GC-MS PROFILING AND MACHINE LEARNING-BASED CHEMINFORMATICS IDENTIFICATION OF BIOACTIVE ANTIBACTERIAL COMPOUNDS FROM JEFFREYCIA ZEYLANICA LEAF EXTRACT

Kasun Chathura Kuruppu[1], Ayesh Madushanka[3], Nisansala Thilakarathne Muthunayake[2]*

[1]Department of Chemical Sciences, Faculty of Applied Sciences, Rajarata University of Sri Lanka, Mihintale, Sri Lanka

[2]Department of Biochemistry, Faculty of Medicine, Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka

[3]Department of Chemistry, Southern Methodist University, Dallas, United States of America

*Corresponding author: nisansalas@wyb.ac.lk

Abstract

Antibiotic resistance is a growing global health threat, driving the urgent need to identify novel plant-derived antimicrobial agents. *Jeffreycia zeylanica*, a Sri Lankan medicinal plant belonging to the family Asteraceae, traditionally used for skin diseases, asthma, and wound healing, remains largely unexplored scientifically. This study aimed to characterize the volatile and semi-volatile phytochemical constituents of the methanolic leaf extract of *J. zeylanica* using Gas Chromatography-Mass Spectrometry (GC-MS) and to develop a cheminformatics model to predict the antibacterial bioactivity of the identified compounds against *Staphylococcus aureus*. Leaf extract of *J. zeylanica* was prepared using methanol as a solvent via Soxhlet extraction. The resulting extract was analyzed on a Shimadzu GC-MS system fitted with a non-polar Rtx-5Sil MS fused silica capillary column, using helium as the carrier gas and electron ionization at 70 eV. Compound identification was performed by matching experimental mass spectra against the NIST library. A total of 30 compounds were identified, with n-hexadecanoic acid (17.83%), 9,12,15-octadecatrienoic acid (9.30%), neophytadiene (8.26%), and lupeol acetate (7.76%) as the dominant constituents. For the cheminformatics pipeline, bioactivity data for *S. aureus* were curated from the ChEMBL and NPASS databases. They were filtered to IC_{50} and MIC endpoints and standardized as pChEMBL values, with compounds classified as active (pChEMBL $\geq$ 6.0) or inactive. Molecular descriptors were computed using RDKit, complemented by Morgan fingerprints (2048-bit) and MACCS keys (167-bit) to form a hybrid feature set. Extreme Gradient Boosting (XGBoost) machine learning classifier was trained on known *S. aureus* antibacterial datasets. The model achieved an AUC of 0.9049 and an accuracy of 86.80%, demonstrating strong predictive performance. Applied to the GC-MS-identified compounds, the model predicted β-amyrone and lupeol acetate as the most promising antibacterial leads. These findings support *J. zeylanica* as a bioactive phytochemical source and validate GC-MS-coupled cheminformatics for antibacterial drug discovery.

Keywords: *Jeffreycia zeylanica*, GC-MS, Phytochemicals, Cheminformatics, Antibacterial