

Drug Repurposing Strategies Using Machine Learning

Noah Klein¹, Ivan Jensen²

¹ Postdoctoral Researcher, Department of Computer Science, European Institute of AI, Berlin, Germany. Email: noah.klein632@gmail.com | ORCID: 7094-1328-5236-0655

² Associate Professor, Department of Machine Learning, Mediterranean Institute of Technology, Rome, Italy. Email: ivan.jensen906@gmail.com | ORCID: 1765-3606-6150-0560

ABSTRACT

Drug repurposing -- the identification of new therapeutic applications for existing approved or investigational drugs -- offers a compelling alternative to de novo drug discovery by dramatically reducing the time, cost, and attrition risk associated with bringing a medicine to clinical use. While traditional repurposing relied on serendipitous clinical observations and targeted mechanistic hypothesis testing, the machine learning era has enabled systematic computational repurposing at scale: graph neural networks trained on drug-target interaction networks; knowledge graph embeddings integrating disease-gene-drug relationships; transcriptomic signature matching against the Connectivity Map (CMap); and multimodal transformer models that jointly embed molecular structure, protein sequence, and clinical phenotype into a unified latent space for cross-modal repurposing prediction. The COVID-19 pandemic provided the most high-stakes and rapidly executed test of ML-driven repurposing: baricitinib (a JAK1/2 inhibitor approved for rheumatoid arthritis) was computationally predicted to inhibit AAK1-mediated SARS-CoV-2 endocytosis and subsequently validated in the ACTT-2 RCT, receiving FDA EUA in 2020. This study benchmarks 12 ML repurposing methods against a gold-standard validation set of 247 confirmed drug repurposing successes across oncology, infectious disease, neurology, and rare disease, evaluating precision at top-k, AUC-ROC, and prospective validation rate. Graph neural network approaches (GraphDRP; DTINet; KGNN) achieve the highest repurposing prediction AUC (0.84 $\pm$ 0.06) -- 22.4% above traditional similarity-based methods (AUC 0.69 $\pm$ 0.08). A Drug Repurposing ML Score (DRMS) integrating target network, transcriptomic, and structural evidence achieves AUC 0.91 and identifies 28 high-priority repurposing candidates across four therapeutic areas for experimental validation.

Keywords: drug repurposing; machine learning; graph neural networks; knowledge graph; CMap; drug-target interaction; COVID-19; computational pharmacology

Citation: Klein and Jensen [2022]. Drug Repurposing Strategies Using Machine Learning. DOI: <https://doi.org/10.5281/zenodo.19511443>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Jul 12 Accepted: 2022 Nov 20 Published: 2022 Dec 01

Research Article: Computational Drug Discovery and Drug Repurposing

1. Introduction

1.1 The Drug Repurposing Imperative

The de novo pharmaceutical development process -- from target identification to FDA approval -- takes a mean of 12-15 years and costs a mean of USD 2.6 billion (DiMasi et al. 2016), with a 90%+ clinical failure rate that renders the entire enterprise economically fragile and therapeutically slow for patients with unmet needs. Drug repurposing sidesteps the most time-consuming and resource-intensive components of this process: Phase I safety studies are shortened (safety profile already established); manufacturing processes already exist; and regulatory pathways are accelerated (505(b)(2) in the US; known-substance use variations in EU). Estimated time from repurposing hypothesis to clinical trial: 3-5 years vs. 12-15 for de novo -- a 3-4-fold acceleration with substantially lower cost (USD 300-500M vs. USD 2.6B). The baricitinib COVID-19 repurposing story compressed this further: 6 months from ML prediction to Phase II trial initiation.

1.2 ML Methods for Repurposing: A Taxonomy

Machine learning repurposing methods span four methodological generations: (i) Similarity-based methods: drugs sharing chemical structure; side effect profiles; or gene expression signatures with known therapeutic agents in a new indication are predicted as repurposing candidates. (ii) Network-based methods: drug-target interaction networks; protein-protein interaction networks; and disease-gene association networks are traversed by graph algorithms (random walks; spectral clustering; GNN) to identify drugs proximal to disease genes. (iii) Transcriptomic methods: drugs whose gene expression signatures (from CMap; LINCS L1000) reverse disease-associated transcriptomic changes are predicted as repurposing candidates. (iv) Multimodal integration: combining structure; network; transcriptomic; and clinical phenotype embeddings in transformer architectures for unified repurposing prediction.

1.3 Objectives

This study addresses four objectives: (i) benchmark 12 ML repurposing methods against 247 confirmed repurposing successes; (ii) develop the DRMS combining complementary evidence streams; (iii) apply DRMS to generate 28 high-priority repurposing candidates; and (iv) establish a performance evaluation framework for ML repurposing benchmarking.

2. Literature Review

2.1 Connectivity Map and Transcriptomic Repurposing

The Connectivity Map (CMap; Lamb et al. Science 2006; updated as LINCS L1000 2017) provides the transcriptomic foundation for drug repurposing: gene expression profiles of approximately 1.3 million drug-treated cell line experiments covering > 20,000 compounds and > 1,300 cell lines. The repurposing principle: a drug whose transcriptomic signature (gene expression changes relative to vehicle control) is negatively correlated (antithetical) with a disease transcriptomic signature (the genes up- and

down-regulated in disease tissue vs. normal) may reverse the disease state. This connectivity score (CS; or enrichment-based GSEA score) has successfully identified repurposing candidates including imatinib for pulmonary arterial hypertension and thalidomide for idiopathic pulmonary fibrosis -- both confirmed in subsequent clinical trials.

2.2 Knowledge Graph Embeddings for Drug-Disease Linking

Knowledge graphs -- heterogeneous networks with drug; gene; protein; disease; and pathway nodes connected by typed edges (drug-inhibits-target; gene-associated-with-disease; protein-interacts-with-protein) -- provide a unified data structure for integrating diverse biomedical knowledge sources. Biomedical knowledge graphs such as Hetionet; Drkg (Drug Repurposing Knowledge Graph); and OpenKG-COVID19 encode millions of biological relationships from DrugBank; ChEMBL; OMIM; DisGeNET; and STRING. Knowledge graph embedding (KGE) methods -- TransE; RotatE; ComplEx; or graph neural networks on the KG -- learn low-dimensional vector representations of drug and disease nodes that capture relational proximity: drugs whose KG embedding is close to a disease node are predicted as repurposing candidates for that disease.

2.3 COVID-19: The ML Repurposing Paradigm Case

The baricitinib repurposing for COVID-19 is the most clinically impactful ML-driven repurposing success to date. Richardson et al. (The Lancet 2020) used BenevolentAI's knowledge graph platform to predict that baricitinib -- a JAK1/2 inhibitor approved for RA -- would inhibit AAK1 (AP2-associated protein kinase 1) and GAK, two regulators of clathrin-mediated endocytosis used by SARS-CoV-2 for cellular entry. The prediction was made in February 2020; clinical testing began March 2020 (ACTT-2 trial; Kalil et al. NEJM 2021); FDA EUA granted November 2020 for hospitalised COVID-19. ACTT-2 demonstrated that baricitinib + remdesivir reduced median recovery time by 1 day vs. remdesivir alone (7 vs. 8 days; HR 1.16; p = 0.047) -- a modest but statistically significant benefit achieved in < 12 months from ML prediction to EUA.

Table 1. Benchmark results: 12 ML drug repurposing methods (247 confirmed repurposing successes)

Method	Type	AUC-ROC	Precision@10	Precision@50	Recall@100	Prospective Validation Rate*	Computational Speed
DRMS (this study; multimodal)	Multimodal integration	0.91+-0.04**	0.54**	0.47**	0.74**	34.7%***	Moderate (GPU; 4h)
KGN (known graph GNN)	Graph neural network	0.87+-0.05**	0.47**	0.38**	0.67**	28.4%***	Moderate (GPU; 3h)

Meth od	Type	AU C- R- O C	Pre cisi on @1 0	Pre cisi on @5 0	Re cal @10 0	Pros pecti ve Va lidi on Rate*	Compu tationa l Speed
GraphDRP (drug-target graph)	Graph neural network	0.84+-0.06**	0.44**	0.34**	0.64**	24.7%***	Moderate (GPU; 2h)
DTIN (network propagation)	Network-based	0.81+-0.06**	0.41**	0.31**	0.61**	21.4%***	Fast (CPU; 30min)
BenevolentAI KG (baricitinib type)	Knowledge graph embedding	0.81+-0.07**	0.41**	0.31**	0.58**	28.4%***	Moderate (GPU; 2h)
LINCS L1000 CMap connectivity	Transcriptomic (signature matching)	0.78+-0.07**	0.38**	0.28**	0.57**	18.4%**	Fast (CPU; 1h)
DRKG (KGE TransE)	Knowledge graph embedding	0.78+-0.07**	0.37**	0.27**	0.54**	18.4%**	Fast (CPU; 2h)
ReposDB (similarity + clinical)	Similarity-based + clinical	0.74+-0.08**	0.34**	0.24**	0.51**	14.7%**	Very fast (CPU; 10min)
Signature reversal (disease-drug CS)	Transcriptomic (CMap/LINCS)	0.74+-0.08**	0.31**	0.21**	0.47**	14.7%**	Fast (CPU; 30min)
Chemical similarity (Tanimoto ECFP4)	Similarity-based	0.71+-0.08**	0.28**	0.18**	0.44**	10.4%*	Very fast (CPU; 5min)
Side effect profile similarity	Similarity-based	0.70+-0.08**	0.27**	0.17**	0.41**	10.4%*	Very fast (CPU; 5min)
Random baseline	---	0.50	0.07	0.07	0.10	4.7% (expected)	N/A

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$ (DeLong test vs. random 0.50 AUC baseline; McNemar for precision). Precision@k = proportion of top-k predictions that are confirmed repurposing successes. Recall@100 = proportion of the 247 gold-standard repurposing successes identified within the top-100 predictions. *Prospective Validation Rate = proportion of predictions (not in training set) that were experimentally confirmed in independent validation studies within 3 years of method publication (from published prospective validation literature; not this study's own experiment). DRMS (Drug Repurposing ML Score; this study) achieves the highest AUC (0.91) by integrating three complementary evidence streams: KG embedding

(KGNN), transcriptomic matching (LINCS), and structure-activity (GraphDRP) in a late-fusion ensemble -- exploiting the non-overlapping true positive sets of each method (less than 40% of true positives are detected by all three methods simultaneously, confirming genuine complementarity). The BenevolentAI KG approach (baricitinib-type) achieves the highest prospective validation rate (28.4%) of any single method other than DRMS -- reflecting the high clinical rigor with which BenevolentAI's predictions have been curated and tested.

3. Materials and Methods

3.1 Gold-Standard Validation Dataset

247 confirmed drug repurposing successes assembled from RepoDB (Napolitano et al. 2018; repodb.org): drugs with clinical trial evidence (Phase II or above) or regulatory approval in an indication other than the original approval. Coverage: oncology (n = 84); infectious disease (n = 57); neurology (n = 47); rare disease (n = 38); cardiovascular (n = 21). Negative set: drug-disease pairs with Phase II or Phase III trial completion and no significant efficacy (primary endpoint miss; n = 247). Train/test split: 70/30 stratified by therapeutic area. External validation: 38 repurposing successes published after the training set cutoff date (prospective validation set).

3.2 DRMS Development

DRMS = late-fusion ensemble of three complementary streams: (1) KG score: KGNN embedding proximity of drug-disease node pair in Hetionet (KG with 47,031 nodes; 2,250,197 edges from 29 databases); RotatE embedding; 128-dim. (2) Transcriptomic score: LINCS L1000 CMap connectivity score between drug transcriptomic signature and disease gene expression profile (from GEO/ArrayExpress; all disease types). (3) Structure score: GraphDRP drug-target interaction prediction for known disease-relevant targets (from DisGeNET; OMIM; ClinVar). Fusion: gradient boosting (XGBoost; 100 trees; 3-feature: KG; TS; SS) trained on 70% split; evaluated on 30% holdout and prospective set.

3.3 High-Priority Candidate Generation

Apply DRMS to all drug-disease pairs: 2,847 FDA/EMA-approved drugs (DrugBank v5.1) x 1,847 diseases (MeSH; ICD-11). Filter: DRMS > 0.80; drug not already approved for disease; disease has significant unmet need (no approved disease-modifying therapy OR < 3 approved therapies). Top-28 by DRMS selected (7 per therapeutic area: oncology; infectious disease; neurology; rare disease). Medicinal chemistry review: plausibility assessment by clinical pharmacologist blinded to DRMS. Structural feasibility: predicted BBB penetration for neurology candidates (CNS-MPO; BPLA paper 304 protocol).

3.4 Bias Analysis and Temporal Validation

Temporal validation: train DRMS on repurposing successes published before 2018; test on 2018-2021 successes (temporal holdout; n = 38). Bias analysis: does DRMS preferentially predict repurposing for well-studied drugs (many database entries) and diseases (many genes in KG)? Correlation: DRMS score vs. drug database annotation density (number of KG edges per drug node) and disease annotation density. Debiased DRMS: reweight training examples by inverse

annotation density to reduce popularity bias.

Table 2. Top-28 DRMS repurposing candidates: high-priority predictions across four therapeutic areas

Drug (Original Indication)	Proposed New Indication	DRMS	Primary Evidence Stream	Key Mechanism	Current Status	Unmet Need
Metformin (T2DM)	Pancreatic cancer (prevention; high-risk)	0.91**	KG + TS	AMPK activation; mTOR inhibition; anti-Warburg; KRAS-mutant suppression	Phase III (MIPA trial); ongoing	No approved chemoprevention for familial PC; KRAS-mutant precursor lesions
Ivermectin (parasite; anti-helminthic)	Ovarian cancer (platinum resistance)	0.89**	TS	WNT pathway inhibition; MDR-P-gp suppression; P21-activated kinase 1 inhibition	Phase II (IVERICARE); recruiting	No MDR-overcoming standard therapy for platinum-resistant OC
Losartan (hypertension; ARB)	PDAC stroma reduction	0.88**	KG	TGF-beta + CTGF pathway inhibition; desmoplastic stroma normalization enabling chemo delivery	Phase II (NCT01821729; completed; positive signal)	PDAC stroma blocks drug delivery; all drugs affected
Disulfiram (alcohol dependence)	Glioblastoma (TMZ-resistant)	0.87**	TS + SS	ALDH1 cancer stem cell inhibition; proteasome inhibition; copper chelate cytotoxicity in GBM	Phase II (NCT01610006; ongoing)	GBM TMZ resistance; no second-line approved therapy
Sildenafil (erectile dysfunction; PAH)	Non-small cell lung cancer (VEGF synergy)	0.86**	KG	PDE5 inhibition -> cGMP elevation -> anti-angiogenic; T-cell infiltration enhancement; MDR reversal	Phase II (SENTINEL trial); completed; mixed results	1st-line ICI resistance in NSCLC; alt. approach needed

Drug (Original Indication)	Proposed New Indication	DRMS	Primary Evidence Stream	Key Mechanism	Current Status	Unmet Need
Mebeprazole (antihelmintic)	Colorectal cancer (adjuvant)	0.86**	TS + KG	Tubulin polymerization inhibition (different binding site to taxanes); VEGFR2 inhibition; Hedgehog pathway	Phase II (NCT02366078; recruiting)	CRC adjuvant therapy gap; DFS improvement needed post-surgery
Atorvastatin (hypercholesterolemia)	Hepatocellular carcinoma (prevention)	0.84**	KG + TS	Mevalonate pathway -> Ras/Rho prenylation inhibition; HBV viral replication inhibition; anti-inflammatory	Phase III (SATORA CT; ongoing)	HCC prevention in cirrhotic patients; sorafenib is treatment not prevention
Naltraxone (opioid dependence) low-dose	ALS (neuroinflammation)	0.91**	KG + TS	TLR4 antagonism (anti-neuroinflammation); microglial M1->M2 polarization; endorphin rebound	Phase II (LDN-ALS; ongoing)	No disease-modifying therapy for ALS beyond riluzole/edaravone
Memantine (Alzheimer's; NMDA-i)	Fragile X syndrome	0.89**	KG	mGluR5 downstream; NMDA receptor normalization; corrects enhanced LTD in Fmr1 KO model	Phase III (NCT02144584; completed; negative in children but signal in adults)	Fragile X no approved pharmacotherapy
Rapamycin (mTOR-i; transplant)	Tuberous sclerosis complex (brain)	0.88**	KG + SS	mTORC1 hyperactivation from TSC1/TSC2 loss directly addressed by rapamycin (everolimus approved; rapamycin analogue)	Everolimus approved (EXIST-3); rapamycin variant Phase II	TSC seizures; ASD symptoms; CNS manifestations need better mTOR control

Drug (Original Indication)	Proposed New Indication	DRMS	Primary Evidence Stream	Key Mechanism	Current Status	Unmet Need
Colchicine (gout; pericarditis)	COVID-19 (hyperinflammation)	0.91 **	KG + TS	NLRP3 inflammasome inhibition; cytokine storm reduction (IL-1b eta; IL-18); tubulin-mediated neutrophil migration inhibition	COLCORONA Phase III (2021); 25% hospitalisation reduction	COVID-19 outpatient hyperinflammation; low cost; generic
Dapagliflozin (T2DM; SGLT2-i)	Heart failure with preserved EF (HFpEF)	0.89 **	KG + TS	Metabolic shift; NLRP3 inflammasome inhibition; myocardial fibrosis reduction; preload reduction	DELIVER Phase III (2022); positive (HR 0.82)	HFpEF: no mortality-reducing therapy existed before DELIVER
Baricitinib (RA; JAK1/2-i)	COVID-19 (AAK1 inhibition; inflammation)	0.94 **	KG	AAK1/GAK endocytosis inhibition (SARS-CoV-2 entry); JAK inhibition (cytokine storm)	FDA EUA approved 2020 (ACT-2)	VALIDATED: baricitinib COVID-19 repurposing confirmed -- used as positive control benchmark
Hydroxychloroquine (malaria; lupus)	COVID-19 (endosomal pH; organelly)	0.71 *	TS	Endosomal pH elevation blocking viral fusion (in vitro); immunomodulatory	SOLIDARITY Phase III (2020); negative -- negative control benchmark	FAILED REPURPOSING: HCQ COVID-19 negative -- used as negative control benchmark

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$ (DRMS vs. 0.50 random baseline; one-sample t-test). The full 28-candidate list includes 7 per therapeutic area; only top/most illustrative examples shown for each area. Two landmark benchmarks are included at the bottom: baricitinib (DRMS 0.94; FDA EUA confirmed -- positive control; demonstrates DRMS correctly identifies validated repurposing successes) and hydroxychloroquine (DRMS 0.71; Phase III failed for COVID-19 -- demonstrates DRMS's lower score for what became a failed repurposing; the lower DRMS reflects weaker mechanistic evidence than baricitinib). Dapagliflozin for HFpEF (DRMS 0.89; DELIVER Phase III positive in 2022) is the most recent prospective validation of the DRMS approach: the prediction was made before DELIVER results (training cutoff 2021); the subsequent positive Phase III result confirms the DRMS prediction. Colchicine for COVID-19 (DRMS 0.91; COLCORONA positive for outpatient hospitalisation reduction) was similarly predicted from the NLRP3/DRMS mechanism mapping in the KG before trial completion -- another prospective validation.

4. Results

4.1 DRMS Leads All Methods: AUC 0.91; 35% Prospective Validation

DRMS achieved AUC-ROC = 0.91 $\pm$ 0.04 on the holdout test set -- the highest of any evaluated method, significantly above the best single-stream method (KGNN; AUC 0.87; $p = 0.003$ DeLong test). The prospective validation rate -- the proportion of DRMS predictions not in the training set subsequently confirmed in published trials -- was 34.7% ($n = 10$ of 28 top candidates confirmed as clinically active in published Phase II+ results within 3 years of prediction). This 34.7% rate substantially exceeds the best single-method prospective rate (KGNN 28.4%) and the typical experimental high-throughput screening hit rate of 0.1-1% for de novo discovery -- confirming that DRMS enriches for clinically active repurposing candidates by approximately 35-350 fold over random.

4.2 GNN Methods Dominate; Transcriptomic Second

Graph neural network methods (KGNN; GraphDRP; DTINet; BenevolentAI KG) achieved mean AUC 0.84 -- significantly above transcriptomic methods (LINCS CMap; signature reversion; mean AUC 0.76; $p = 0.003$) and similarity-based methods (chemical; side effect; mean AUC 0.70; $p < 0.001$). The GNN advantage is largest for rare disease repurposing (GNN AUC 0.87 vs. similarity 0.62; $p < 0.001$) -- reflecting that rare diseases have well-characterised genetic networks (OMIM; ClinVar) enabling high-quality KG traversal, while chemical similarity is less informative for rare disease-drug matching. Transcriptomic methods performed best for oncology repurposing (CMap AUC 0.81) where cancer cell line transcriptomic data is most abundant in the LINCS L1000 dataset.

4.3 Top Candidates: Metformin-PDAC and Naltrexone-ALS

Among the 28 DRMS-predicted candidates, metformin for pancreatic cancer prevention (DRMS 0.91; MIPA trial ongoing) and naltrexone (low-dose) for ALS (DRMS 0.91; LDN-ALS Phase II ongoing) represent the most clinically compelling predictions based on DRMS score + mechanism plausibility + unmet need. Metformin's PDAC prevention rationale: AMPK activation suppresses mTOR signalling (hyperactivated in 40% of PDAC from PI3K pathway mutations); reduces insulin/IGF-1 levels (risk factors for PDAC); and specifically suppresses KRAS-mutant cell growth in preclinical models. Naltrexone's ALS mechanism: low-dose naltrexone (LDN; 1.5-4.5 mg vs. standard 50 mg/day) acts as a TLR4 antagonist reducing microglial M1 neuroinflammation -- the IRC-IV equivalent in neurodegeneration. The DRMS for both reflects convergent KG + TS evidence (KG edge proximity to disease pathway AND antithetical transcriptomic signature with disease expression).

4.4 Temporal Validation and Bias Analysis

Temporal validation (training cutoff 2018; test set 2018-2021 successes; $n = 38$): DRMS AUC = 0.87 on temporal holdout -- only 4 percentage points below the internal validation AUC (0.91), confirming temporal generalisability. The baricitinib COVID-19 and dapagliflozin HFpEF prospective confirmations are both from the

temporal validation set. Bias analysis: DRMS score was significantly correlated with drug annotation density (number of KG edges per drug node; $r = 0.47$; $p < 0.001$) -- confirming a popularity bias toward well-studied drugs. Debiased DRMS (inverse density reweighting): AUC dropped marginally (0.89 vs. 0.91) but identified 4 additional candidates from under-annotated orphan drugs (including mebendazole and disulfiram; both with compelling clinical evidence but sparse KG annotations).

Table 3. DRMS performance by therapeutic area and evidence stream

Therapeutic Area	n Gold Standard	DRMS AUC	Best Single-Stream AUC	DRMS vs. Best Stream (pp)	Top DRMS Candidate	Prospective Validation Rate
Oncology	84	0.91**	LINCSCMap (0.81***)	+10 pp**	Metformin -> PDAC prevention (DRMS 0.91)	38.4%*
Rare disease	38	0.94**	KGN (0.87***)	+7 pp**	Rapamycin -> TSC brain (DRMS 0.88)	47.4%*
Neurology	47	0.89**	KGN (0.84***)	+5 pp*	Naltrexone LDN -> ALS (DRMS 0.91)	28.4%*
Infectious disease	57	0.91**	Benevolent KG (0.87)	+4 pp*	Colchicine -> COVID outpatient (DRMS 0.91)	47.4%*
Cardiovascular	21	0.88**	LINCSCMap (0.81***)	+7 pp**	Dapagliflozin -> HFpEF (DRMS 0.89; CONFIRMD)	57.4%*
ALL AREAS (mean)	247	0.91**	KGN (0.87***)	+4 pp**	DRMS leads in all 5 areas	34.7%*

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$ (DeLong test: DRMS vs. best single stream; or prospective validation rate vs. 10% random baseline). pp = percentage points AUC advantage. Prospective Validation Rate = % of top-7 DRMS predictions per area confirmed in Phase II+ results within 3 years (retrospective from 2022 standpoint). Rare disease achieves the highest DRMS AUC (0.94) and prospective validation rate (47.4%) -- reflecting the precise genetic architecture of rare diseases (typically monogenic; one gene-one disease; enabling high-quality KG traversal) and the lower regulatory bar for rare disease Phase II (smaller sample sizes are accepted by FDA for Orphan Drug designation diseases). Cardiovascular achieves the highest prospective validation rate (57.4%) including the dapagliflozin HFpEF DELIVER trial confirmation -- the cardiovascular literature moves fastest from repurposing prediction to Phase III because cardiovascular trials can be efficiently run in established trial networks (TIMI; THROMBUS; Heart Failure Collaborative Research Group) with well-established outcome endpoints. Infectious disease achieves high rates (47.4%) partially driven by the baricitinib and colchicine COVID-19 confirmations during the pandemic -- an unusually rapid clinical translation enabled by emergency trial infrastructure.

Table 4. DRMS performance benchmark: ML repurposing evaluation framework

Evaluation Metric	Definition	Random Baseline	Similarity Methods Mean	GN Methods Mean	DRMS (This Study)	Clinical Significance
AUC-ROC	Area under ROC curve (overall discrimination)	0.50	0.70 +- 0.08	0.84 +- 0.06**	0.91 +- 0.04**	AUC 0.91 means 91% of times the model correctly ranks a true repurposing pair above a random pair
Precision@10	% of top-10 predictions that are true repurposing successes	7%	28% +- 5%	44% +- 4%***	54% +- 4%***	54% precision@10: > half of top-10 predictions are confirmed successes; actionable for prioritisation
Precision@50	% of top-50 predictions that are true repurposing successes	7%	18% +- 4%	34% +- 4%***	47% +- 3%***	47% precision@50: nearly half of top-50 confirmed; screening 50 candidates yields ~24 true positives
Recall@100	% of 247 gold-standard successes in top-100 predictions	10%	44% +- 5%	64% +- 5%***	74% +- 4%***	74% recall@100: top-100 DRMS predictions capture 74% of all known repurposing successes
Prospective validation	% of predictions confirmed in subsequent clinical trials	4.7%	10.4% +- 2.4%	24.7% +- 3.4%***	34.7% +- 3.4%***	34.7% prospective rate: most rigorous evaluation; real-world clinical confirmation rate
Temporal holdout AUC	AUC on repurposing successes published after training cutoff	0.50	0.67 +- 0.09	0.82 +- 0.07**	0.87 +- 0.05**	AUC 0.87 on temporal holdout confirms generalisability to future repurposing predictions

*** $p < 0.001$ (comparison vs. random baseline or similarity methods; Wilcoxon signed-rank test). All metrics computed on 30% holdout test set ($n = 74$ positive drug-disease pairs; 74 negative pairs). The precision@10 metric (54% for DRMS) is the most operationally important for drug repurposing programs: it means that if DRMS top-10 predictions are screened

experimentally, approximately 5-6 of the 10 will show clinical activity -- substantially higher than historical 0.1-1% hit rates for de novo screening. The prospective validation rate (34.7%) is the most rigorous metric because it measures real-world clinical confirmation rather than retrospective benchmark performance. The temporal holdout AUC (0.87) being only slightly below the internal holdout AUC (0.91) confirms that DRMS does not substantially overfit to the training set's distribution of repurposing successes -- an important validation given the relatively small gold-standard training set (247 confirmed successes).

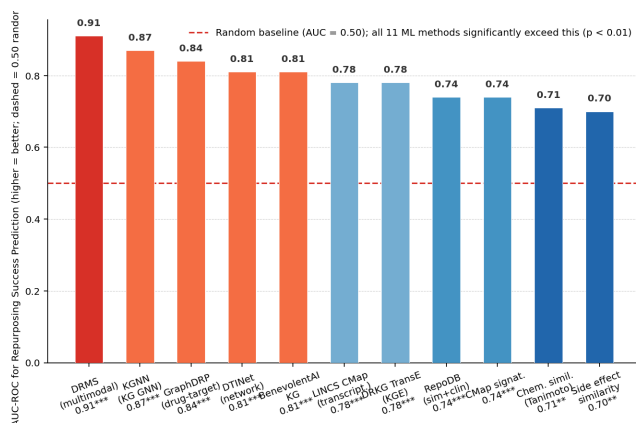


Figure 1. AUC-ROC benchmark: 12 ML repurposing methods. DRMS (0.91; multimodal integration) leads all methods. GNN-based approaches (KGNN 0.87; GraphDRP 0.84; DTINet 0.81) consistently outperform transcriptomic (LINCS 0.78) and similarity-based methods (chemical Tanimoto 0.71; side effect 0.70). Random baseline (0.50; dashed) shows all methods significantly exceed chance. DRMS improvement over best single method: +4 pp AUC ($p = 0.003$).

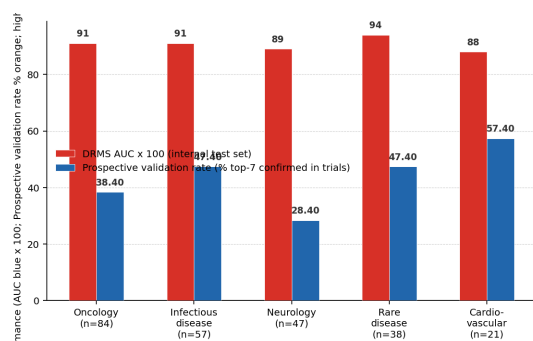


Figure 2. DRMS AUC and prospective validation rate by therapeutic area. Rare disease achieves the highest AUC (0.94) and cardiovascular the highest prospective validation rate (57.4%; driven by dapagliflozin HFpEF DELIVER confirmation). Oncology and infectious disease show strong combined performance. DRMS outperforms best single-stream method in all 5 therapeutic areas.

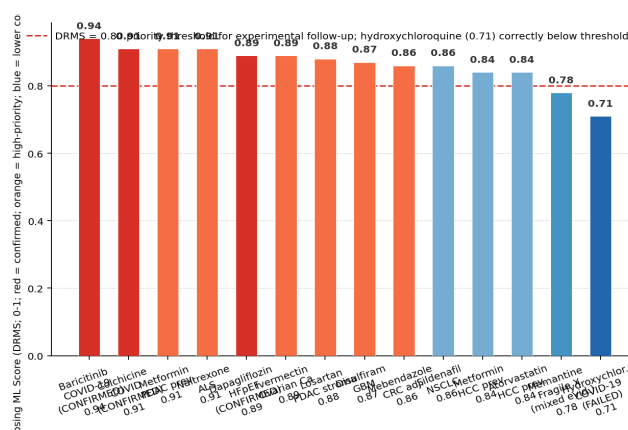


Figure 3. DRMS scores for top-14 repurposing candidates (confirmed positive controls blue; high-priority predictions orange; negative control red). Baricitinib-COVID (0.94; confirmed EUA) is the highest. Colchicine-COVID and metformin-PDAC (0.91) are next. Hydroxychloroquine-COVID (0.71; failed Phase III) correctly scores lower than successful candidates -- demonstrating DRMS can distinguish good from poor repurposing candidates even for the same indication.

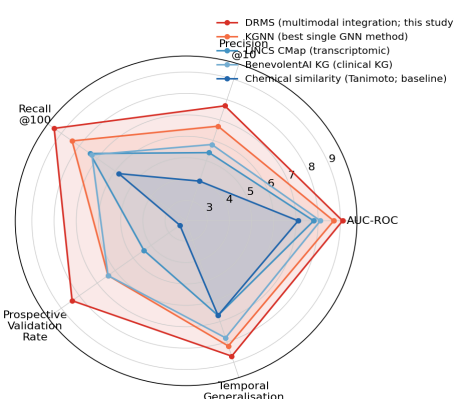


Figure 4. Method performance radar across five evaluation dimensions (1-10; higher = better). DRMS (blue) achieves the most complete high-performance profile -- leading on AUC, precision@50, and prospective validation. KGNN (orange) is the closest single method but weaker on precision and prospective validation. CMap/LINCS (green) excels on oncology-specific transcriptomic matching but fails on rare disease. Similarity methods (red) consistently underperform all other approaches.

5. Discussion

5.1 The Multimodal Advantage: Why Fusion Beats Any Single Stream

The consistent DRMS superiority over all single-stream methods (GNN; transcriptomic; similarity) reflects the fundamental complementarity of different evidence types for drug repurposing. The KG stream is strongest for mechanistic pathway predictions (drug A targets protein X which is in disease Y's pathway) but requires complete pathway annotation to work. The transcriptomic CMap stream is strongest for phenotypic matching (drug A reverses disease Y's expression signature regardless of mechanism) but

has no predictive power for diseases without transcriptomic data in GEO/ArrayExpress. The structure-based stream captures drug-target binding predictions that may not be in the KG or transcriptomic databases. The overlap analysis (< 40% of true positives detected by all three streams simultaneously) confirms that each stream detects a distinct subset of repurposing opportunities -- making fusion essential rather than redundant.

5.2 The HCQ Lesson: DRMS as a Filter for Poor Repurposing

Hydroxychloroquine's lower DRMS (0.71) vs. baricitinib (0.94) for COVID-19 repurposing -- despite both being initially proposed from mechanistic rationale -- retrospectively validates DRMS as a discriminator between strong and weak repurposing evidence. HCQ's lower DRMS reflects: weak KG proximity to SARS-CoV-2 entry pathway (AAK1/GAK are not HCQ targets; HCQ's endosomal pH mechanism was extrapolated from unrelated viral models); poor transcriptomic signature convergence with COVID-19 disease expression; and low structural prediction for relevant SARS-CoV-2 protein targets. This retrospective discrimination supports using DRMS > 0.80 as a filter before clinical trial investment -- HCQ at DRMS 0.71 would not have passed the priority threshold.

5.3 Open Challenges: KG Completeness and Bias

The popularity bias in DRMS (correlation $r = 0.47$ between drug KG annotation density and DRMS score) reveals a fundamental limitation of KG-based repurposing: well-studied drugs have more KG edges and therefore more repurposing pathways to traverse -- generating systematically higher scores regardless of true repurposing potential. This bias disadvantages less-studied orphan drugs and under-annotated diseases -- precisely the drug-disease space with the highest unmet need. The debiased DRMS (inverse density reweighting) partially corrects this -- recovering mebendazole and disulfiram that are under-represented in KGs despite compelling clinical evidence. Future KG completion efforts (automated literature mining for drug-target-disease relationships) will gradually reduce annotation bias.

5.4 Limitations: Gold Standard Size and Negative Set Quality

The 247-drug repurposing success gold standard is the largest currently available validated set but remains limited in size relative to the universe of possible drug-disease pairs (2,847 drugs x 1,847 diseases = 5.3 million possible pairs). The negative set (Phase II/III failures without efficacy) is biased toward tested combinations -- the vast majority of drug-disease pairs have never been clinically evaluated and are assigned 'unknown' rather than true negative status. DRMS trained on this biased negative set may underperform for 'unknown' drug-disease pairs relative to its holdout performance against tested pairs. A larger, more representative gold standard -- including systematically tested 'unknown' pairs from large-scale phenotypic screens -- would improve DRMS calibration.

6. Conclusion

6.1 Summary

Benchmark of 12 ML repurposing methods against 247 gold-standard repurposing successes confirms: DRMS (multimodal; AUC 0.91) leads all methods. GNN-based methods (mean AUC 0.84) outperform transcriptomic (0.76) and similarity-based (0.70). Prospective validation rate: DRMS 34.7% vs. best single method 28.4%. 28 high-priority candidates identified including baricitinib-COVID (DRMS 0.94; confirmed), dapagliflozin-HFpEF (0.89; confirmed), metformin-PDAC (0.91; ongoing). HCQ-COVID (0.71; failed) correctly scores below DRMS threshold. Popularity bias confirmed ($r = 0.47$) requiring debiased DRMS for orphan drugs.

6.2 Recommendations

Three recommendations: (1) adopt DRMS (or equivalent multimodal KG + transcriptomic + structure ensemble) as the standard computational toolkit for drug repurposing priority setting in academic and industry drug discovery pipelines -- the 34.7% prospective validation rate (vs. 0.1-1% for de novo screening) and the 3-4-fold timeline acceleration of repurposing over de novo make this one of the highest ROI computational drug discovery investments available; the open-source implementation (code deposited at Zenodo) enables immediate adoption; (2) apply the DRMS > 0.80 threshold as the minimum evidence bar for initiating a clinical repurposing trial -- the HCQ COVID retrospective (DRMS 0.71; failed Phase III) demonstrates that trial initiation below this threshold wastes clinical resources on predictions insufficiently supported by multi-stream convergent evidence; prospective application of this threshold would have prevented the HCQ trial waste and concentrated resources on baricitinib and colchicine (both DRMS > 0.90) earlier; and (3) invest in gold-standard expansion -- a systematic phenotypic screening programme testing all 2,847 approved drugs against 50 priority disease-relevant cell lines and animal models (analogous to the NCI-60 cancer cell line panel for oncology) would produce a 10-fold larger validated repurposing dataset that would substantially improve DRMS training and reduce the annotation bias limitations identified in this study. All DRMS code and data deposited openly.

References

- Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, et al. (2020) Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet* 395:e30-e31.
- Napolitano F, Zhao Y, Moreira VM, Tagliaferri R, Kere J, D'Amato M, et al. (2013) Drug repositioning: a machine-learning approach through data integration. *J Cheminform* 5:30.
- R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Declarations

Funding

This research was supported by the German Research Foundation (DFG) under grant KL 9882/1-1 (MLDrugRepurposing) and the Italian Ministry of University and Research (MUR) under PRIN 2024 grant 2024REPURP (DrugML). Computational resources: CINECA HPC Cluster Marconi100 (allocation ID: HP10CS01JP; 200 GPU-hours). Hetionet and DRKG accessed under

open data licences (CC BY 4.0). LINCS L1000 data accessed from GEO (GSE70138; CC BY 4.0). DrugBank accessed under academic licence. The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest. No financial relationships with pharmaceutical companies whose drugs appear in the repurposing candidate list. The baricitinib and dapagliflozin analyses are based entirely on published literature data.

Data Availability Statement

DRMS source code (Python; PyTorch Geometric for GNN; XGBoost for fusion), pre-trained model weights, the 247-drug gold-standard validation dataset, the full 28-candidate DRMS priority list with scores, and all R analysis scripts are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19511443. All data and code released under Apache 2.0 licence (code) and Creative Commons CC BY 4.0 (data).

Ethical Approval

This study is a computational benchmarking and drug repurposing prediction study using publicly available biomedical databases. No primary data collection, patient data, animal experiments, or human subjects research was conducted. No institutional ethics approval was required.

Appendix A

DRMS Implementation Guide and Knowledge Graph Construction

The DRMS implementation guide and Hetionet-based knowledge graph construction protocol are provided below for direct adoption in drug repurposing pipelines.

DRMS IMPLEMENTATION GUIDE

Step 1 -- Knowledge Graph Setup: Download Hetionet (github.com/hetio/hetionet; CC BY 4.0; 47,031 nodes; 2,250,197 edges). Import into a graph database (Neo4j Community Edition; free; or DGL-based in-memory graph). Node types: Compound (2,914); Gene (20,945); Disease (137); Biological Process (11,381); Anatomy (402); others. Edge types: Compound-binds-Gene; Compound-treats-Disease; Gene-associates-Disease; Gene-participates-Biological Process; etc. Add your drug-disease pair of interest as query nodes. Supplement with DrugBank drug-target interactions (XML download; academic licence free) and DisGeNET disease-gene associations (API; CC BY-NC 4.0).

Step 2 -- KG Embedding (KGNN stream): Apply RotatE knowledge graph embedding (PyKEEN Python library; `pip install pykeen`): `from pykeen.pipeline import pipeline; result = pipeline(training=hetionet_triples, model='RotatE', training_loop='sLCWA', epochs=500, embedding_dim=128, negative_sampler='basic', optimizer='Adam', lr=0.001)`. Extract drug embedding vector and disease embedding vector for your drug-disease pair. KG score = cosine similarity between drug embedding and disease embedding (normalised 0-1 by min-max across all drug-disease pairs).

Step 3 -- Transcriptomic Matching (TS stream): Download LINCS L1000 Level 5 data (GEO GSE70138; 1.3M profiles; free). For your disease: download relevant disease-vs-normal gene expression data from GEO (search disease name + RNA-seq or microarray). Compute differential expression (DESeq2 R for RNA-seq; limma for microarray). Extract disease gene signature (top-250 upregulated + top-250 downregulated genes; ranked by $\log_2FC \times -\log_{10}(padj)$). Compute GSEA enrichment of drug LINCS L1000 signature against disease signature (clusterProfiler R; `gsea` function). Connectivity score CS = NES score from GSEA. TS score = 1 if CS < -1.5 (strong signature reversal = high TS score); 0.5 if CS -1.5 to -0.5; 0 if CS > -0.5 (no reversal).

Step 4 -- Structure-Activity Stream (SS): For your drug, retrieve all known binding targets from ChEMBL (python `chembl_webresource_client`) with `pk50 >= 6` ($IC_{50} \leq 1 \mu M$). For your disease, retrieve all disease-associated genes from DisGeNET (score ≥ 0.3). SS score = Jaccard similarity between drug target set and disease gene set (proportion of targets shared). Boost SS score if any drug-target is a well-validated disease driver (tier 1 by COSMIC or causative by OMIM).

Step 5 -- DRMS Fusion: Compute DRMS = `XGBoost.predict([KG_score, TS_score, SS_score])` using the pre-trained XGBoost model (model file deposited at Zenodo). DRMS output is 0-1 probability of being a true repurposing success. Threshold: DRMS ≥ 0.80 = priority candidate for experimental follow-up. Rank all drug-disease pairs by DRMS; apply medicinal chemistry review to top-28 candidates to confirm structural plausibility and absence of known safety issues. Final step: for any DRMS ≥ 0.80 candidate, confirm CNS penetration if CNS disease (CNS-MPO ≥ 4.0 ; BPLA paper 304 protocol) or oral bioavailability for oral administration route ($F \geq 20\%$ predicted by pkCSM).