combinations. Using MeDIC as the approved drug source, along with 22K diseases and 4.4K human genes from ROBOKOP KG, we conducted inference on approximately 194 billion drug-target-disease triplets, excluding those used in training and testing sets. We further filtered the nominations using a manually curated drug list from Every-Cure (EC) (Every Cure, 2026) which is also the subset of MeDIC. This list is manually curated by medical expert of EC and helps us to filter out the nominations made using combination therapies, and excludes any non-therapeutic approved drugs such as vaccines, allergens and diagnostics during model inference.

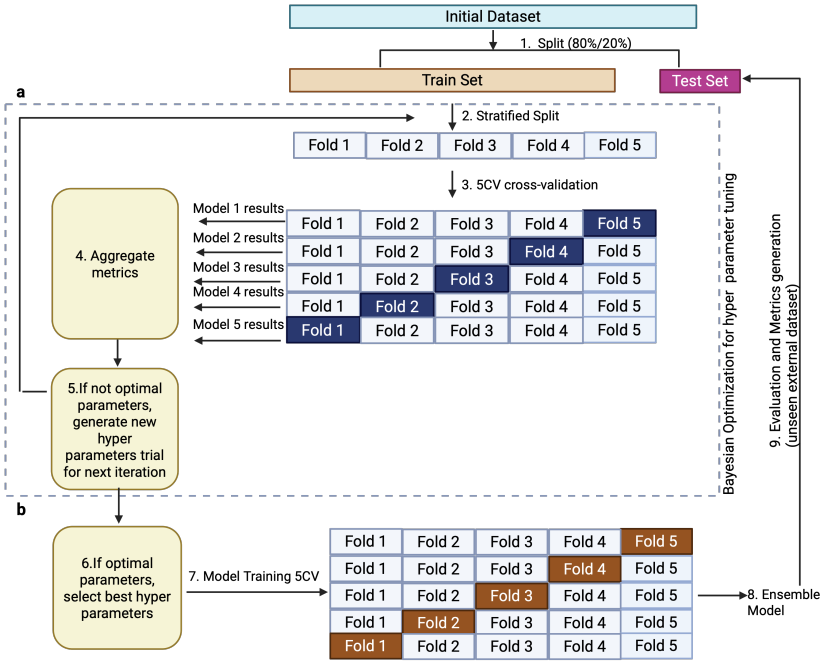


Figure 2: Cross-validation framework with Bayesian hyperparameter optimization. (a) Hyperparameter search via BO during model training. (b) Final model training using the optimal hyperparameters identified in (a).

2.9 LLM-Based Validation Of Drug Repurposing Predictions.

To validate and provide mechanistic explanations for the top 50 drug-disease hypotheses associated with a specific disease, using two models: ChatGPT-5.6-Sol (OpenAI) (OpenAI, 2026) and Claude-Sonnet-5 (Anthropic) (AI, 2025). To minimize hallucinations, models were constrained to retrieve evidence exclusively from PubMed (National Library of Medicine (US), 2026b) and ClinicalTrials.gov (National Library of Medicine (US), 2026a). The LLM-model generates rationales for predicted hypotheses and helps avoid false positives, which are a major issue in drug repurposing approach. We removed any drug-disease association without evidence and if flagged contraindicated by the LLM models.

3 Experiments and Results.

To assess the predictive performance of the model trained with target information, we compared the DTD' triplet model against a matched DD' doublet model trained on same drug-disease pairs but with target embeddings withheld, as described in the training set generation section above. The DTD' model outperformed DD' model across all the classification metrics (Table 1). We further evaluated both models across varying positive-to-negative ratios. As the number of negative samples increased, both the DTD' and DD' model shows decreasing trends for classification metrics, however, the early enrichment factors gradually increase. (Appendix Table A1).

Given the high experimental costs associated with validating potential drug-disease predictions, metrics such as an enrichment factor are particularly valuable for assessing model accuracy. The DTD' model demonstrated superior performance in recovering known indications of existing drugs within the held-out test set. Notably, as negative sampling increased, the enrichment factor at 1% also increases for DTD' model (Appendix Table A1) indicating that the model became more conservative in predicting positive labels while its positive predictions were increasingly likely to be true. Previous studies have shown that tree-based models such as gradient boosting often perform well with

Table 1: Performance of DTD' triplet models versus DD' doublet models (mean ± standard deviation) evaluated on the test set.

Model	AUROC	Precision	Recall	F1-score	MCC	EF1%	EF3%
DTD′	**0.98** ± 0.00	**0.61** ± 0.06	**0.76** ± 0.03	**0.68** ± 0.04	**0.67** ± 0.04	**27.54** ± 0.68	**21.87** ± 1.02
DD′	0.91 ± 0.00	0.41 ± 0.05	0.35 ± 0.02	0.38 ± 0.01	0.36 ± 0.01	15.56 ± 1.02	12.11 ± 0.63

feature embeddings(Xuan et al., 2019; Liu et al., 2019). Thus, we compared the performance of the MLP-model with other baseline approaches. The MLP-based DTD' model outperformed other baseline models. Both precision and recall are better for MLP-based DTD' model compared to another baselines. (Table 2).

Table 2: Performance of MLP and baseline classifiers on DTD′ triplet embeddings (mean ± standard deviation). Best value per metric in bold.

Model	AUROC	Precision	Recall	F1 score	MCC	EF1%	EF3%
DTD'-MLP	**0.98** ± 0.00	**0.61** ± 0.06	0.76 ± 0.03	**0.68** ± 0.04	**0.67** ± 0.04	**27.54** ± 0.68	**21.87** ± 1.02
DTD'-XGBoost	0.93 ± 0.00	0.52 ± 0.01	**0.79** ± 0.00	0.63 ± 0.01	0.63 ± 0.01	27.33 ± 0.15	21.41 ± 0.15
DTD'-Random Forest	0.95 ± 0.00	0.47 ± 0.01	0.68 ± 0.01	0.56 ± 0.00	0.56 ± 0.00	24.81 ± 0.11	18.32 ± 0.04
DTD'-CatBoost	0.97 ± 0.00	0.51 ± 0.00	0.77 ± 0.00	0.62 ± 0.00	0.62 ± 0.00	27.15 ± 0.00	21.10 ± 0.07
DTD'-LSTM	0.96 ± 0.00	0.51 ± 0.01	0.75 ± 0.01	0.61 ± 0.01	0.60 ± 0.01	26.54 ± 0.21	10.47 ± 0.79

To assess the contribution of KG-derived feature embeddings, we compared DTD'-MLP model trained on N2V embeddings with results from PubMedBERT, BioBERT, and text-embedding-3-small (OpenAI). Models trained on PubMedBERT and BioBERT embeddings yielded weaker classification metrics and AUC-PR (Table 3, 4), indicating that, despite being pretraining on large corpora, these embeddings are limited in distinguishing true drug-disease associations. In contrast, N2V embeddings, produced the strongest overall performance, with higher precision (0.61 ± 0.060), F1-score (0.68 ± 0.04). These findings highlight the significance of graph-derived embeddings from curated KGs for task-specific applications such as drug hypotheses generation, and the utility of such KGs as open-science resources. Notably, at the 1:10 ratio, all models showed comparable early enrichment of true positives, with the N2V-based model performing slightly better(Table 2). This suggests that while all three embeddings rank top candidates with similar effectiveness, N2V's graph-derived representations better discriminate positive from negative associations across the full prediction space. Further for or the imbalanced dataset, AUC-PR is one of the key metrics of comparison, and we observed MLP-based DTD' outperformed DD'-based model (0.75 ± 0.05 vs. 0.33 ± 0.03) and all other models (Table 4). This demonstrates inclusion of target improved model's ability to identify positives among imbalanced negative samples.

Table 3: Performance of MLP classifiers using different embedding types (mean ± standard deviation). N2V denotes Node2Vec. Best value per metric in bold.

Model	AUROC	Precision	Recall	F1-score	MCC	EF1%	EF3%
N2V+MLP	**0.98** ± 0.00	**0.61** ± 0.06	**0.76** ± 0.03	**0.68** ± 0.04	**0.67** ± 0.04	**27.54** ± 0.68	**21.87** ± 1.02
PubMedBERT+MLP	0.96 ± 0.00	0.54 ± 0.05	0.69 ± 0.04	0.60 ± 0.04	0.60 ± 0.05	25.65 ± 1.01	19.31 ± 1.20
BioBERT+MLP	0.96 ± 0.01	0.54 ± 0.05	0.67 ± 0.04	0.60 ± 0.05	0.59 ± 0.05	25.63 ± 1.05	19.12 ± 1.20
OpenAI+MLP	0.97 ± 0.01	0.60 ± 0.07	0.75 ± 0.04	0.66 ± 0.06	0.66 ± 0.06	27.34 ± 1.08	21.35 ± 1.47

The MLP-based model outperformed baseline approaches in ranking-based metrics, demonstrating improved ranking capability of true drug-target-disease triplets from randomly generated candidates (Table 1, 2 3). For the imbalanced dataset, AUC-PR is one of the key metrics of comparison, and we observed MLP-based DTD' outperformed DD'-based model (0.75 ± 0.05 vs 0.33 ± 0.03) (Table 4). To isolate the contribution of target information, we compared triplet-trained and doublet-trained models on identical test split (Table 4). Since doublet models lack target information, we evaluated both models on their ability to rank true drug-disease pairs, effectively removing the target component from the comparison. The triplet model's superior performance confirms that explicit target information improves ranking metrics. We saw similar trend of performance for triplet-based

Table 4: Ranking performance of DTD' and DD' models across classifiers and embedding types (mean ± standard deviation). N2V denotes Node2Vec. Best value per metric in bold.

Model	AUC-PR	MRR	Hit@1	Hit@3
N2V+MLP (DTD')	**0.75** ± 0.05	**0.28** ± 0.00	**0.14** ± 0.01	**0.31** ± 0.02
N2V+MLP (DD')	0.33 ± 0.03	0.17 ± 0.01	0.07 ± 0.01	0.16 ± 0.01
XGBoost	0.73 ± 0.00	0.27 ± 0.00	0.13 ± 0.00	0.30 ± 0.00
Random Forest	0.60 ± 0.00	0.25 ± 0.00	0.12 ± 0.02	0.30 ± 0.00
CatBoost	0.71 ± 0.00	0.27 ± 0.05	0.13 ± 0.05	0.30 ± 0.05
LSTM	0.67 ± 0.01	0.25 ± 0.00	0.12 ± 0.00	0.27 ± 0.01
PubMedBERT+MLP	0.65 ± 0.06	0.26 ± 0.02	0.12 ± 0.01	0.27 ± 0.02
BioBERT+MLP	0.64 ± 0.06	0.25 ± 0.01	0.12 ± 0.01	0.27 ± 0.02
OpenAI+MLP	0.73 ± 0.07	**0.28** ± 0.00	0.13 ± 0.01	**0.31** ± 0.01

baseline models. Finally, we applied the trained MLP-based triplet model to generate novel hypotheses for human diseases in ROBOKOP KG. Once the model predicted the probability score for each drug-target-disease association, we utilized ChatGPT 5.6-Sol and Claude-Sonnet-5 to conduct a detailed investigation of the top 50 nominated drugs for a given disease and generated LLM-based feasibility reports for potential drug-disease candidate pairs. We further reviewed the literature to examine the role of top-ranked targets in supporting the drug-disease association. To demonstrate the usefulness of the model, we present potential hypotheses for two neurodegenerative case-study diseases.

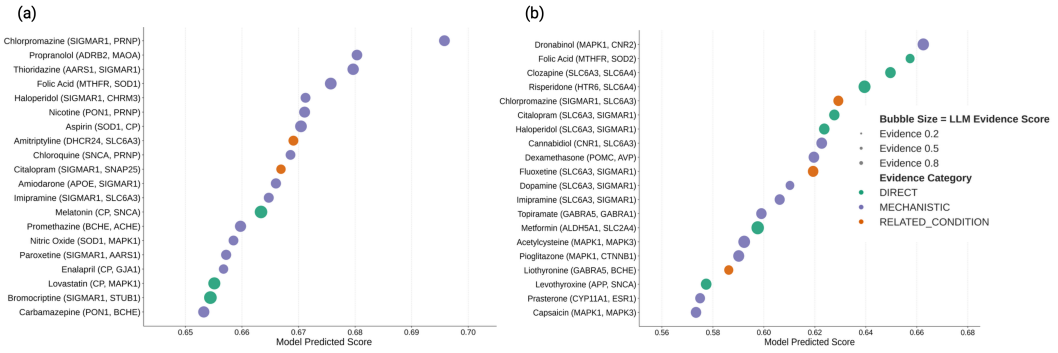


Figure 3: Top 20 LLM-based evidence scores for potential drug candidates against two neurodegenerative diseases: (a) Amyotrophic Lateral Scelerosis (ALS) and (b) Huntington's Disease (HD). The x-axis shows the predicted score from the triplet model, and the y-axis lists the top 20 candidate drugs. The parentheses next to each drug name show its top two predicted targets. Bubble color is coded by evidence category, and bubble size reflects the LLM-based evidence score. Details on the LLM evidence score and category are explained in the supplemental file.

3.1 Case Study 1: Amyotrophic Lateral Sclerosis (ALS)

Amyotrophic lateral sclerosis (ALS), commonly known as Lou Gehrig's disease, is a rare motor neuron disorder (MND) characterized by the progressive loss of both upper and lower motor neurons (LMNs)(Masrori & Van Damme, 2020). Despite a growing understanding of the disease and its genetic markers, there is currently no cure against ALS.

Among the top 20 ranked drugs predicted by our model(Fig. 3a), the LLMs identified several drugs whose mechanism strongly align with the disease's underlying biology. For instance, chlorpromazine binds the sigma-1 receptor (S1R), an endoplasmic reticulum chaperone concentrated at the mitochondria-associated membrane (MAM) in motor neurons (Hayashi & Su, 2007; Mancuso et al., 2012), where loss-of-function SIGMAR1 mutations are an established genetic cause of ALS via MAM collapse, ER stress, and Ca2+ dysregulation (Hayashi & Su, 2007; Watanabe et al., 2016). Since restoring Ca2+ handling and ER-stress signaling rescues motor neuron degeneration in pre-

clinical SIGMAR1-knockout mice, a compound engaging this receptor could plausibly slow motor neuron loss and show mechanistic effect against ALS (Mancuso et al., 2012).

Thioridazine was also recommended against ALS and has been previously demonstrated to induce TDP-43 aggregate clearance in a drosophila model (Cragnaz et al., 2021). While it has not been proven that TDP-43 is causative of ALS, it is a pathological hallmark; under the hypothesis that TDP-43 clearance would improve disease state, this prediction warrants further investigation.

Folic acid was predicted by the model as another candidate drug against ALS, acting via methylenetetrahydrofolate reductase (MTHFR) pathway. The MTHFR C677T polymorphism reduces MTHFR enzymatic activity, and its TT genotype has been associated with increased sporadic ALS risk across multiple case-control studies and a Caucasian-population meta-analysis(Żur Wyrozumska et al., 2017; Kühnlein et al., 2011). Reduced MTHFR activity impairs folate metabolism and raises homocysteine levels, which are elevated in ALS patients and is neurotoxic to motor neurons via NMDA receptor (GluN2A) overstimulation, calcium influx, and downstream excitotoxic and oxidative motor neuron injury (Deep et al., 2024). Because folic acid is the substrate that MTHFR-impaired individuals process poorly, supplementation may lower homocysteine and mitigate this excitotoxic pathway, offering a plausible genotype-linked rationale in ALS.

3.2 Case Study 2: Huntington's Disease (HD)

Huntington's disease (HD) is a rare autosomal dominant neurodegenerative disorder caused by an expansion of CAG repeats in HTT gene located on chromosome 4p16.3, resulting in an abnormal polyglutamine expansion in the huntingtin protein (Nopoulos, 2016). Despite advances in understanding disease pathology, no approved disease-modifying therapies currently exist for HD.

Our model predicted dronabinol as one of the potential nominations against HD via the cannabinoid receptor type 1 (CB1) pathway (Fig 3b). Dronabinol, activates CB1, which engages ERK/MAPK1-dependent pathways involved in neuronal survival in mouse models of Huntington's disease.(Scotter et al., 2010) Because CB1-linked ERK signaling capacity is diminished in the Huntington's disease state, cannabinoid receptor agonism represents a plausible mechanism for restoring this pro-survival pathway. (Laprairie et al., 2016)

Metformin presents another potential candidate via AMPK activation, which reduces mutant huntingtin aggregation and improves motor and cognitive phenotypes.(Sanchis et al., 2019) Aldehyde dehydrogenase 5 family member A1 (ALDH5A1) controls the breakdown of gamma-aminobutyric acid (GABA) and GABA signaling has been found to be disrupted in the striatum in HD(Rosas-Arellano et al., 2018). Since metformin separately promotes GABA type A receptor trafficking to the cell surface, its benefit in HD may be associated with role in the GABAergic pathway, however, this mechanism has not been directly tested.

4 Discussion

In our framework, we utilize feature embeddings generated from KGs to train an MLP model to predict the treatment probabilities. Our results demonstrate that incorporating target embeddings during model training improves classification performance compared to models trained on drug-disease pairs alone (Table 1). The inclusion of target information provides mechanistic context that enhances both prediction accuracy and explainability. Additionally, integrating LLM-based filtering and feasibility reports helps prioritizing predictions by identifying supporting evidence in published clinical and research literature that may assist medical professionals, including physicians and other licensed practitioners, in making a more informed and rational decision.

Although KGs have previously been used for link prediction for drug-disease association, many existing approaches rely on relatively small-sized graphs and often struggle to scale to larger graphs with more complex biomedical relationships. In this study, we addressed this limitation by employing embedding generation techniques such as N2V to derive feature representation from BKGs. We observed better model performance compared to feature vectors generated using pretrained LLMs. This superior performance can be attributed to several factors: (1) N2V embeddings are trained specifically on the graph structure and traverse through relationships relevant to drug repurposing, whereas language models capture general biomedical text patterns; and (2) N2V captures multi-hop

relationships and network topology that text-based models cannot easily learn. Furthermore, the MLP architecture outperformed alternative approaches, including XGBoost, Random Forest, and LSTM models in our experiments (Table 2). Unlike LSTMs, which are designed for sequential data, MLPs are well-suited for the static embedding representations used in our framework. This framework enables systematic generation and validation of drug repurposing hypotheses for any disease, with feasibility assessments provided prior to presenting candidates to medical professionals.

5 LIMITATIONS

Despite these promising results, we acknowledge several limitations. We rely on ground truth extracted from ROBOKOP KG. Like other biomedical knowledge graphs, ROBOKOP suffers from a lack of completeness in drug-target-disease associations and is error-prone during curation steps, which may result in erroneous connections. Since genes and proteins are conflated in ROBOKOP KG, targets may not necessarily represent genes only, which is a more specific node type for drug-target-disease associations. As the data are ingested from known literature and databases, widely studied drug-disease associations can dominate the triplets with multiple target associations, while understudied associations could be underrepresented, which could influence the model's ability to generalize.

Drug repurposing remains a complex and challenging problem in biomedical research, and application of machine learning or artificial intelligence in this domain should be done with caution. It should be noted that the top-ranked drugs do not constitute a medical recommendation and require extensive investigation and validation prior to considering clinical use. Continued improvements in the completeness and curation of KGs is expected to further enhance representation quality. In the current implementation, edge-type semantics are not explicitly incorporated during the embedding generation process, meaning that N2V treats all edges equally regardless of their type. This can limit the model's ability to capture relationship-specific information, which may be critical in drug repurposing. Future work will explore the integration of heterogeneous graph embedding methods, such as Heterogeneous Graph Transformer ((Hu et al., 2020)) and Hierarchical Heterogeneous Graph Transformer Zhu et al. (2024), to incorporate relationship-specific semantics and capture richer biological context.

6 CONCLUSION

In this work, we introduced a target-aware drug repurposing framework that models drug-target-disease triplets rather than conventional drug-disease pairs, demonstrating that explicit incorporation of molecular target information improves both classification performance and mechanistic interpretability. By combining N2V embeddings derived from the ROBOKOP biomedical knowledge graph with a multi-head MLP classifier and LLM-assisted mechanistic validation, the framework enables scalable generation of target-informed, interpretable, and biologically grounded repurposing hypotheses across thousands of human diseases for the first time.

AI USE STATEMENT

We used generative AI tools (OpenAI and Anthropic Claude Code) used them in LLM-based feasibility report generation and to assist in refactoring the code. We used OpenAI "text-embedding-3-small" via API calls to generate the embeddings for entities used in model performance comparison with Node2Vec model. The AI tools were not used for the generation of research ideas or experimental designs. Generative tools were used in editing the manuscript text for language refinement,grammatical errors and improving clarity and readability. We take full responsibility for the accuracy and integrity of the paper's content.

ETHICS STATEMENT

This study uses publicly available ROBOKOP KG, and open-source PubmedBERT and Biobert models. No human-subject data were used. The nominated drug-disease associations in two case study neurodegenerative diseases are potential drug hypotheses; they are not clinical recommen-

dations and require extensive investigation and validation from domain expert prior to considering clinical use.

Reproducibility Statement

The ground truth, embeddings are uploaded in Harvard Dataverse (https://dataverse.harvard.edu/previewurl.xhtml?token=a9accd77-12cb-4e20-893c-7e1054b11914). The codebase and benchmark assets are ready for release. At submission, anonymized versions will be provided to reviewers, along with instructions to reproduce the main results. These resources are not yet public but will be released upon acceptance.

References

Anthropic AI. Claude. *Sonnet-5,*, 2025.

Sorin Avram, Cristian G. Bologa, Jayme Holmes, Giovanni Bocci, Thomas B. Wilson, Dac-Trung Nguyen, Ramona Curpan, Liliana Halip, Alina Bora, Jeremy J. Yang, Jeffrey Knockel, Suman Sirimulla, Oleg Ursu, and Tudor I. Oprea. DrugCentral 2021 supports drug discovery and repositioning. *Nucleic Acids Research*, 49(D1):D1160–D1169, January 2021. ISSN 1362-4962. doi: 10.1093/nar/gkaa997.

Chris Bizon, Steven Cox, James Balhoff, Yaphet Kebede, Patrick Wang, Kenneth Morton, Karamarie Fecho, and Alexander Tropsha. ROBOKOP KG and KGB: Integrated Knowledge Graphs from Federated Sources. *Journal of Chemical Information and Modeling*, 59(12):4968–4973, December 2019. ISSN 1549-9596. doi: 10.1021/acs.jcim.9b00683. URL https://doi.org/10.1021/acs.jcim.9b00683.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating Embeddings for Modeling Multi-relational Data. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html.

Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

Payal Chandak, Kexin Huang, and Marinka Zitnik. Building a knowledge graph to enable precision medicine. *Scientific Data*, 10(1):67, February 2023. ISSN 2052-4463. doi: 10.1038/s41597-023-01960-3. URL https://www.nature.com/articles/s41597-023-01960-3.

Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, August 2016. doi: 10.1145/2939672.2939785. URL http://arxiv.org/abs/1603.02754. arXiv:1603.02754 [cs].

Davide Chicco and Giuseppe Jurman. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Mining*, 16:4, February 2023. ISSN 1756-0381. doi: 10.1186/s13040-023-00322-4. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9938573/.

Lucia Cragnaz, Greta Spinelli, Laura De Conti, Emilie A. Bureau, Janet Brownlees, Fabian Feiguin, Valentina Romano, Natasa Skoko, Raffaella Klima, Catherine A. Kettleborough, Francisco E. Baralle, and Marco Baralle. Thioridazine reverts the phenotype in cellular and *Drosophila* models of amyotrophic lateral sclerosis by enhancing TDP-43 aggregate clearance. *Neurobiology of Disease*, 160:105515, December 2021. ISSN 0969-9961. doi: 10.1016/j.nbd.2021.105515. URL https://www.sciencedirect.com/science/article/pii/S0969996121002643.

Satya Narayan Deep, Sarah Seelig, Surojit Paul, and Ranjana Poddar. Homocysteine-induced sustained GluN2A NMDA receptor stimulation leads to mitochondrial ROS generation and neurotoxicity. *The Journal of Biological Chemistry*, 300(5):107253, April 2024. ISSN 0021-9258. doi: 10.1016/j.jbc.2024.107253. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC11081806/.

Marcello DeLuca, Nico Matentzoglu, Elliott Sharp, Jane Li, Charlie Hempstead, May Lim, Piotr Kaniewski, E Kathleen Carter, Kushal Koirala, Elvin Ding, Laurens Vijnck, Pascal Brokmeier, Sabrina Toro, Kevin Schaper, Jacques Vergine, Olivia Li, Tudor I Oprea, David C Fajgenbaum, Christopher Bizon, Melissa Haendel, and Alexander Tropsha. Medicines, Diseases, Indications, and Contraindications (MeDIC): a foundational resource to support drug repurposing. *Nucleic Acids Research*, pp. gkaf1312, December 2025. ISSN 1362-4962. doi: 10.1093/nar/gkaf1312. URL https://doi.org/10.1093/nar/gkaf1312.

Every Cure. Every cure drug list. https://huggingface.co/datasets/everycure/drug-list, 2026. Hugging Face dataset. Accessed: 2026-09-23.

Karamarie Fecho, Evan Morris, Jon-Michael Beasley, E. Kathleen Carter, Cheng-Han Chung, Steve Edwards, David Glymph, Daniel Korn, Olawumi Olasunkanmi, Carrie Pasfield, Michael Ramon, Joey Richardson, Nyssa Tucker, David Williams, Weifan Zheng, Alexander Tropsha, and Chris Bizon. The ROBOKOP v1.0 knowledge graph system for exploring relationships between biomedical entities. *Scientific Reports*, 16:17368, June 2026. ISSN 2045-2322. doi: 10.1038/s41598-026-53036-y. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC13237073/.

Nadime Francis, Alastair Green, Paolo Guagliardo, Leonid Libkin, Tobias Lindaaker, Victor Marsault, Stefan Plantikow, Mats Rydberg, Petra Selmer, and Andrés Taylor. Cypher: An Evolving Query Language for Property Graphs. In *Proceedings of the 2018 International Conference on Management of Data*, SIGMOD '18, pp. 1433–1445, New York, NY, USA, May 2018. Association for Computing Machinery. ISBN 978-1-4503-4703-7. doi: 10.1145/3183713.3190657. URL https://dl.acm.org/doi/10.1145/3183713.3190657.

Amy K Glen, Chunyu Ma, Luis Mendoza, Finn Womack, E C Wood, Meghamala Sinha, Liliana Acevedo, Lindsey G Kvarfordt, Ross C Peene, Shaopeng Liu, Andrew S Hoffman, Jared C Roach, Eric W Deutsch, Stephen A Ramsey, and David Koslicki. ARAX: a graph-based modular reasoning tool for translational biomedicine. *Bioinformatics*, 39(3):btad082, March 2023. ISSN 1367-4811. doi: 10.1093/bioinformatics/btad082. URL https://doi.org/10.1093/bioinformatics/btad082.

Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Trans. Comput. Healthcare*, 3(1):2:1–2:23, October 2021. doi: 10.1145/3458754. URL https://dl.acm.org/doi/10.1145/3458754.

Teruo Hayashi and Tsung-Ping Su. Sigma-1 Receptor Chaperones at the ER- Mitochondrion Interface Regulate Ca2+ Signaling and Cell Survival. *Cell*, 131(3):596–610, November 2007. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2007.08.036. URL https://www.cell.com/cell/abstract/S0092-8674(07)01099-9.

Daniel Scott Himmelstein, Antoine Lizee, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, and Sergio E Baranzini. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife*, 6:e26726. ISSN 2050-084X. doi: 10.7554/eLife.26726. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5640425/.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997.

Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous Graph Transformer, March 2020. URL http://arxiv.org/abs/2003.01332. arXiv:2003.01332 [cs].

Yi Hua, Xiaowen Dai, Yuan Xu, Guomeng Xing, Haichun Liu, Tao Lu, Yadong Chen, and Yanmin Zhang. Drug repositioning: Progress and challenges in drug discovery for various diseases. *European Journal of Medicinal Chemistry*, 234:114239, April 2022. ISSN 0223-5234. doi: 10.1016/j.ejmech.2022.114239. URL https://www.sciencedirect.com/science/article/pii/S0223523422001416.

Peter Kühnlein, Hans Jung, Melinda Farkas, Salla Keskitalo, Benjamin Ineichen, Ilijas Jelcic, Jens Petersen, Alexander Semmler, Michal Weller, Albert Christian Ludolph, and Michael Linnebank. The thermolabile variant of 5,10-methylenetetrahydrofolate reductase is a possible risk factor for amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*, 12(2):136–139, March 2011. ISSN 1471-180X. doi: 10.3109/17482968.2010.536985.

Robert B. Laprairie, Amina M. Bagher, Melanie E. M. Kelly, and Eileen M. Denovan-Wright. Biased Type 1 Cannabinoid Receptor Signaling Influences Neuronal Viability in a Cell Culture Model of Huntington Disease. *Molecular Pharmacology*, 89(3):364–375, March 2016. ISSN 1521-0111. doi: 10.1124/mol.115.101980.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, February 2020. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btz682. URL http://arxiv.org/abs/1901.08746. arXiv:1901.08746 [cs].

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal Loss for Dense Object Detection, February 2018. URL http://arxiv.org/abs/1708.02002. arXiv:1708.02002 [cs].

Hui Liu, Wenhao Zhang, Lixia Nie, Xiancheng Ding, Judong Luo, and Ling Zou. Predicting effective drug combinations using gradient tree boosting based on features extracted from drug-protein heterogeneous network. *BMC Bioinformatics*, 20:645, December 2019. ISSN 1471-2105. doi: 10.1186/s12859-019-3288-1. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6902475/.

Renzo Mancuso, Sara Oliván, Amaya Rando, Caty Casas, Rosario Osta, and Xavier Navarro. Sigma-1R Agonist Improves Motor Function and Motoneuron Survival in ALS Mice. *Neurotherapeutics*, 9(4):814–826, October 2012. ISSN 1933-7213. doi: 10.1007/s13311-012-0140-y. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3480575/.

P. Masrori and P. Van Damme. Amyotrophic lateral sclerosis: a clinical review. *European Journal of Neurology*, 27(10):1918–1929, 2020. ISSN 1468-1331. doi: 10.1111/ene.14393. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ene.14393.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space, September 2013. URL http://arxiv.org/abs/1301.3781. arXiv:1301.3781 [cs].

John H Morris, Karthik Soman, Rabia E Akbas, Xiaoyuan Zhou, Brett Smith, Elaine C Meng, Conrad C Huang, Gabriel Cerono, Gundolf Schenk, Angela Rizk-Jackson, Adil Harroud, Lauren Sanders, Sylvain V Costes, Krish Bharat, Arjun Chakraborty, Alexander R Pico, Taline Mardirossian, Michael Keiser, Alice Tang, Josef Hardi, Yongmei Shi, Mark Musen, Sharat Israni, Sui Huang, Peter W Rose, Charlotte A Nelson, and Sergio E Baranzini. The scalable precision medicine open knowledge engine (SPOKE): a massive knowledge graph of biomedical information. *Bioinformatics*, 39(2):btad080, February 2023. ISSN 1367-4811. doi: 10.1093/bioinformatics/btad080. URL https://doi.org/10.1093/bioinformatics/btad080.

Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pp. 807–814, Madison, WI, USA, June 2010. Omnipress. ISBN 978-1-60558-907-7.

National Library of Medicine (US). ClinicalTrials.gov. https://clinicaltrials.gov/, 2026a. Bethesda (MD): National Library of Medicine. Accessed: September 2026.

National Library of Medicine (US). PubMed. https://pubmed.ncbi.nlm.nih.gov/, 2026b. Bethesda (MD): National Center for Biotechnology Information. Accessed: September 2026.

Peggy C. Nopoulos. Huntington disease: a single-gene degenerative disorder of the striatum. *Dialogues in Clinical Neuroscience*, 18(1):91–98, March 2016. ISSN 1294-8322. doi: 10.31887/DCNS.2016.18.1/pnopoulos. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4826775/.

OpenAI. GPT-5.6 Sol [large language model]. https://developers.openai.com/api/docs/models/gpt-5.6-sol, 2026. Accessed via the OpenAI API. Accessed: 2026-08-23.

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. CatBoost: unbiased boosting with categorical features, January 2019. URL http://arxiv.org/abs/1706.09516. arXiv:1706.09516 [cs.LG].

Sudeep Pushpakom, Francesco Iorio, Patrick A. Eyers, K. Jane Escott, Shirley Hopper, Andrew Wells, Andrew Doig, Tim Guilliams, Joanna Latimer, Christine McNamee, Alan Norris, Philippe Sanseau, David Cavalla, and Munir Pirmohamed. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery*, 18(1):41–58, January 2019. ISSN 1474-1784. doi: 10.1038/nrd.2018.168. URL https://www.nature.com/articles/nrd.2018.168.

Justin Reese, Deepak Unni, Tiffany J. Callahan, Luca Cappelletti, Vida Ravanmehr, Seth Carbon, Tommaso Fontana, Hannah Blau, Nicolas Matentzoglu, Nomi L. Harris, Monica C. Munoz-Torres, Peter N. Robinson, Marcin P. Joachimiak, and Christopher J. Mungall. KG-COVID-19: a framework to produce customized knowledge graphs for COVID-19 response. *bioRxiv*, pp. 2020.08.17.254839, August 2020. ISSN 2692-8205. doi: 10.1101/2020.08.17.254839. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC7444288/.

Abraham Rosas-Arellano, Carlos Tejeda-Guzmán, Enrique Lorca-Ponce, Lourdes Palma-Tirado, Carola A. Mantellero, Patricio Rojas, Fanis Missirlis, and Maite A. Castro. Huntington's disease leads to decrease of GABA-A tonic subunits in the D2 neostriatal pathway and their relocalization into the synaptic cleft. *Neurobiology of Disease*, 110:142–153, February 2018. ISSN 0969-9961. doi: 10.1016/j.nbd.2017.11.010. URL https://www.sciencedirect.com/science/article/pii/S0969996117302681.

Ana Sanchis, María Adelaida García-Gimeno, Antonio José Cañada-Martínez, María Dolores Sequedo, José María Millán, Pascual Sanz, and Rafael P. Vázquez-Manrique. Metformin treatment reduces motor and neuropsychiatric phenotypes in the zQ175 mouse model of Huntington disease. *Experimental & Molecular Medicine*, 51(6):1–16, June 2019. ISSN 2092-6413. doi: 10.1038/s12276-019-0264-9. URL https://www.nature.com/articles/s12276-019-0264-9.

EL Scotter, CE Goodfellow, ES Graham, M Dragunow, and M Glass. Neuroprotective potential of CB1 receptor agonists in an in vitro model of Huntington's disease. *British Journal of Pharmacology*, 160(3):747–761, June 2010. ISSN 0007-1188. doi: 10.1111/j.1476-5381.2010.00773.x. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC2931573/.

Chang Su, Yu Hou, Manqi Zhou, Suraj Rajendran, Jacqueline R. M. A. Maasch, Zehra Abedi, Haotan Zhang, Zilong Bai, Anthony Cuturrufo, Winston Guo, Fayzan F. Chaudhry, Gregory Ghahramani, Jian Tang, Feixiong Cheng, Yue Li, Rui Zhang, Steven T. DeKosky, Jiang Bian, and Fei Wang. Biomedical discovery through the integrative biomedical knowledge hub (iBKH). *iScience*, 26(4):106460, April 2023. ISSN 2589-0042. doi: 10.1016/j.isci.2023.106460.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space, February 2019. URL http://arxiv.org/abs/1902.10197. arXiv:1902.10197 [cs].

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex Embeddings for Simple Link Prediction, June 2016. URL http://arxiv.org/abs/1606.06357. arXiv:1606.06357 [cs].

Deepak R. Unni, Sierra A. T. Moxon, Michael Bada, Matthew Brush, Richard Bruskiewich, J. Harry Caufield, Paul A. Clemons, Vlado Dancik, Michel Dumontier, Karamarie Fecho, Gustavo Glusman, Jennifer J. Hadlock, Nomi L. Harris, Arpita Joshi, Tim Putman, Guangrong Qin, Stephen A. Ramsey, Kent A. Shefchek, Harold Solbrig, Karthik Soman, Anne E. Thessen, Melissa A. Haendel, Chris Bizon, Christopher J. Mungall, Liliana Acevedo, Stanley C. Ahalt, John Alden, Ahmed Alkanaq, Nada Amin, Ricardo Avila, Jim Balhoff, Sergio E. Baranzini, Andrew Baumgartner, William Baumgartner, Basazin Belhu, MacKenzie Brandes, Namdi Brandon, Noel Burtt, William Byrd, Jackson Callaghan, Marco Alvarado Cano, Steven Carrell, Remzi Celebi, James Champion, Zhehuan Chen, Mei-Jan Chen, Lawrence Chung, Kevin Cohen, Tom Conlin, Dan Corkill, Maria Costanzo, Steven Cox, Andrew Crouse, Camerron Crowder, Mary E. Crumbley, Cheng Dai, Vlado Dančík, Ricardo De Miranda Azevedo, Eric Deutsch, Jennifer Dougherty, Marc P. Duby, Venkata Duvvuri, Stephen Edwards, Vincent Emonet, Nathaniel Fehrmann, Jason Flannick, Aleksandra M. Foksinska, Vicki Gardner, Edgar Gatica, Amy Glen, Prateek Goel, Joseph Gormley, Alon Greyber, Perry Haaland, Kristina Hanspers, Kaiwen He, Kaiwen He, Jeff Henrickson, Eugene W. Hinderer, Maureen Hoatlin, Andrew Hoffman, Sui Huang, Conrad Huang, Robert Hubal, Kenneth Huellas-Bruskiewicz, Forest B. Huls, Lawrence Hunter, Greg Hyde, Tursynay Issabekova, Matthew Jarrell, Lindsay Jenkins, Adam Johs, Jimin Kang, Richa Kanwar, Yaphet Kebede, Keum Joo Kim, Alexandria Kluge, Michael Knowles, Ryan Koesterer, Daniel Korn, David Koslicki, Ashok Krishnamurthy, Lindsey Kvarfordt, Jay Lee, Margaret Leigh, Jason Lin, Zheng Liu, Shaopeng Liu, Chunyu Ma, Andrew Magis, Tarun Mamidi, Meisha Mandal, Michelle Mantilla, Jeffrey Massung, Denise Mauldin, Jason McClelland, Julie McMurry, Philip Mease, Luis Mendoza, Marian Mersmann, Abrar Mesbah, Matthew Might, Kenny Morton, Sandrine Muller, Arun Teja Muluka, John Osborne, Phil Owen, Michael Patton, David B. Peden, R. Carter Peene, Bria Persaud, Emily Pfaff, Alexander Pico, Elizabeth Pollard, Guthrie Price, Shruti Raj, Jason Reilly, Anders Riutta, Jared Roach, Ryan T. Roper, Greg Rosenblatt, Irit Rubin, Sienna Rucka, Nathaniel Rudavsky-Brody, Rayn Sakaguchi, Eugene Santos, Kevin Schaper, Charles P. Schmitt, Shepherd Schurman, Erik Scott, Sarah Seitanakis, Priya Sharma, Ilya Shmulevich, Manil Shrestha, Shalki Shrivastava, Meghamala Sinha, Brett Smith, Noel Southall, Nicholas Southern, Lisa Stillwell, Michael " Michi" Strasser, Andrew I. Su, Casey Ta, Anne E. Thessen, Jillian Tinglin, Lucas Tonstad, Thi Tran-Nguyen, Alexander Tropsha, Gaurav Vaidya, Luke Veenhuis, Adam Viola, Marcin von Grotthuss, Max Wang, Patrick Wang, Paul B. Watkins, Rosina Weber, Qi Wei, Chunhua Weng, Jordan Whitlock, Mark D. Williams, Andrew Williams, Finn Womack, Erica Wood, Chunlei Wu, Jiwen Kevin Xin, Hao Xu, Colleen Xu, Chase Yakaboski, Yao Yao, Hong Yi, Arif Yilmaz, Marissa Zheng, Xinghua Zhou, Eric Zhou, Qian Zhu, and Tom Zisk. Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. *Clinical and Translational Science*, 15(8):1848–1855, August 2022. ISSN 1752-8054. doi: 10.1111/cts.13302. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9372416/.

Nicole A Vasilevsky, Sabrina Toro, Nicolas Matentzoglu, Joseph E Flack, Kathleen R Mullen, Harshad Hegde, Sarah Gehrke, Patricia L Whetzel, Yousif Shwetar, Nomi L Harris, Mee S Ngu, Gioconda L Alyea, Megan S Kane, Paola Roncaglia, Eric Sid, Courtney L Thaxton, Valerie Wood, Roshini S Abraham, Maria Isabel Achatz, Pamela Ajuyah, Joanna S Amberger, Lawrence Babb, Jasmine Baker, James P Balhoff, Jonathan S Berg, Amol Bhalla, Xavier Bofill-De Ros, Ian R Braun, Eleanor C Broeren, Blake K Byer, Alicia B Byrne, Tiffany J Callahan, Leigh C Carmody, Lauren E Chan, Amanda R Clause, Julie S Cohen, Marcello DeLuca, Natalie T Deuitch, May Flowers, Jamie Fraser, Toyofumi Fujiwara, Vanessa Gitau, Jennifer L Goldstein, Dylan Gration, Tudor Groza, Benjamin M Gyori, William Hankey, Jason A Hilton, Daniel S Himmelstein, Stephanie S Hong, Charles T Hoyt, Robert Huether, Eric Hurwitz, Julius O B Jacobsen, Atsuo Kikuchi, Sebastian Köhler, Daniel R Korn, David Lagorce, Bryan J Laraway, Jane Y Li, Adriana J Malheiro, James McLaughlin, Birgit H M Meldal, Shruthi Mohan, Sierra A T Moxon, Monica C Munoz-Torres, Tristan H Nelson, Frank W Nicholas, David Ochoa, Daniel Olson, Tudor I Oprea, Tomiko T Oskotsky, David Osumi-Sutherland, Kelley Paris, Helen E Parkinson, Zoë M Pendlington, Xiao P Peng, Amy Pizzino, Sharon E Plon, Bradford C Powell, Julie C Ratliff, Heidi L Rehm, Lyubov Remennik, Erin R Riggs, Sean Roberts, Peter N Robinson, Justyne E Ross, Kevin Schaper, Brian M Schilder, Johanna L Schmidt, Elliott W Sharp, Morgan N Similuk, Damian

Smedley, Tam P Sneddon, Rachel Sparks, Ray Stefancsik, Gregory S Stupp, Shilpa Sundar, Terue Takatsuki, Imke Tammen, Kezang C Tshering, Deepak R Unni, Eloise Valasek, Adeline Vanderver, Alex H Wagner, Ryan F Webb, Danielle Welter, Doron Yaya-Stupp, Andreas Zankl, Xingmin Aaron Zhang, Julie A McMurry, Christopher G Chute, Ada Hamosh, Christopher J Mungall, and Melissa A Haendel. Mondo: Integrating Disease Terminology Across Communities. *Genetics*, pp. iyaf215, October 2025. ISSN 1943-2631. doi: 10.1093/genetics/iyaf215. URL https://doi.org/10.1093/genetics/iyaf215.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge Graph Embedding by Translating on Hyperplanes. *Proceedings of the AAAI Conference on Artificial Intelligence*, 28 (1), June 2014. ISSN 2374-3468. doi: 10.1609/aaai.v28i1.8870. URL https://ojs.aaai.org/index.php/AAAI/article/view/8870.

Seiji Watanabe, Hristelina Ilieva, Hiromi Tamada, Hanae Nomura, Okiru Komine, Fumito Endo, Shijie Jin, Pedro Mancias, Hiroshi Kiyama, and Koji Yamanaka. Mitochondria-associated membrane collapse is a common pathomechanism in SIGMAR1- and SOD1-linked ALS. *EMBO Molecular Medicine*, 8(12):7, November 2016. ISSN 1757-4684. doi: 10.15252/emmm.201606403. URL https://doi.org/10.15252/emmm.201606403.

E. C. Wood, Amy K. Glen, Lindsey G. Kvarfordt, Finn Womack, Liliana Acevedo, Timothy S. Yoon, Chunyu Ma, Veronica Flores, Meghamala Sinha, Yodsawalai Chodpathumwan, Arash Termehchy, Jared C. Roach, Luis Mendoza, Andrew S. Hoffman, Eric W. Deutsch, David Koslicki, and Stephen A. Ramsey. RTX-KG2: a system for building a semantically standardized knowledge graph for translational biomedicine. *BMC Bioinformatics*, 23(1):400, September 2022. ISSN 1471-2105. doi: 10.1186/s12859-022-04932-3. URL https://doi.org/10.1186/s12859-022-04932-3.

Ping Xuan, Chang Sun, Tiangang Zhang, Yilin Ye, Tonghui Shen, and Yihua Dong. Gradient Boosting Decision Tree-Based Method for Predicting Interactions Between Target Genes and Drugs. *Frontiers in Genetics*, 10, May 2019. ISSN 1664-8021. doi: 10.3389/fgene.2019.00459. URL https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2019.00459/full.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding Entities and Relations for Learning and Inference in Knowledge Bases, August 2015. URL http://arxiv.org/abs/1412.6575. arXiv:1412.6575 [cs].

Ling Yue, Sixue Xing, Yingzhou Lu, and Tianfan Fu. BioMamba: A Pre-trained Biomedical Language Representation Model Leveraging Mamba, August 2024. URL http://arxiv.org/abs/2408.02600. arXiv:2408.02600 [cs].

Shuangjia Zheng, Jiahua Rao, Ying Song, Jixian Zhang, Xianglu Xiao, Evandro Fei Fang, Yuedong Yang, and Zhangming Niu. PharmKG: a dedicated knowledge graph benchmark for bomedical data mining. *Briefings in Bioinformatics*, 22(4):bbaa344, July 2021. ISSN 1477-4054. doi: 10.1093/bib/bbaa344. URL https://doi.org/10.1093/bib/bbaa344.

Qiuyu Zhu, Liang Zhang, Qianxiong Xu, Kaijun Liu, Cheng Long, and Xiaoyang Wang. Hhgt: Hierarchical heterogeneous graph transformer for heterogeneous graph representation learning, 2024. URL https://arxiv.org/abs/2407.13158.

Kamila Żur Wyrozumska, Joanna Pera, Anna Dziubek, Małgorzata Sado, Aleksandra Golenia, Agnieszka Słowik, and Tomasz Dziedzic. Association between C677T polymorphism of MTHFR gene and risk of amyotrophic lateral sclerosis: Polish population study and a meta-analysis. *Neurologia i Neurochirurgia Polska*, 51(2):135–139, March 2017. ISSN 0028-3843. doi: 10.1016/j.pjnns.2017.01.008. URL https://www.sciencedirect.com/science/article/pii/S0028384317300415.

A Appendix

A.1 List of Predicates Used

A list of predicates used to define DTD' "ground truth" DTD' triplets.

Drug-disease edges included predicates such as *"treats"*, *"treats_or_applied_or_studied_to_treat"*, *"applied_to"*, *"ameliorates_conditions"*, *"positively_correlated_to"*,*"correlated_to"*, *"contributes_to"*, and *"contraindicated_in"*; drug-target edges includes *"affects"*, *"regulates"*, *"related_to"*, and *"directly_physically_interacts_with"*; target-disease edges included *"gene_associated_with_condition"*, *"genetically_associated_with"*, *"has_phenotypes"*, *"causes"*,*"regulates"*, and *"has_adverse_event"*.

A.2 Architecture of DTD' Model

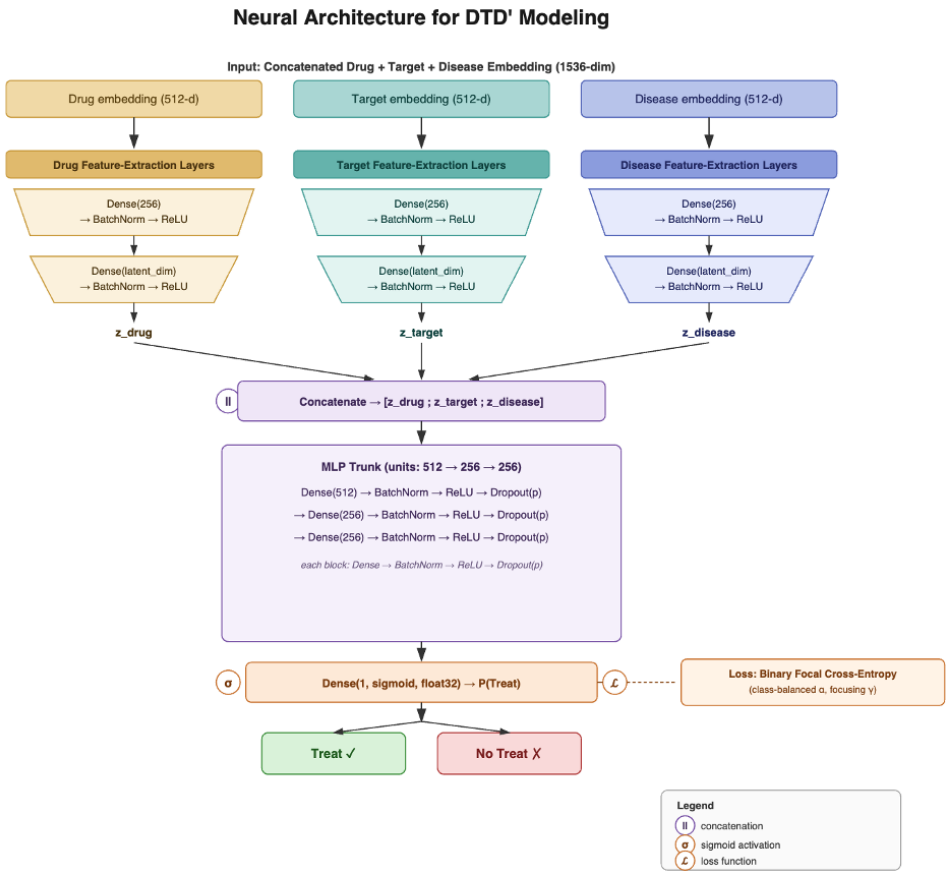


Figure A1: Architecture of the MLP model based on DTD′. Image generated by Claude (Sonnet 5).

A.3 Classification Metrics

Threshold metrics (Precision, Recall, F1, MCC):

The predicted label is $\hat{y} = 1[s \geq \tau]$. By default, τ is the threshold that gives the highest F_1 on the out-of-fold (OOF) validation scores.

$$\text{Precision} = \begin{cases} \dfrac{TP}{TP+FP}, & TP+FP>0 \\ 0, & TP+FP=0 \end{cases} \tag{1}$$

$$\text{Recall} = \begin{cases} \dfrac{TP}{TP+FN}, & TP+FN>0 \\ 0, & TP+FN=0 \end{cases} \tag{2}$$

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precison + Recall} = \frac{2TP}{2TP+FP+FN} \tag{3}$$

$$\text{MCC} = \begin{cases} \dfrac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, & \text{denominator} \neq 0 \\ 0, & \text{denominator} = 0 \end{cases} \tag{4}$$

Threshold-free metrics (AUROC, AUC-PR):

AUROC is computed with `roc_auc_score`. AUCPR uses `average_precision_score`.

A.4 Ranking Metrics

Enrichment factor calculation, EF@k:

All pairs are ranked globally by score. The cutoff n_k is defined as The enrichment factor at k is

$$\text{EF@}k = \frac{\text{TP}(n_k) \,/\, n_k}{P \,/\, N},$$

where $\text{TP}(n_k)$ is the number of true positives among the top n_k ranked pairs, P is the total number of positives, and N is the total number of pairs. Hit@K and MRR:

Let R be the list of ranks of the true-positive drug–disease pairs in the test set, where each pair is ranked among all test pairs for the same disease by predicted score (raw ranks, highest score first).

Mean reciprocal rank (MRR) is the average inverse rank of the true-positive pairs:

$$\text{MRR} = \frac{1}{|R|} \sum_{r \in R} \frac{1}{r}$$

Hits@K is the proportion of true-positive pairs whose rank is at most K:

$$\text{Hits@}K = \frac{1}{|R|} \sum_{r \in R} |r \leq K|$$

A.5 Table of Comparison

A.6 Prompt for LLM Feasibility Report for Drug–Disease Association

```
You are a biomedical expert evaluating drug repurposing potential. The
↪  {drug} was likely NOT developed for this {disease}. However, it may
↪  have being studies for this disease in the past or have mechanistic
↪  rationale for repurposing.

CRITICAL INSTRUCTIONS TO PREVENT HALLUCINATION:
1. Base ALL scores ONLY on the provided evidence - do NOT use external
↪  knowledge
2. If evidence is weak/absent, you MUST assign LOW scores (0.0-0.3)
```

Table A1: Performance comparison across different class imbalance ratios.

Ratio	Model	AUROC	Precision	Recall	F1	MCC	EF@1%	EF3%
1:3	DTD′	0.97 ± 0.00	0.63 ± 0.01	$\mathbf{0.85 \pm 0.02}$	$\mathbf{0.72 \pm 0.00}$	$\mathbf{0.71 \pm 0.00}$	11.93 ± 0.05	11.53 ± 0.05
	DD′	0.90 ± 0.00	0.47 ± 0.02	0.49 ± 0.04	0.49 ± 0.01	0.43 ± 0.01	8.33 ± 0.36	7.52 ± 0.28
1:5	DTD′	0.97 ± 0.00	$\mathbf{0.64 \pm 0.04}$	0.79 ± 0.04	0.71 ± 0.05	0.69 ± 0.05	17.45 ± 0.41	16.04 ± 0.82
	DD′	0.90 ± 0.00	0.45 ± 0.01	0.44 ± 0.02	0.44 ± 0.01	0.41 ± 0.01	11.56 ± 0.08	9.80 ± 0.12
1:10	DTD′	$\mathbf{0.98 \pm 0.00}$	0.61 ± 0.06	0.76 ± 0.03	0.68 ± 0.04	0.67 ± 0.04	$\mathbf{27.54 \pm 0.68}$	$\mathbf{21.87 \pm 1.02}$
	DD′	0.91 ± 0.00	0.41 ± 0.05	0.35 ± 0.02	0.38 ± 0.01	0.36 ± 0.01	15.56 ± 1.02	12.11 ± 0.63

```
3. Be conservative and honest about uncertainty

CONTEXT:
This is a REPURPOSING analysis. The drug "{drug}" was likely NOT
↪  originally developed for "{disease}".
Disease: {disease}
Drug: {drug}
Evidence from PubMed (up to 15 abstracts): {pubmed_text}
Clinical Trials Data (up to 100 trials): {trials_text}

TASK:
Score the following dimensions (0.0 to 1.0):
1. mechanistic_plausibility_score: Does the drug's mechanism align with
↪  disease pathophysiology?
   - 0.0-0.3: No clear mechanistic link
   - 0.4-0.6: Plausible indirect mechanisms
   - 0.7-1.0: Strong direct mechanistic rationale
2. literature_support_score: Quality and quantity of published evidence
   - 0.0-0.3: Minimal or no relevant publications
   - 0.4-0.6: Some relevant basic science or case reports
   - 0.7-1.0: Multiple high-quality studies
3. retrospective_validation_score: Clinical evidence from any source
   - 0.0-0.3: No clinical data
   - 0.4-0.6: Observational data or related conditions
   - 0.7-1.0: Controlled trials or strong real-world evidence
4. novelty_score: Is this a novel repurposing hypothesis?
   - 0.0-0.3: Already well-established
   - 0.4-0.6: Some prior investigation
   - 0.7-1.0: Novel hypothesis

Calculate overall score as:
overall score = (0.30 × mechanistic_plausibility_score) + (0.25 ×
↪  literature_support_score) + (0.25 × retrospective_validation_score)
↪  + (0.20 × novelty_score)

Classify as:
- INDICATION: Evidence suggests potential therapeutic benefit
↪  (overall_score ≥ 0.5)
- CONTRAINDICATION: Evidence suggests harm or worsening of disease
- NONE: Insufficient evidence (overall_score < 0.5, no harm signals)

Evidence categories:
- MECHANISTIC: Strong mechanistic connection, even without direct trials
- RELATED_CONDITION: Evidence from related diseases
- DIRECT: Direct evidence for this specific drug-disease combination
- NO_STRONG_EVIDENCE: Minimal or no relevant evidence

Return ONLY valid JSON (no markdown, no explanation):
{
  "drug": "{drug}",
  "scores": {
```

```
    "Mechanistic_plausibility_score": 0.0,
    "Literature_support_score": 0.0,
    "Retrospective_validation_score": 0.0,
    "Novelty_score": 0.0
  },
  "overall_score": 0.0,
  "confidence_level": "High|Medium|Low|Very Low",
  "evidence_category":
  ↪  "MECHANISTIC|RELATED_CONDITION|DIRECT|NO_STRONG_EVIDENCE",
  "clinical_recommendation": "INDICATION|CONTRAINDICATION|NONE",
  "recommendation_reasoning": "Brief explanation of
  ↪  indication/contraindication/none classification",
  "summary": "2-3 sentences explaining the repurposing potential based
  ↪  on mechanisms and evidence quality"
}
```

Figure A2: Prompt used for generating the LLM-based feasibility report for a putative candidate drug against a disease. Placeholders in braces are filled in for each drug–disease pair.

A.7 Bubble Plot for Top 50 Nominations of Case-Study Deneurogenerative Diseases

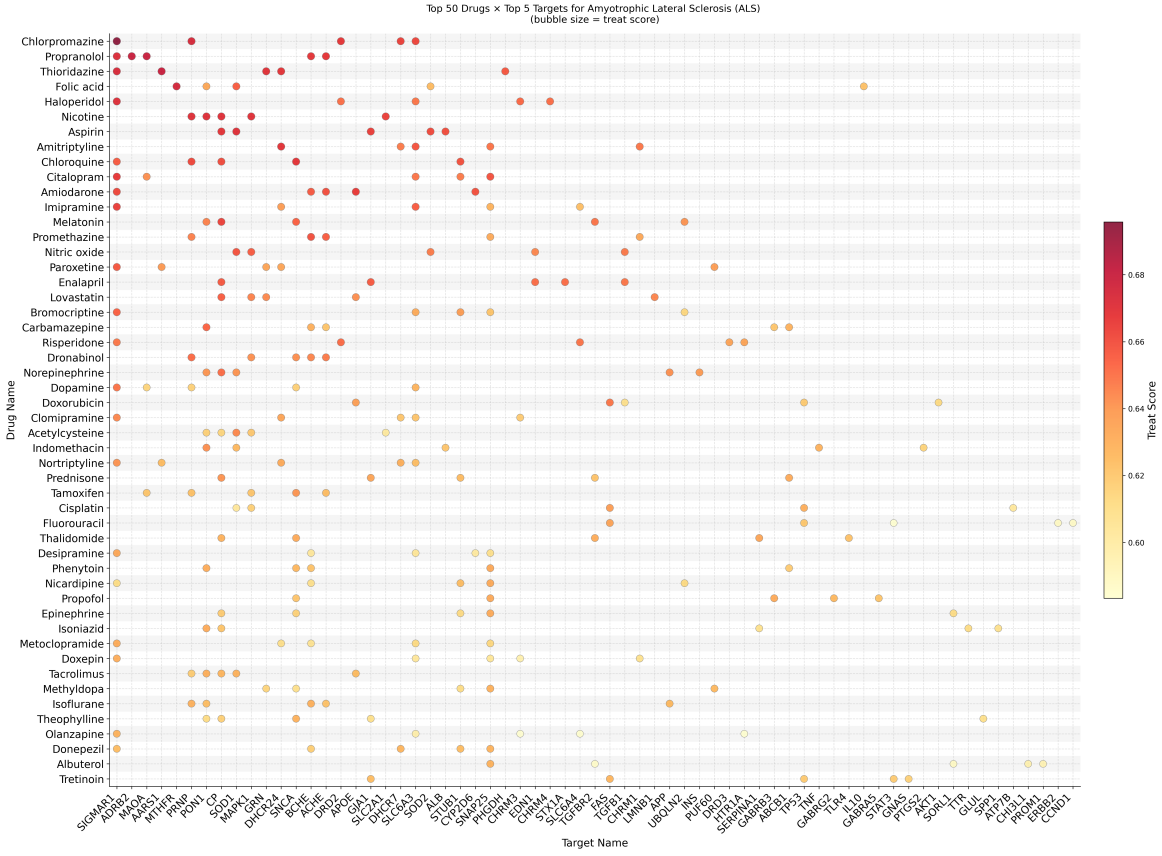


Figure A3: Top 50 FDA approved drugs against Amyotrophic Lateral Sclerosis (ALS) with top 5 targets predicted association per drugs using curated EveryCure (EC) drug list. Each bubble shows the associated target and intensity of the color shows the predicted treat score by the DTD' triplet model for drug-target-disease association.

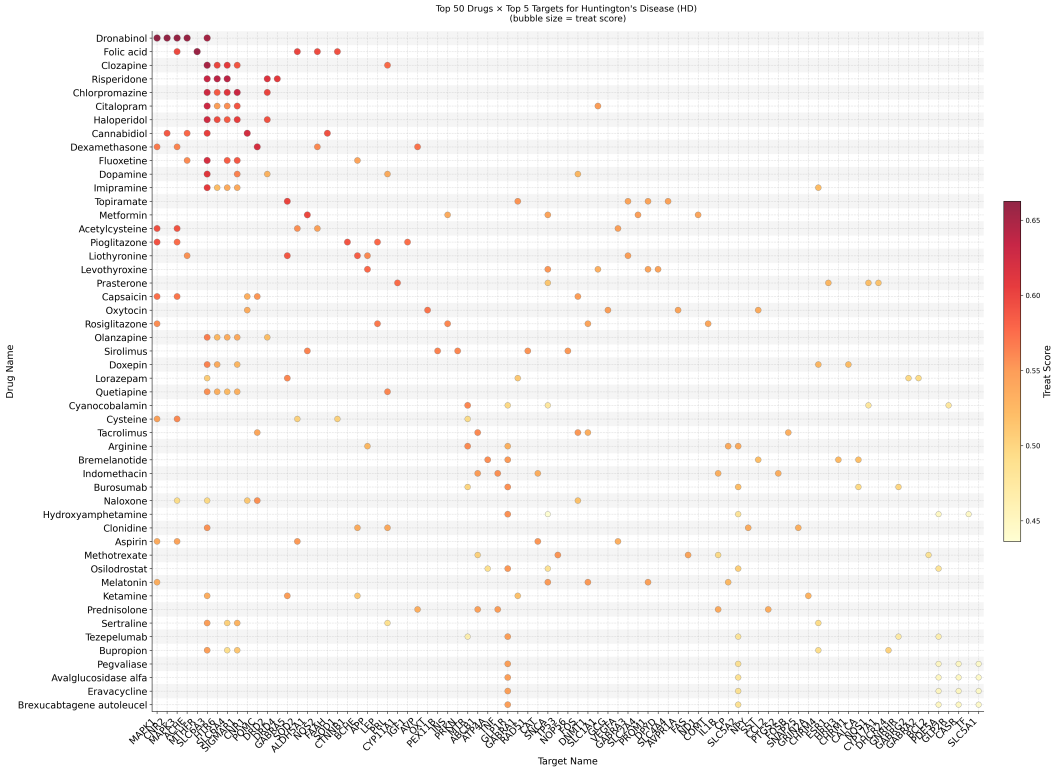


Figure A4: Top 50 FDA approved drugs against Huntington's Disease (HD) with top 5 targets predicted association per drugs using curated EveryCure (EC) drug list. Each bubble shows the associated target and intensity of the color shows the predicted treat score by the DTD' triplet model for drug-target-disease association.