H.E.P.A.R (Hepatic Ensemble Predictive Analysis Resource): A Convergent Machine Learning Architecture for Drug-Induced Liver Injury Prediction

Agarwal, Laksh (School: Carmel High School)
Dong, Neo (School: Carmel High School)

Drug-induced liver injury (DILI) accounts for 13% of acute toxicity cases in the U.S. and remains a primary barrier to pharmaceutical development. While most machine learning models predict DILI via chemical structure alone, few utilize bimodal approaches incorporating biological indicators. This study introduces HEPAR (Hepatotoxic Ensemble Predictive Analytics Resource), a novel framework combining chemical structure analysis with bioindicator data for holistic DILI prediction. The research utilized SMILES strings for 1,199 pharmaceutical compounds, analyzing topological fingerprints via the MACCS bit system and sourcing enzyme-compound interaction data through bioinformatics APIs. The predictive engine employs a diverse cohort of base learners, including deep neural architectures, gradient-boosted decision trees, Graph Neural Networks (GNNs), and Genetic Algorithms to process metabolic data. HEPAR identified a unique, understudied chemical bit significantly contributing to DILI, which was subsequently verified through hydrogen peroxide oxidation testing. An ensemble-based voting system demonstrated that combining base learners significantly improved performance over individual models. Metrics for Accuracy, MCC, and AUC-ROC consistently met or exceeded established hepatotoxicity benchmarks. HEPAR validates the effectiveness of bimodal analysis in toxicity prediction. By enabling preemptive DILI identification, this approach offers a pathway to accelerate drug development timelines while mitigating the public health risks associated with liver toxicity.

Awards Won:
Third Award of $1,200