

## Virtual screening of anti-mycobacterium tuberculosis based on machine learning approaches

Naidu Subbarao<sup>\*\*</sup>, Madhulata Kumari<sup>1</sup>, NeerajTiwari<sup>2</sup> and Subhash Chandra<sup>3</sup>

<sup>1</sup>Department of Information Technology, Kumaun University, India

<sup>2</sup>Department of Statistics, Kumaun University, India

<sup>3</sup>Department of Botany, Kumaun University, India

<sup>4</sup>School of Computational and Integrative Sciences, Jawaharlal Nehru University, India

In present work, we used three machine learning classifiers: Random Forest, Decision Tree, and Support Vector Machine to build three predictive models of an anti-mycobacterial dataset obtained from ChEMBL database and evaluated for their predictive capability. Based on the statistical evaluation, Random forest model showed the slightly higher area under the curve (AUC), better accuracy, sensitivity, and specificity in the cross-validation tests as compared to Decision Tree and Support Vector Machine. The statistical results indicated that the Random Forest model was the best predictive model with 93.83% accuracy, 98.4% ROC, Decision tree model came out to be the second with 92.8% accuracy, 95.2% ROC and the SVM model was found to be last with 90.64% accuracy, 93.8% ROC. Additionally, we built toxicity prediction models on the SingleCellIDSSTox carcinogenicity database (AID1189) using machine learning classifiers. The result showed that the Random Forest model was the best predictive model based on zero toxicity tolerance. We deployed the anti-mycobacterial and toxicity prediction models were on ChEMBL anti-malarial dataset containing 1.5 lakh compounds and a dataset of 1554 approved drugs resulting 27705 and 1554 compounds respectively as nontoxic anti-mycobacterial compounds. Thus machine learning models present highly efficient methods to find out novel hit anti-mycobacterial compounds. We suggest that such machine learning techniques could be very useful to screen drug candidates not only for tuberculosis but also for other diseases.