Machine Learning Strategies for Drug Discovery in AML: Focus on RUNX1 Bioactivity

Deepesh Kumar Verma
Contact: deepeshvashu@gmail.com Institution: ExcelR Solutions, Bangalore, Karnataka 560068, India

Introduction

- RUNX1 transcription factor, a critical gene for hematopoiesis, is highly prevalent in Acute myeloid leukemia (AML). Mutations within this gene are associated with poor patient outcomes.
- In the current study, we utilized a machine learning approach based on quantitative structure-activity relationships (QSAR) model to virtually design and predict versatile inhibitors of RUNX1.

Methods

- Data Preprocessing: Bioactivity data for ID CHEMBL2093862 was retrieved from the CHEMBL database.
- EDA & Lipinski descriptors: Chemical Space Analysis and Mann-Whitney U Test were performed to assess the drug-likeness of the compounds.
- Descriptor Calculation: PubChem fingerprints were generated for the compounds and assigned as the X variable, while the Y variable was set to the pIC50 values, representing their bioactivity.
- Machine Learning Models: Low-variance features were removed, followed by an 80/20 train-test split and application of 41 machine learning models for analysis.
- Deployment in web app: A machine learning model was deployed as a web app using the Streamlit framework.

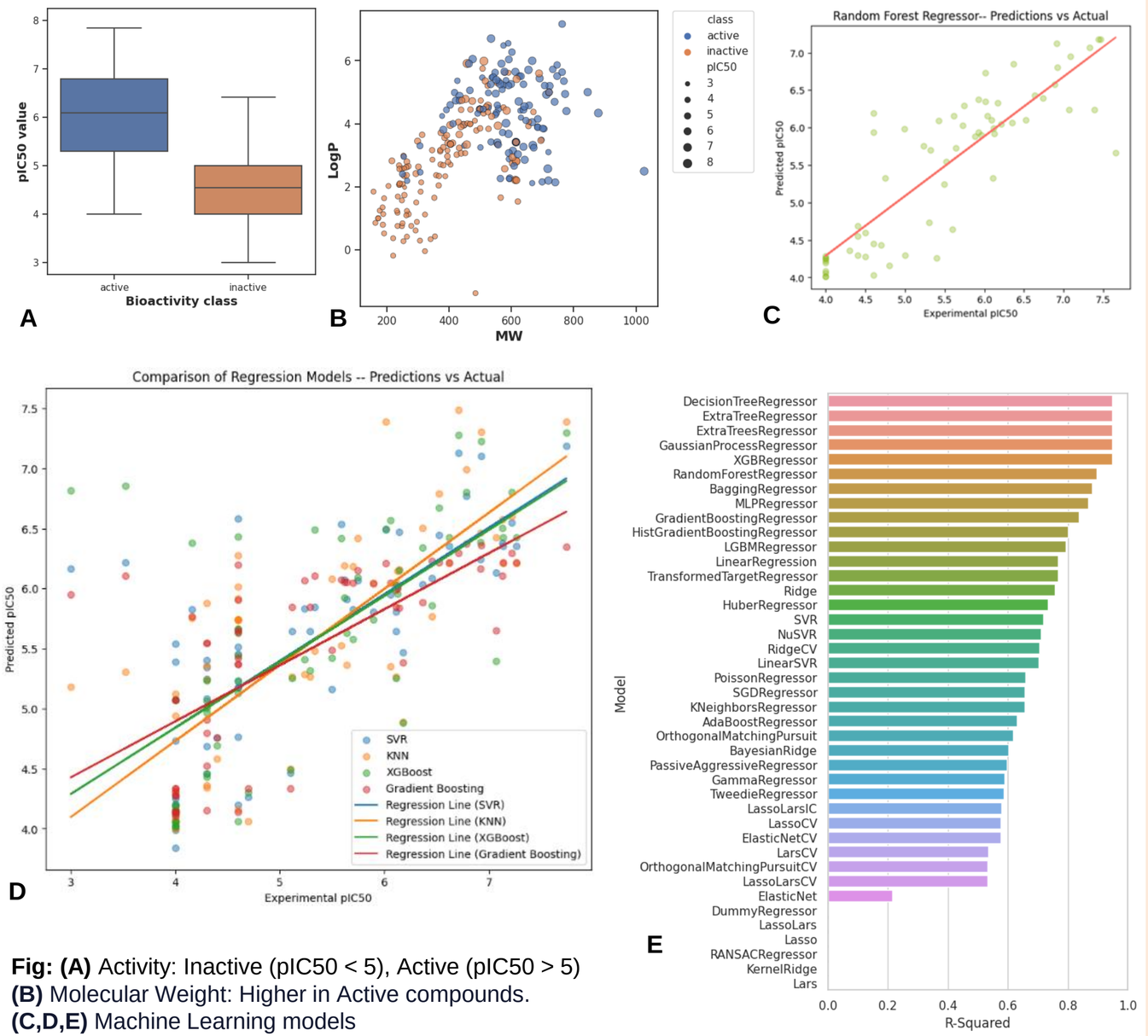

Fig: (A) Activity: Inactive (pIC50 < 5), Active (pIC50 > 5)
(B) Molecular Weight: Higher in Active compounds.
(C,D,E) Machine Learning models

Results

- Tree-based (Decision Tree, Random Forest) and boosting (XGBoost) models achieved superior performance (R^2 > 0.925).
- The web application predicts bioactivity, represented by pIC50 (a measure of inhibitory concentration at 50%), for various compounds targeting RUNX1.
- A web app predicts the bioactivity (pIC50) of multiple RUNX1-targeting compounds using their chemical structure (SMILES or composition) and name, achieving ~90% accuracy through cross-validation.
- The app is currently available only locally and needs further experimental validation.

Conclusion

- This study demonstrates the potential of machine learning for early drug discovery. By analyzing RUNX1-targeting compounds, the work highlights the ability of ML to identify key features for designing potent drugs against RUNX1.