SynDRep: A Knowledge Graph-Enhanced Tool based on Synergistic Partner Prediction for Drug Repurposing

Karim S. Shalaby[1,2], Sathvik Guru Rao[1], Bruce Schultz[3], Martin Hofmann-Apitius[1,4], Alpha Tom Kodamullil[1,5,*,+], Vinay Srinivas Bharadhwaj[1,4,+]

1. Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing, 53757 Sankt Augustin, Germany
2. Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmacy, Ain Shams University, 11566 Cairo, Egypt
3. Institute of Biomedical Informatics, University Hospital Cologne, 50937 Cologne, Germany
4. Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, 53115 Bonn, Germany
5. Causality Biomodels, Kinfra Hi-Tech Park, Kalamassery, Cochin, 683503 Kerala, India

***Corresponding Author**: Alpha Tom Kodamullil, Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin 53757, Germany. Telephone details: +49 2241 14-4050. Email: alpha.tom.kodamullil@scai.fraunhofer.de

[+] These authors contributed equally to this work.

Abstract

Drug repurposing is gaining interest due to its high cost-effectiveness, low risks, and improved patient outcomes. However, most drug repurposing methods depend on drug-disease-target semantic connections of a single drug rather than insights from drug combination data. In this study, we propose SynDRep, a novel drug repurposing tool based on enriching knowledge graphs (KG) with drug combination effects. It predicts the synergistic drug partner with a commonly prescribed drug for the target disease, leveraging graph embedding and machine learning techniques. This partner drug is then repurposed as a single agent for this disease by exploring pathways between them in KG. Some of our selected candidates, such as miconazole and albendazole for Alzheimer's disease, have been validated, while others lack either a clear pathway or literature evidence for their use for the disease of interest. Therefore, complementing SynDRep with more specialized KG, and additional training data, would enhance its efficacy and offer cost-effective and timely solutions for patients.

Keywords

Knowledge graph embedding, Machine learning, Drug repurposing, Synergism, Alzheimer's disease, Covid-19.

Code availability

SynDRep is available as an open-source Python package at https://github.com/SynDRep/SynDRep under the Apache 2.0 License.

Introduction

Despite tremendous technological, regulatory, and scientific advances that increase the efficiency of drug research and development, the resulting therapeutic outcomes need to catch up with the corresponding spending on these advances[1,2]. Additionally, the rising cost and time required to develop new drugs have resulted in lower profits for the pharmaceutical sector and a longer response time to disease outbreaks[3]. Conversely, drug repurposing, i.e., finding novel indications for current drugs, has advantages over de-novo drug development, including shorter development time and lower cost risk[4,5], since compounds already investigated and approved by regulatory bodies, incorporating safety and efficacy profiles, can be reassessed critically in a new therapeutic context[6].

In recent years, drug repurposing research has greatly benefited from the exploding growth of biomedical databases. Therefore, plenty of computational techniques have been devised to analyze different biomedical data systematically to hypothesize new indications for a drug or to find new drugs for a specific disease[7–9]. Computational drug repurposing approaches are mostly data-driven; they encompass the systematic analysis of data from various modalities, e.g., chemical structure, proteomic data, gene expression, genotype, or electronic health records, which can then drive the repurposing hypotheses[10,11]. For practical analysis of such vast data types, measures for appropriately aggregating them in an informative manner need to be taken. One of these measures is the organization and representation of data into a knowledge graph (KG), which aids in identifying semantic connections between multiple resources and allows for knowledge reasoning[12,13]. Extending these mechanistic KGs with drug-related data to form drug-target-mechanism-oriented data models results in so-called PHARMACOMES [6].

Pharmacomes with their integration of pathophysiology mechanisms, drug targets, and drugs/compounds offer the possibility to look at dual targeting strategies and combinatorial targeting of different pathophysiology mechanisms through combinations of drug repurposing candidates. Drug combinations offer excellent efficacy in treating multifactorial diseases involving more than one genetic pathway, such as cancer[14], diabetes[15], Alzheimer's disease (AD)[16], and cardiovascular diseases[17]. In principle, they also offer the option to specifically target co-morbidity pathways. Therefore, incorporating new links among drugs into pharmacomes, indicating drug combinations, paves the way for developing new synergistic drug combinations. It warns of potential drug-drug interactions in a more comprehensive way that depends on direct as well as indirect links between

drugs. Applying some link prediction algorithms afterward will predict new drug relationships, gain more insights into drug mechanisms, and eventually repurpose drug candidates for various diseases.

For a long time, drug synergy studies depended on trial and error, which suffers from high labor and time costs and exposes patients to ineffective treatment or undesirable side effects[18,19]. This was then replaced by high-throughput screening (HTS), where many measurements can be produced reasonably fast and at a lower cost[20]. During HTS, different concentrations of two or more drugs are applied to a cell line. However, the high genomic correlation between the original tissues and the derived cell lines remains imperfect[21]. Moreover, HTS cannot cover the whole combination space for drugs[22]. Computational methods such as deep and machine learning (DL/ML) models can efficiently explore the vast synergistic space using the available HTS synergy data. Recent methods range from systems biology[23], kinetic models[24], mixed integer linear programming[18], computational methods based on Drug-induced gene expression profile and dose-response curves[22], to ML approaches including Random Forests and Naive Bayes methods[25,26], and DL approaches such as deep neural networks, graph autoencoder, and convolutional neural network[27–32]. However, these methods are restricted to predicting synergistic combinations and not consider drug synergy prediction as an intermediate step in the drug repurposing process. In our approach, we leverage the synergistic prediction as a foundation for the repurposing process.

We propose a new drug repurposing tool (SynDRep), which depends on enriching knowledge graphs with drug combination effects. Our approach selects repurposing candidates, by predicting synergistic drug partners of a commonly prescribed drug for the target disease. This is followed by the selection of "safe drug partners" as a single-agent therapy for the disease. The drug's candidacy for repurposing is confirmed by exploring the pathway within the KG between the drug and the target disease. Additionally, experimental evidence about the beneficial effect of the candidate on target disease supports the repurposing profile. Therefore, this approach combines the speed and cost reduction of the computational approach with the accuracy and certainty of manual curation and expands the current drug repurposing landscape with a new concept relying not only on drug-disease-target semantic connections but also on the drug-drug synergy effect.

Results

1. Data collection

The primary objective entailed the integration of drug-drug relationships into an established KG to serve as a foundational framework for subsequent computational methodologies (Fig. 1). Consequently, the initial phase involved the careful selection of a comprehensive KG such as the Human Brain Pharmacome (HBP) and appropriate sources for drug-drug combinations including drug synergy databases e.g., DrugcombDB, DrugcombPortal, and SYNERGxDB.

a. Human Brain Pharmacome

The KG selected for this study was the Human Brain pharmacome (https://graphstore.scai.fraunhofer.de/, by selecting the pharmacome option under Database)[6], which is a comprehensive KG consisting of 136838 nodes and 731974 edges that combines knowledge from various sources with a focused drug-target-mechanism-oriented data model. It contains information curated from Bibliographic databases such as PubMed, Pathway databases such as Reactome, KEGG, and Pathway Commons, Protein-protein interaction databases such as IntAct, BioGRID, and StringDB, and Drug databases such as DrugBank, Clinical Trials, Sider, and ChEMBL. The data of this pharmacome has been extracted from the online source, stored locally using neo4j, and formed the base for the next step of KG enrichment.

b. Drug Synergy databases

There are many databases for drug combinations and their synergism or antagonism. To accommodate our expansive pharmacome we selected databases that contain the highest number of drugs. Drug combination effects have been gathered from DrugcombDB, DrugcombPortal, and SYNERGxDB. The scores for synergism models, such as HSA[33], Bliss[34], Loewe[35], and ZIP[36], were used to supplement the new edges created in the next step. These models consider in their calculation the different effects of drug combinations at different drug concentrations. Moreover, the effect of different cell types were also avoided by selecting only combinations that produced synergism, antagonism, or additive effect on all cell types.

2. KG enrichment

The KG enrichment was done over several steps as described in Methods section 1. The drug combination dataset contained 23171 pairs from 882 unique drugs, which were used to enrich the pharmacome. These combinations were further converted into edges and then added to the neo4j instance of the pharmacome by converting the values of ZIP scores into synergistic, antagonistic, or additive effects. To avoid class imbalance and model overfitting due to insufficient additive effect relations, additive edges have not been added to the pharmacome. The enriched pharmacome was then completely extracted as triples of source, relation, and target and was used to train and validate ML and KG embedding models.

3. Classical machine learning

In order to predict new synergistic relations between drugs, we started with the classical ML approaches, based on graph properties and drugs' physicochemical attributes, to assess their ability and efficiency for link prediction compared to KG embedding. Four ML models, namely: logistic regression, elastic net, gradient boosting, and random forest, were selected along with the feature space of each pair of drugs to classify their combination either into synergism or antagonism. Various features were employed in training and evaluating the models. These features encompassed aspects related to the HBP, as well as physicochemical attributes of the drugs as depicted in Table 1. The cross-validation and synergy prediction results are detailed in Supplementary Results section 1. Due to the lack of virtually validating studies and because this classical ML approach does not take relation type into consideration, we conducted a graph-embedding-based approach, as explained in the next section.

Table 1. The features and labels used to train the different ML models.

Features and metrics		Labels	Models
KG-related	Physicochemical		
Drug node degree Drug node clustering coefficient Drug node page rank Shortest path length Cosine Similarity	Molecular weight Log P Total polar surface area Number of hydrogen bond donors Number of hydrogen bond acceptors Rotatable bond count Tanimoto coefficient Morgan fingerprint	Synergism Antagonism	Logistic regression Elastic net Gradient Boosting Random Forest

4. KG Embedding

Lacking a virtual validation by confirming literature, we turned our focus to novel combination prediction using graph embeddings. To do so we performed pharmacome embedding using different algorithms to model and predict novel drug-drug links in the KG. We selected five models for creating embeddings of the pharmacome: TransE, TransR, RotatE, ComplEx, and HolE, as explained in Methods section 3b and Supplementary Results section 2. The results demonstrated that RotatE consistently outperformed other models in producing true predictions at the lowest rank (Fig. 2). Based on these results, the RotatE model was selected as the model of choice for predicting relationships between drugs that do not have a direct link within the pharmacome.

5. Synergy prediction

Synergy predictions were generated, following the methodology outlined in Methods section 4, leveraging the trained RotatE model. The prediction set was further processed as outlined in Supplementary Results section 3. The five highest-scoring synergy combinations were subjected to a thorough literature review to validate the reliability of the model's predictions, as detailed in Supplementary Table S2. The findings revealed that most of these combinations either exhibited documented synergy or were being utilized in combination for the treatment of the specific diseases for which they were intended.

6. Drug repurposing

The synergy effect frequently stems from different drugs having influences on the same, parallel or even different pathways essential for an observed phenotype, and synergy is induced by targets aggregating at specific pathways that control the state of the disease[37,38]. Consequently, if the predicted synergy combination has a shared pathway in our pharmacome, then it is highly probable that the synergistic partner to a commonly prescribed drug for the target disease can be used solely for the management of this disease. Therefore, we assessed the predicted synergistic combination that includes drugs prescribed for AD, schizophrenia, and bipolar disorder to determine the plausibility of repurposing their synergistic partner for these diseases. Based on our predictions, we selected a list of drugs that exhibited the highest-scoring synergistic combinations for each drug. It is noteworthy that our selection criteria excluded drugs with cytotoxic or severe side effects, such as anti-cancer or carcinogenic drugs, ensuring that the chosen repurposing candidates prioritize safety considerations. This approach was initially followed by a meticulous search for a possible common pathway in the pharmacome between the two drugs in the combination and the disease. Subsequently, we reviewed the literature for possible studies about using these repurposed candidates as single agents for the disease of interest. Detailed explanations of these candidates are elaborated in Supplementary Results section 4.

Leveraging a broad and highly connected KG such as HBP, with both casual and non-causal relations like association and complexity, can lead to suboptimal model training and prediction, as well as to less explainable pathways between drugs and diseases. In addition, the presence of nodes with high degrees in KG will lead to inadequate training and inaccurate predictions as well. We believe these factors contributed to the discrepancy between our model's predictions for schizophrenia repurposing candidates and existing literature evidence. To address this, we conducted a subsequent trial using a causal-only pharmacome, where we removed non-causal relationships and hub nodes. The hub nodes

were removed before embedding, but we retained them in another copy of the causal-only pharmacome used for pathway confirmations.

7. Causal-only pharmacome

After the removal of non-causal relations and disease nodes that form hubs in the HBP, the same KG embedding models were used to model the causal-only version of the HBP. HolE was the best model to produce true predictions at the lowest rank (74.90% for all relations), as shown in Supplementary Fig. S1. Although it was also the best-performing model for drug-drug relations, the percentage of true predictions at the lowest rank was low (54.65%), indicating nearly random predictions (Supplementary Fig. S2). Therefore, we changed the design of the experiment to incorporate KG embeddings followed by ML model training and prediction using the extracted embedding vectors as input features. In this approach, we utilized the vector embeddings of the causal-only pharmacome without enrichment with drug-drug relations. The synergistic data was then used as labels for training and testing the ML models. In this run, HolE again proved to be the best-performing embedding model (84.58% for all relations), and random forest emerged as the best ML model with an ROC value of 0.796 (Fig. 3 and Fig. 4). We have made the causal-only pharmacome, enriched with drug-drug relations (predicted and from databases), available at: https://doi.org/10.5281/zenodo.12806409. Consistent with prior methods, we validated the top-scoring synergistic combinations through a literature review (detailed in Supplementary Table S3). This analysis found that most combinations (three out of five) either demonstrated synergy between drug pairs or one of the drugs enhanced the effect of the other. Subsequently, we selected safe drugs predicted to be synergistic partners with the previously chosen drugs for AD and schizophrenia, or bipolar disorder.

a. Alzheimer's candidates

Based on our research, miconazole and albendazole have emerged as promising candidates for repurposing to treat AD. They were predicted to act synergistically with three and two of the selected AD drugs, respectively. By tracing their pathways to AD in the copy of the causal-only pharmacome, where we retained the disease nodes, we found that they share pathways with their synergistic partners to AD.

Miconazole is a broad-spectrum antifungal with some antibacterial activity[39]. On the other hand, it offers a potential therapeutic approach for early intervention in AD by promoting myelination of the medial prefrontal cortex and ameliorating neuroinflammation-mediated AD progression in different

mice models[40,41]. A prominent common pathway of miconazole with donepezil, rivastigmine, and galantamine to AD was found in pharmacome, as shown in Fig. 5. On the other hand, Albendazole is primarily employed as an anthelmintic to treat helminth infections[42]. However, research on H4 neuroglioma cells has shown that albendazole can reduce Tau levels, suggesting a beneficial effect on AD[43]. In the causal-only pharmacome, it has a shared pathway with donepezil and rivastigmine, in which it intersects with them in decreasing the levels of phosphorylated Microtubule-associated protein tau (MAPT), which is a hallmark of AD (Fig. 5). Comprehensive description of miconazole and albendazole pathways is elaborated in Supplementary Results section 5.

Other candidates, such as disulfiram, auranofin, and finafloxacin, were also predicted as synergistic partners with AD drugs and studies, in cell and animal models, showed their beneficial effects for the management of AD[44–49]. However, no supporting pathway in the pharmacome could be detected for these drugs. Additionally, prochlorperazine was consistently predicted as a synergistic partner with the three selected AD drugs. However, it has anticholinergic properties that in higher doses might worsen AD-associated dementia[50]. Moreover, the explored pathways between prochlorperazine and AD were controversial, with some suggesting it might have a beneficial effect for the management of AD while others suggest it might exacerbate the condition.

b. Schizophrenia and bipolar disorder candidates

Upon detecting repurposing candidates for schizophrenia or bipolar disorder, we found no connecting pathways between any drug in the causal-only pharmacome and these diseases, even for the drugs that are typically prescribed for these conditions. Therefore, we couldn't identify any repurposing candidates for schizophrenia or bipolar disorder based on the causal-only pharmacome.

c. COVID-19 candidates

We wanted to challenge our model further by checking its ability to predict synergistic drug combinations for diseases, which the KG was not built for, to test its ability to be used as a fast tool for repurposing drugs to new disease outbreaks. We selected a drug, baricitinib, used for COVID-19 and extracted the synergistic combinations from those predicted by our model. From these combinations, Prochlorperazine and disulfiram were identified as repurposing candidates based on their predicted synergy with baricitinib and their beneficial effects in COVID-19 management. These effects include the inhibition of SARS-CoV-2 entry by targeting the spike protein and ACE2, which

were confirmed computationally by molecular docking and experimentally in VeroE6 and HEK293T-hACE2 cell cultures[51,52].

Discussion

Pursuing a new drug candidate for a disease has been exhaustively overwhelming. Therefore, drug repurposing has recently gained significant interest. Here, we present our study of enriching existing KGs with drug synergy data to achieve a primary goal: repurposing predicted drug synergy partners as single agents for the disease of interest. The alignment between the model's predictions and the real-world literature highlights the model's effectiveness in identifying clinically relevant and potentially impactful drug candidates. Although synergy predictions followed by safe drug selection, and pathway in pharmacome tracing, have reduced the number of selected drug candidates, they form a strong foundation for further research on these candidates.

Some of our selected candidates such as albendazole and miconazole, mefloquine, ciprofloxacin, and moxifloxacin for AD, have a robust profile of clear pathways in our pharmacome with the disease of interest as well as experimental (in cell and animal models) and clinical studies that support their willingness to be repurposed for that disease. Another portion of selected drugs, including disulfiram, auranofin, finafloxacin, taribavirin, and ivermectin for AD, has strong literature evidence but unclear pathways in our pharmacome, which require further analysis and understanding of the relation encompassed in their pathway in the pharmacome. Finally, drugs that have no literature evidence or clear pathways were marked as the least suitable for repurposing including atovaquone for AD and pyrimethamine for Schizophrenia. In addition to these groups exists a controversial group in which literature supports their harmful effect on the selected disease; however, they appear many times in our predictions as a valuable agent for controlling that disease. This group includes mefloquine, chloroquine, and albendazole for schizophrenia. We highly recommend further clinical and experimental investigation of these drugs for that disease before the commencement of their repurposing procedures.

To challenge our model's applicability, we selected a drug for COVID-19 even knowing that there is no disease node for COVID-19 in our pharmacome. The results showed its ability to predict synergy and repurposing candidates, which a strong literature profile confirmed. This vast ability underscores that the model does not rely on a single node or relation but on the overall interaction within the network. The potential of this approach to repurposing drugs for diseases that are out of the scope of

the used pharmacome gives insights into comorbidity pathways that exist between These diseases. Specifically, we refer to the possible comorbidity between COVID-19 and neurodegenerative diseases (NDD). The ability of SynDRep to find repurposing candidates for COVID-19 may be attributed to these underlying comorbidity pathways. Therefore, this work paves the way for further research already being conducted for detecting such comorbidities through the COMMUTE project[53]. It also shows the model's ability to extend to other diseases beyond the primary focus of the used KG and the potential for repurposing drugs to new emergent diseases. Hence, we provide a tool that can be utilized, after enrichment with related entities and relations, as a rapid selector of drugs to be used against future outbreaks and pandemics.

Our study first took a broad approach that relied on graph topology metrics as well as the physicochemical properties of the drugs. ML models were then used to classify combinations into synergistic or antagonistic categories. However, this approach primarily neglected the "relationship type" factor within the KG and relied solely on the data associated with the drug nodes. In contrast, KG embedding models take these relations into account when embedding all the nodes of the KG into vectors. This consideration improves the performance and predictions of KG embedding compared to ML modeling. Therefore, this approach emphasizes the beneficial effect of organizing data into KGs and the further extraction of this data using graph embedding techniques over the classical ML approaches. Analyzing biomedical data using network structures requires a thorough understanding of network topology. Therefore, we used the topological features along with the physicochemical features of the drugs for the training and prediction in the classical ML approach. However, these methods often demand high computational and space costs[54] and result in lower performance than the graph embedding method as evidenced by the lack of literature studies for predicted top scorer partners. On the contrary, organizing the data into a graph that can describe the complex structure of data and enables the characterization of high-order geometric patterns for the networks, improves the performance of various data analysis tasks[55]. Graph embedding techniques are able to convert sparse high-dimensional graphs into continuous low-dimensional vectors that maximally preserve the graph structure properties[56]. The generated highly informative and nonlinear embeddings can be subsequently used for different downstream analytic tasks such as node classification and link prediction. We applied these graph embedding techniques for the prediction of the link between pairs of drugs. Unlike the ML approach, these predictions were confirmed by published scientific studies. Consequently, utilizing data represented as graphs and incorporating their embeddings represents the future direction for pharmacome data mining.

In the context of drug repurposing, maintaining a clear chain of causality from drugs to disease targets is critical. Using a broad and highly connected KG such as HBP, which contains both cause-and-effect relations and less explicit relations such as association and complexity, can lead to suboptimal model training and prediction. Although selecting a cause-and-effect subgraph is optimal, the extensive relations pool in the pharmacome captures complex protein interactions that may not be strictly cause-and-effect. Additionally, the graph needs to be large enough for effective link prediction; otherwise, performance may be compromised, which was observed with the causal-only pharmacome trial, hence pathway reviews and post-prediction literature validation were essential steps to compensate for the lack of a pure cause-and-effect subgraph by focusing on promising candidates. Another important consideration in the use of HBP is the presence of so-called ”super-hubs”, which are nodes with extremely high node degrees, whose presence in the KG dilutes information and hinders learning[57] . The topological imbalance in KGs has negative effects on learning using KG embedding models, where low-degree nodes embed at a much lower quality relative to high-degree nodes[58]. Moreover, high-degree nodes are mostly predicted as answers simply due to their higher degree, not their domain relevance[58,59]. Based on these considerations, we performed a pruning of the HBP to remove non-causal relations between entities and super-hub nodes, which were mainly disease nodes. The disease nodes were removed before embedding, but we retained them in another copy of the causal-only pharmacome used for pathway confirmations. The results showed more promising repurposing candidates for AD. However, some candidates lacked pathways in the causal-only pharmacome. Additionally, we couldn’t find any connections between schizophrenia or bipolar disorder and any drugs in the pharmacome, including those usually prescribed for these diseases. Their connecting relations might have been removed during the causal relation selection step. This indicates the incompleteness of the causal-only pharmacome, which significantly impacts the repurposing approach we undertook in this study. Consequently, we recommend more manual curation of certain relation types in the HBP to enhance and update their causality comprehension. This could help maintain the relation pool present in the HBP, which is crucial for effective embedding and repurposing.

In contrast to our approach’s advantages, it exhibits some shortcomings. First, it cannot model all drug relations due to the imbalance in relation type. Most databases and studies focus on synergistic or antagonistic combinations, while scarce data about additive effect combinations are available. For instance, our work had only three additive combinations compared to tens of thousands of antagonistic or synergistic ones. Therefore, after many trials, we decided to omit these additive relations to avoid the class imbalance problem. Moreover, the embedding model’s prediction for drug combinations is not, in all cases, a drug-drug relation but may predict any other relation available in the pharmacome.

Consequently, an amount of the input data might get neither synergistic nor antagonistic prediction, resulting in the loss of some combinations. Third, the controversy between some predictions and the published data about these drugs and diseases necessitates thorough investigations. Lastly, the diversity of drugs in the combination databases is limited; most are cancer-related and measure only cytotoxicity. Therefore, our approach must be extended to other KGs and drug combination databases to broaden our search landscape and produce more guided synergistic predictions.

Moreover, this study highlights the hugely beneficial effect of computational methods not only in reducing the chemical, energy, and resource waste required to conduct thousands of wet-lab investigations but also by helping in sustainability through re-using the same drugs for more diseases and the reduction of the capital required to set up new production plans. Therefore, tons of hours, labor, and costs have been spared, which can foster further projects and speed up the pace by which treatment plans can be exploited. While this study is based on the HBP, its scope extends beyond drugs specifically prescribed for neuro and psychiatric illnesses. Accordingly, we present an approach to drug repurposing based on the prediction of drug-drug synergies and the elucidation of conformational pathways. When complemented with more specialized and highly curated KGs, such as cause-and-effect subgraphs, and additional synergistic training data, this methodology would hold monumental potential as a robust tool for the pharmaceutical sector. Therefore, we present SynDRep as an open-source library to be used by researchers for various KGs and drug combination datasets, hoping it will enrich scientific research with an advanced tool for drug repurposing. It can assist in strategically selecting drug candidates for existing diseases or potential disease outbreaks. The simplicity and adaptability of this approach to different diseases and KGs consolidate its position as a reliable tool for drug repositioning, providing valuable insights for drug development and therapeutic discovery.

Methods

1. KG enrichment

The KG enrichment was done over several steps. Data extracted from drug combination databases were deduplicated. The synergism scores from the same combinations with different scores were averaged. Some combinations of drugs produce synergism in one cell line and antagonism in another. Therefore, to remove the effect of different cell types, we selected only combinations that produced either synergism or antagonism on all cell types. Moreover, to go with a standardized approach, we selected only one synergism score (ZIP score), as the ZIP model encompasses the Loewe additivity

and the Bliss independence. In addition, it is more accurate at detecting potency changes in drug combinations compared to HSA and Bliss independence models[36].

2. Machine learning modeling

We used Scikit-learn, a Python package for ML purposes, to train and validate our classification models. Features related to KG depend on the network structure and topological features. These features were extracted from the pharmacome using NetworkX, a Python package for the creation and study of the structure of complex networks. Physicochemical attributes of the drugs were extracted from PubChem or computed using the RDKit Python package. The features, labels, and models used are listed in Table 1. The classification was carried out over 10-fold nested cross-validation using Grid Search optimizer, and the model performance was assessed based on areas under the receiver operating characteristic curve (ROC-AUC) values for all the models.

Following a comprehensive assessment of all models and the calculation of ROC-AUC, we utilized the elastic net model to predict the synergistic interactions between each pair of drugs in the pharmacome. Subsequently, combinations predicted as synergistic were chosen to constitute the predicted synergism set. A literature check was conducted to validate the top five synergistic combinations based on their predicted probabilities during the initial prediction process. This validation step aimed to ensure the credibility and accuracy of the model's predictions by cross-referencing them with existing scientific literature.

3. KG Embedding

To embed the enriched pharmacome, we used PyKEEN (Python KnowlEdge EmbeddiNgs)[60], a Python package designed for training and evaluating KG embedding models. We worked under stochastic local closed world assumption (SLCWA), where a randomized subset is drawn from the combination of head and tail generation strategies, initially defined in local closed world assumption, and these selected triples are treated as negatives. This approach offers several advantages, including the lower load of computation and the flexibility to include new negative sampling strategies.

a. Data splitting

To prevent overfitting, the set of triples obtained from the extraction of enriched pharmacome was then stratified using the PyKEEN into a training set (80%) and a test set (20%). To prevent the

dissemination of the test set into the training set during the HPO or training of the model, we isolated the test set, and the training set was further split into training (80%) and validation (20%) sets. We checked that each split contained the corresponding percentage of triples and that the training set contained all the relation types in the pharmacome to ensure that the test and validation sets did not contain any relation type new to the model after training. To further assess the model's efficiency in predicting drug-drug relations, one more test set was formed from the original test set, the drug-drug test set, which contained only the drug-drug relations from the test set. These two sets were used to evaluate model performance.

b. Model selection

We selected five models for embedding the pharmacome: TransE, TransR, RotatE, ComplEx, and HolE. We afterward used the best-performing model to predict the new synergistic or antagonistic relations. Predictions were made using the two entities as head and tail, and the model predicts the relation type between them. The output of the prediction model is a ranking of the possible relations between these two entities according to a score produced by the model. Therefore, we selected the first three predictions to assess the model's performance by calculating the percentage of True prediction in each rank compared to all predictions in this rank (equation):

$$Percentage\ of\ true\ predictions\ = \frac{Number\ of\ true\ predictions\ in\ a\ rank\ \times\ 100}{Total\ number\ of\ prediction\ in\ the\ same\ rank}$$

We calculated the multi-class ROC-AUC using the highest-ranked prediction for each pair of drugs in the test set. This involved converting the actual and predicted relation types into binary form and then averaging the ROC-AUC values for all relation types, as shown in the following equation:

$$Multiclass\ ROC - AUC\ = \frac{\Sigma\ ROC - AUC\ for\ each\ relation\ type}{Total\ number\ of\ relation\ types}$$

Based on the values of the percentage of true predictions and multi-class ROC-AUC, we selected the best-performing model for further prediction of the drug-drug relations that are not in the pharmacome.

4. Synergy prediction

After assessing all models and calculating the percentage of true prediction, we selected RotatE to predict all drug-drug relations further. Utilizing the trained RotatE model, we predicted the relations between each drug pair within the pharmacome. Both the forward case (with drug A as a head and

drug B as a tail) and the reverse case (with drug B as a head and drug A as a tail) were predicted. The predicted relationship which ranked as the first was then extracted for each drug pair. To refine our dataset, we eliminated cases where nonmutual synergism or antagonism was observed. Then, we segregated the predicted dataset into synergism and antagonism categories, focusing on selecting data that showed synergistic interactions for subsequent in-depth analysis. A comprehensive literature review was conducted to validate the model's predictive power, particularly for the predictions with the highest scores.

5. Drug repurposing

The assessment of the plausibility of repurposing candidates for the target disease was done by revising the literature for scientific data about their use in the selected disease and determining the common pathways between the synergistic drugs and the disease in our pharmacome using the Python package NetworkX. This comprehensive approach enhances our understanding of potential therapeutic applications and facilitates informed decision-making regarding drug repurposing candidates.

6. Causal-only pharmacome

Due to the lack of clarity in certain relations within the pathways explored in the HBP between repurposing candidates and the target diseases, we repeated the entire trial using a causal-only version of the HBP. In this version, we retained only the relations indicating direct causality between entities, such as increases, decreases, or causes no effect. Additionally, the presence of hubs within the pharmacome's structure, primarily stemming from disease nodes, could introduce their overrepresentation during embedding and ML analysis. Consequently, these nodes and their connecting relationships were also removed before KG embedding. The disease nodes were removed before embedding, but we retained them in another copy of the causal-only pharmacome used for pathway confirmations. However, the performance of KG embedding models, in predicting drug-drug relations, was found to be suboptimal. Therefore, we introduced an additional step. We extracted vector embeddings of the KG and utilized them as inputs for training and testing ML models. By leveraging the best-performing ML model, we obtained the final predictions of drug-drug relations, which were then used for drug repurposing. We further challenged our repurposing approach for COVID-19, which has no node in our pharmacome, to explore the capabilities of our model, as a fast tool for drug repurposing for disease outbreaks.

Authors' Contributions

KSS and VSB conceived and designed the study. KSS analyzed the datasets, interpreted the results, and wrote the manuscript. BS collected data and proofread the manuscript. ATK and MHA acquired the funding. MHA, SGR, and ATK reviewed the manuscript. All authors have read and approved the final manuscript.

Funding

This work was partly supported by the German Federal Ministry of Education and Research (BMBF, grant 01ZX1904C). This work was developed in the Fraunhofer Cluster of Excellence "Cognitive Internet Technologies".

Competing interests

The authors report no conflict of interest.

Acknowledgment

The authors gratefully acknowledge Andrea Zaliani for his insights into the manuscript.

References

1. Ashburn, T. T. & Thor, K. B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov. 2004 38* **3**, 673–683 (2004).
2. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov. 2012 113* **11**, 191–200 (2012).
3. Pushpakom, S. *et al.* Drug repurposing: Progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **18**, 41–58 (2018).
4. Choudhury, C., Arul Murugan, N. & Priyakumar, U. D. Structure-based drug repurposing: Traditional and advanced AI/ML-aided methods. *Drug Discov. Today* **27**, 1847–1861 (2022).
5. Hua, Y. *et al.* Drug repositioning: Progress and challenges in drug discovery for various diseases. *Eur. J. Med. Chem.* **234**, 114239 (2022).
6. Lage-Rupprecht, V. *et al.* A hybrid approach unveils drug repurposing candidates targeting an Alzheimer pathophysiology mechanism. *Patterns* **3**, (2022).
7. Jarada, T. N., Rokne, J. G. & Alhajj, R. A review of computational drug repositioning: Strategies, approaches, opportunities, challenges, and directions. *J. Cheminformatics* **12**, (2020).
8. Luo, H. *et al.* Biomedical data and computational models for drug repositioning: A comprehensive review. *Brief. Bioinform.* **22**, 1604–1619 (2021).
9. Pan, X. *et al.* Deep learning for drug repurposing: Methods, databases, and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **12**, e1597 (2022).
10. Hurle, M. R. *et al.* Computational drug repositioning: from data to therapeutics. *Clin. Pharmacol. Ther.* **93**, 335–341 (2013).
11. Zong, N. *et al.* Computational drug repurposing based on electronic health records: a scoping review. *Npj Digit. Med.* **5**, 1–8 (2022).
12. Chen, H., Cheng, F. & Li, J. IDrug: Integration of drug repositioning and drug-target prediction via cross-network embedding. *PLoS Comput. Biol.* **16**, e1008040 (2020).

13. Gao, Z., Ding, P. & Xu, R. KG-Predict: A knowledge graph computational framework for drug repurposing. *J. Biomed. Inform.* **132**, 104133 (2022).
14. Zhou, Z., Edil, B. H. & Li, M. Combination therapies for cancer: challenges and opportunities. *BMC Med.* **21**, (2023).
15. Dubourg, J., Fouqueray, P., Quinslot, D., Grouin, J. M. & Kaku, K. Long-term safety and efficacy of imeglimin as monotherapy or in combination with existing antidiabetic agents in Japanese patients with type 2 diabetes (TIMES 2): A 52-week, open-label, multicentre phase 3 trial. *Diabetes Obes. Metab.* **24**, 609–619 (2022).
16. Knorz, A. L. & Quante, A. Alzheimer's Disease: Efficacy of Mono- and Combination Therapy. A Systematic Review. *J. Geriatr. Psychiatry Neurol.* **35**, 475–486 (2022).
17. Lombardi, N. *et al.* Adherence to Triple-Free-Drug Combination Therapies Among Patients With Cardiovascular Disease. *Am. J. Cardiol.* **125**, 1429–1435 (2020).
18. Pang, K. *et al.* Combinatorial therapy discovery using mixed integer linear programming. *Bioinformatics* **30**, 1456–1463 (2014).
19. Day, D. & Siu, L. L. Approaches to modernize the combination drug development paradigm. *Genome Med.* **8**, 115 (2016).
20. He, L. *et al.* Methods for High-throughput Drug Combination Screening and Synergy Scoring. *Methods Mol. Biol. Clifton NJ* **1711**, 351–398 (2018).
21. Ferreira, D. *et al.* The Importance of Cancer Cell Lines as in vitro Models in Cancer Methylome Analysis and Anticancer Drugs Testing. in *Oncogenomics and Cancer Proteomics - Novel Approaches in Biomarkers Discovery and Therapeutic Targets in Cancer* (IntechOpen, 2013). doi:10.5772/53110.
22. Goswami, C. P., Cheng, L., Alexander, P. S., Singal, A. & Li, L. A New Drug Combinatory Effect Prediction Algorithm on the Cancer Cell Based on Gene Expression and Dose-Response Curve. *CPT Pharmacomet. Syst. Pharmacol.* **4**, e9 (2015).

23. Feala, J. D. *et al.* Systems approaches and algorithms for discovery of combinatorial therapies. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 181–193 (2010).

24. Sun, X., Bao, J., You, Z., Chen, X. & Cui, J. Modeling of signaling crosstalk-mediated drug resistance and its implications on drug combination. *Oncotarget* **7**, 63995–64006 (2016).

25. Li, P. *et al.* Large-scale exploration and analysis of drug combinations. *Bioinformatics* **31**, 2007–2016 (2015).

26. Wildenhain, J. *et al.* Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning. *Cell Syst.* **1**, 383–395 (2015).

27. Preuer, K. *et al.* DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics* **34**, 1538–1546 (2018).

28. Kuenzi, B. M. *et al.* Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell* **38**, 672-684.e6 (2020).

29. Sun, Z., Huang, S., Jiang, P. & Hu, P. DTF: Deep Tensor Factorization for predicting anticancer drug synergy. *Bioinformatics* **36**, 4483–4489 (2020).

30. Kim, Y. *et al.* Anticancer drug synergy prediction in understudied tissues using transfer learning. *J. Am. Med. Inform. Assoc.* **28**, 42–51 (2021).

31. Liu, Q. & Xie, L. TranSynergy: Mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLOS Comput. Biol.* **17**, e1008653 (2021).

32. Li, H. *et al.* Predicting Drug Synergy and Discovering New Drug Combinations Based on a Graph Autoencoder and Convolutional Neural Network. *Interdiscip. Sci. Comput. Life Sci.* **15**, 316–330 (2023).

33. Berenbaum, M. C. What is synergy. *Pharmacol. Rev.* **41**, 93–141 (1989).

34. Bliss, C. I. The Toxicity of Poisons Applied Jointly1. *Ann. Appl. Biol.* **26**, 585–615 (1939).

35. Loewe, S. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung.* **3**, 285–290 (1953).

36. Yadav, B., Wennerberg, K., Aittokallio, T. & Tang, J. Searching for Drug Synergy in Complex Dose–Response Landscapes Using an Interaction Potency Model. *Comput. Struct. Biotechnol. J.* **13**, 504–513 (2015).

37. Chen, D., Liu, X., Yang, Y., Yang, H. & Lu, P. Systematic synergy modeling: understanding drug synergy from a systems biology perspective. *BMC Syst. Biol.* **9**, 56 (2015).

38. Cokol, M. *et al.* Systematic exploration of synergistic drug pairs. *Mol. Syst. Biol.* **7**, 544 (2011).

39. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

40. Yeo, I. J., Yun, J., Son, D. J., Han, S.-B. & Hong, J. T. Antifungal drug miconazole ameliorated memory deficits in a mouse model of LPS-induced memory loss through targeting iNOS. *Cell Death Dis.* **11**, 623 (2020).

41. Wang, Z. *et al.* Miconazole Promotes Cooperative Ability of a Mouse Model of Alzheimer Disease. *Int. J. Neuropsychopharmacol.* **25**, 951–967 (2022).

42. Sungkar, S., Putri, K. Q., Taufik, M. I. S., Gozali, M. N. & Sudarmono, P. The Effectiveness of Triple Dose Albendazole in Treating Soil Transmitted Helminths Infection. *J. Parasitol. Res.* **2019**, 6438497 (2019).

43. Dickey, C. A. *et al.* Pharmacologic reductions of total tau levels; implications for the role of microtubule dynamics in regulating tau expression. *Mol. Neurodegener.* **1**, 6 (2006).

44. Madeira, J., Renschler, C., Gill, N. & Klegeris, A. P2-373: The anti-rheumatic gold drug auranofin could be beneficial in Alzheimer's disease by inhibiting astrocyte-mediated neuroinflammation. *Alzheimers Dement.* **8**, P394–P394 (2012).

45. Roder, C. & Thomson, M. J. Auranofin: Repurposing an Old Drug for a Golden New Age. *Drugs RD* **15**, 13–20 (2015).

46. Reinhardt, S. *et al.* Identification of disulfiram as a secretase-modulating compound with beneficial effects on Alzheimer's disease hallmarks. *Sci. Rep.* **8**, 1329 (2018).

47. Upīte, J., Kadish, I., van Groen, T. & Jansone, B. Subchronic administration of auranofin reduced amyloid-β plaque pathology in a transgenic APPNL-G-F/NL-G-F mouse model. *Brain Res.* **1746**, 147022 (2020).

48. Jun, C. & Fang, B. Current progress of fluoroquinolones-increased risk of aortic aneurysm and dissection. *BMC Cardiovasc. Disord.* **21**, 470 (2021).

49. Guo, W., Chen, S., Li, C., Xu, J. & Wang, L. Application of Disulfiram and its Metabolites in Treatment of Inflammatory Disorders. *Front. Pharmacol.* **12**, (2022).

50. Obara, K. *et al.* Characterization of binding of antipsychotics to muscarinic receptors using mouse cerebral cortex. *J. Pharmacol. Sci.* **140**, 197–200 (2019).

51. Chen, H.-F. *et al.* Disulfiram blocked cell entry of SARS-CoV-2 via inhibiting the interaction of spike protein and ACE2. *Am. J. Cancer Res.* **12**, 3333–3346 (2022).

52. Liang, T. *et al.* Phenothiazines Inhibit SARS-CoV-2 Entry through Targeting Spike Protein. *Viruses* **15**, 1666 (2023).

53. COMMUTE. *Comorbidity Mechanisms Utilized in Healthcare* https://www.commute-project.eu/en/about (2024).

54. Su, C., Tong, J., Zhu, Y., Cui, P. & Wang, F. Network embedding in biomedical data science. *Brief. Bioinform.* **21**, 182–197 (2020).

55. Xu, M. Understanding graph embedding methods and their applications. Preprint at https://doi.org/10.48550/arXiv.2012.08019 (2020).

56. Cai, H., Zheng, V. W. & Chang, K. C.-C. A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications. Preprint at https://doi.org/10.48550/arXiv.1709.07604 (2018).

57. Sardina, J., Costabello, L. & Guéret, C. Veni, Vidi, Vici: Solving the Myriad of Challenges before Knowledge Graph Learning. Preprint at http://arxiv.org/abs/2402.06098 (2024).
58. Bonner, S., Kirik, U., Engkvist, O., Tang, J. & Barrett, I. P. Implications of topological imbalance for representation learning on biomedical knowledge graphs. *Brief. Bioinform.* **23**, bbac279 (2022).
59. Ratajczak, F., Joblin, M., Ringsquandl, M. & Hildebrandt, M. Task-driven knowledge graph filtering improves prioritizing drugs for repurposing. *BMC Bioinformatics* **23**, 84 (2022).
60. Ali, M. *et al.* PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings. *J. Mach. Learn. Res.* **22**, 1–6 (2021).

Figures

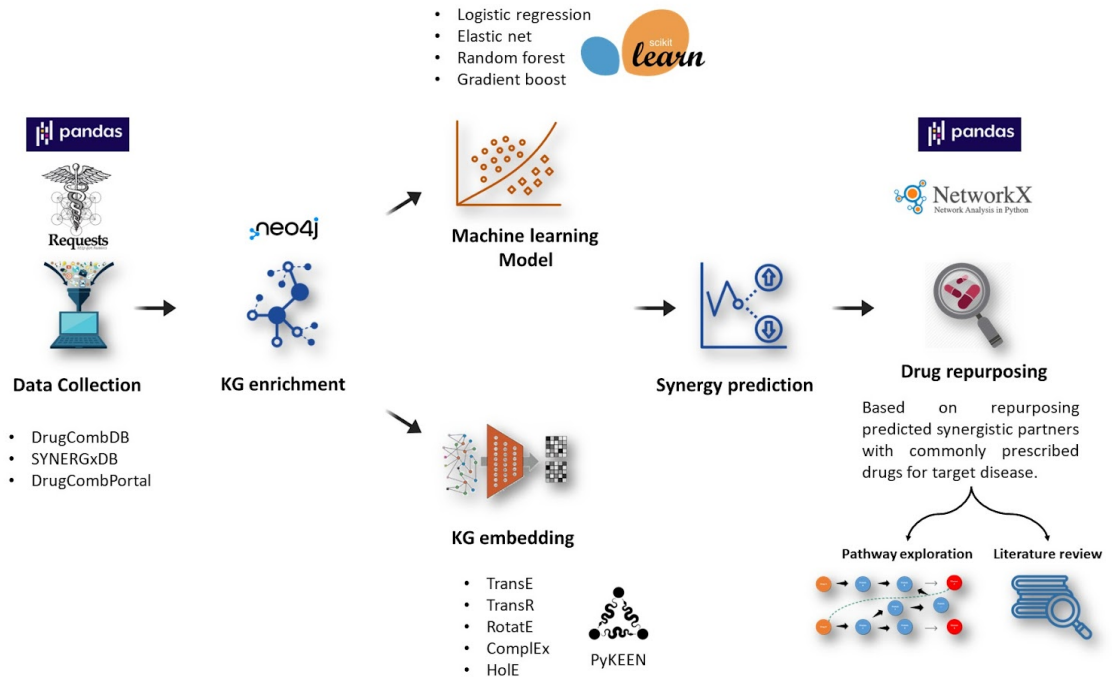

Figure 1. The overall workflow of the study, including the Python packages and data sources used. The work starts with data collection and refinement from combination databases using Pandas and Requests, then the synergy data is fed into a neo4j instance of the KG. Third, many ML and embedding models were used to model and predict the synergies using scikit-learn and PyKEEN, respectively. Finally, the identification of repurposing candidates by predicting synergistic drug partners for commonly prescribed drugs for the target disease and repurposing safe partners as a single agent for this disease. Candidate profiles are confirmed by examining the existence of pathways within the knowledge graph between the candidate drug and the target disease using Pandas and NetworkX. Experimental evidence from the literature supporting the candidate's beneficial effect on the target disease further validates the repurposing profile.

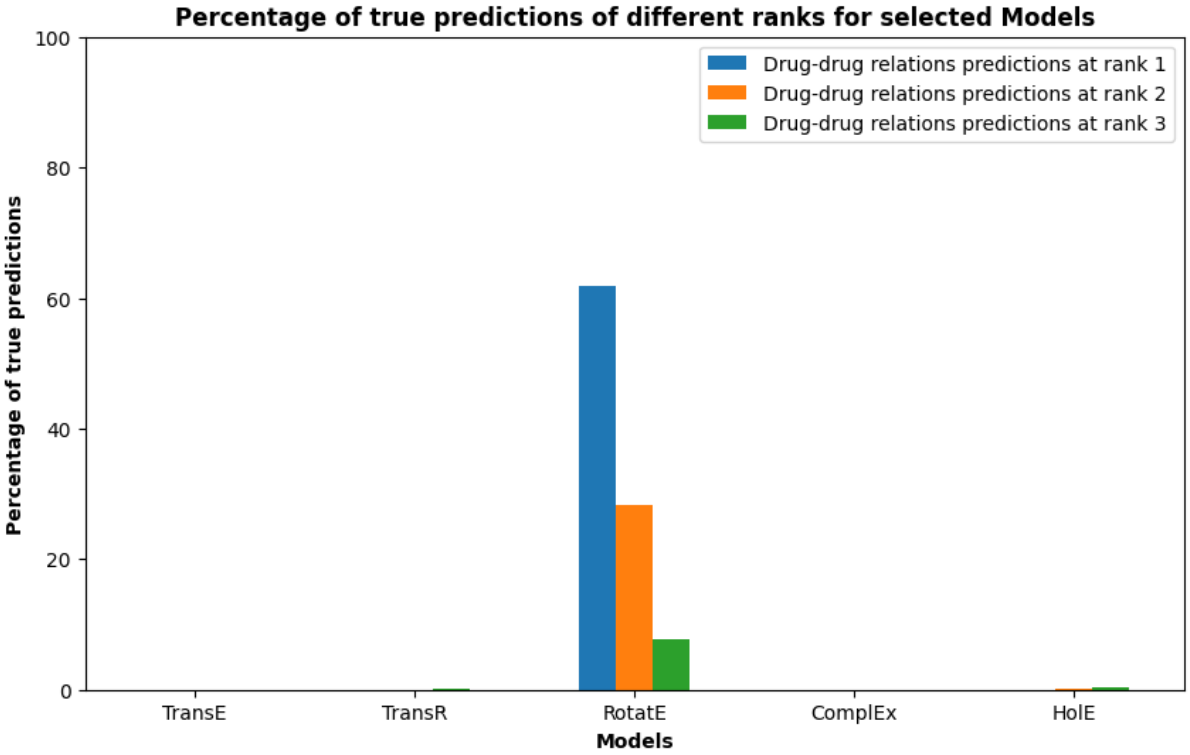

Figure 2. Percentage of true drug-drug relation predictions at different ranks for selected models. Optimum models were used to predict the drug-drug relations in the test set from the original HBP. Then, the predicted relations were compared to the actual relations to calculate the percentage of true predictions.

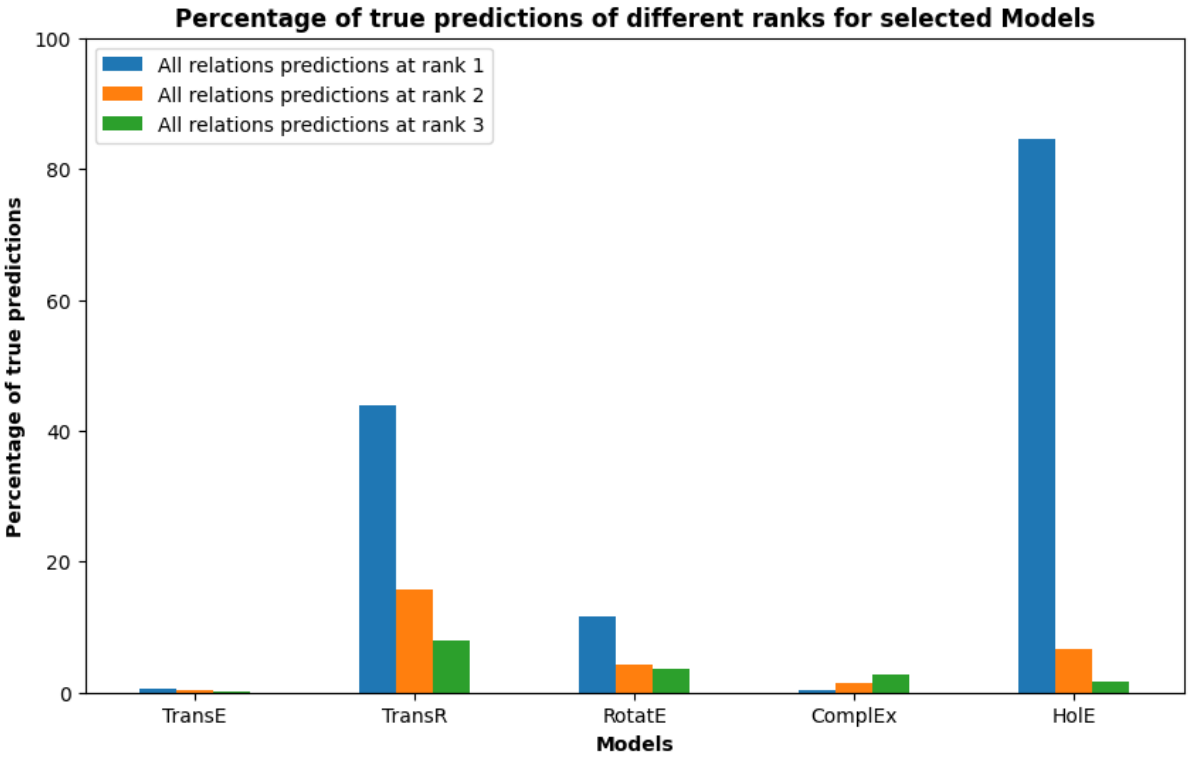

Figure 3. Percentage of true all relations predictions at different ranks for selected models used to embed Casual-only pharmacome before ML. Optimum models were used to predict the test set, and then the predicted relations were compared to the actual relations to calculate the percentage of true predictions. Embedding vectors of HolE were extracted and used as input features for the training and testing of ML models.

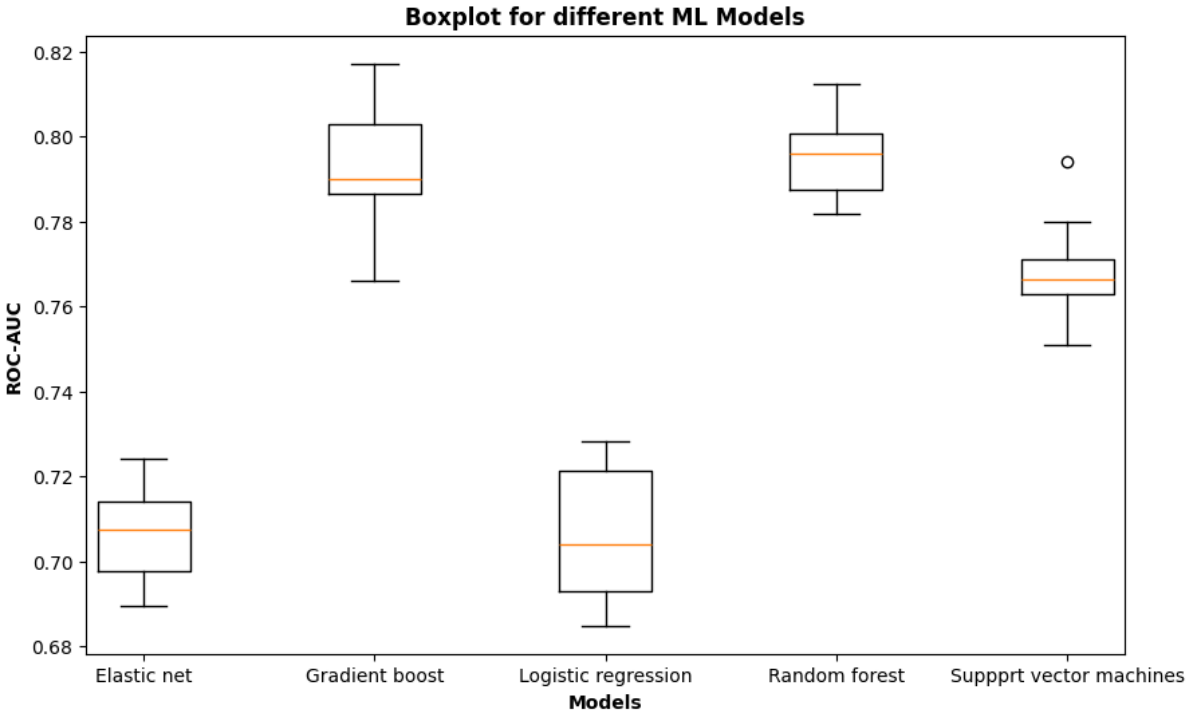

Figure 4. Benchmarking of ML models trained to classify between synergism and antagonism using the embedding vectors from causal-only pharmacome. Each boxplot shows the distribution of the ROC-AUC values over ten repeats of the ten-fold nested cross-validation procedure.

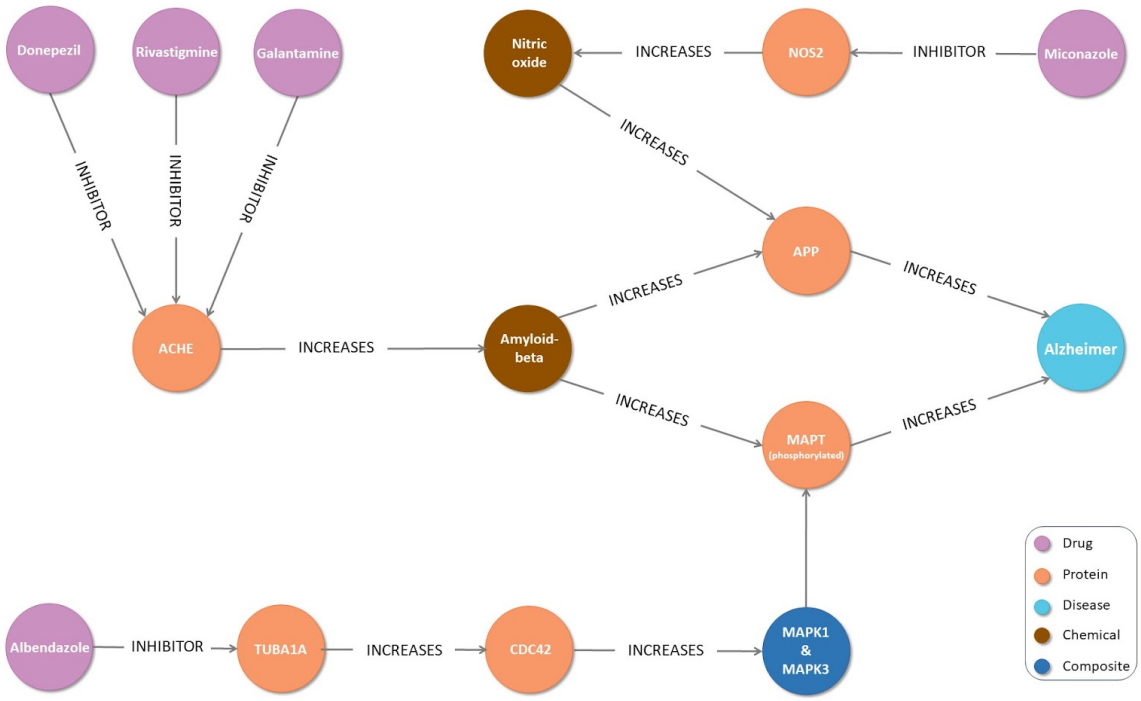

Figure 5. The shared pathways between miconazole, albendazole, donepezil, rivastigmine, and galantamine to Alzheimer's disease. Disease nodes were retained in a copy of the causal-only pharmacome used for pathway confirmations. (ACHE: acetylcholinesterase, NOS2: nitric oxide synthase 2, APP: amyloid-beta precursor protein, MAPT: microtubule-associated protein tau, TUBA1A: tubulin alpha-1A protein, CDC42: cell division control protein 42 homolog, MAPK1 and MAPK3: mitogen-activated protein kinase 1 and 3).