Side Information in Drug–Target Interaction Prediction

MSc (Tech) Thesis
University of Turku
Department of Computing
Biomedical Engineering
and Health Technology
May 2025
Mikko Happonen

Supervisors:
Assoc. Prof. Antti Airola
MSc Riikka Numminen

UNIVERSITY OF TURKU
Department of Computing

Mikko Happonen: Side Information in Drug–Target Interaction Prediction

MSc (Tech) Thesis, 48 p.
Biomedical Engineering
and Health Technology
May 2025

Artificial intelligence constantly offers new ways to enhance drug discovery. The computing power of advanced models helps in predicting potential interactions of drugs and targets non-experimentally and on a large scale. This thesis presents a drug–target interaction prediction model that utilises binary identifiers of drugs and targets. Side information is also added to the model to determine whether it improves the model's prediction performance compared to using only binary identifiers. The model is based on a supervised learning method, factorisation machine, that captures and predicts the pairwise interactions of drugs and targets. Side information refers to additional data that describes the chemical properties of drug and target molecules. The side information is obtained from external data, whereas binary identifiers and the interaction labels are derived from interaction matrix describing known interactions of drugs and targets. The proposed model tries to find patterns that map drug–target data to labels. The drug–target data has three feature options: using only binary identifiers, using only side information, and using both binary identifiers and side information. The results indicate that the proposed model's performance, measured with C-index, is relatively high, and the addition of the side information does not systematically enhance the predictions. Enhancement caused by side information can be seen only in some datasets. Potential model improvements and future work possibilities are also discussed.

Keywords: drug–target interaction prediction, matrix factorisation, side information

Contents

List of Acronyms

DTI Drug–Target Interaction

ML Machine Learning

FM Factorisation Machine

DL Deep Learning

CV Cross-Validation

GPCR G Protein–Coupled Receptor

KI Kinase Inhibitor

OHE One-Hot Encoding

AUPRC Area Under Precision–Recall Curve

AUC Area Under Receiver Operating Characteristic Curve

SMILES Simplified Molecular Input Line Entry System

1 Introduction

Defining interactions of drugs and targets plays a key role in drug discovery processes. In traditional ways—using biological experiments in a laboratory—it is challenging, time-consuming, and expensive. Drug–target interaction (DTI) prediction is a way to predict potential interactions non-experimentally and on a large scale with computational models. (Chen et al. 2016)

Main components in general DTI prediction can be represented, for example, as drugs $\mathbf{x}_d$, targets $\mathbf{x}_t$, drug–target pairs $\mathbf{x} = (\mathbf{x}_d, \mathbf{x}_t)$, and their interactions as labels y (Pahikkala et al. 2015). Both $\mathbf{x}$ and y are derived from interaction matrix $\mathbf{Y}$ describing known interactions. In this thesis, the interaction predictions are made inside the $\mathbf{Y}$, and with regression, so the y is real-valued.

This kind of prediction is made with paired data, and the interaction predictions can be made with supervised machine learning (ML) models that generalise predictions based on known interactions as training data (Marsland 2014, pp. 5–6). In this thesis, a drug and a target form a pair $\mathbf{x}$, predictions are the interactions y between them, and the algorithm that the predictions are made with, is matrix factorisation. It is implemented with a factorisation machine (FM)—an ML model that combines the advantages of support vector machines with factorisation models, and can predict parameters on large datasets and with any real-valued feature vectors as input (Rendle 2010). Thus, DTI prediction can be seen as a subset of ML—objective is to learn a mapping from drug–target pairs to their corresponding interaction values.

DTI prediction is based on feature representations describing drugs, targets, and interactions derived from interaction matrix. Additionally, side information can be included in the process. Side information describes additional chemical properties of drugs and targets, and is used to obtain potentially more accurate predictions. It is obtained from external source, and added to feature representations.

Identifying interactions between drugs and targets offers numerous possibilities. In addition to finding new drugs, it can also help in repurposing existing ones for therapeutic applications. However, the complex molecular and biological mechanisms underlying these interactions make it hard to develop perfect predictive models. Incorporating side information of drugs and targets can provide better, and more accurate representations that capture the interaction patterns—this may improve the generalisability of the models. Determining whether side information consistently enhances prediction performance could lead towards more robust and realistic DTI prediction models. The use of side information in computational drug discovery tools has the potential to streamline drug discovery processes, reduce reliance on costly experimental methods, and accelerate the identification of effective treatments.

Thus, the purpose of this thesis is to find out if using side information instead of only the interaction matrix based features improves prediction. In particular, the thesis presents a way to implement a FM based DTI prediction model with incorporation of side information. This new, enhanced model is then used with existing drug–target datasets. The thesis aims to find an answer to the following research question.

Research question: Does the incorporation of side information improve drug–target interaction prediction compared to using only factorisation of an interaction matrix?

In general, DTI prediction is a technology that helps to automate the research of new drugs and their effectiveness. It accelerates the drug discovery process that is typically time- and resource-intensive. In addition to acceleration of drug discovery,

the possibility to leverage new technological advancements—such as ML and deep learning (DL)—is an advantage of DTI prediction and can complement laboratory methods (Shi et al. 2024). Overall, DTI prediction can serve as a useful tool in the fields of biomedical engineering and health technology to promote health.

In this thesis, Chapter 2 presents background of the paired data prediction in ML with FM, as well as ML based DTI prediction including details of side information. Chapter 3 takes a closer look at related work where FM and side information are used as part of DTI prediction. The actual model and datasets used for DTI prediction are introduced in Chapter 4. Chapters 5, 6, and 7 report the results, discuss them, and conclude the whole thesis.

2 Background

The goal of this thesis is to determine whether using side information improves DTI prediction. In order to incorporate side information, it is necessary to understand the basic concepts of the DTI prediction. Background introduced in this chapter lays the foundation for understanding how drugs and targets are structured into data, and used in the DTI prediction model.

In this chapter, the essential concepts needed to understand DTI prediction are presented. Supervised ML, paired data prediction, FM, ML based DTI prediction, and side information are covered.

2.1 Paired Data in Machine Learning

When original data is labelled, and the goal of the ML model is to predict the labels of new, unseen, and unlabelled data, this is called supervised ML (Nelson 2023, pp. 53). In this thesis, the input data is concatenated vector $\mathbf{x}^{(i,j)} = (\mathbf{x}_d^{(i)}, \mathbf{x}_t^{(j)})$. Interactions are real-valued labels $y^{(i,j)}$, and the ML model tries to identify a general pattern to map $\mathbf{x}^{(i,j)}$ to $y^{(i,j)}$. All presented vectors are by default assumed to be row vectors, unless otherwise specified—this deviates from the conventional assumption of vectors but aligns better with the structure of the used feature matrices. Paired data models in supervised ML are used in different domains as recommender systems. An example is the Netflix Prize competition, where the goal was to create a

recommender system based on paired data: users' ratings of different movies. With a supervised ML approach, it is possible to model user-object interactions to predict user preferences. (Koren, Bell & Volinsky 2009) This same high-level idea can be applied in DTI prediction. Now the context is different, but principles are the same—to model interactions between drugs and targets using paired data.

2.1.1 Supervised Machine Learning

Briefly, building of supervised ML model consists of creating the model, training the model and then evaluating the model (Alzubi, Nayyar & Kumar 2018). Before these steps, the data is acquired, cleaned, and organised before feeding it to the model—and this part actually takes the majority of the time (Nelson 2023, p. 54). The model is created so that the original data serves as input to a chosen algorithm that produces then the label of it as an output. Most of ML models have model parameters that are based on the original data and optimised during training. (Alzubi, Nayyar & Kumar 2018) The ready, trained model is deployed to production environment where it can be used as a predictive model (Géron 2019, pp. 80–81).

To estimate realistic performance on unseen data, the original dataset can be split into training and testing purposes (Berthold et al. 2010, p. 102). This so-called train-test split is done with predetermined ratio, and the training set is used only for training the model and the testing set is used only for evaluating the model.

Once the model is trained, the evaluation of it is done with testing data and certain metrics. The metrics to evaluate the model depend on task type—for classification tasks, confusion matrix and receiver operating characteristic are popular, and for regression tasks root mean square error and mean absolute error are popular (Géron 2019, pp. 90–98, 39–41). Concordance Index, or C-index, is a metric used to evaluate a model's ability to correctly rank predictions, so it can be used in DTI prediction.

In model validation, the original data set is split into three: train set, test set, and validation set. The validation set is used separately for hyperparameter selection—the model performing best on the validation set is chosen. Hyperparameters are preset parameters of the training algorithm, whereas model parameters are learned during training. Validation is done before training and evaluation of the model. (Berthold et al. 2010, p. 103) With limited amount of data, the hold-out method—where separate validation set is used—can be replaced with cross-validation (CV): the entire dataset is divided to subsets of equal size, from which repeatedly one subset is used as a validation set, and the others as training set (Berthold et al. 2010, p. 104). When using CV, there is more data to use for training to avoid overfitting: the model performs well on the training data, but does not generalise. (Berthold et al. 2010, pp. 27–29)

The main part of the ML model is the algorithm determining the training. There are several basic algorithms that are common in supervised learning: K-nearest neighbours, decision tree, and random forest for classification tasks, and linear, polynomial, and ridge regression for regression tasks (Sarker 2021). FM is an algorithm that works with a training set consisting of paired data and matching labels. The algorithm learns patterns from the training data and produces the predicted value for each label based on original data. FM can be used for regression, classification, or ranking tasks. (Rendle 2010) Thus, FM can also be used as an algorithm in DTI prediction tasks.

2.1.2 Factorisation Machine in Paired Data Prediction

A general pipeline for paired data prediction task based on supervised ML consists of a definition of original data and matching labels, creation of model that uses some algorithm to predict correct label based on original data, as well as model training and evaluation based on train-test split and evaluation metrics.

The dataset of paired data used in a supervised ML task can be defined as

$$\left\{(\mathbf{x}^{(1,1)}, y^{(1,1)}), (\mathbf{x}^{(1,2)}, y^{(1,2)}), \ \ldots\right\}, \tag{2.1}$$

where $\mathbf{x}^{(i,j)}$ represents the original paired data vector, and $y^{(i,j)}$ represents the matching real-valued label (Rendle 2010). Then, the actual model for supervised ML task is created with an algorithm f that best maps input data to predicted label as

$$\hat{y}^{(i,j)} = f\left(\mathbf{x}^{(i,j)}\right),$$

where $\mathbf{x}^{(i,j)}$ is the input data vector and $y^{(i,j)}$ is the matching target label (Vieira et al. 2019). Next, the dataset is split into train and test sets using train-test split with a predetermined ratio. Then, the algorithm f is fitted using the train set: this can be described as learning of the f. Evaluation of the model is done using the test set. In practice, now the f is used with unseen test set data, which the model did not use for training. Evaluation of model's performance is done with a chosen metric $m(y, \hat{y})$ that compares real label $y^{(i,j)}$ and predicted label $\hat{y}^{(i,j)}$. Validation of the model is possible before training and evaluation, if needed. (Vieira et al. 2019)

When FM is used as the algorithm f, the prediction function is defined as

$$\hat{y}\left(\mathbf{x}\right) = w_0 + \sum_{i=1}^{d} w_i x_i + \sum_{i=1}^{d} \sum_{j=i+1}^{d} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j, \tag{2.2}$$

where $\hat{y}(\mathbf{x})$ is the predicted label and the term $\sum_{i=1}^{d} \sum_{j=i+1}^{d} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$ captures the pairwise interactions of the input features x_i and x_j. The model parameters to be estimated are $w_0 \in \mathbb{R}$, $\mathbf{w} \in \mathbb{R}^d$, and $\mathbf{V} \in \mathbb{R}^{d \times k}$, where d is the number of features. (Rendle 2010; Rendle 2012)

Here, w_0 is the global bias and w_i represents the strength of the i-th variable, x_i. The $\mathbf{V}$ is a matrix containing all latent vectors—a latent vector $\mathbf{v}_i$ captures the hidden interactions of x_i with other variables. Formally, the $\mathbf{V}$ is a matrix where each row $\mathbf{v}_i \in \mathbb{R}^k$ represents the latent vector of the x_i as

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_d \end{bmatrix} = \begin{bmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,k} \\ v_{2,1} & v_{2,2} & \dots & v_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ v_{d,1} & v_{d,2} & \dots & v_{d,k} \end{bmatrix}.$$

In the $\mathbf{V}$, a row $\mathbf{v}_i$ represents the latent factorisation of the corresponding feature with k factors—$k \in \mathbb{N}_0^+$ is a hyperparameter defining the dimensionality of the factorisation. In other words, it is the number of latent factors in $\mathbf{V}$. Choosing k in the validation phase is crucial, as it affects model complexity and performance. The dot product $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ is also computed from $\mathbf{V}$. (Rendle 2010; Rendle 2012)

When using FM, instead of learning a separate parameter for every pairwise interaction, it is possible to model interactions—such as those between drugs and targets—by factorising them through latent representations. This factorisation enables efficient parameter estimation even in sparse data settings, making FM effective for high-dimensional problems. The estimates of model parameters are found by optimising them with a learning algorithm, like stochastic gradient descent

$$\frac{\partial}{\partial \theta} \hat{y}(\mathbf{x}) = \begin{cases} 1, & \text{if } \theta = w_0 \\ x_i, & \text{if } \theta = w_i \\ x_i \sum_{j=1}^{n} v_{j,f} x_j - v_{i,f} x_i^2 & \text{if } \theta = v_{i,f} \end{cases}, \tag{2.3}$$

where each parameter's gradients are computed separately. (Rendle 2010; Rendle 2012) Alternatively, other learning algorithms can also be used.

Figure 2.1 illustrates the main steps of general pipeline for paired data prediction task based on supervised ML method—the accurate implementation varies according to task type and requirements of the task.

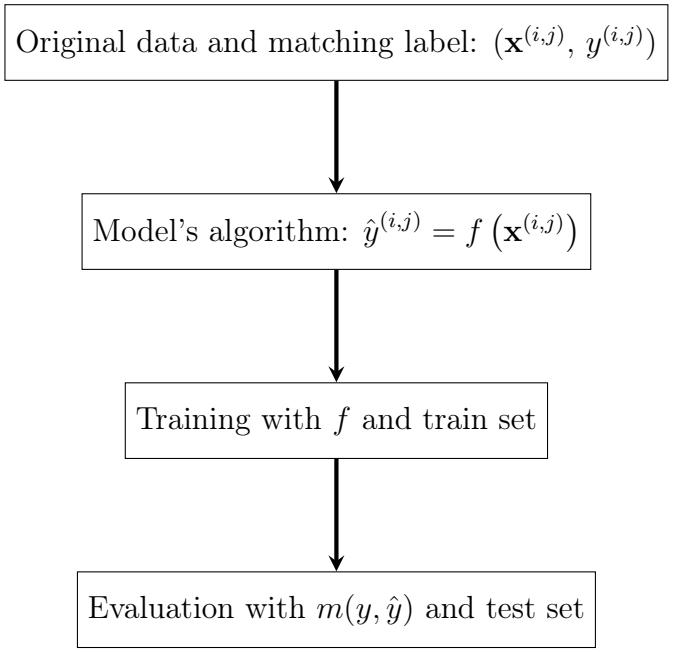

Figure 2.1: General pipeline for paired data prediction using supervised ML consists of defining orignal data and labels, using an algorithm f to learn to map data to label, model training with f, and model evaluation with a metric $m(y, \hat{y})$.

2.2 Drug–Target Interaction Prediction

Drugs produce different kinds of effects at different functional levels—including biochemical, cellular, physiological, and structural level. Direct effects can also lead to secondary effects that can be relevant in clinical situations. (Rang et al. 2020, p. 21) The goal in the drug discovery and development is the identification of a single, novel compound that can alter the outcome of the effects as compound's biological properties are a function of its chemical structure (Blass 2015, p. 203–204).

Generally, drugs are useful when acting selectively on specific cells—drug molecules must get so close to target molecules that these two can interact chemically in a way where the function of the latter is altered. Because the target molecules are much bigger than drug molecules, the specific binding sites within the target are critical. Ligand is a term for a molecule that binds to a larger molecule—like drug to its target. There are various different effects that ligands can cause to targets: agonist ligands activate target and antagonist block them. Proteins that function as targets usually have a high degree of ligand specificity. This means that they bind only drug molecules of a certain, precise type. This very exact complementary specificity of ligands and binding sites is central in explaining the basic phenomena of pharmacology. (Rang et al. 2020, pp. 6–16)

The dissociation constant K_d represents the affinity of a ligand for its target. In other words, the interaction of a drug and a target to form a complex is defined by its K_d. Because of this, the K_d is defined as a ratio of free ligand concentration, free target concentration, and the ligand-target complex concentrations. (Blass 2015, p. 146) A higher K_d value means a weaker drug–target binding. On the contrary, stronger binding produces lower K_d value. (Salahudeen & Nishtala 2017) This occurs, because a smaller concentration of the ligand is needed to reach equilibrium with the target. For example, compounds with K_d values in the nanomolar range bind more tightly than those in the micromolar range. (Blass 2015, p. 147)

Most of the targets are based on four classes of macromolecule proteins: enzymes, G protein–coupled receptors (GPCR), ion channels, and transporters (Blass 2015, p. 87). All targets have a complex, three-dimensional structure. The configurations of the molecule controlling the functionality are affected by other molecular interactions, like hydrogen bonds and salt bridges. A ligand interacting with a binding site of a target molecule activates these same principles. (Blass 2015, p. 205) For example, a kinase inhibitor (KI) inhibits the activity of a target kinase enzyme.

When drugs bind to targets, it provokes a series of biochemical and physiological changes to produce the wanted effect. The maximum drug effect is achieved once all the targets are occupied, and so K_d is the central factor when modelling interactions of drugs and target. (Salahudeen & Nishtala 2017) In the Figure 2.2, the interaction of a drug and a target is demonstrated.

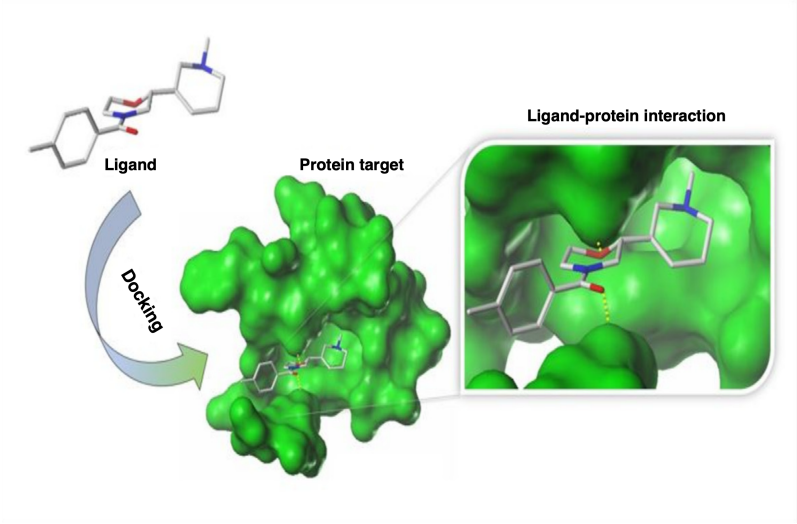

Figure 2.2: An example molecular modelling of DTI, where a ligand is docked to interact properly with a protein (D'Souza, Prema & Balaji 2020).

The molecular information is needed to be represented in a form that can be used in DTI prediction model. Each cell in the interaction matrix $\mathbf{Y}$ represents a K_d value between a drug and a target. Because of this, the labels y are acquired directly from these values. The original data $\mathbf{x} = (\mathbf{x}_d, \mathbf{x}_t)$ is also derived from $\mathbf{Y}$ with one-hot encoding (OHE). Like this, the ML based DTI prediction models works solely based on information from $\mathbf{Y}$, and the model learns to map $\mathbf{x}$ to y.

With OHE, the original data $\mathbf{x}$ is formed by creating identifiers for the drug d_i and the target t_j. Each unique category—one drug or one target—is represented as a binary vector with a length of total number of categories. The position corresponding to the specific category in the vector is set to 1, and other positions are set to

0. (Hancock & Khoshgoftaar 2020) In the context of DTI prediction, OHE can be used as a tool to represent the drugs or targets. This kind of encoding can be used in scenarios where representation of drugs and proteins are needed—also with other biological sequences than drugs or targets. In general, OHE expands the dimensionality of the representations and so enables the model to capture interactions of drugs and targets. (Zhai et al. 2023) In biochemical applications, like DTI prediction, OHE is used for converting categorical data into a binary form usable in ML model due to its simplicity and interpretability.

When original data is viewed in the same form as in equation 2.1, it is done by concatenating representation of a drug d_i, $\mathbf{x}_d^{(i)}$, and representation of a target t_j, $\mathbf{x}_t^{(j)}$, into a drug–target pair $\mathbf{x}^{(i,j)} = (\mathbf{x}_d^{(i)}, \mathbf{x}_t^{(j)})$. The cells of $\mathbf{Y}$, $y^{(i,j)}$, correspond to a K_d value between drug d_i and target t_j. Sometimes, depending on drugs and targets, entries in $\mathbf{Y}$ might be scaled to make it comparable to other datasets (Davis et al. 2011). Now, each drug–target pair can be defined as

$$\mathbf{x}^{(i,j)} = \left(\mathbf{e}_d^{(i)}, \mathbf{e}_t^{(j)}\right),$$

where $\mathbf{x}^{(i,j)}$ is the original data used as input in an ML model, $\mathbf{e}_d^{(i)}$ is the binary encoding of a drug d_i, and $\mathbf{e}_t^{(j)}$ is the binary encoding of a target t_j.

2.2.1 Machine Learning Based Prediction

In the context of DTI prediction, one way to build a supervised ML pipeline for paired data prediction is to first define the original data and matching labels. Then, create an ML model that uses FM as algorithm to predict correct labels based on original data. Finally train the model with FM, as well as evaluate it based on C-index as an evaluation metric. The original drug–target data is now $\mathbf{x}^{(i,j)} = \left(\mathbf{e}_d^{(i)}, \mathbf{e}_t^{(j)}\right)$. The labels are represented as $y^{(i,j)}$.

When using FM as algorithm, predictions are made with equation 2.2, where x_i and x_j present the scalar values from the feature vector $\mathbf{x}^{(i,j)}$ used as input. The training of the model is done with a function 2.3. FM is used for regression: predicted value $\hat{y}(\mathbf{x})$ is a continuous real number. Model parameters w_0, $\mathbf{w}$, and $\mathbf{V}$, are estimated during training, whereas the hyperparameter k is set with CV. Before training, the dataset's train-test split is done, and the train set is used for model's training.

The model's evaluation is done with the test set, and C-index as evaluation metric, denoted as $m(y, \hat{y})$. The C-index is suitable as it quantifies how well the model ranks the interaction strengths between drug–target pairs. In DTI prediction—where some interactions are known and others are completely unknown—the C-index assesses how well the model preserves the relative ordering of known interactions. The C-index is a measure that calculates the probability that two predicted values $\left(\hat{y}^{(a)}, \hat{y}^{(b)}\right)$ maintain the same relative order as their true values $\left(y^{(a)}, y^{(b)}\right)$. This can be expressed as $P\left(\hat{y}^{(b)} > \hat{y}^{(a)} \mid y^{(b)} > y^{(a)}\right)$. C-index has value 1 if the pair is concordant and value 0 if the pairs is discordant. The final C-index value lies within $[0, 1]$, where a value of 1 indicates perfect ranking, a value of 0.5 indicates random ranking, and a value of 0 indicates complete misordering. (Alabdallah et al. 2024)

When using pipeline like this, the predictions are made only with the data acquired directly from the $\mathbf{Y}$, since the $\mathbf{x}^{(i,j)}$ is only a binary representation of $\mathbf{Y}$. In DTI prediction, there are two possible sources of information: the interaction matrix based information describing known interactions, and side information describing chemical properties of the drug and the target. The OHE based identifiers represent the first one. The latter—side information—can be added into $\mathbf{x}^{(i,j)}$ to possibly improve predictions. In general, side information cannot be deduced from the interaction matrix, and the incorporation of it is done with an information source that is separate from interaction matrix.

2.2.2 Utilisation of Side Information

In DTI prediction, side information refers to additional data of drugs and targets added to model's input data to capture more accurate interactions. The side information can consist of different properties of drugs and targets, and it can be based on, for example, chemical databases or imaging data that can capture different cellular and chemical features (Trapotsi et al. 2021). In the pipeline that is presented in chapter 2.2.1, the drug–target data used as input is based only on identifiers configured with OHE. One way to add side information to this kind of drug–target representation is by similarity matrices that include chemical properties of drugs and targets.

Drugs and targets can be both represented in similarity matrices. In a drug–drug similarity matrix, each row presents a drug and each column presents a drug. The same logic applies to targets: the first row and first column correspond to the same molecule, and so on. Like this, each entry of the matrix represents the similarity between the two molecules. (Davis et al. 2011) Similarity score can be calculated with several methods. However, the similarity matrix only notes the similarity between two molecules, and not within the global structure; this leads to some level of information loss (C. Wang et al. 2023). Another side information can also be added—this is handled in chapter 3.2.

Let $\mathbf{X}_D \in \mathbb{R}^{m \times m}$ represent the drug–drug similarity matrix, where m is the number of drugs d. Each entry $X_{D,i,j}$ denotes the similarity between drug d_i and drug d_j. Similarly, let $\mathbf{X}_T \in \mathbb{R}^{n \times n}$ represent the target–target similarity matrix, where n is the number of targets t. Each entry $X_{T,i,j}$ denotes the similarity between target t_i and target t_j. Each drug d_i can be represented as a feature vector $\mathbf{x}_d^{(i)} \in \mathbb{R}^m$, corresponding to the i-th row of $\mathbf{X}_D$. Likewise, each target t_i can be represented as a feature vector $\mathbf{x}_t^{(i)} \in \mathbb{R}^n$, corresponding to the i-th row of $\mathbf{X}_T$. (Davis et al. 2011)

Like this, side information can be added to the data, maintaining a valid format. The complete representation of a drug–target pair with side information is

$$\mathbf{x}^{(i,j)} = \left(\mathbf{e}_d^{(i)}, \mathbf{e}_t^{(j)}, \mathbf{x}_d^{(i)}, \mathbf{x}_t^{(j)}\right),$$

where $\mathbf{x}^{(i,j)}$ illustrates the input data, $\mathbf{e}_d^{(i)}$ is the binary encoding of drug d_i, $\mathbf{e}_t^{(j)}$ is the binary encoding of target t_j, $\mathbf{x}_d^{(i)}$ is the side information representation of drug d_i, and $\mathbf{x}_t^{(j)}$ is the side information representation of target t_j.

The original data serving as an input to the model is a concatenation of drug and target identifiers, as well as similarity based drug and target vectors. Overall, there is now information derived from the interaction matrix, and also additional side information for both drugs and targets. Let $\mathbf{x}^{(i,j)}$ in the complete input feature matrix $\mathbf{X}$ denote the element at row i and column j of the input feature matrix, where $i = 1, \ldots, m$ and $j = 1, \ldots, n$. The complete input feature matrix $\mathbf{X}$ consists of all input feature vectors $\mathbf{x}^{(i,j)}$, and it can be presented as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^{(1,1)} \\ \mathbf{x}^{(1,2)} \\ \vdots \\ \mathbf{x}^{(i,j)} \end{bmatrix}.$$

Similarly, the interaction matrix $\mathbf{Y}$, where the entries are representing interaction between a drug and a target, $y^{(i,j)}$, for a drug d_i and a target t_j, can be presented as

$$\mathbf{Y} = \begin{bmatrix} y^{(1,1)} & y^{(1,2)} & \ldots & y^{(1,j)} \\ y^{(2,1)} & y^{(2,2)} & \ldots & y^{(2,j)} \\ \vdots & \vdots & \ddots & \vdots \\ y^{(i,1)} & y^{(i,2)} & \ldots & y^{(i,j)} \end{bmatrix}.$$

The feature matrix $\mathbf{X}$ serves as an input to a ML model, where each row represents a unique drug–target pair with added side information. The label matrix $\mathbf{Y}$ represents known interactions of drugs and targets. In the datasets used in this thesis, the $\mathbf{Y}$ includes sparse values—because of unknown interactions between a drug and a target—and FM is used to predict these sparse values inside the $\mathbf{Y}$. Figures 2.3 and 2.4 illustrate formation of the input data that is based on information derived from interaction matrix, and on side information derived from similarity matrices.

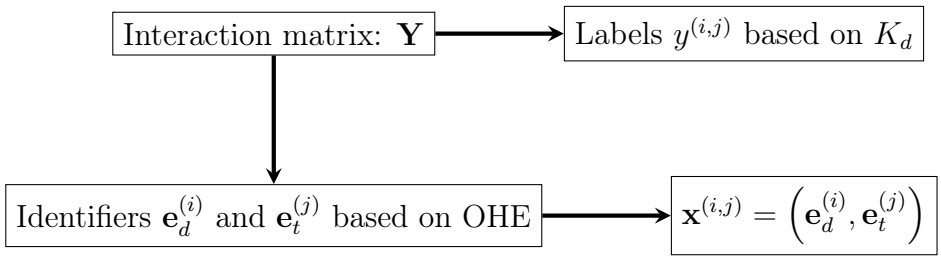

Figure 2.3: Labels of interactions between drugs and targets and identifiers for drugs and targets used in the input data are derived from the interaction matrix.

While binary drug and target identifiers derived from the interaction matrix are easily available, the use of side information—such as similarity matrices—relies on the availability and content of external databases. Sometimes, side information can be missing for newly discovered molecules. The binary identifier based information may have limitations in cases where the known interactions are really sparse. Also, if some drug–target pairs are well-studied and some others are less-studied, the predictions might be biased. In these cases, relying only on binary identifiers might fail to capture the true interactions.

Figure 2.5 illustrates the completion of the input data that can be used in a DTI prediction model. The completed input data includes information from interaction matrix and side information from similarity matrices. The DTI model identifies patterns to map data to labels.

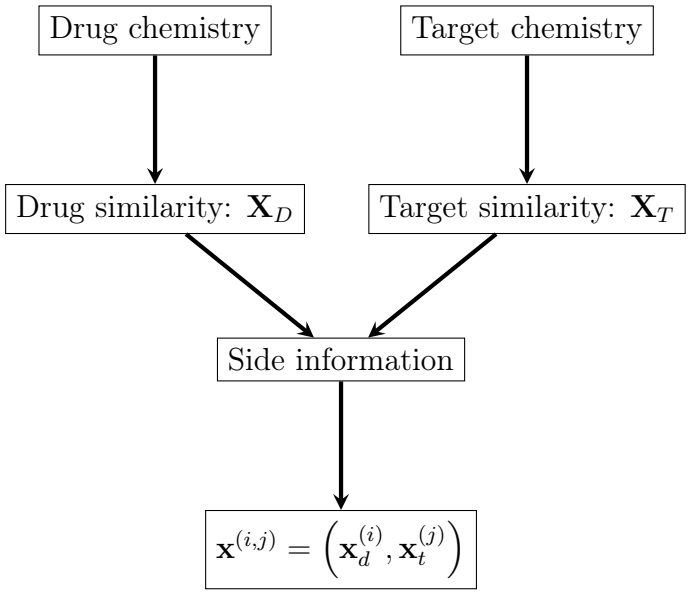

Figure 2.4: Side information used in input data is derived from similarity matrices that are based on chemical properties of drugs and targets.

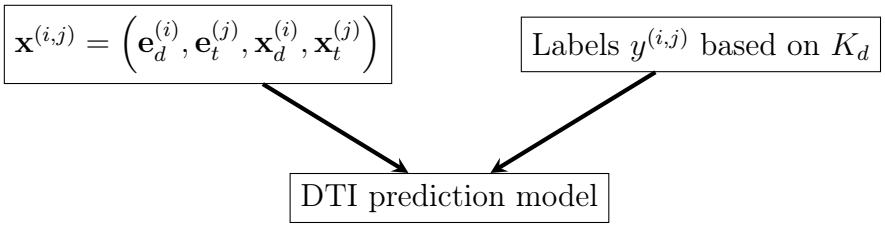

Figure 2.5: The complete input data consists of identifier information and side information describing chemical properties. This data and matching labels can be used in a DTI prediction model.

3 Literature Review

The DTI prediction pipeline presented in detail in chapter 2.2.1—definition of the original data and matching labels, creation of an FM model to predict correct labels based on original data, training the model with FM, and evaluating it with C-index—can be modified also to serve in other DTI prediction cases. The two big entities are the ML algorithm in model and the addition of side information: these both can be done in various ways.

In this chapter, there is review of literature related to the use of matrix factorisation as an ML algorithm and different variations of the side information in DTI prediction. The literature review of these topics was conducted from IEEE Xplore, ScienceDirect, and PubMed databases. Keeping in mind the research question, especially the effect of side information to the prediction results is reviewed.

3.1 Matrix Factorisation

The general theory of FM, its main advantages, and its functioning as an ML model are explored in chapters 1, 2.1.1, and 2.1.2. When considering matrix factorisation based methods as a part of DTI prediction, it is important to understand the additional features they offer specifically to DTI prediction. It is reviewed why matrix factorisation is used over other ML models in DTI prediction, and what results does one get using it in DTI prediction.

One reason why matrix factorisation is preferred in ML based DTI prediction models, is the ability to handle high sparsity and high dimensionality in dataset. (Ye et al. 2021; Sajadi et al. 2023) This is a key factor, because many drug–target datasets that are used are characterised by high sparsity and high dimensionality. In practice, this means that many of the interactions are unknown, and the complete interaction matrix is large due to the size of drug–target data. Unlike with many basic ML models, it is possible to handle these kinds of datasets and capture the interactions between variables efficiently—like shown in chapter 2.1.2.

Because of this ability to capture interactions well despite the sparsity issue, matrix factorisation based models are a great basis when using side information in the input data. It is common to use matrix factorisation when incorporating side information into drug–target input data. (Sajadi et al. 2023) Different aspects of the usage of side information are looked more closely in chapter 3.2.

Using matrix factorisation in DTI prediction has displayed advantages in the prediction results. In many cases, it outperforms other ML models. When using a hybrid model that combines matrix factorisation with denoising autoencoders, high accuracy is achieved when using area under precision–recall curve (AUPRC) and area under receiver operating characteristic curve (AUC) as classification metrics (Sajadi et al. 2023). AUPRC shows the trade-off between precision and recall, and AUC shows the area underneath a curve that compares the performance of classifiers by with false positives or negatives (J. Kim et al. 2021). With a novel dual-network integrated logistic matrix factorisation DTI prediction scheme, AUPRC and AUC values are in most cases higher than with other state-of-the-art methods (J. Li et al. 2022). A model combining neural neighbourhood regularised matrix factorisation and tangent kernel model achieves also better AUPRC values than other existing methods (Y. Wang et al. 2023).

In general, the matrix factorisation based methods get systematically accurate and robust results on DTI predictions on gold standard datasets (Sajadi et al. 2023; J. Li et al. 2022; Y. Wang et al. 2023). Gold standard refers to common target datasets that include information of enzymes, GPCRs, ion channels, and nuclear receptors (Cao et al. 2012).

3.2 Side Information

The general theory of side information based on similarity matrices as well as its incorporation to DTI pipeline with matrix factorisation are explored in chapters 2.2, 2.2.1, and 2.2.2. Side information can be incorporated in DTI prediction also in many other—usually computationally heavier—ways. In addition to exploring different side information types, it is also reviewed how side information is incorporated in DTI prediction and what results does one get using side information in DTI prediction.

Other side information types than similarity matrices are for example chemical structure and side effects of drugs, as well as protein sequences of targets. Also, DTI networks, multi-source biochemical and pharmacological information, and combinations of different types of side information can be used.

Simplified molecular input line entry system (SMILES) is one way to present structure of a drug (Zhao et al. 2024). It is a string that describes the three-dimensional structure of the drug, and it can be used for encoding the drug in DTI prediction. For example, C1NC2=CC(=C(C=C2S(=O)(=O)N1)S(=O)(=O)N)CI) is the SMILES for a drug hydrochlorothiazide (Chen et al. 2024). Hence, the SMILES string is usually needed to convert into numerical form before using in any model. The process of embedding a drug's chemical structure into a drug vector is presented in Figure 3.1.

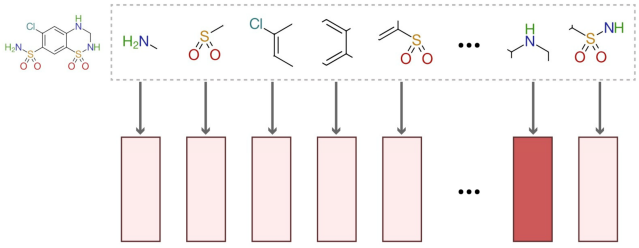

Figure 3.1: An illustration of how the structure of hydrochlorothiazide can be embedded and transformed into a vector (Chen et al. 2024).

Side effects of drugs can be used as side information by calculating side effect similarities of drugs, as well as side effect based protein–protein interaction similarities of targets. This information is then enhanced when predicting interactions based on fulfilled drug and target similarities and known interactions of drugs and targets. (Lu et al. 2021) Side effects of drugs can also be used as side information by utilising databases like SIDER—database of interactions between drugs and side effects—and STRING—database used to model protein–protein interactions in genes—that can be used to build network useful in predicting interactions of drug and targets (Xu, Zhou & Chen 2019). An example network is presented in Figure 3.2.

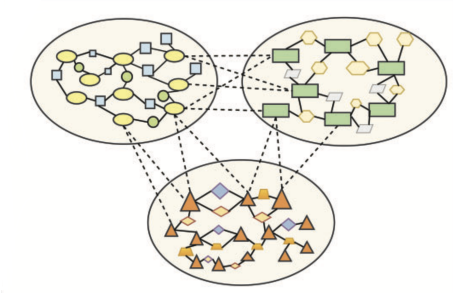

Figure 3.2: Drug–side effect context-sensitive network connecting drugs' structures, side effects, and effect on genes (Xu, Zhou & Chen 2019).

The structure of targets is usually described with amino acid sequences. This can be presented, for example, as a vector consisting of letters based on 22 possible amino acids to obtain the feature matrix of the target (He et al. 2024). This so-called sequence-view data can capture the structure of a target protein, and thus it can be utilised when predicting interactions (Yang et al. 2024). When encoded into a numerical form, protein sequence can be used as input in DTI prediction pipeline, like Figure 3.3 shows. Overall, Figure 3.4 illustrates the mapping of a target structure to a vector form that can be used in a DTI prediction pipeline.

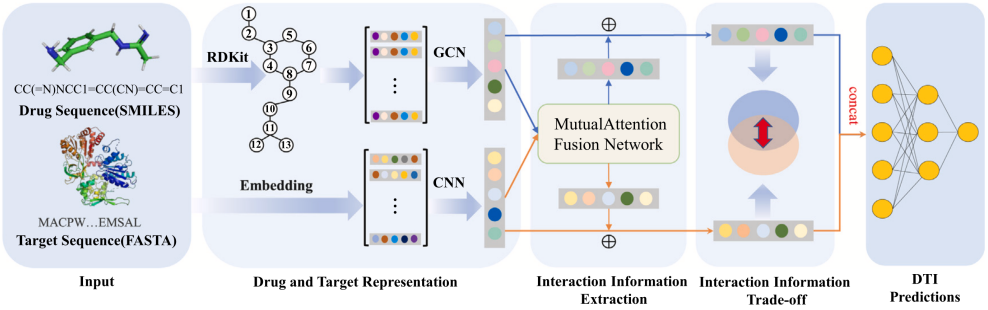

Figure 3.3: An example DTI prediction pipeline with structure based drug representation and sequence based target representation (He et al. 2024).

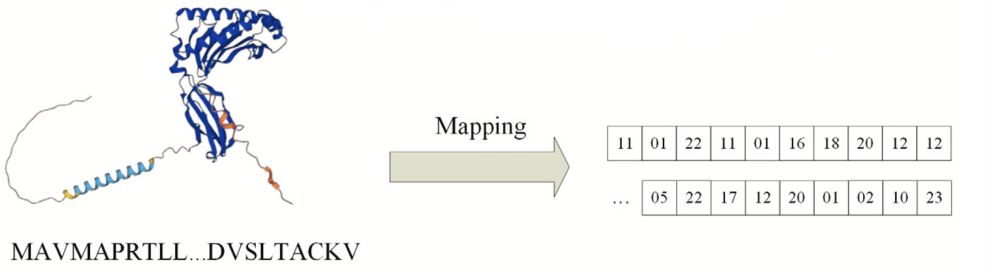

Figure 3.4: A generalisation of mapping protein's amino acid sequence to a vector of integers (Zhu et al. 2023).

Most of the DTI prediction methods use only one side information type in their pipelines, but it is possible to also utilise multi-view side information got from public databases. For example, it is possible to use drug response data and drug structure data to form drug vectors. Similarly, it is possible to use target gene expression data and target sequence data to form target vectors. (Hao, Cai & L. Li 2019) Like this, it is possible to take advantage of both structural and chemical properties of drugs and targets. One possible side information combination and incorporation is presented in Figure 3.5. Generally, similarity of drugs and targets is one of the most widespread sources of side information in DTI prediction, and it is typically supplemented with more accurate information about drugs' chemical properties as well as targets' sequence and gene ontology (Bolgár & Antal 2017).

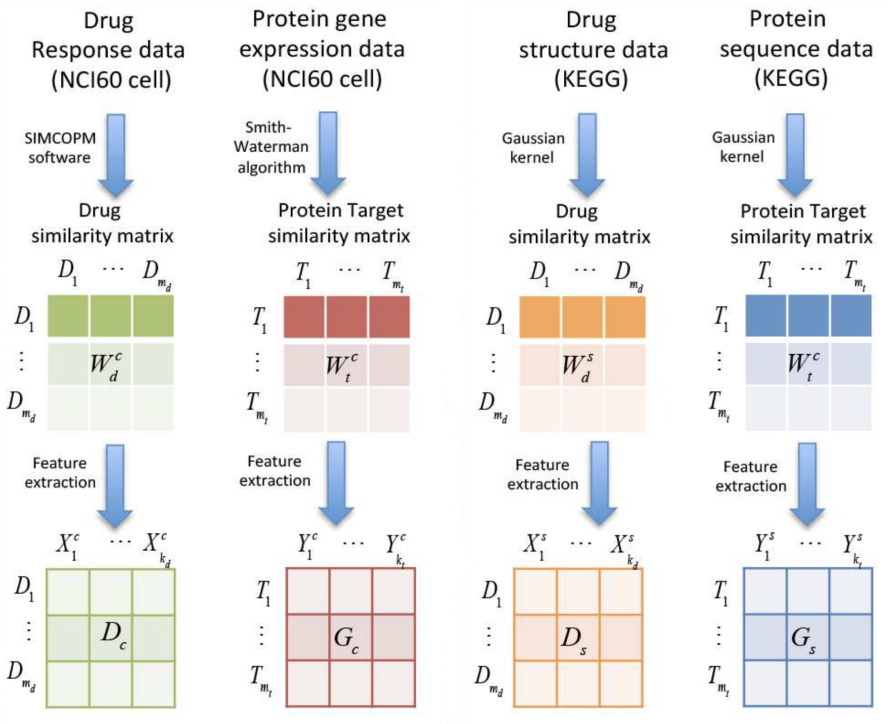

Figure 3.5: Multi-view side information example from KEGG and NCI databases made by constructing similarity matrices in both chemical and structural view for both drugs and targets (Hao, Cai & L. Li 2019).

It is also possible to fuse larger combinations of side information to be used in DTI prediction. For example, side information data can be gathered from various public datasets about drugs, targets, diseases, and side effects, and then forming several interaction networks that can be used together (Qiao, G. Wang & Y. Li 2024). Figure 3.6 shows an example of extraction and gathering of multi-source information. Also, it is typical to merge various biochemical and pharmacological datasets from different public databases to be used in DTI prediction. For instance, interaction information from ChEMBL, side effects from SIDER, and other interaction information from STITCH can be integrated to form one big dataset (S. Kim, Jin & Lee 2013).

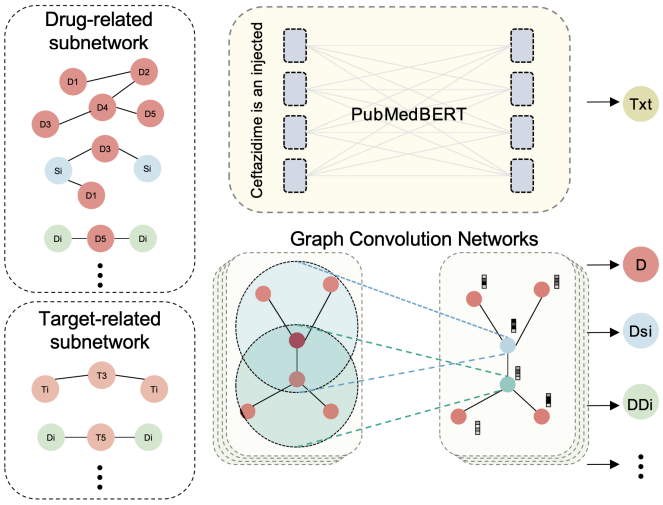

Figure 3.6: Side information network fusion can be made by compounding subnetworks of drugs, targets, diseases, and side effects. Also, PubMedBERT model is used to analyse drugs' text descriptions. (Qiao, G. Wang & Y. Li 2024)

Other ways to incorporate side information to DTI prediction pipeline than matrix factorisation are, for example, graph based methods and DL approaches. These methods are also ML based, but they are more computationally complex. A graph based method in Figure 3.7 is based on generation and utilisation of embeddings from knowledge graph (Liu, Zhang & X. Wang 2024).

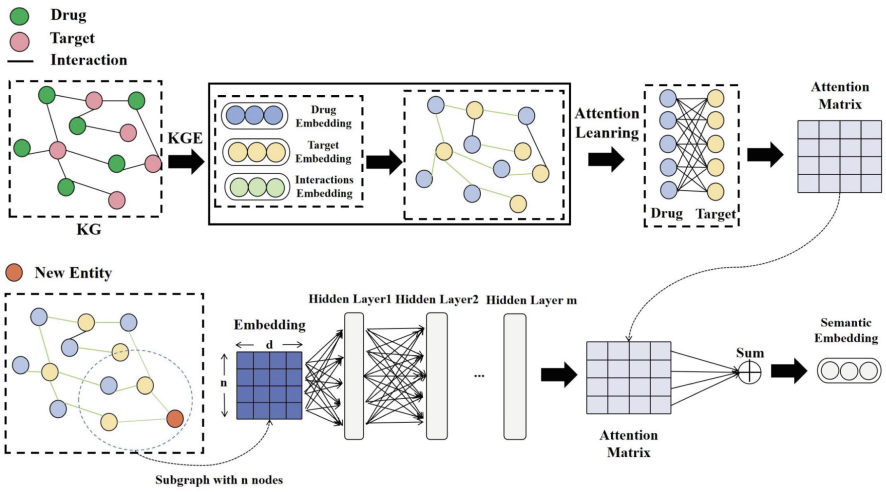

Figure 3.7: DTI prediction with a knowledge graph (Liu, Zhang & X. Wang 2024).

DL based method offers the benefits of neural networks—ability to learn complex patterns, high performance, and generalisation on big data. A network in Figure 3.8 consists of the input module that handles drug and target data, the molecular representation optimisation learning module that includes the actual DL neural network, and the output module that combines feature representations and makes prediction with fully connected layers (Zhu et al. 2023).

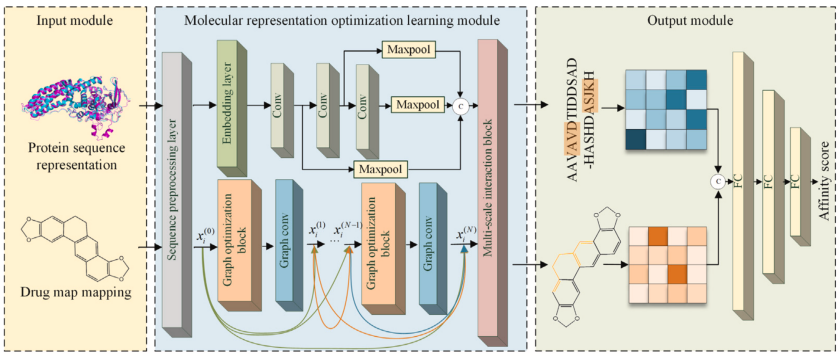

Figure 3.8: A framework of a DL based DTI prediction model consisting of input and output models, as well as neural network based module build with embedding, optimising, convolutional, and pooling layers (Zhu et al. 2023).

It is quite hard to compare the direct effects of different side information types, since all the models utilising side information are built in different ways. Overall, the models leveraging side information can successfully identify new interactions between drugs and targets, and thus improve prediction accuracy (Hao, Cai & L. Li 2019; Qiao, G. Wang & Y. Li 2024). When using only one side information type, drug–drug similarity, AUC and AUPRC scores improve on almost all molecule types when using gold standard dataset with different advanced DTI prediction models (Hassanzadeh & Shabani-Mashcool 2022).

4 Proposed Prediction Model

A new DTI prediction model is built based on the theory introduced in chapter 2. The binary identifiers of drug–target data—derived from an interaction matrix—are used as input to the model that predicts the interactions between drugs and targets. Then, similarity matrix based side information is used as additional data to potentially enhance the model. The proposed model uses FM as a supervised ML algorithm to predict the interactions of drug–target pairs.

In this chapter, there is presented a way to implement this DTI prediction model with FM and incorporation of the side information. The basic model presented in chapter 4.2 does not include side information, and the enhanced model presented in chapter 4.3 includes side information. The enhanced model can use only binary identifiers, only side information, or both. The basic model and the enhanced model are then used on different drug–target datasets to find out the improvement caused by the side information. All the datasets used on the models are openly available, and their contents and sizes vary.

Detailed architectures of models are presented, and specifics of how side information is obtained and how it is exactly added to architectures are also described. The results of the models are viewed in the context of the research question: Does the incorporation of side information improve drug–target interaction prediction compared to using only factorisation of an interaction matrix?

4.1 Drug–Target Datasets

There are a total of seven datasets that are used in the DTI prediction model (Davis et al. 2011; Metz et al. 2011; Tang et al. 2014; Merget et al. 2017; Yamanishi et al. 2008). The datasets have different amounts of drugs and targets, as well as different types of drugs and targets. However, each dataset includes an interaction matrix for labels and binary identifiers, as well as drug and target similarity matrices for side information. The type of each dataset can be seen from Tables 5.1 and 5.2, where KIBA refers to KI bioactivity, KW to kinome-wide, IC to ion channels, and E for enzymes.

The datasets partially overlap since they include some same drugs or targets (Tang et al. 2014). In the interaction matrix describing known interactions of drugs and targets, the interaction value types used as labels vary depending on the dataset. Also, there are sparse values in the interaction matrices due to unknown interactions. In the similarity matrices, drug and target similarities are calculated with different kinds of methods. Despite these variations, all datasets share the same basic structure, enabling their use in evaluating the effect caused by side information. The exact content, details of set-up and scaling of values vary between datasets.

The first dataset has biochemical selectivity assays for clinically relevant KI drugs. The interaction matrix consists of K_d values between a drug and a target. The drug feature matrix is based on the chemical properties of the drug, containing structural fingerprint similarities between two drugs. Drug similarity values are computed with 2D Tanimoto coefficients. (Davis et al. 2011) The target feature matrix is based on genomic data of the targets, and it contains the normalised version of the Smith–Waterman scores that are calculated between two targets (Smith & Waterman 1981). An example of KI drugs and matching targets of this dataset are presented in Table 4.1.

Table 4.1: A list of a few KI drugs and matching target kinases—a drug and a target on the same row interact together to form a drug–target pair (Davis et al. 2011). Information is gathered from PubChem database.

Drug	Target
Imatinib	ABL1
Crizotinib	ALK
Lapatinib	ERBB2
Sunitinib	KIT
Staurosporine	PLK1
Vandetanib	RET

The second dataset is quite similar, it contains biochemical selectivity assays for clinically relevant KI drugs. Also, the drug similarity matrix and the target similarity matrix are calculated the same way. The labels in the interaction matrix however represent inhibition constant between a drug and a target, instead of dissociation constant. (Metz et al. 2011)

The third dataset has KI bioactivity data integrated from multiple different databases. The gathered information contains various bioactivity types, like IC50, kinase inhibition constants, and kinase dissociation constants. The binding affinities of drugs and targets in the interaction matrix are used as labels. The drug and target similarity matrices have similar structure as in the first and second datasets. (Tang et al. 2014)

The fourth dataset includes a comprehensive kinome-wide map of interactions between drugs and targets. In this dataset, the labels in the interaction matrix are processed with the ChEMBL bioactivities. The similarity matrix for drugs contains a fingerprint of the shortest paths between atoms, taking into account ring systems and charges. The similarity matrix for targets contains amino acid sub-sequences of ATP binding pockets as well as amino acid properties. (Merget et al. 2017)

The last three datasets are from a gold standard dataset—they include GPCR, ion channel, and enzyme targets. These datasets also include the drugs targeting these pharmaceutically relevant targets. The interaction matrices in these datasets have binary interaction values, and the information in these datasets is obtained from KEGG BRITE, BRENDA, SuperTarget, and DrugBank databases. (Yamanishi et al. 2008) The drug similarity matrix is computed using the SIMCOMP score (Hattori et al. 2003). The similarity scores in the target similarity matrix are computed using the normalised Smith–Waterman score (Smith & Waterman 1981).

The datasets have different kinds of drug and target types, and also the numbers of interactions vary from tens of thousands to hundreds of thousands. Only the effect of side information to the overall DTI prediction model is reviewed, and the effect of different datasets is not reviewed as closely. An example of a KI drug and the theory behind how it binds to a target is demonstrated in Figure 4.1.

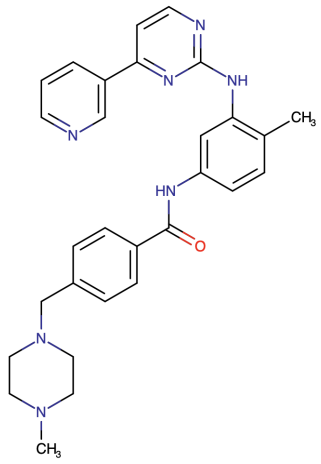

Figure 4.1: An example of a chemical structure of a KI drug, Imatinib (Davis et al. 2011). With the 2-phenylaminopyrimidine core, it binds to and inhibits the ABL1 target that is implicated in cancer cell growth—like this, it is used to treat cancer. Information is gathered from DrugBank database.

4.2 Basic Model

To evaluate the effect of side information, the datasets presented in chapter 4.1 are used in a basic DTI prediction model and in an enhanced DTI prediction model. The basic model does not utilise side information, whereas the enhanced model utilises it. The architecture of the basic model consists of the following steps: loading the dataset, CV to replace train–test split of the dataset, formation of the drug–target data with only binary identifiers, CV to validate the model, training of the model, and evaluation of the model. The training is done with FM (Rendle 2012). The evaluation is done with C-index metric (Pahikkala et al. 2015).

A fixed train-test split of the dataset is replaced with CV to reduce variability caused by random seed selection. The chosen train-test split ratio would be 67/33, so 3-fold CV is used. This means that one of the three folds is used as a test set at a time, whereas the two others form the train set. Formation of the drug–target data is done by concatenating only the OHE based identifiers of drugs and targets. The input data is solely binary and derived from the interaction matrix. Thus, the basic model works solely based on interaction matrix. The side information is not required to execute the model, but it is used in addition to enhance predictions.

The k—determining the dimensionality of interactions, or in other words, the number of latent factors—is optimised during the validation phase using 3-fold CV. This CV module finds the optimal $k \in \{4, 8, 16\}$, with $k = 8$ as the initial choice (Rendle 2012). There are a total of 2 CV modules: the outer CV replaces train–test split, and the inner CV is for validation. Together they form nested CV procedure. In both CV modules, it is necessary to ensure that the drug and target components in the test data are also in the train data. The k and the other tuning settings of FM are the hyperparameters of the model. These are chosen to balance computational cost and model's predictive performance.

During the training, FM based model learns to map the input data $\mathbf{x}^{(i,j)}$ representing drug–target pairs to labels $y^{(i,j)}$ representing interaction predictions. The model is trained using the best k value. The model parameters of FM—global bias w_0, variable strengths $\mathbf{w}$, and latent vectors $\mathbf{V}$—are optimised during the learning. The learning algorithm of FM is alternating least squares, the number of iterations is 100, and the predictions of the model are based on regression. Overfitting of the model is controlled using L2 regularisation, with a penalty term set to $\lambda = 1$. C-index is used to evaluate the performance of the model. The predicted values $\hat{y}(\mathbf{x})$ from the trained model are compared with actual values $y(\mathbf{x})$ using the C-index. Evaluation of the model is done across all folds. Thus, the final C-index score is obtained by averaging the C-index values across all folds.

4.3 Enhanced Model

Compared to the basic model, the only difference in the enhanced model is the incorporation of the side information. There are a total of three scenarios when it comes to feature options of the input data: using only binary identifiers, using only side information, and using both binary identifiers and side information. The feature option is chosen manually. This possibility to use different feature options is the key to answer to the research question, and it allows a direct comparison between different input feature options while ensuring standardised conditions.

In the first scenario, the input data—that has only OHE based binary identifiers—has the form $\mathbf{x}^{(i,j)} = \left(\mathbf{e}_d^{(i)}, \mathbf{e}_t^{(j)}\right)$. In the second scenario with input data consisting of only similarity matrix based side information, the input data has the form $\mathbf{x}^{(i,j)} = \left(\mathbf{x}_d^{(i)}, \mathbf{x}_t^{(j)}\right)$. In the third scenario with both of these features used, the input has the form $\mathbf{x}^{(i,j)} = \left(\mathbf{e}_d^{(i)}, \mathbf{e}_t^{(j)}, \mathbf{x}_d^{(i)}, \mathbf{x}_t^{(j)}\right)$. These scenarios allow a detailed set-up to research the impact of the side information on the DTI prediction.

4.3.1 Incorporation of Side Information

The incorporation of the side information—that enables the two latter feature options—is done with an additional algorithm added to the architecture of the basic model. The concept of deriving the side information from similarity matrices is handled in chapter 2.2.2. The similarity matrices are represented as $\mathbf{X}_D$ and $\mathbf{X}_T$, and the actual side information incorporated into drug–target feature vectors is represented as $\mathbf{x}_d$ and $\mathbf{x}_t$. The similarity matrices are used as side information method, since the drug and target similarity matrices are available in all the datasets.

In the proposed DTI prediction model, side information values are min–max scaled. This ensures that no particular feature dominates due to larger values. Scaling of side information features is also important when using them together with binary identifiers.

When using the feature option with both features, binary identifiers and side information, the input data is a feature matrix consisting of both binary and real-valued features. An example of the input data can be presented as

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0.0358 & 0.0732 & 0.0807 \\ 1 & 0 & 0 & \dots & 0.0153 & 0.0501 & 0.0485 \\ 1 & 0 & 0 & \dots & 0.0230 & 0.0606 & 0.0570 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0.0268 & 0.0751 & 0.0611 \\ 0 & 0 & 0 & \dots & 0.0395 & 0.0803 & 0.0780 \\ 0 & 0 & 0 & \dots & 0.0271 & 0.0692 & 0.0602 \end{bmatrix}, \tag{4.1}$$

where a row represents one drug–target pair as $\mathbf{x}^{(i,j)} = \left(\mathbf{e}_d^{(i)}, \mathbf{e}_t^{(j)}, \mathbf{x}_d^{(i)}, \mathbf{x}_t^{(j)}\right)$. Former values correspond to binary identifiers of drugs and targets, whereas the latter values correspond to real-valued features representing side information of drugs and targets.

This structure of the input data enables the model to leverage both sources of information effectively. An example of the corresponding labels can be presented as

$$\mathbf{Y} = \begin{bmatrix} 11.1 \\ 11.1 \\ 11.1 \\ \vdots \\ 10.4979 \\ 10.4979 \\ 10.4979 \end{bmatrix}, \tag{4.2}$$

where each component of the vector represents the label of a drug–target pair in the feature matrix. The proposed DTI prediction model attempts to learn a mapping from drug–target pairs in matrix $\mathbf{X}$ to the respective interaction labels in $\mathbf{Y}$.

4.3.2 Complete Architecture

The incorporation of the side information can be added to the architecture of the basic model. This is done with a possibility to manually choose the features of input data. Then, the complete architecture of the proposed DTI prediction model includes six main modules: dataset loading, CV to replace train-test split, feature concatenation, CV for validation, FM for training, and C-index for evaluation.

Dataset loading is done for one of the datasets presented in the chapter 4.1. All the datasets include interaction matrix $\mathbf{Y}$ containing known interactions of drugs and targets as labels, drug–drug similarity matrix $\mathbf{X}_D$, and target–target similarity matrix $\mathbf{X}_T$. The interaction matrix might be scaled to get it to match all the other datasets. The binary identifiers of both drugs and targets are calculated based on the interaction matrix, allowing the model to use structural information of the data. Also, both of the similarity matrices are scaled.

CV to replace train-test split is implemented with 3-fold CV in the same way as in the basic model. Feature concatenation handles the input data according to the chosen feature option. According to the choice, the model creates wanted input data $\mathbf{x}^{(i,j)}$ for each drug–target pair. The same concatenation is applied to both, train and test sets. CV for validation selects the best hyperparameter k for the FM model. Validation, as well as training and evaluation of the model, are done in the same way as in the basic model. The C-index values for this architecture on all the seven datasets and on all the three feature options are presented in the Table 5.1.

All six modules of the architecture work as a pipeline to form the proposed DTI prediction model. The ready, evaluated model can be deployed to predict the interactions of drugs and targets. Figure 4.2 shows a simplified pipeline of building the model. Algorithm 1 demonstrates the model's architecture as pseudocode.

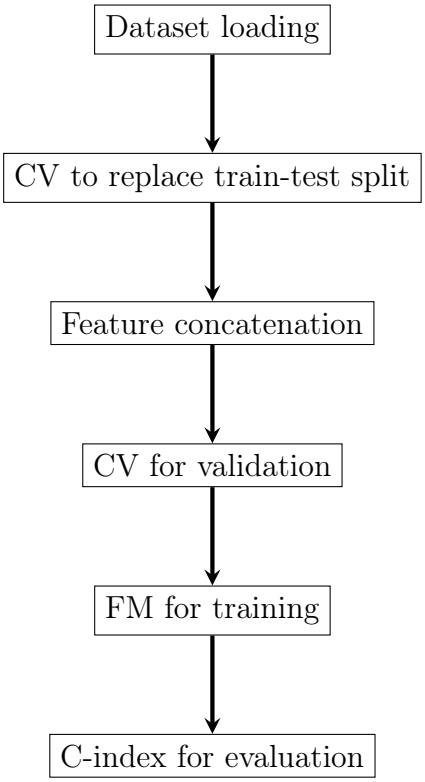

Figure 4.2: A simple illustration of modules used to build the DTI prediction model.

Algorithm 1 Architecture of the proposed DTI prediction model

Input: Dataset S, feature option $f \in \{\text{bi}, \text{si}, \text{both}\}$

Output: C-index of the model C_{final}

$S = \{\mathbf{Y}, \mathbf{X}_D, \mathbf{X}_T\}$ ▷ Dataset loading

$\mathbf{X}_D \leftarrow \frac{\mathbf{X}_D - \min(\mathbf{X}_D)}{\max(\mathbf{X}_D) - \min(\mathbf{X}_D)}$

$\mathbf{X}_T \leftarrow \frac{\mathbf{X}_T - \min(\mathbf{X}_T)}{\max(\mathbf{X}_T) - \min(\mathbf{X}_T)}$

$\mathbf{e}_d^{(i)} = \mathbf{e}_i, \quad (\mathbf{e}_i)_r = \begin{cases} 1, \text{ if } r = i \\ 0, \text{ if } r \neq i \end{cases}$

$\mathbf{e}_t^{(j)} = \mathbf{e}_j, \quad (\mathbf{e}_j)_r = \begin{cases} 1, \text{ if } r = j \\ 0, \text{ if } r \neq j \end{cases}$

for $S_p \subset S$, $p \in \{1, \ldots, 3\}$ **do** ▷ CV to replace train–test split

$\mathbf{x}^{(i,j)} \leftarrow \begin{cases} \left(\mathbf{e}_d^{(i)}, \mathbf{e}_t^{(j)}\right), & \text{if } f = \text{bi} \\ \left(\mathbf{x}_d^{(i)}, \mathbf{x}_t^{(j)}\right), & \text{if } f = \text{si} \\ \left(\mathbf{e}_d^{(i)}, \mathbf{e}_t^{(j)}, \mathbf{x}_d^{(i)}, \mathbf{x}_t^{(j)}\right), & \text{if } f = \text{both} \end{cases}$ ▷ Feature concatenation

  for $k \in \{4, 8, 16\}$ **do** ▷ CV for validation

    for $S_q \subset S_p$, $q \in \{1, \ldots, 3\}$ **do**

      $\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$

      $C_k = P\left(\hat{y}^{(b)} > \hat{y}^{(a)} \mid y^{(b)} > y^{(a)}\right)$

    end for

  end for

  $k^* = \arg\max_k C_k$

  $\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$ ▷ FM for training

  $C_p = P\left(\hat{y}^{(b)} > \hat{y}^{(a)} \mid y^{(b)} > y^{(a)}\right)$

end for

$C_{\text{final}} = \frac{1}{3} \sum_{p=1}^{3} C_p$ ▷ C-index for evaluation

5 Results

The proposed DTI prediction model, described in chapter 4, was evaluated with three different feature options—only binary identifiers, only side information, and both binary identifiers and side information—across seven distinct datasets. Table 5.1 shows the C-index values of the model across all scenarios. Table 5.2 shows the improvement in C-index values caused by the different feature options on all the different scenarios. As the model architecture remains constant across scenarios—differing only in feature inputs—the impact of each feature option can be directly assessed.

Table 5.1: C-index values of the proposed DTI prediction model with different feature options and on different datasets. There are scenarios where only binary identifiers—referred to as BI—are used, where only side information—referred to as SI—is used, and where both of these are used.

Dataset	C-index with BI	C-index with SI	C-index with BI and SI
KI (Davis et al. 2011)	0.855	0.841	0.856
KI (Metz et al. 2011)	0.830	0.753	0.815
KIBA (Tang et al. 2014)	0.831	0.740	0.822
KW (Merget et al. 2017)	0.851	0.800	0.847
GPCR (Yamanishi et al. 2008)	0.894	0.914	0.921
IC (Yamanishi et al. 2008)	0.959	0.961	0.967
E (Yamanishi et al. 2008)	0.955	0.927	0.947

The C-index values are relatively high, ≥ 0.830, when using only binary identifiers as the feature, as shown in the Table 5.1. When using only side information as the feature, the C-index values increase when using GPCR and ion channel datasets and decrease when using the other datasets. When both feature types are combined, the C-index values remain largely similar to those obtained using only binary identifiers—showing slight improvements for the GPCR and ion channel datasets, and slight deterioration or stability within the margin of error for the other datasets.

Table 5.2: Improvements of C-index values when using the proposed DTI prediction models in the scenarios with using only SI, and with using both BI and SI, relative to the scenario with using only BI.

Dataset	Improvement on using SI	Improvement on using BI and SI
KI (Davis et al. 2011)	-1.74 %	0.05 %
KI (Metz et al. 2011)	-9.28 %	-1.76 %
KIBA (Tang et al. 2014)	-10.94 %	-1.08 %
KW (Merget et al. 2017)	-5.97 %	-0.43 %
GPCR (Yamanishi et al. 2008)	2.19 %	3.01 %
IC (Yamanishi et al. 2008)	0.15 %	0.76 %
E (Yamanishi et al. 2008)	-2.93 %	-0.87 %

Table 5.2 shows that using only side information increases C-index 2.19 % in the GPCR dataset and 0.15 % in the ion channel dataset. In other datasets, C-index is reduced by at least 1.74 % and up to 10.94 % compared to using only binary identifiers. When combining side information with binary identifiers, C-index is improved by 3.01 % in the GPCR dataset and 0.76 % in the ion channel dataset. In other datasets, C-index is improved by up to 0.05 % in the best case, but decreased by as much as 1.76 % in the worst case. Using only side information seems to reduce performance, and using both features combined does not consistently improve results. Only the GPCR and ion channel datasets show improvement in both scenarios.

These results suggest that the side information does not capture the underlying structure of interactions between drugs and targets very effectively. When using the proposed DTI prediction model, it seems that incorporation of the side information alone does not systematically improve DTI prediction compared to using only matrix factorisation of the interaction matrix. Moreover, combining side information with binary identifiers also fail to yield a consistent and significant performance improvement compared to using only matrix factorisation of the interaction matrix. Only in the GPCR and ion channel datasets side information improves the model's performance in both these cases. These results answer the research question regarding the effect of side information over only factorisation of an interaction matrix: when using the proposed model, incorporation of side information does not systematically enhance predictive performance beyond of what was achieved with binary identifiers, whether used independently or together. Small enhancement can be only seen in individual datasets.

6 Discussion

The results presented in Tables 5.1 and 5.2 show that incorporation of similarity matrix based side information does not consistently improve the FM based DTI prediction model. Of all the seven datasets, only the GPCR and ion channel datasets benefit from side information, likely because their binary interaction matrices align well with the FM model. In contrast, the other datasets—with real-valued interaction labels—contain additional complexity that FM may be too simple to capture. The enzyme dataset, despite being binary, shows no improvement—possibly due to its larger size and chemical heterogeneity, which is not generalised enough with a single side information type.

When using only binary identifiers, the feature values are always either 0 or 1, meaning that they are at the extremes of the interval [0,1]. Instead, the scaled real-valued side information features can be somewhere in the middle of that interval, resulting in consistently weaker predictive performance. Scaling of side information features is necessary, since without scaling, large feature values cause the model weights to remain close to their initialisation values. This causes low predictive performance and low C-index values. This is supported by the values of model parameters when scaling of side information is not used—global bias w_0 is large, and variable strengths $\mathbf{w}$ have large variations since they do not update during learning. Thus, without scaling, the model fails to learn meaningful patterns, leading to nearly random predictions.

Notably, when using binary identifiers alone, the C-index values are high enough to build a working prediction model. This suggests that the structural information of drugs and targets captured by binary identifiers may already encode essential interaction patterns, making side information redundant—or even to add noise.

Another factor why side information did not improve the model is the simplicity of the FM. Unlike the advanced models reviewed in chapter 3.2, which employ graph or DL techniques, FM lacks the capacity to exploit complex interactions. Within graph or DL networks, there are many different operations for feature extraction and data processing. Also, the input data in the models presented in chapter 3.2 consisted frequently of many different side information types that led to better capturing of the interactions between drugs and targets. For instance, the results obtained with MSI-DTI model—that obtains feature representations from different views by integrating them with knowledge graph representations—consistently outperforms basic ML, graph, and transformer models (Zhao et al. 2024). This underscores the importance of combining various feature representation and prediction methods in DTI models.

Thus, the DTI prediction model presented in this thesis could be improved with multifaceted datasets and more powerful computing methods, since complex interactions between drugs and targets might require nonlinear transformations and higher-order feature interactions. When utilising big data in DTI prediction model, for example, by combining different side information types, the interactions of drugs and targets would potentially be represented more comprehensively and accurately. In this thesis, only one side information type is used: drug–drug and target–target similarity matrices. Alternatively, more complex and precise feature representations—such as chemical structures, drug and target sequences, or side effect profiles—could be used. These multiple side information types could be collected from various databases, and they could also be fused to form a side information network.

However, increasing the size of the side information also significantly increases the computational demands of the DTI prediction model. Computational power can be leveraged by using DL neural networks or graph based models that leverage parallel processing. Both DL and graph based models rely on networks of entities—such as neurons or nodes, and edges or connections—that can handle information through the structured relationships. This enables high-dimensional feature extraction that can lead to more accurate processing of data compared to a single FM model. Like this, the model could potentially capture the interaction patterns better due to the increased capacity to learn complex and non-linear relationships within the data.

All the used datasets to train and evaluate the model were based on matrices: drug–drug similarity matrix, target–target similarity matrix, and interaction matrix. However, some of these matrices could alternatively be replaced with tensors to provide more dimensions to the data—like in Figure 6.1. This would mean that there is more information in the original datasets, leading to better outcomes.

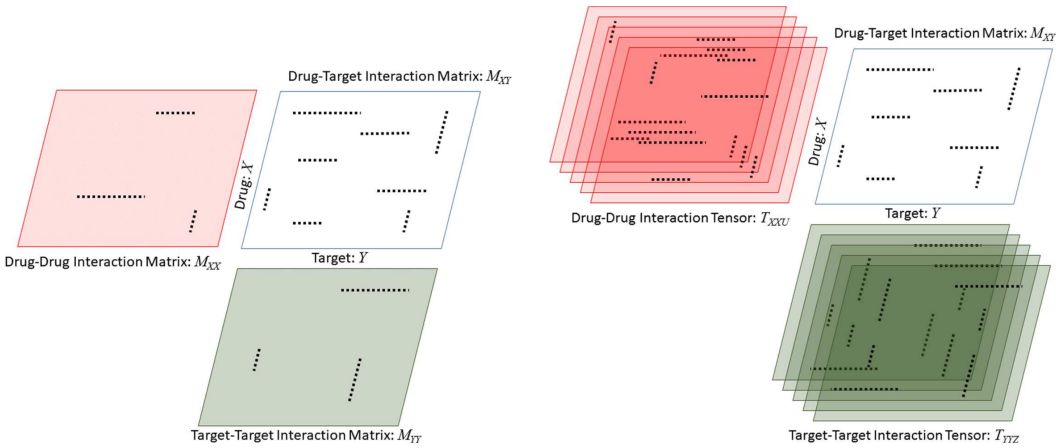

Figure 6.1: If some of the matrices used in DTI prediction would be replaced with tensors, there would be more dimensions in data—leading to a better representation of complex interactions between drugs and targets (Bagherian et al. 2021).

In this thesis, the interaction predictions are made inside the $\mathbf{Y}$, meaning that both drugs and targets are present in the training data. In this setting, the model can rely heavily on the $\mathbf{Y}$ alone, and simple binary identifiers seem to provide reasonable predictive performance. However, this setup does not generalise to cases where drug, target, or both do not have known interactions—this is common in real-world scenarios where novel drugs or novel targets emerge.

When one or both of the components of a drug–target pair are not in the training data, the model cannot predict meaningful interactions solely from the $\mathbf{Y}$. In such cases, the best the model can do is to predict some trivial heuristic, such as average interaction. In setups like this, side information becomes important, as it enables generalisation outside the observed interactions in the $\mathbf{Y}$.

While the predictions in this thesis were made only inside the $\mathbf{Y}$, it is important to contextualise the limitations of side information and applicability of the results in light of the scenarios where predictions are made outside the $\mathbf{Y}$. The observation made in the chapter 5, that side information did not consistently enhance the performance of the proposed model does not diminish its relevance in other applications. Future work could adapt this model to handle also the scenarios involving novel drugs or targets, where side information is essential. Addressing these scenarios can enhance the model's generalisability and practical utility in future drug discovery.

The possible directions of research in the field of DTI prediction will utilise the rapid development of ML computing—including progress of architectures, algorithms, and chips. In the future, it will be possible to use even larger datasets and more computationally efficient models. In the Figure 6.2, there is modelling of interactions of drugs and targets related to Alzheimer's disease—this illustrates the complexity of interactions in just one disease. When modelling multiple diseases and their effect on overall health, the need for efficient models becomes evident.

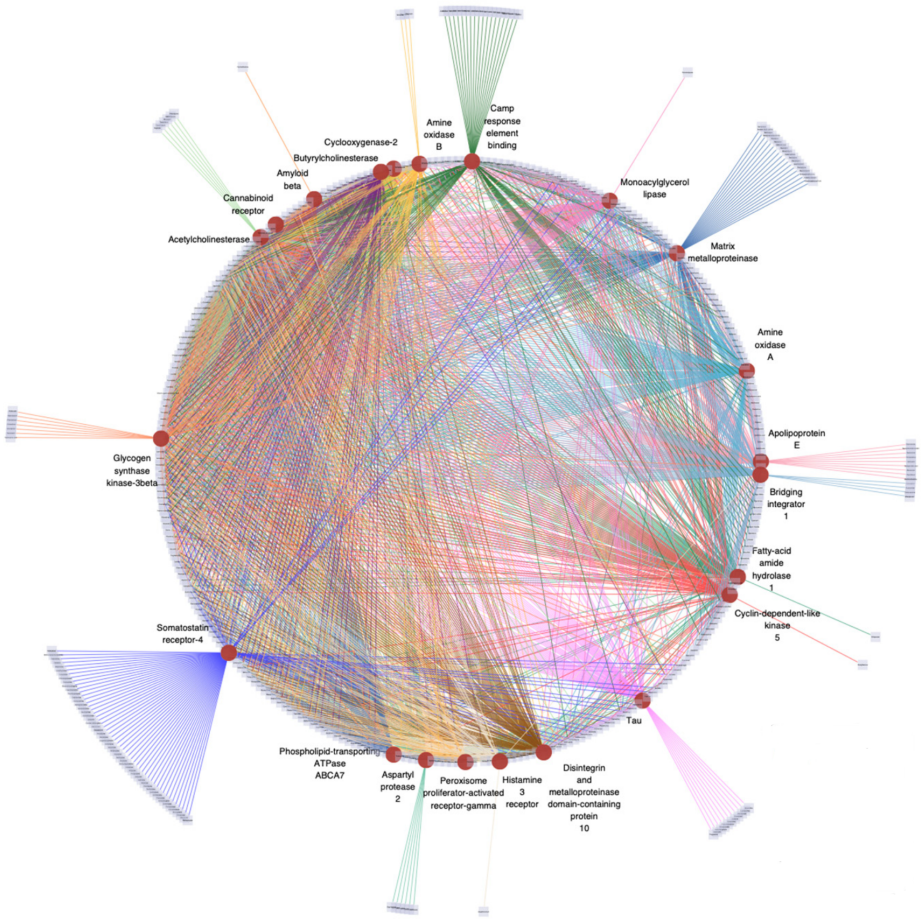

Figure 6.2: All the interactions in the drug–target data related to Alzheimer's disease. The red dots represent the targets, the purple squares represent drugs, and the lines represent the interactions between them. (Chen et al. 2024)

More broadly, the whole drug discovery field has possibilities of utilising developed computational models. In general, the computational power could offer a possibility to utilise large and complex medical datasets gathered from patients. The potential of integrating computational power with vast medical datasets could be a paradigm shift. When even bigger datasets can be collected and explored, the scope of drug discovery can be expanded beyond the drugs and targets to larger and more complicated entities.

For example, viruses cause complex effects in human body. New, computational models could help to understand how viruses affect health. By leveraging advanced computational models, it could be possible to first map the viruses that inhabit the human body, and then elucidate the complex interactions between viruses and human health. This kind of approach could give insights for new drug discoveries, new therapeutic strategies, and a better understanding of human body mechanisms.

Even though the development is fast, significant challenges remain. For instance, the connections between microbes, cells, human virome, and diseases, are extremely complicated, making it difficult to construct an accurate and interpretable system (Wu & Peng 2024). Interdisciplinary fields that combine these different research areas, such as biomedical engineering and health technology, are essential when facing these issues.

7 Conclusions

This thesis presented a DTI prediction model that is based on FM and is capable of utilising drug's and target's binary identifiers, as well as additional side information. The goal was to determine whether the side information improves the model. The research question was: Does the incorporation of side information improve drug–target interaction prediction compared to using only factorisation of an interaction matrix?

The proposed DTI prediction model is presented in chapter 4. Its complete architecture consists of six main modules: dataset loading, CV to replace train-test split, feature concatenation, CV for validation, FM for training, and C-index for evaluation. The architecture is illustrated in the Figure 4.2 and in the Algorithm 1.

To answer the research question, the results of the proposed DTI prediction model are presented in the chapter 5. The performance is evaluated with C-index with seven datasets and three feature options: using only binary identifiers, using only side information, and using both of these. The results presented in Tables 5.1 and 5.2 indicate that using only binary identifiers gives relatively high C-index values, using only side information improves the C-index values only in GPCR and ion channel datasets, and using both features enhances performance only in GPCR and ion channel datasets. In chapter 6, there is outlined the possibility to develop a better ML method instead of FM, as well as using multiple side information sources instead of only similarity matrices. These findings suggest that simpler models may suffice for DTI prediction when computational resources are limited.

Overall, the two most important components in the DTI prediction are feature representations and computational models. There are numerous databases for creating different feature representations of drugs and targets. These include different resources of drug space, target space, DTI networks, and side effect networks. When it comes to model's performance, particularly models based on various ML methods are found to be effective. (Chen et al. 2016)

The ML methods used in DTI prediction models can be categorised, for example, into six main categories: similarity based methods, DL methods, matrix factorisation methods, feature based methods, network based methods, and hybrid methods. The latter includes all the combinations of the methods. (Bagherian et al. 2021) The proposed model in this thesis is built with matrix factorisation. The performance of the model could be improved by combining it with other methods. The other—more computationally demanding—models could also possibly outperform it alone.

The information in the datasets used in this thesis is gathered from sources such as KEGG BRITE, BRENDA, SuperTarget, and DrugBank databases. These are commonly used, but there are also alternative sources to get drug–target data. Databases like STITCH, BindingDB, and ASDCD have other information about interactions, whereas ChemBank and ChEMBL databases contain more than 1.5 million molecular entries (Chen et al. 2016). Utilising a larger number of diverse databases as input data or as side information can improve the model's performance—requiring, however, more powerful computing.

Making full use of different heterogeneous data sources with computational models could enable more effective identification of interactions between drugs and targets. Also, developing effective computational DTI prediction models from heterogeneous datasets could aid in understanding human biological interactions, biological processes, drug discovery, and medicine improvement. (Chen et al. 2016)

DTI prediction offers multiple ways to accelerate and enhance traditional drug discovery processes. With access to more data and increasingly powerful computational models, DTI prediction is an valuable tool in the fields of biomedical engineering and health technology to promote health. Health technology and biomedical engineering are inherently interdisciplinary fields that leverage artificial intelligence in life sciences. These fields provide the necessary interdisciplinary framework to bridge computational advancements with real-world medical applications. As the research in this area continues to evolve, even better and more personalised treatment strategies could become available—ultimately improving individual and public health. DTI prediction, as demonstrated in this thesis, exemplifies how computational tools can accelerate drug discovery, paving the way for personalised medicine.

References

Alabdallah, A., Ohlsson, M., Pashami, S. & & Rögnvaldsson, T. (2024). "The Concordance Index decomposition: A measure for a deeper understanding of survival prediction models". *Artificial Intelligence in Medicine* 148, p. 102781.

Alzubi, J., Nayyar, A. & & Kumar, A. (2018). "Machine Learning from Theory to Algorithms: An Overview". *Journal of Physics: Conference Series* 1142.1, p. 12012.

Bagherian, M., Sabeti, E., Wang, K., Sartor, M. A., Nikolovska-Coleska, Z. & & Najarian, K. (2021). "Machine learning approaches and databases for prediction of drug–target interaction: a survey paper". *Briefings in Bioinformatics* 22.1, pp. 247–269.

Berthold, M., Borgelt, C., Höppner, F. & & Klawonn, F. (2010). *Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data.* 1st edition. Springer.

Blass, B. E. (2015). *Basic Principles of Drug Discovery and Development.* 1st edition. Academic Press.

Bolgár, B. & & Antal, P. (2017). "VB-MK-LMF: Fusion of drugs, targets and interactions using variational Bayesian multiple kernel logistic matrix factorization". *BMC Bioinformatics* 18.1.

Cao, D. S., Liu, S., Xu, Q. S., Lu, H. M., Huang, J. H., Hu, Q. N. & & Liang, Y. Z. (2012). "Large-scale prediction of drug–target interactions using protein

sequences and drug topological structures". *Analytica Chimica Acta* 752, pp. 1–10.

Chen, X., Yan, C. C., Zhang, X., Zhang, X., Dai, F., Yin, J. & & Zhang, Y. (2016). "Drug-target interaction prediction: Databases, web servers and computational models". *Briefings in Bioinformatics* 17.4, pp. 696–712.

Chen, Y., Liang, X., Du, W., Liang, Y., Wong, G. & & Chen, L. (2024). "Drug–Target Interaction Prediction Based on an Interactive Inference Network". *International Journal of Molecular Sciences* 25.14, p. 7753.

D'Souza, S., Prema, K. V. & & Balaji, S. (2020). "Machine learning models for drug–target interactions: current knowledge and future directions". *Drug Discovery Today* 25.4, pp. 748–756.

Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., Hocker, M., Treiber, D. K. & & Zarrinkar, P. P. (2011). "Comprehensive analysis of kinase inhibitor selectivity". *Nature Biotechnology* 29.11, pp. 1046–1051.

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd edition. O'Reilly Media.

Hancock, J. T. & & Khoshgoftaar, T. M. (2020). "Survey on categorical data for neural networks". *Journal of Big Data* 7.1, pp. 1–41.

Hao, Y., Cai, M. & & Li, L. (2019). "Drug repositioning via matrix completion with multi-view side information". *IET Systems Biology* 13.5, pp. 267–275.

Hassanzadeh, R. & & Shabani-Mashcool, S. (2022). "Does adding the drug-drug similarity to drug-target interaction prediction methods make a noticeable improvement in their efficiency?" *BMC Bioinformatics* 23.1.

Hattori, M., Okuno, Y., Goto, S. & & Kanehisa, M. (2003). "Development of a Chemical Structure Comparison Method for Integrated Analysis of Chemical and Genomic Information in the Metabolic Pathways". *Journal of the American*

Chemical Society 125.39, pp. 11853–11865.

He, Y., Sun, C., Meng, L., Zhang, Y., Mao, R. & & Yang, F. (2024). "Flexible drug-target interaction prediction with interactive information extraction and trade-off". *Expert Systems With Applications* 249, p. 123821.

Kim, J., Park, S., Min, D. & & Kim, W. (2021). "Comprehensive Survey of Recent Drug Discovery Using Deep Learning". *International Journal of Molecular Sciences* 22.18, p. 9983.

Kim, S., Jin, D. & & Lee, H. (2013). "Predicting drug-target interactions using drug-drug interactions". *Plos One* 8.11.

Koren, Y., Bell, R. & & Volinsky, C. (2009). "Matrix Factorization Techniques for Recommender Systems". *Computer* 42.8, pp. 30–37.

Li, J., Yang, X., Guan, Y. & & Pan, Z. (2022). "Prediction of Drug–Target Interaction Using Dual-Network Integrated Logistic Matrix Factorization and Knowledge Graph Embedding". *Molecules* 27.16, p. 5131.

Liu, Y., Zhang, Y. & & Wang, X. (2024). "SAGS-DynamicBio: Integrating Semantic-Aware and Graph Structure-Aware Embedding for Dynamic Biological Data with Knowledge Graphs". *Lecture Notes in Computer Science* 14949, pp. 287–302.

Lu, X., Liu, F., Li, J., He, K., Jiang, K. & & Gu, C. (2021). "An Efficient Computational Method to Predict Drug-Target Interactions Utilizing Matrix Completion and Linear Optimization Method". *Intelligent Computing Theories and Application* 12838, pp. 603–614.

Marsland, S. (2014). *Machine learning: An Algorithmic Perspective*. 2nd edition. CRC Press.

Merget, B., Turk, S., Eid, S., Rippmann, F. & & Fulle, S. (2017). "Profiling Prediction of Kinase Inhibitors: Toward the Virtual Assay". *Journal of Medicinal Chemistry* 60.1, pp. 474–485.

Metz, J. T., Johnson, E. F., Soni, N. B., Merta, P. J., Kifle, L. & & Hajduk, P. J. (2011). "Navigating the kinome". *Nature Chemical Biology* 7.4, pp. 200–202.

Nelson, H. (2023). *Essential Math for AI: Next-Level Mathematics for Efficient and Successful AI Systems*. 1st edition. O'Reilly Media.

Pahikkala, T., Airola, A., Pietilä, S., Shakyawar, S., Szwajda, A., Tang, J. & & Aittokallio, T. (2015). "Toward more realistic drug-target interaction predictions". *Briefing in Bioinformatics* 16.2, pp. 325–337.

Qiao, G., Wang, G. & & Li, Y. (2024). "Causal enhanced drug–target interaction prediction based on graph generation and multi-source information fusion". *Bioinformatics* 40.10.

Rang, H. P., Ritter, J. M., Flower, R., Henderson, G., Loke, Y. K. & & MacEvan, D. (2020). *Rang and Dale's Pharmacology*. 9th edition. Elsevier.

Rendle, S. (2010). "Factorization Machines". *IEEE International Conference on Data Mining*, pp. 995–1000.

Rendle, S. (2012). "Factorization Machines with libFM". *ACM Transactions on Intelligent Systems and Technology* 3.3, pp. 1–22.

Sajadi, S. Z., Chahooki, M. A. Z., Tavakol, M. & & Gharaghani, S. (2023). "Matrix factorization with denoising autoencoders for prediction of drug–target interactions". *Molecular Diversity* 27.3, pp. 1333–1343.

Salahudeen, M. S. & & Nishtala, P. S. (2017). "An overview of pharmacodynamic modelling, ligand-binding approach and its application in clinical practice". *Saudi Pharmaceutical Journal* 25.2, pp. 165–175.

Sarker, I. H. (2021). "Machine Learning: Algorithms, Real-World Applications and Research Directions". *SN Computer Science* 2.3, p. 160.

Shi, W., Yang, H., Xie, L., Yin, X. X. & & Zhang, Y. (2024). "A review of machine learning-based methods for predicting drug–target interactions". *Health Information Science and Systems* 12.1, p. 30.

Smith, T. F. & & Waterman, M. S. (1981). "Identification of common molecular subsequences". *Journal of Molecular Biology* 147.1, pp. 195–197.

Tang, J., Szwajda, A., Shakyawar, S., Xu, T., Hintsanen, P., Wennerberg, K. & & Aittokallio, T. (2014). "Making Sense of Large-Scale Kinase Inhibitor Bioactivity Data Sets: A Comparative and Integrative Analysis". *Journal of Chemical Information and Modeling* 54.3, pp. 735–743.

Trapotsi, M.-A., Mervin, L. H., Afzal, A. M., Sturm, N., Engkvist, O., Barrett, I. P. & & Bender, A. (2021). "Comparison of Chemical Structure and Cell Morphology Information for Multitask Bioactivity Predictions". *Journal of Chemical Information and Modeling* 61.3, pp. 1444–1456.

Vieira, S., Lopez, P., Walter, H. & & Mechelli, A. (2019). "Main concepts in machine learning". *Machine Learning: Methods and Applications to Brain Disorders*. Ed. by Academic Press, pp. 21–44.

Wang, C., Zhao, Z., Gu, P., Chen, C., Hu, Y., Xu, H., Sun, L. & & Shao, H. (2023). "Prediction of Drug-target Interactions Based on Multi-Modal Autoencoder". *The 16th International Conference on Advanced Computer Theory and Engineering*, pp. 75–79.

Wang, Y., Zhang, Y., Wang, J., Xie, F., Zheng, D., Zou, X., Guo, M., Ding, Y., Wan, J. & & Han, K. (2023). "Prediction of drug-target interactions via neural tangent kernel extraction feature matrix factorization model". *Computers in Biology and Medicine* 159, p. 106955.

Wu, Y. & & Peng, Y. (2024). "Ten computational challenges in human virome studies". *Virologica Sinica* 39.6, pp. 845–850.

Xu, R., Zhou, M. & & Chen, Y. (2019). "A Drug-Side Effect Context-Sensitive Network approach for drug target prediction". *Bioinformatics* 35.12, pp. 2100–2107.

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W. & & Kanehisa, M. (2008).

"Prediction of drug–target interaction networks from the integration of chemical and genomic spaces". *Bioinformatics* 24.13, pp. 232–240.

Yang, H., Chen, Y., Zuo, Y., Deng, Z., Pan, X., Shen, H. B., Choi, K. S. & & Yu, D. J. (2024). "MINDG: a drug–target interaction prediction method based on an integrated learning algorithm". *Bioinformatics* 40.4.

Ye, Y., Chen, Y., Zhang, Z., Wen, Y., He, S. & & Bo, X. (2021). "Drug-target interaction prediction based on non-negative and self-representative matrix factorization". *IEEE International Conference on Bioinformatics and Biomedicine*, pp. 2352–2359.

Zhai, H., Hou, H., Luo, J., Liu, X., Wu, Z. & & Wang, J. (2023). "DGDTA: dynamic graph attention network for predicting drug–target binding affinity". *BMC Bioinformatics* 24.1.

Zhao, W., Yu, Y., Liu, G., Liang, Y., Xu, D., Feng, X. & & Guan, R. (2024). "MSI-DTI: predicting drug-target interaction based on multi-source information and multi-head self-attention". *Briefings in Bioinformatics* 25.3.

Zhu, Z., Yao, Z., Zheng, X., Qi, G., Li, Y., Mazur, N., Gao, X., Gong, Y. & & Cong, B. (2023). "Drug–target affinity prediction method based on multi-scale information interaction and graph optimization". *Computers in Biology and Medicine* 167, p. 107621.