



CHARITÉ

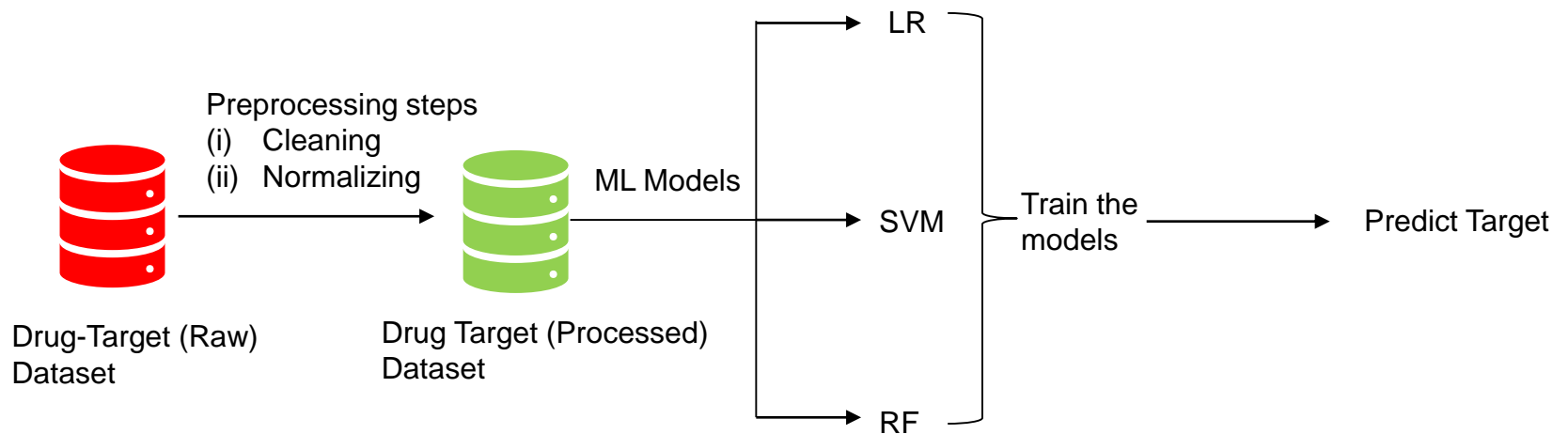
UNIVERSITÄTSMEDIZIN BERLIN

AG Preißner

1.2.2019

Software Praktikum 2019 – Target prediction I

4.3.2019 – 30.4.2019



- Building machine learning pipeline(preprocessing, normalizing, training and testing) for drug-target prediction
- ML models like Logistic Regression (**LR**), Support Vector Machine (**SVM**) and Random forest (**RF**) will be used to train the models
- Chemical fingerprints, structural similarity, chemical-chemical interaction will be used as input parameters in the dataset
- Train the models based on the training dataset and use the test dataset to predict the target
- Which model predicts better (higher accuracy and sensitivity)?

1. About 80% of people worldwide rely on herbs for some part of primary health care.

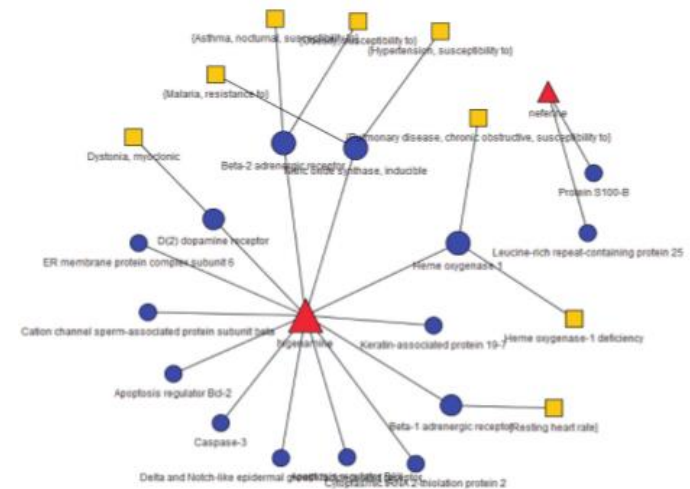
For example, herbal medicines are used by Chinese people over thousands of years, and are supported as effective by many research studies.

2. Any medicinal herb is a complex system with multiple ingredients (small molecules).

3. Most of drugs can cure the diseases because they can bind to the target proteins and perturb the proteins functions.

4. In order to understand the molecular mechanisms of herbal medicine:

- Herbs and their corresponding ingredients are collected.
- The potential targets proteins for the herbal ingredients are predicted.
- Through the proteins the connections between the herbal ingredients and the known drugs or diseases can be built.

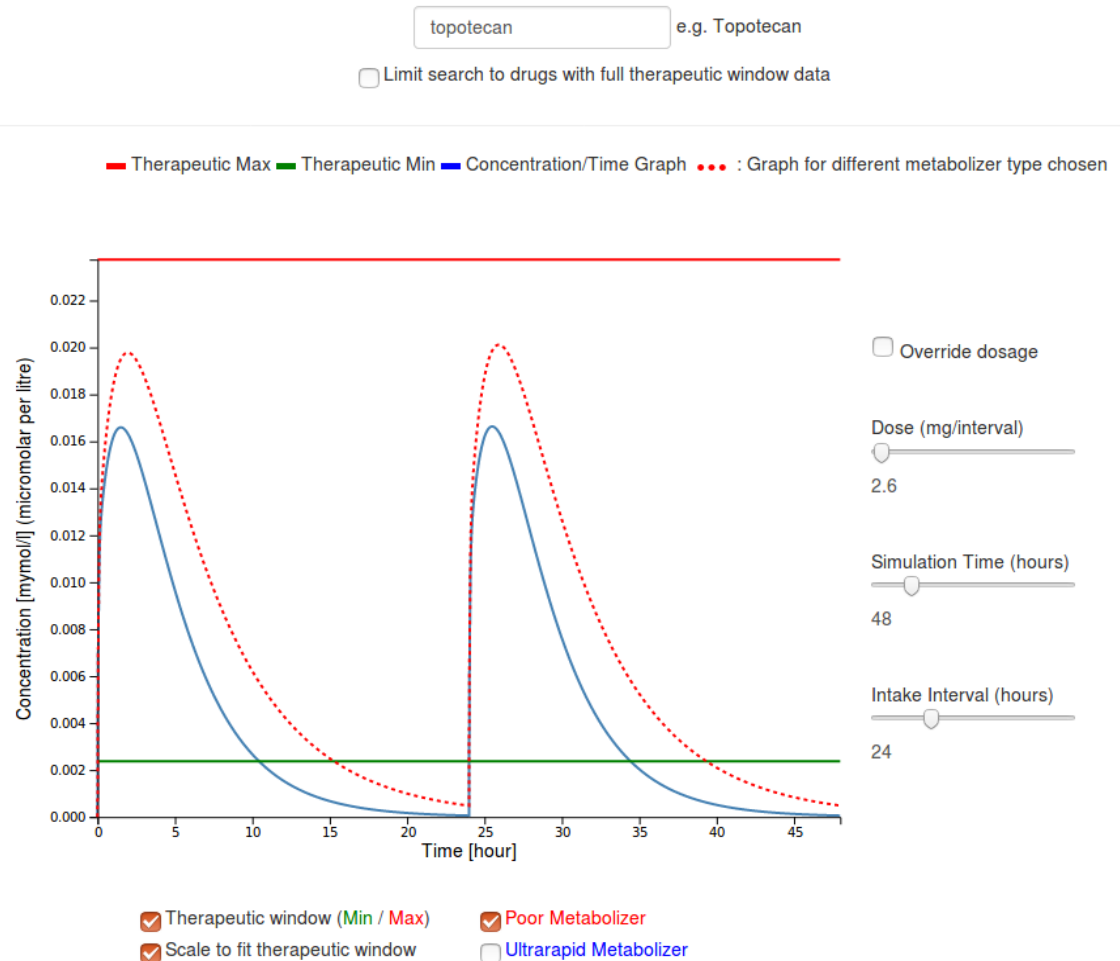


Herbal ingredient-target-disease network

- Pharmacokinetische Simulation zur Vorhersage der Plasmakonzentration von Medikamenten in Abhängigkeit von ihrer Dosis und Halbwertszeit

- Einbeziehung von Gabekriterien und Patientenphysiologie:

- Geschlecht
- Größe
- Gewicht
- Metabolische Variation
- Gabeintervall
- Gabedauer



Aufgaben:

- Entwicklung eines Javascript-Standalone-Prototypen zur Pharmacokinetiksimulation (auf Basis bestehenden Codes)
- Weiterentwicklung des Interfaces auf Basis von bestehendem Interface und Beispielcode (D3-Visualisierungsbibliothek)
- Implementierung einer alternativen Berechnung der Plasmakonzentration durch numerische Lösung des Differentialgleichungssystems (vs explizit wie bisher)
 - Motivation : Sättigungsverhalten führt zu nur numerisch lösbaren Situationen
- Simulation der Akkumulation über längere Zeiträume (Anzeige von "Tag n" statt Tag 1, frei wählbares $n > 0$)

- MySQL-Datenbank mit Angaben zur Häufigkeit des Co-Auftretens verschiedener Krankheiten, aus verschiedenen Quellen
- Interaktive Visualisierung:
- Weiter- oder Neuentwicklung einer existierenden JavaScript Heatmap mit Tooltips
- Filterung nach (statistischen) Kriterien
- Implementierung verschiedener Normalisierungsmethoden
- Vergleich Daten aus verschiedenen Quellen

