



# Utilizing machine learning techniques to predict the blood-brain barrier permeability of compounds detected using LCQTOF-MS in Malaysian Kelulut honey

R. Edros <sup>a,b</sup>, T.W. Feng <sup>a</sup> and R.H. Dong <sup>c</sup>

<sup>a</sup>Faculty of Chemical and Process Engineering Technology, Universiti Malaysia Pahang, Gambang, Malaysia; <sup>b</sup>Centre for Advanced Research in Fluid Flow, Universiti Malaysia Pahang, Gambang, Malaysia; <sup>c</sup>The Insight Centre for Data Analytics, School of Computer Science, University College Dublin, Dublin, Ireland

## ABSTRACT

Current *in silico* modelling techniques, such as molecular dynamics, typically focus on compounds with the highest concentration from chromatographic analyses for bioactivity screening. Consequently, they reduce the need for labour-intensive *in vitro* studies but limit the utilization of extensive chromatographic data and molecular diversity for compound classification. Compound permeability across the blood–brain barrier (BBB) is a key concern in central nervous system (CNS) drug development, and this limitation can be addressed by applying cheminformatics with codeless machine learning (ML). Among the four models developed in this study, the Random Forest (RF) algorithm with the most robust performance in both internal and external validation was selected for model construction, with an accuracy (ACC) of 87.5% and 86.9% and area under the curve (AUC) of 0.907 and 0.726, respectively. The RF model was deployed to classify 285 compounds detected using liquid chromatography quadrupole time-of-flight mass spectrometry (LCQTOF-MS) in Kelulut honey; of which, 140 compounds were screened with 94 descriptors. Seventeen compounds were predicted to permeate the BBB, revealing their potential as drugs for treating neurodegenerative diseases. Our results highlight the importance of employing ML pattern recognition to identify compounds with neuroprotective potential from the entire pool of chromatographic data.

## ARTICLE HISTORY

Received 20 April 2023  
Accepted 24 June 2023


## KEYWORDS

Machine learning; Kelulut honey; blood–brain barrier; chemometrics; random forest

## Introduction

Drug discovery, a critical element in pursuing precision medicine, frequently requires the longest time in the medicinal product development and commercialization process [1,2]. To date, machine learning (ML) has been used in drug design to better understand molecular diversity and develop precise drugs by extracting useful molecular descriptors from chemical analysis data in pharmacophore prediction [3]. Chemical graph theory mathematically defines atomic connections in compounds from the aspects of topology and combinatorics, thereby generating descriptors for

**CONTACT** R. Edros  [rzahirah@ump.edu.my](mailto:rzahirah@ump.edu.my)

 Supplemental data for this article can be accessed at: <https://doi.org/10.1080/1062936X.2023.2230868>

© 2023 Informa UK Limited, trading as Taylor & Francis Group