Prediction of anti-bacterial class specificity of compounds using machine learning

Yojana Gadiya[1,2], Andrea Zaliani[1] and Philip Gribbon[1]
[1]Fraunhofer ITMP ScreeningPort, Hamburg; [2]University of Bonn, Bonn

Motivation for approach

Current data repositories collect:
- Bacterial target or genetic data
- Antibiotic compound libraries
- Survilleance dataset

No effort to systematically collect preclinical data and leverage the data for guided drug discovery

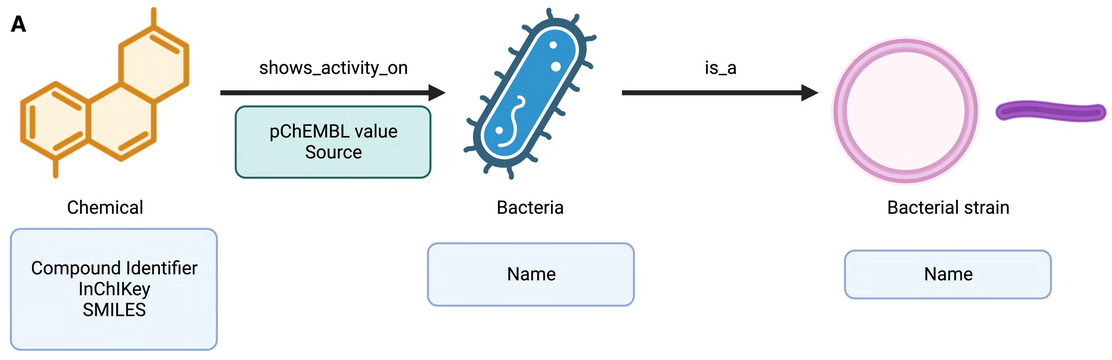
Preclinical databases
- CO-ADD
- PubChEM
- ChEMBL
- BindingDB
- DrugCentral and more

Existing ML pipelines focus on
- Virtual library generation
- Resistance mechanisms
- Resistant strain prediction
- Drug optimization among others

AMR Knowledge Graph

Collected, harmonized and generated a **graph database** for **publicly available bioactivity data**

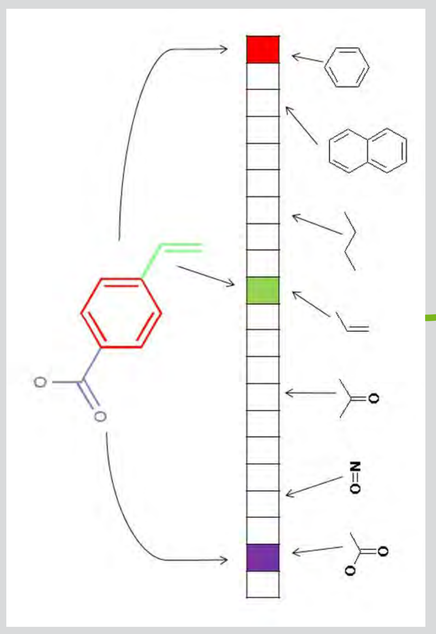
AMR-KG consist of:
- ~54,000 compounds
- 382 bacterial strains across 3 classes (Gram+/- and acid-fast)



AMR-KG Schema

Machine learning model evaluation and predictions

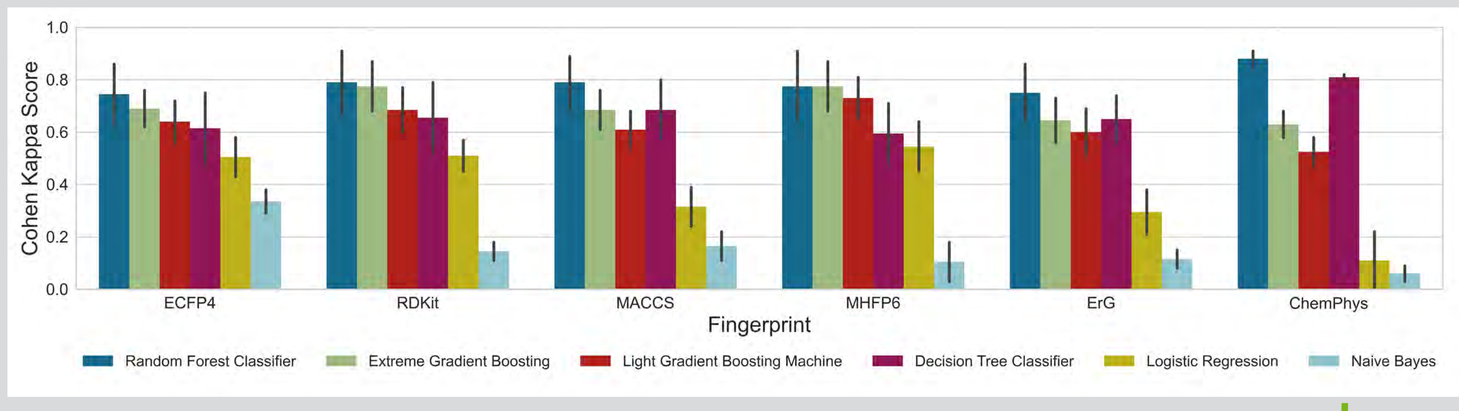
Model trained with five fingerprints: RDKit, MACCS, ECFP4, ErG, MHFP6 and ChemPhys



Compound fingerprint vector

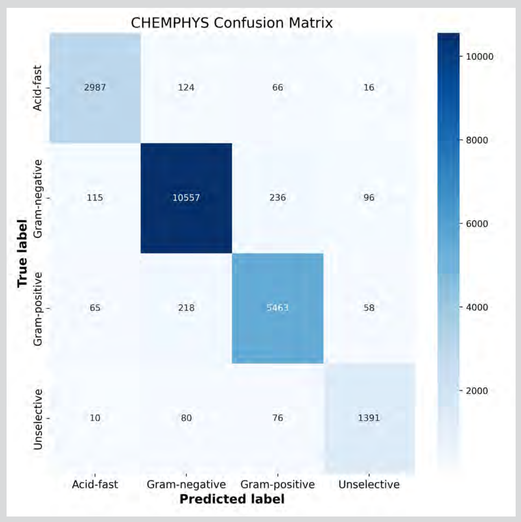
Find the best model

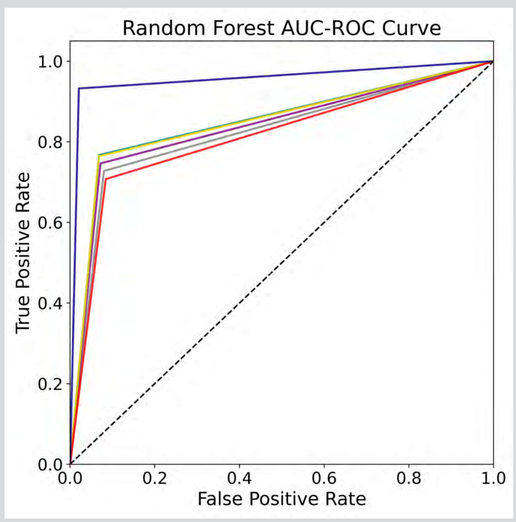
Choosing best model from a cohort based on Kappa score



Fingerprint	Accuracy	Kappa	Macro F1
ErG	0.80	0.67	0.72
ECFP4	0.82	0.70	0.76
RDKit	0.82	0.72	0.76
MACCS	0.82	0.72	0.77
MHFP6	0.84	0.74	0.79
ChemPhys	**0.95**	**0.92**	**0.94**

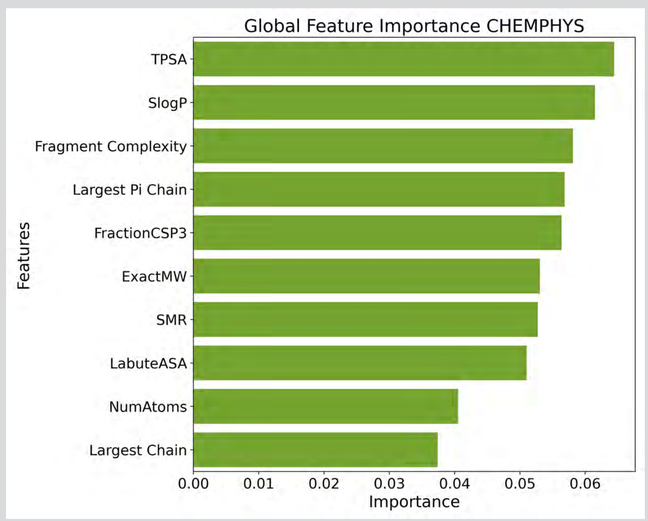
Optimizing the best model





Finalizing the best model with hyperparameter optimization using Optuna

Metrics shown here are on the test dataset (21,000 compounds)



Applicability of model

Use the model to select antibacterial compounds in:
- Prioritizing compounds in screening
- Filtering virtual screening for potential active compounds
- Drug repositioning
- In complement with antibacterial active vs inactive model

Where can you find us?

Contact: Yojana.Gadiya@itmp.fraunhofer.de
GitHub: https://github.com/IMI-COMBINE/broad_spectrum_prediction
Model will be deposited on BioModel with manuscript submission.

www.imi.europa.eu

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under Grant Agreement No 853967. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA companies' in kind contribution.