I CONGRESSO INTERNACIONAL DE BIOLOGIA CELULAR E MOLECULAR
X CURSO DE INVERNO

MACHINE LEARNING-BASED VIRTUAL SCREENING APPLIED TO ANTI-TUBERCULOSIS DRUG DISCOVERY

Diego de Souza Lima[1]*, Flávio Augusto Vicente Seixas[1]
Departamento de Tecnologia, UEM, Paraná, Brasil[1]
E-mail para contato*: diegodslima92@gmail.com

PBC Programa de Pós-Graduação em Ciências Biológicas

Introduction

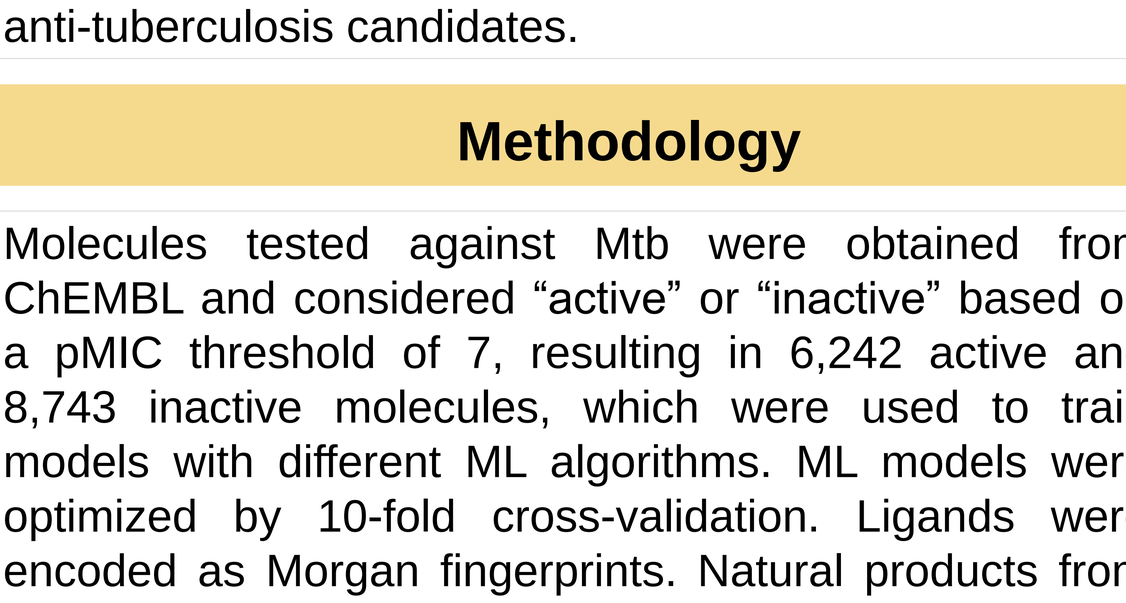
Tuberculosis (TB), a disease caused by *Mycobacterium tuberculosis* (Mtb), remains a major public health problem, especially in developing nations. Approximately 1.3 million deaths were reported in 2020, which more than 150,000 were caused by drug-resistant strains. Therefore, novel treatments are needed. Computational methods are an interesting approach due to their relatively low-cost and higher speed in comparison with traditional screening methods. Recently, machine learning (ML) became a promising method in early stages of drug discovery.

Objectives

Our goal was to develop an ML model able to classify molecules as either active or inactive, based on its two-dimensional structures, and to use it to virtually screen thousands of natural compounds in the search for novel anti-tuberculosis candidates.

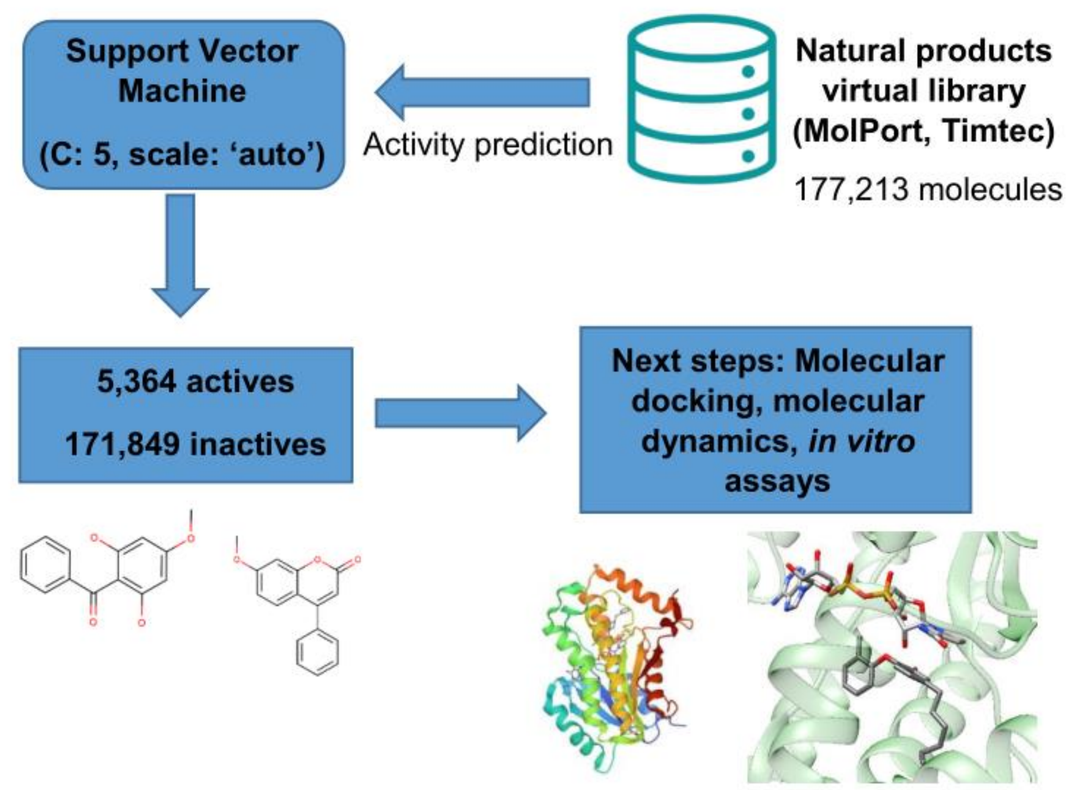
Methodology

Molecules tested against Mtb were obtained from ChEMBL and considered “active” or “inactive” based on a pMIC threshold of 7, resulting in 6,242 active and 8,743 inactive molecules, which were used to train models with different ML algorithms. ML models were optimized by 10-fold cross-validation. Ligands were encoded as Morgan fingerprints. Natural products from Molport and Timtec, totaling 177,213 compounds, were used for virtual screening.



Results

Support Vector Machines was the chosen algorithm, achieving 89% accuracy, 86% precision, 82% recall, and 85% F_1-score on the test set. After screening, 5,364 natural compounds were classified as potential actives.



Conclusions

Natural compounds were virtually screened in search of novel anti-tuberculosis candidates. Further studies will be conducted to validate our results, such as molecular docking and growth inhibition assays, contributing to anti-tuberculosis drug discovery.

Acknowledgements

Authors thank CNPq, CAPES, and Fundação Araucária for the financial support.

References

LO, Y. C.; RENSI, S. E.; TORNG, W.; ALTMAN, R. B. Machine learning in chemoinformatics and drug discovery. **Drug Discovery Today**, 2018.

WORLD HEALTH ORGANISATION. **World Health Organisation Global Tuberculosis Report 2014**. Geneva: World Health Organization, 2014.

UEM Universidade Estadual de Maringá