Yojana Gadiya et al., Fraunhofer ITMP

The development of novel therapeutics in antibiotic drug discovery has gradually slowed down in the past years. This is attributed to the lack of financial incentives supporting antimicrobial (AMR) drug discovery (ADD) and the decline in the efficacy of known drugs on bacteria, giving rise to drug-resistant pathogens. In the era of advanced computational methods, approaches like machine learning (ML) could be one potential solution to address the latter issues. In our work, we specifically developed one of the largest AMR knowledge graphs (KG), a data repository, for collecting and visualizing public bioassay data available in the field of AMR. Utilizing the data in AMR-KG, we build ML models to efficiently scan compound libraries for the identification of compounds demonstrating antibiotic activity. Moreover, our model assists in identifying the selectivity of these identified compounds to a class of bacterial strains such as Gram-positive, Gram-negative and Tuberculosis. Our strategy involved training seven classic ML models across six compound fingerprints: ECFP4, MACCS, RDKit, ErG, MHFP6 and chemical-physical (ChemPhys) properties. To counter imbalances in the dataset, which could affect model robustness, we performed a minority over-sampling technique allowing for an increment in the performance of our models over existing ML models. We found that XGBoost trained on ChemPhys fingerprint outperformed, demonstrating an accuracy of 87.1% and Cohen's Kappa score of 0.8. These results thus provide means for accelerating research in AMR drug discovery by filtering out non-antimicrobial compound scaffolds from screening collections in early ADD projects.

[1] Mongia, M., Guler, M., & Mohimani, H. (2022). An interpretable machine learning approach to identify the mechanism of action of antibiotics. Scientific Reports, 12(1), 10342.