A Unified View of Relational Deep Learning for Drug Pair Scoring

Benedek Rozemberczki[1] , **Stephen Bonner**[2] , **Andriy Nikolov**[1] , **Michaël Ughetto**[1] , **Sebastian Nilsson**[1] and **Eliseo Papa**[1]
[1]Research Data & Analytics, Research & Development IT, AstraZeneca
[2]Data Sciences and Quantitative Biology, Discovery Sciences R&D, AstraZeneca
benedek.rozemberczki@astrazeneca.com

Abstract

In recent years, numerous machine learning models which attempt to solve polypharmacy side effect identification, drug-drug interaction prediction, and combination therapy design tasks have been proposed. Here, we present a unified view of relational machine learning models which can address these tasks. We provide fundamental definitions, compare existing model architectures and discuss performance metrics, datasets, and evaluation protocols. In addition, we emphasize possible high-impact applications and important future research directions in this domain.

1 Introduction

Relational deep learning has an unprecedented potential for revolutionizing the drug discovery process and pharmaceutical industry [Gaudelet *et al.*, 2021]. A number of high value use cases for relational deep learning in the pharmaceutical domain involve answering questions about what happens when two drugs are administered at the same time. These potential applications might want to answer questions such as: Will a combination of two drugs be more effective at destroying a specific type of lung cancer cells [Preuer *et al.*, 2018]? Is there an unexpected (polypharmacy) side effect [Zitnik *et al.*, 2018a] of using these two drugs together? Is there an unwanted chemical interaction [Sunyoung *et al.*, 2017] that these drug molecules can have?

All of these previously mentioned questions can be answered by what we see as *drug pair scoring*, a machine learning task that involves a set of drugs and the task of predicting the behaviour of pairs in a specific context of interest. Given an incomplete database of drug pairs, drug administration contexts and outcomes, the goal is to train a model to accurately make probabilistic predictions for unseen entries. The reasons for answering these questions via algorithmic methods are multi-fold. Firstly, testing all drug pairs in all of the contexts is not feasible due to time and financial constraints such as drug prices and labour costs [Preuer *et al.*, 2018]. Secondly, certain pair scoring tasks such as polypharmacy side effect prediction can only be validated in human-based trials. Finally, laboratory testing of drug pairs is prone to human errors [Liu *et al.*, 2020].

Traditional supervised machine learning methods which solve the drug pair scoring task use handcrafted molecular features to predict the outcome of administering the drugs together in a specific context [Sidorov *et al.*, 2019; Chiang *et al.*, 2020]. Another group of techniques uses an unsupervised approach which diffuses the profile of the drug pair on a heterogeneous biological graph [Zhang *et al.*, 2017; Li *et al.*, 2018; Huang *et al.*, 2019] in order to find potential polypharmacy, synergy or interaction indications. Deep learning techniques which solve the drug pair scoring task can be seen as a fusion and extension of these traditional methods. Such models first generate drug representations based either on molecular structure or the heterogeneous graph based neighbourhood context. In the second optional step, these representations are propagated in the biological graph and aggregated. Finally, drug pair representations are formed and probability scores are outputted in the specific drug administration contexts. We present a high level summary of the drug pair scoring task idea in Figure 1.

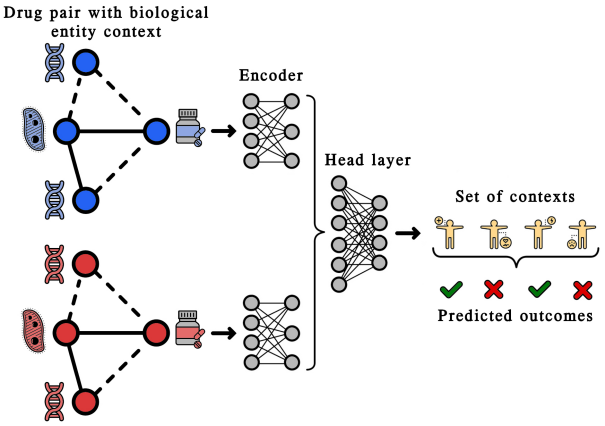

Figure 1: Drug-drug interaction, polypharmacy side effect, and pair combination therapy design prediction tasks follow the same template. Given a pair of drugs with optional biological context, the task is to predict an outcome in a specific application domain. Relational machine learning models which solve these task can exploit molecular features, knowledge graph based neighbourhoods or both.

Our main contributions can be summarized as:

1. We provide a unification of drug-drug interaction, polypharmacy side effect and synergistic combination prediction tasks.
2. We present an overview on the design of relational machine learning models which can address these predictive tasks.

3. We highlight the publicly available datasets used to train and test the models on these tasks and survey the literature for the most commonly used evaluation metrics.
4. We review the most important applications of these techniques and discuss directions for future research in the domain.

The remainder of this survey is structured as follows. In Section 2 we establish the foundations of a unified view of discriminative machine learning tasks defined on pairs of drugs. Section 3 discusses the architectural details of models that can solves these tasks. The evaluation metrics, protocols and datasets used in the literature are detailed in Section 4. Several important key application areas are highlighted in Section 5. We discuss the limitations of current approaches and future research directions in Section 6. The paper concludes with Section 7. The survey is supported by a collection of related work under the https://github.com/AstraZeneca/awesome-drug-pair-scoring repository.

2 Background

Our discussion of drug pair scoring models requires the introduction of a drug set $\mathcal{D} = \{d_1, \ldots, d_n\}$ that describes compounds of interest and a context set $\mathcal{C} = \{c_1, \ldots, c_k\}$ that contains contexts where two drugs are used in a pair combination.

Definition 1. ***Labeled drug pair.*** *A labeled drug pair defined on drug set $\mathcal{D}$ and context set $\mathcal{C}$ is the tuple $(d, d', c, y_{d,d',c})$, where the binary indicator $y_{d,d',c} \in \{0, 1\}$ is the outcome for drug pair $d, d' \in \mathcal{D}$ in context $c \in \mathcal{C}$.*

A labeled drug pair is a known fact about the drug pair having an effect in a context such as a specific polypharmacy side effect, interaction or synergistic relationship at treating a disease. The purpose of pair scoring models is to learn from these tuples to predict the labels for unlabeled drug pairs and contexts.

Definition 2. ***Database of labeled drug pairs.*** *A database of labeled drug pairs defined on drug and context sets $\mathcal{D}$ and $\mathcal{C}$ is the set $\mathcal{S}$ containing labeled drug pairs $(d, d', c, y_{d,d',c})$ where $d, d' \in \mathcal{D}$ and $c \in \mathcal{C}$.*

Pair scoring models are trained on databases of labeled drug pairs and the trained models are used to predict the label of pairs for which we do not know the outcome in certain contexts.

Definition 3. ***Heterogeneous interaction graph with drug entities.*** *We denote with $G(\mathcal{V}, \mathcal{R}, \mathcal{E})$ the heterogeneous interaction graph with drug entities, where $\mathcal{V}$ and $\mathcal{R}$ are the entity and relation sets, it holds that the drug set $\mathcal{D} \subset \mathcal{V}$ and $\mathcal{E}$ is formed by typed edges of the form $(v, r, u) \in \mathcal{V} \times \mathcal{R} \times \mathcal{V}$.*

We consider a heterogeneous graph where the drug set is a subset of the vertex set. This definition of heterogeneous (biological) knowledge graph helps to create knowledge graph based representation for the compounds of interest.

Definition 4. ***Neighbourhood encoder.*** *A neighbourhood encoder is the function:*

$$\mathbf{h}_d = \text{AGGREGATE}(\Theta_u, \forall u \in \mathcal{N}(d)). \tag{1}$$

In Equation (1) Θ_u is a parametric vector representation of $u \in \mathcal{V}$ and $\mathcal{N}(\cdot)$ is a neighbourhood set.

The neighbourhood encoder function [Hamilton *et al.*, 2017] creates a vector representation of drug vertices of the graph based on the aggregation of trainable parameter vectors in the neighbourhood of the source node. Neighbourhoods of a drug can be defined based on arbitrary notions of proximity and the aggregation itself could be a parametric transformation.

Definition 5. ***Molecular encoder.*** *A molecular encoder is the function $\boldsymbol{h}_d = h_\Theta(\mathcal{M}_d)$, parametrized by Θ where $\boldsymbol{h}_d$ is the learned vector representation and $\mathcal{M}_d$ is a generic notation of molecular features describing the drug d.*

A molecular encoder is a neural network which generates a vector representation from the features of the molecule - these molecular features can be derived from generic features (e.g. hydrophilicity), a string representation, molecular graph or geometry.

Definition 6. ***Neighbourhood informed molecular encoder.*** *This encoder is the function :*

$$\boldsymbol{h}_d = \text{AGGREGATE}(h_\Theta(\mathcal{M}_u), \forall u \in \mathcal{N}(d));$$

where $h_\Theta(\mathcal{M}_u)$ and $\text{AGGREGATE}(\cdot, \forall u \in \mathcal{N}(d))$ are molecular and neighbourhood encoders respectively.

This encoder combines the layers described in Definitions 4 and 5. It is essentially a neighbourhood encoder parametrized by representations outputted by a molecular encoder – molecular representations learned by the molecular encoder are aggregated in the neighbourhood of source drug nodes in the knowledge graph which has drug entities.

Definition 7. ***Molecular representation combiner.*** *Given the drugs $d, d' \in \mathcal{D}$ with vector representations $\boldsymbol{h}, \boldsymbol{h}_{d'}$ the molecular representation combiner is the function $\boldsymbol{h}_{d,d'} = g_\Theta(\boldsymbol{h}_d; \boldsymbol{h}_{d'})$ that outputs $\boldsymbol{h}_{d,d'}$ a vector representation of the drug pair.*

The representation output by this combiner function can be drug orchestration order dependent. This way the temporal order of drug orchestration can be expressed by the pair scoring model. For example the concatenation of drug vectors results in order dependent representations of pairs, while a bilinear transformation of drug representations with a diagonal matrix does not.

Definition 8. ***Scoring head layer.*** *The scoring head layer is the function $\hat{y}_{d,d',c} = k_\Theta(\boldsymbol{h}_{d,d'}, c)$, where the predicted label proability satisfies that $\{\hat{y}_{d,d',c} \in \mathbb{R} \mid 0 \leq \hat{y}_{d,d',c} \leq 1\}$.*

Given a drug pair representation and a context, the scoring head layer outputs a probability score for the outcome.

Definition 9. ***Drug pair scoring loss and cost functions.*** *Given the drug pair $d, d' \in \mathcal{D}$, context $c \in \mathcal{C}$, ground-truth label $y_{d,d',c}$ and predicted score $\hat{y}_{d,d',c}$ the loss is defined as the function $\ell(y_{d,d',c}; \hat{y}_{d,d',c})$. The cost on the whole drug pair database $\mathcal{S}$ is defined by Equation (2).*

$$\mathcal{L} = \sum_{(d,d',c) \in \mathcal{S}} \ell(y_{d,d',c}; \hat{y}_{d,d',c}). \tag{2}$$

In practical settings, drug pair scoring models are trained by the minimization of the binary cross-entropy summed over the labeled drug pair, context triples.

					Entity Types			Drug Features			
Task	**Method**	**Reference**	**Model view**	**Induction**	**Drug**	**Protein**	**Disease**	**SMILES**	**Graph**	**Geometry**	**Generic**
Polypharmacy	DECAGON	[Zitnik *et al.*, 2018a]	Higher		•	•					
	KBLRN	[Malone *et al.*, 2018]	Higher		•	•					
	SDHINE	[Hu *et al.*, 2018]	Higher		•	•	•				
	ESP	[Burkhardt *et al.*, 2019]	Higher		•	•					
	MHCADDI	[Deac *et al.*, 2019]	Lower	•					•		
	TIP	[Xu *et al.*, 2020]	Higher		•	•					
Interaction	DeepCCI	[Sunyoung *et al.*, 2017]	Lower	•				•			•
	MVGAE	[Ma *et al.*, 2018]	Hierarchical	•	•	•	•	•			
	DeepDDI	[Ryu *et al.*, 2018]	Lower	•				•			
	D^3I	[Peng *et al.*, 2019]	Hierarchical		•						•
	MR-GNN	[Xu *et al.*, 2019]	Lower	•					•		
	SkipGNN	[Huang *et al.*, 2020b]	Higher		•	•	•				
	CASTER	[Huang *et al.*, 2020a]	Lower	•				•			
	DeepDrug	[Cao *et al.*, 2020]	Lower	•				•	•	•	
	GoGNN	[Wang *et al.*, 2020]	Hierarchical		•				•		
	DPDDI	[Feng *et al.*, 2020]	Lower	•							•
	KGNN	[Lin *et al.*, 2020]	Higher		•	•	•				
	BiGNN	[Bai *et al.*, 2020]	Hierarchical	•	•	•			•		
	MIRACLE	[Wang *et al.*, 2021b]	Hierarchical	•	•				•		
	EPGCN-DS	[Sun *et al.*, 2020a]	Lower	•							
	SumGNN	[Yu *et al.*, 2021]	Higher		•	•	•				
	DANN-DDI	[Liu *et al.*, 2022]	Higher		•	•					
	SSI-DDI	[Nyamabo *et al.*, 2021]	Lower						•		
	MTDDI	[Feng *et al.*, 2021]	Hierarchical		•	•	•				•
	MUFFIN	[Chen *et al.*, 2021]	Hierarchical		•	•	•	•			
	DDIAAE	[Dai *et al.*, 2021]	Higher		•	•	•				
	RWGCN	[Feeney *et al.*, 2021]	Higher		•	•	•				
	SmileGNN	[Han *et al.*, 2022]	Hierarchical		•	•		•			
	GCN-BMP	[Chen *et al.*, 2020]	Lower	•	•			•			
Synergy	DeepSynergy	[Preuer *et al.*, 2018]	Lower	•							•
	MCDC	[Chen *et al.*, 2019]	Hierarchical		•	•		•			
	DTF	[Sun *et al.*, 2020b]	Higher		•						
	DeepSignalFlow	[Zhang *et al.*, 2021a]	Hierarchical		•	•	•				
	DeepDDS	[Wang *et al.*, 2021a]	Lower	•				•			
	GraphSynergy	[Yang *et al.*, 2021]	Higher		•	•					
	TranSynergy	[Liu *et al.*, 2021]	Hierarchical	•	•	•					•
	MatchMaker	[Brahim *et al.*, 2021]	Lower	•				•			•
	DNNSynergy	[Zhang *et al.*, 2021b]	Lower	•				•			
	AID	[Kim *et al.*, 2021]	Hierarchical	•	•	•		•			
	MOOMIN	[Rozemberczki *et al.*, 2021a]	Hierarchical	•	•	•			•		

Table 1: A machine learning task, model view level, induction, interaction graph node type (entity) and drug feature based comparison of drug pair scoring machine learning models. Machine learning models that solve a specific pair scoring task are ordered chronologically in the table.

3 Drug Pair Scoring Models

Our discussion of the drug pair scoring models introduces our unified view about the general architecture of these models and compares state-of-the-art architecture designs.

3.1 Unified View: The Drug Pair Scoring Model

Based on the definitions outlined in Section 2 we propose a unified view of drug pair scoring models. We postulate that the abstract design of drug pair scoring models irrespective of the specific subtask solved always has the following architecture:

1. An encoder to generate drug representations - this can be one of the functions described by Definitions 4, 5 and 6.
2. A molecular representation combiner function to generate a drug pair representation – see Definition 7.
3. The scoring head layer to predict the probability of a context dependent outcome proposed by Definition 8.
4. The loss function of Definition 9 which depends on ground-truth labels and the probabilities output by the head layer.

This architecture and design allows for the joint end-to-end training of the individual model components – gradient descent based update of the layer weights.

3.2 Specific Architecture Designs

We compare state-of-the-art model architectures in Table 1 that can solve pair scoring tasks. Our comparison considers the model level, induction capabilities, specific subtask, node types of the heterogeneous graph and the molecular features exploited by the model. Model attributes used for comparison were the following:

- *Model level:* A model operates at the following levels based on the encoder architecture used for generating the drug representations: (a) higher-view – *neighbourhood encoder*, (b) lower-view – *molecular encoder*, (c) hierarchical-view – *neighbourhood informed molecular encoder*.
- *Machine learning task:* The drug pair scoring task of interest solved by the dedicated model architecture proposed in the research paper. It has to be one of interaction, polypharmacy or synergy prediction.
- *Induction:* A model is inductive if it can predict the label of drug pairs where at least one of the drugs was not in the training set drug pairs.
- *Entities:* The types of hetereogeneous graph entities (*drugs*, *proteins*, *diseases*) used by the model to solve the task.
- *Drug features:* Molecular features and information about the compound encoded by the molecular encoder function.

Our comparison highlights that there is a hard trade-off between induction and exclusion of compound features. It is also evident that there is a connection between the subtask and architecture design: for example, polypharmacy side effect prediction models are mostly high level transductive neighbourhood encoders with a scoring layer on top. Synergy scoring models are mostly inductive techniques which exploit the molecular information about the drugs. Currently, there is no single pair scoring model which includes all of the considered biological modalities.

4 Evaluation

The evaluation of machine learning models requires performance metrics, train-test split strategies and publicly accessible datasets.

4.1 Performance Metrics

The predictive performance of drug pair scoring models is evaluated by metrics tailored to binary classification tasks. We summarise how these metrics are used for the evaluation of state-of-the-art drug pair scoring architectures in Table 2.

		Evaluation metric					
Model	**Reference**	**AUPRC**	**AUROC**	**Precision**	**Recall**	**Accuracy**	**F_1**
DECAGON	[Zitnik *et al.*, 2018a]	•	•	•			
KBLRN	[Malone *et al.*, 2018]	•	•	•			
SDHINE	[Hu *et al.*, 2018]		•	•			
ESP	[Burkhardt *et al.*, 2019]	•	•	•			
MHCADDI	[Deac *et al.*, 2019]		•				
TIP	[Xu *et al.*, 2020]	•	•	•			
DeepCCI	[Sunyoung *et al.*, 2017]		•			•	•
MVGAE	[Ma *et al.*, 2018]	•	•				
DeepDDI	[Ryu *et al.*, 2018]			•	•	•	•
D^3I	[Peng *et al.*, 2019]			•	•	•	•
MR-GNN	[Xu *et al.*, 2019]		•		•		•
SkipGNN	[Huang *et al.*, 2020b]	•					•
CASTER	[Huang *et al.*, 2020a]	•	•				•
DeepDrug	[Cao *et al.*, 2020]	•	•				
GoGNN	[Wang *et al.*, 2020]		•	•			
DPDDI	[Feng *et al.*, 2020]	•	•	•	•	•	•
KGNN	[Lin *et al.*, 2020]	•	•			•	•
BiGNN	[Bai *et al.*, 2020]		•	•			•
MIRACLE	[Wang *et al.*, 2021b]	•	•				•
EPGCN-DS	[Sun *et al.*, 2020a]		•				•
SumGNN	[Yu *et al.*, 2021]	•	•	•			•
DANN-DDI	[Liu *et al.*, 2022]	•	•	•	•	•	•
SSI-DDI	[Nyamabo *et al.*, 2021]	•	•			•	
MTDDI	[Feng *et al.*, 2021]	•	•	•	•	•	•
MUFFIN	[Chen *et al.*, 2021]		•	•	•	•	•
DDIAAE	[Dai *et al.*, 2021]	•	•	•			
RWGCN	[Feeney *et al.*, 2021]	•	•				
SmileGNN	[Han *et al.*, 2022]		•			•	•
GCN-BMP	[Chen *et al.*, 2020]	•	•				•
DeepSynergy	[Preuer *et al.*, 2018]	•	•	•		•	
MCDC	[Chen *et al.*, 2019]	•	•				
DTF	[Sun *et al.*, 2020b]	•	•	•		•	
DeepSignalFlow	[Zhang *et al.*, 2021a]					•	
DeepDDS	[Wang *et al.*, 2021a]	•	•	•		•	
GraphSynergy	[Yang *et al.*, 2021]	•	•		•	•	•
TranSynergy	[Liu *et al.*, 2021]	•	•				
MatchMaker	[Brahim *et al.*, 2021]	•	•				
DNNSynergy	[Zhang *et al.*, 2021b]	•	•	•		•	
AID	[Kim *et al.*, 2021]	•					
MOOMIN	[Rozemberczki *et al.*, 2021a]	•	•				•

Table 2: Predictive performance evaluation metrics used by the research papers which proposed novel drug pair scoring techniques. Models are grouped by the pair scoring task and ordered chronologically.

Looking at Table 2 it is evident that the evaluation metrics used in the literature can be grouped into two categories:

- *Score based metrics:* These quantify predictive performance based over the whole domain of discrimination thresholds. The precision-recall area under the curve (AUPRC) considers the precision-recall trade off under the whole domain of discrimination thresholds while the receiver operating characteristic area under the curve (AUROC) considers false and true positive rates.
- *Hard cut off evaluation metrics:* These performance metrics (accuracy, F_1 score, precision, recall) apply a hard discrimination threshold to assign a label to the data points based on the scores output by the pair scoring model. In order to calculate these, one needs to set a discrimination threshold.

Our findings demonstrate that pair scoring models are predominantly evaluated by score based metrics (AUPRC and AUROC) which do not require manual setting of a discrimination threshold. It is also evident that seminal research works which defined the key pair scoring tasks influenced the later evaluation metric choices – polypharmacy prediction models adapted the evaluation metrics from [Zitnik *et al.*, 2018a] for example.

4.2 Train-Test Split Strategies

The evaluation of drug pair scoring tasks allows for the use of various train-test split strategies [Preuer *et al.*, 2018] to test the performance of the model under cold-start and inductive scenarios [Dewulf *et al.*, 2021]. Given a labeled drug pair-context database $\mathcal{S}$, defined on the drug and context sets $\mathcal{D}$ and $\mathcal{C}$, we assume that one can create the randomized splits $\mathcal{S}_{Train}$ and $\mathcal{S}_{Test}$. We summarized these splitting strategies in Figure 2.

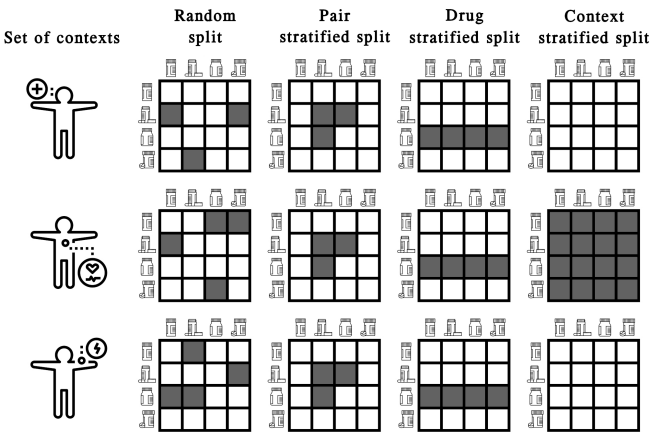

Figure 2: The train and test split of drug pair scoring datasets allows for various stratified splits. Stratification of the pairs can happen across the evaluated drugs or the outcomes being tested. Drug pairs used for training the pair scoring model are denoted with white, pairs used for testing are denoted with gray.

Using the formalism established to describe the pair scoring models, the splitting strategies are defined as:

- **Random split:** labeled drug pair - context entries of $\mathcal{S}$ are randomly split between $\mathcal{S}_{Train}$ and $\mathcal{S}_{Test}$.
- **Drug pair stratified split:** A drug pair $d, d' \in \mathcal{D}$ that appears in entries of S_{Train} does not appear in entries of $\mathcal{S}_{Test}$. This split requires a pair scoring model which is inductive with respect to drugs.
- **Drug stratified split:** A drug $d \in \mathcal{D}$ that appears in entries of $\mathcal{S}_{Train}$ does not appear in entries of $\mathcal{S}_{Test}$. Like the drug pair stratified split this requires the model to be inductive with respect to new drugs.
- **Context stratified split:** A context $c \in \mathcal{C}$ that appears in entries of $\mathcal{S}_{Train}$ does not appear in entries of $\mathcal{S}_{Test}$. This requires that the pair scoring model is inductive with respect to the set of contexts.

4.3 Datasets

We detail public sources for drug pair data which have been used by the approaches in this review in Table 3. Datasets are listed chronologically according to subtask and the licence and any restrictions for commercial use are detailed where available. It

can be seen that the majority of datasets contain a small number of drugs, indicating most focus on approved drugs rather than all possible compounds, with the interactions captured in drug pairs being much more numerous.

It should be noted that established resources such as TWOSIDES and DrugBank are frequently filtered, cleaned and split into new datasets. For example the Therapeutics Data Commons (TDC) resource contains filtered versions of both of these datasets designed for benchmark use [Huang *et al.*, 2021]. It is also common for datasets to be named differently in publications, for example the split of TWOSIDES contained in TDC is also called ChChSe-Decagon in some works [Zitnik *et al.*, 2018b].

5 Applications

In this section we introduce three key, yet currently largely, unexplored applications for the methods detailed in this review.

5.1 Combination Therapy for COVID-19

One topical application of these methods is in relation to COVID-19 pandemic. Patients affected by polypharmacy of certain drug types (anti-psychotics and opiates being prominent examples) had a significantly higher chance of a negative clinical outcome from COVID-19 [Iloanusi *et al.*, 2021; Jin *et al.*, 2021]. Using methods covered in this review to predict which combinations may have a negative effect for COVID-19 patients, could enable high risk groups to seek alternative treatments, reducing the risk of a negative outcome.

5.2 Antibiotic Evolutionary Pressure

The prevalent use of antibiotics has resulted in microbes evolving resistance to the drugs, reducing efficacy and potentially eliminating cost effective ways of treating severe bacterial-related diseases such as Tuberculosis. Interestingly, it has been shown that the combination of different antibiotics can slow, and even reverse, this evolutionary resistance [Singh *et al.*, 2017]. However discovering these suppressive interactions using traditional methods is a complex and slow process, yet one currently unexplored using the methods covered in this review.

5.3 Reducing Toxicity

Although drug combinations can result in an increase of unwanted side effects, one promising application is that the combination of two or more drugs can actually lead to a reduced level of toxicity for patients. This is due to the fact that synergistic drugs, which together posses a higher level of efficacy at targeting a certain condition, means that the levels of each individual compound can actually be lowered, reducing toxicity issues associated with higher doses [Ianevski *et al.*, 2020]. Thus, accurate prediction of synergistic drug combinations can reduce the impact of toxicity resulting from the individual compounds.

6 Discussion and Future Directions

The body of work regarding relational machine learning for drug pair scoring primarily focuses on the design of novel architectures and applications. Our unification survey identified a number of potential shortcomings of existing approaches and venues for novel research in the domain.

6.1 Encoding Molecular Geometry

Our summary on the design of relational machine learning architectures for drug pair scoring tasks in Table 1 highlighted that molecular geometry and spatial structure of the molecules is rarely encoded by existing models. Recent advances in geometric deep learning applied to chemistry [Qi *et al.*, 2017; Xie *et al.*, 2018; Fey *et al.*, 2018] would allow the inclusion of geometric information which could lead to better predictive performance on the pair scoring tasks. The modularity of existing architectures makes replacing the molecular encoders with state-of-the-art geometric encoder layers a possibility.

6.2 Higher Order Drug Combinations

Existing research about the interactions, unwanted side effects and synergy of drugs is primarily focused on the evaluation of binary pair combinations. This is driven by the lack of datasets focused on the outcomes of using higher order drug combinations and the lack of architectures designed specifically for these higher order combinations. By using set based representation aggregation layers [Vinyals *et al.*, 2016; Baek *et al.*, 2020], the existing pair scoring models could be adapted to generate drug subset representations.

6.3 Transfer Learning

Self-supervised and unsupervised learning for pretraining molecular encoders is already widely used for single molecule machine learning tasks [Dewulf *et al.*, 2021]. This provides an opportunity for pretraining the molecular encoders on single molecule tasks and fine-tuning them on the data scarce, pair scoring tasks. Another opportunity for transfer learning comes from the fact that certain pair scoring tasks have a greater quantity of labeled data available. The summary of drug pair scoring datasets in Table 2 demonstrated that the *drug-drug interaction prediction* task has datasets such as STITCH-CCI-5 which covers a large number of pair combinations, while the polypharmacy side effect and synergy prediction tasks have smaller databases. Pretraining models by performing drug-drug interaction prediction and fine-tuning these models for other tasks seems to be an important future research direction for training accurate, and therefore useful, models.

6.4 Multimodal Learning

A heterogeneous graph based representation of drugs allows for the fusion of multiple data modalities. Our survey of existing models in Table 1 has demonstrated that only a handful of existing architectures integrates multimodal data effectively [Rozemberczki *et al.*, 2021a; Liu *et al.*, 2021] without losing induction. Integrating multi-omics data such as proteomics, molecular structure and biological pathway information could be an important venue for designing novel pair scoring architectures.

6.5 Software for Drug Pair Scoring

Currently there is no dedicated open-source machine learning library which was specifically designed for solving the drug pair scoring task. Developing a dedicated relational machine learning framework on top of existing geometric deep learning [Fey *et al.*, 2019; Wang *et al.*, 2019; Rozemberczki *et al.*, 2021b] and deep chemistry frameworks [Ramsundar *et al.*, 2019; Korshunova *et al.*, 2021] could be an important contribution to the domain. This

Dataset	Reference	Subtask	Compounds	Pairs	Contexts	Licence	Restricted
TWOSIDES	[Tatonetti *et al.*, 2012]	Polypharmacy	1,918	211,990	12,726	Not Specified	
DrugBank 5	[Wishart *et al.*, 2018]	Polypharmacy	14,575	365,000+	86+	CC BY-NC 4.0	•
DeepDDI	[Ryu *et al.*, 2018]	Polypharmacy	1,710	191,995	86	Not Specified	
TDC (TWOSIDES)	[Huang *et al.*, 2021]	Polypharmacy	645	63,473	1,317	CC BY 4.0	
TDC (DrugBank)	[Huang *et al.*, 2021]	Polypharmacy	1,706	191,519	86	CC BY-NC 4.0	•
STITCH-CCI 5	[Szklarczyk *et al.*, 2016]	Interaction	389,393	17,705,799	4	CC BY-NC-SA 4.0	•
ZhangDDI	[Zhang *et al.*, 2017]	Interaction	548	48,548	-	Not Specified	
ChCh-Miner	[Zitnik *et al.*, 2018b]	Interaction	1,514	48,514	-	Not Specified	
DCDB	[Liu *et al.*, 2010]	Synergy	485	499	–	Not Specified	
DCDB 2.0	[Liu *et al.*, 2014]	Synergy	904	1,363	–	Not Specified	
ASDCD	[Chen *et al.*, 2014]	Synergy	105	215	–	Not Specified	
O'Neil	[O'Neil *et al.*, 2016]	Synergy	38	583	39	Not Specified	
NCI-ALMANAC	[Holbeck *et al.*, 2017]	Synergy	104	5,000+	60	CC BY 4.0	
DrugComb	[Zagidullin *et al.*, 2019]	Synergy	2,276	437,932	93	CC BY-NC-SA 4.0	
SynergyXDB	[Seo *et al.*, 2020]	Synergy	1,977	22,507	151	Not Specified	
DrugCombDB	[Liu *et al.*, 2020]	Synergy	2,887	448,555	124	Not Specified	•
TDC (OncoPolyPharm)	[Huang *et al.*, 2021]	Synergy	38	583	39	CC BY 4.0	
DrugComb 2.0	[Zheng *et al.*, 2021]	Synergy	8,397	751,498	2,320	CC BY-NC-SA 4.0	
TDC (DrugComb)	[Huang *et al.*, 2021]	Synergy	129	5,628	29	CC BY-NC-SA 4.0	

Table 3: Public drug pair scoring datasets ordered chronologically with the subtask, number of compounds, count of tested compound pairs, cardinality of the context set, licence and if commercial use is explicitly restricted.

would require curated datasets and the architectural design of encoder, combiner, scoring layers and drug pair iterators.

7 Conclusion

We have provided an exhaustive overview of relational machine learning models designed to solve drug pair scoring tasks. We outlined a general theoretical framework which unifies the drug-drug interaction, polypharmacy side effect and drug synergy prediction tasks and created a taxonomy of models which address these. By surveying the literature, considering the architecture and evaluation of existing models, we identified key real world application areas and important directions for future research.

Acknowledgements

The authors would like to thank Peizhen Bai, Piotr Grabowski, Haiping Lu, Rocío Mercado, Christopher Morris, and Paul Scherer for feedback throughout the preparation of this manuscript. Stephen Bonner is a fellow of the AstraZeneca postdoctoral program.

References

[Baek *et al.*, 2020] Jinheon Baek et al. Accurate Learning of Graph Representations with Graph Multiset Pooling. In *International Conference on Learning Representations*, 2020.

[Bai *et al.*, 2020] Yunsheng Bai et al. Bi-Level Graph Neural Networks for Drug-Drug Interaction Prediction. *ICML 2020 Graph Representation Learning and Beyond (GRL+) Workshop*, 2020.

[Brahim *et al.*, 2021] Kuru Halil Brahim et al. MatchMaker: A Deep Learning Framework for Drug Synergy Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021.

[Burkhardt *et al.*, 2019] Hannah A. Burkhardt et al. Predicting Adverse Drug-Drug Interactions with Neural Embedding of Semantic Predications. In *American Medical Informatics Association Annual Symposium*, 2019.

[Cao *et al.*, 2020] Xusheng Cao, Rui Fan, and Wanwen Zeng. DeepDrug: A General Graph-Based Deep Learning Framework for Drug Relation Prediction. *bioRxiv*, 2020.

[Chen *et al.*, 2014] Xing Chen et al. ASDCD: Antifungal Synergistic Drug Combination Database. *PloS one*, 9(1):e86499, 2014.

[Chen *et al.*, 2019] Huiyuan Chen et al. Large-Scale Analysis of Drug Combinations by Integrating Multiple Heterogeneous Information Networks. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '19, page 67–76, 2019.

[Chen *et al.*, 2020] Xin Chen et al. GCN-BMP: Investigating Graph Representation Learning for DDI Prediction Task. *Methods*, 179:47–54, 2020.

[Chen *et al.*, 2021] Yujie Chen et al. MUFFIN: Multi-Scale Feature Fusion For Drug–Drug Interaction Prediction. *Bioinformatics*, 2021.

[Chiang *et al.*, 2020] Wen-Hao Chiang et al. Drug-Drug Interaction Prediction Based on Co-Medication Patterns and Graph Matching. *International Journal of Computational Biology and Drug Design*, 13(1):36–57, 2020.

[Dai *et al.*, 2021] Yuanfei Dai et al. Drug–Drug Interaction Prediction with Wasserstein Adversarial Autoencoder-Based Knowledge Graph Embeddings. *Briefings in Bioinformatics*, 22(4):bbaa256, 2021.

[Deac *et al.*, 2019] Andreea Deac et al. Drug-Drug Adverse Effect Prediction with Graph Co-Attention. *ICML Workshop on Computational Biology*, 2019.

[Dewulf *et al.*, 2021] Pieter Dewulf et al. Cold-Start Problems in Data-Driven Prediction of Drug–Drug Interaction Effects. *Pharmaceuticals*, 14(5):429, 2021.

[Feeney *et al.*, 2021] Arthur Feeney et al. Relation Matters in Sampling: A Scalable Multi-Relational Graph Neural Network for Drug-Drug Interaction Prediction. *Graph Representation Learning and Beyond (GRL+) Workhop at ICML*, 2021.

[Feng *et al.*, 2020] Yue-Hua Feng, Shao-Wu Zhang, and Jian-Yu Shi. DPDDI: A Deep Predictor for Drug-Drug Interactions. *BMC Bioinformatics*, 21(1):1–15, 2020.

[Feng *et al.*, 2021] YueHua Feng et al. MTDDI: A Graph Convolutional Network Framework for Predicting Multi-Type Drug-Drug Interactions. *Research Square*, 2021.

[Fey *et al.*, 2018] Matthias Fey et al. SplineCNN: Fast Geometric Deep Learning with Continuous B-Spline Kernels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 869–877, 2018.

[Fey *et al.*, 2019] Matthias Fey et al. Fast Graph Representation Learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

[Gaudelet *et al.*, 2021] T Gaudelet et al. Utilizing Graph Machine Learning within Drug Discovery and Development. *Briefings in Bioinformatics*, 2021.

[Hamilton *et al.*, 2017] William L Hamilton et al. Inductive Representation Learning on Large Graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 1025–1035, 2017.

[Han *et al.*, 2022] Xueting Han et al. Smilegnn: Drug-drug interaction prediction based on the smiles and graph neural network. *Life*, 12(2), 2022.

[Holbeck *et al.*, 2017] Susan L Holbeck et al. The National Cancer Institute ALMANAC: A Comprehensive Screening Resource for the Detection of Anticancer Drug Pairs with Enhanced Therapeutic Activity. *Cancer research*, 77(13):3564–3576, 2017.

[Hu *et al.*, 2018] Baofang Hu et al. Adverse Drug Reaction Predictions Using Stacking Deep Heterogeneous Information Network Embedding Approach. *Molecules*, 23(12):3193, 2018.

[Huang *et al.*, 2019] Lei Huang et al. Driver Network as a Biomarker: Systematic Integration and Network Modeling of Multi-Omics Data. *Bioinformatics*, 35(19):3709–3717, 2019.

[Huang *et al.*, 2020a] Kexin Huang et al. CASTER: Predicting Drug Interactions with Chemical Substructure Representation. *AAAI*, 2020.

[Huang *et al.*, 2020b] Kexin Huang et al. SkipGNN: Predicting Molecular Interactions with Skip-Graph Networks. *Scientific reports*, pages 1–16, 2020.

[Huang *et al.*, 2021] Kexin Huang et al. Therapeutics Data Commons: Machine Learning Datasets and Tasks for Drug Discovery and Development. In *35th Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2021.

[Ianevski *et al.*, 2020] Aleksandr Ianevski et al. SynToxProfiler: An Interactive Analysis of Drug Combination Synergy, Toxicity and Efficacy. *PLoS computational biology*, 16(2):e1007604, 2020.

[Iloanusi *et al.*, 2021] Sorochi Iloanusi et al. Polypharmacy Among COVID-19 Patients: A Systematic Review. *Journal of the American Pharmacists Association*, 61(5):e14–e25, 2021.

[Jin *et al.*, 2021] Wengong Jin et al. Deep Learning Identifies Synergistic Drug Combinations for Treating COVID-19. *Proceedings of the National Academy of Sciences*, 118(39), 2021.

[Kim *et al.*, 2021] Yejin Kim et al. Anticancer Drug Synergy Prediction in Understudied Tissues Using Transfer Learning. *Journal of the American Medical Informatics Association*, 28(1):42–51, 2021.

[Korshunova *et al.*, 2021] Maria Korshunova et al. OpenChem: A Deep Learning Toolkit for Computational Chemistry and Drug Design. *Journal of Chemical Information and Modeling*, 61(1):7–13, 2021.

[Li *et al.*, 2018] Hongyang Li et al. Network Propagation Predicts Drug Synergy in Cancers. *Cancer research*, pages 5446–5457, 2018.

[Lin *et al.*, 2020] Xuan Lin et al. KGNN: Knowledge Graph Neural Network for Drug-Drug Interaction Prediction. In *IJCAI*, pages 2739–2745, 2020.

[Liu *et al.*, 2010] Yanbin Liu et al. DCDB: Drug Combination Database. *Bioinformatics*, 26(4):587–588, 2010.

[Liu *et al.*, 2014] Yanbin Liu et al. DCDB 2.0: A Major Update of the Drug Combination Database. *Database*, 2014, 2014.

[Liu *et al.*, 2020] Hui Liu et al. DrugCombDB: A Comprehensive Database of Drug Combinations Toward the Discovery of Combinatorial Therapy. *Nucleic acids research*, 48(D1):D871–D881, 2020.

[Liu *et al.*, 2021] Qiao Liu et al. TranSynergy: Mechanism-Driven Interpretable Deep Neural Network for the Synergistic Prediction and Pathway Deconvolution of Drug Combinations. *PLoS computational biology*, 17(2):e1008653, 2021.

[Liu *et al.*, 2022] S. Liu, Y. Zhang, et al. Enhancing drug-drug interaction prediction using deep attention neural networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, (01), 2022.

[Ma *et al.*, 2018] Tengfei Ma et al. Drug Similarity Integration Through Attentive Multi-View Graph Auto-Encoders. In *IJCAI*, pages 3477–3483, 2018.

[Malone *et al.*, 2018] Brandon Malone, Alberto García-Durán, and Mathias Niepert. Knowledge Graph Completion to Predict Polypharmacy Side Effects. In *International Conference on Data Integration in the Life Sciences*, pages 144–149. Springer, 2018.

[Nyamabo *et al.*, 2021] Arnold K Nyamabo et al. SSI–DDI: Substructure–Substructure Interactions for Drug–Drug Interaction Prediction. *Briefings in Bioinformatics*, 2021.

[O'Neil *et al.*, 2016] Jennifer O'Neil et al. An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies. *Molecular Cancer Therapeutics*, 15(6):1155–1162, 2016.

[Peng *et al.*, 2019] Bo Peng and Xia Ning. Deep Learning for High-Order Drug-Drug Interaction Prediction. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '19, page 197–206, 2019.

[Preuer *et al.*, 2018] Kristina Preuer et al. DeepSynergy: Predicting Anti-Cancer Drug Synergy with Deep Learning. *Bioinformatics*, 34(9):1538–1546, 2018.

[Qi *et al.*, 2017] Charles R Qi et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[Ramsundar *et al.*, 2019] Bharath Ramsundar et al. *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More*. O'Reilly Media, Inc., 2019.

[Rozemberczki *et al.*, 2021a] Benedek Rozemberczki et al. MOOMIN: Deep Molecular Omics Network for Anti-Cancer Drug Combination Therapy Design. 2021.

[Rozemberczki *et al.*, 2021b] Benedek Rozemberczki et al. PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, 2021.

[Ryu *et al.*, 2018] Jae Yong Ryu et al. Deep Learning Improves Prediction of Drug–Drug and Drug–Food Interactions. *Proceedings of the National Academy of Sciences*, 2018.

[Seo *et al.*, 2020] Heewon Seo et al. SYNERGxDB: An Integrative Pharmacogenomic Portal to Identify Synergistic Drug Combinations for Precision Oncology. *Nucleic Acids Research*, 48(W1):W494–W501, 2020.

[Sidorov *et al.*, 2019] Pavel Sidorov et al. Predicting Synergism of Cancer Drug Combinations Using NCI-ALMANAC Data. *Frontiers in chemistry*, 7:509, 2019.

[Singh *et al.*, 2017] Nina Singh et al. Suppressive Drug Combinations and Their Potential to Combat Antibiotic Resistance. *The Journal of Antibiotics*, 70(11):1033–1042, 2017.

[Sun *et al.*, 2020a] Mengying Sun et al. Structure-Based Drug-Drug Interaction Detection via Expressive Graph Convolutional Networks and Deep Sets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13927–13928, 2020.

[Sun *et al.*, 2020b] Zexuan Sun et al. DTF: Deep Tensor Factorization for Predicting Anticancer Drug Synergy. *Bioinformatics*, 36(16):4483–4489, 2020.

[Sunyoung *et al.*, 2017] Kwon Sunyoung and Yoon Sungroh. DeepCCI: End-to-End Deep Learning for Chemical-Chemical Interaction Prediction. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 203–212, 2017.

[Szklarczyk *et al.*, 2016] Damian Szklarczyk et al. STITCH 5: Augmenting Protein–Chemical Interaction Networks with Tissue and Affinity Data. *Nucleic Acids Research*, 44(D1):D380–D384, 2016.

[Tatonetti *et al.*, 2012] Nicholas P Tatonetti et al. Data-Driven Prediction of Drug Effects and Interactions. *Science translational medicine*, 4(125), 2012.

[Vinyals *et al.*, 2016] Oriol Vinyals et al. Order Matters: Sequence to Sequence for Sets. In *4th International Conference on Learning Representations*, 2016.

[Wang *et al.*, 2019] Minjie Wang et al. Deep Graph Library: Towards Efficient and Scalable Deep Learning on Graphs. 2019.

[Wang *et al.*, 2020] Hanchen Wang et al. GoGNN: Graph of Graphs Neural Network for Predicting Structured Entity Interactions. In Christian Bessiere, editor, *IJCAI*, pages 1317–1323, 2020.

[Wang *et al.*, 2021a] Jinxian Wang and othersi. DeepDDS: Deep Graph Neural Network with Attention Mechanism to Predict Synergistic Drug Combinations. *Briefings in Bioinformatics*, 09 2021.

[Wang *et al.*, 2021b] Yingheng Wang et al. Multi-view Graph Contrastive Representation Learning for Drug-Drug Interaction Prediction. In *Proceedings of the Web Conference 2021*, pages 2921–2933, 2021.

[Wishart *et al.*, 2018] David S Wishart et al. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic acids research*, 46(D1):D1074–D1082, 2018.

[Xie *et al.*, 2018] Tian Xie et al. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters*, 120(14), 2018.

[Xu *et al.*, 2019] Nuo Xu et al. MR-GNN: Multi-Resolution and Dual Graph Neural Network for Predicting Structured Entity Interactions. *IJCAI*, 2019.

[Xu *et al.*, 2020] Hao Xu et al. Tri-Graph Information Propagation for Polypharmacy Side Effect Prediction. 2020.

[Yang *et al.*, 2021] Jiannan Yang et al. GraphSynergy: Network Inspired Deep Learning Model for Anti–Cancer Drug Combination Prediction. 2021.

[Yu *et al.*, 2021] Yue Yu et al. SumGNN: Multi-Typed Drug Interaction Prediction via Efficient Knowledge Graph Summarization. *Bioinformatics*, 2021.

[Zagidullin *et al.*, 2019] Bulat Zagidullin et al. DrugComb: An Integrative Cancer Drug Combination Data Portal. *Nucleic acids research*, 47(W1):W43–W51, 2019.

[Zhang *et al.*, 2017] Wen Zhang et al. Predicting Potential Drug-Drug Interactions by Integrating Chemical, Biological, Phenotypic and Network Data. *BMC Bioinformatics*, pages 1–12, 2017.

[Zhang *et al.*, 2021a] Heming Zhang et al. Mining Signaling Flow to Interpret Mechanisms of Synergy of Drug Combinations Using Deep Graph Neural Networks. *bioRxiv*, 2021.

[Zhang *et al.*, 2021b] Tianyu Zhang et al. Synergistic Drug Combination Prediction by Integrating Multiomics Data in Deep Learning Models. In *Translational Bioinformatics for Therapeutic Development*, pages 223–238. Springer, 2021.

[Zheng *et al.*, 2021] Shuyu Zheng et al. DrugComb Update: A More Comprehensive Drug Sensitivity Data Repository and Analysis Portal. *Nucleic Acids Research*, 2021.

[Zitnik *et al.*, 2018a] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling Polypharmacy Side Effects with Graph Convolutional Networks. *Bioinformatics*, 34(13):i457–i466, 06 2018.

[Zitnik *et al.*, 2018b] Marinka Zitnik et al. BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. http://snap.stanford.edu/biodata, August 2018.