An in silico Scaffold-Based Machine Learning Approach to Analyze the Influence of Molecular Descriptors and Substructures on the Blood-Brain Barrier Permeability of Neurological Drugs

Chakravarty, Ujesh (School: Haynes Academy for Advanced Studies)

The blood-brain barrier (BBB) remains a formidable challenge in pharmacology, blocking ~98% of small-molecule drugs from entering the central nervous system (CNS). To address this challenge, this project developed an interpretable in-silico framework to understand the structural drivers of BBB permeability. By combining the B3DB and BBBP datasets, a library of 6,590 unique molecules was used for training/testing. Structural features were extracted from each molecule using RDKit, encompassing ten physicochemical descriptors such as topological polar surface area, hydrophobicity, and molecular weight, alongside 2048-bit Morgan fingerprints and 166-bit MACCS keys representing molecular substructures. The implementation of Murcko scaffold-based splitting splits data based on core molecular structure, allowing for a true assessment of unseen chemical scaffolds in the test set. The best-performing Gradient Boosting model achieved 81.4% accuracy, Receiver Operating Characteristic-AUC of 0.903, and 89.5% recall on the scaffold-based test set. Feature importance and strip plot analysis identified TPSA, hydrogen bond donors, and molecular weight as significant predictors, while specific substructures, such as MACCS key 71, contributed as well. Scaffold-wise analysis revealed that 30 individual scaffolds (n = 5) achieved 100% prediction accuracy, highlighting the impact that core molecular scaffolds have on predictability. Furthermore, t-SNE clustering revealed a structural separation between permeable and non-permeable compounds in latent space. This investigation demonstrates that machine learning can effectively decipher structure-permeability relationships, which could provide a scalable, low-cost tool for identifying promising chemical features and predicting permeability.

Awards Won: