

Application of Machine Learning Algorithms in Predictive Analysis of the Antimicrobial Activity of Imidazolium Derivatives and Lung Cancer Recurrence

Andželika Lorenc, MSc

Supervisors: prof. Adam Buciński, dr Anna Romaszko-Wojtowicz

Abstract

The dynamic development of machine learning methods creates new opportunities for modeling complex biological and clinical phenomena. The objective of this doctoral dissertation was to evaluate the utility of Multilayer Perceptron (MLP) artificial neural networks in modeling two distinct medical problems: the antimicrobial activity of imidazolium compounds and the prediction of lung cancer recurrence based on clinical data. The dissertation comprises three thematically related publications, demonstrating a methodological evolution from binary classification to multidimensional regression and, ultimately, clinical modeling.

In the first study, an MLP model was developed, supported by decision tree methods and Principal Component Analysis (PCA), to classify the antimicrobial activity of 140 newly synthesized imidazolium derivatives against *Klebsiella pneumoniae*. Based on molecular descriptors determined via 2D and 3D methods, a model was constructed that successfully distinguished between active and inactive compounds, as confirmed by high classification performance metrics.

The second study focused on predicting the Minimum Inhibitory Concentration (MIC) values for 126 imidazolium derivatives against *Staphylococcus aureus* and *Candida albicans*. A hybrid approach was employed, incorporating predictor selection via decision trees, dimensionality reduction using PCA, and modeling with a Multilayer Perceptron. The model enabled the simultaneous prediction of $\ln\text{MIC}$ values for both microorganisms, confirming the efficacy of integrating machine learning algorithms even in complex regression-based problems.

The third study involved a retrospective analysis of clinical data from 2,296 lung cancer patients hospitalized between 2012 and 2017. An MLP model was developed and trained in two stages using the backpropagation and conjugated gradient algorithms. While the model achieved high specificity and identified the most significant predictors of recurrence, its sensitivity was assessed as moderate.

The results obtained in this dissertation confirm that artificial neural networks are an effective tool for modeling both chemometric and clinical data. They can significantly support the design of new biologically active compounds and enhance the medical decision-making process.

Key words: machine learning; artificial neural networks; antimicrobial activity; lung cancer recurrence;

