VenomPred: A Machine Learning Based Platform for Molecular Toxicity Predictions

Miriana Di Stefano [1,2] , Salvatore Galati [1], Elisa Martinelli [1], Marco Macchia [1], Adriano Martinelli [1], Giulio Poli [1], and Tiziano Tuccinardi [1,3]

[1] Department of Pharmacy, University of Pisa, 56126 Pisa, Italy; [2] Department of Life Sciences, University of Siena, 53100 Siena, Italy
[3] Center for Biotechnology, Sbarro Institute for Cancer Research and Molecular Medicine, College of Science and Technology, Temple University, Philadelphia, PA 19122, USA

mmvsl.pisa@virgilio.it

mmvsl.pisa

RDKit

Open-Source Cheminformatics
and Machine Learning

**RDKit UGM
12-14 October 2022
Berlin, Germany**

Università di Pisa
Molecular Modeling & Virtual Screening Laboratory
Dipartimento di Farmacia

VenomPred Platform

VenomPred is a user-friendly online platform for the evaluation of potential **mutagenic**, **hepatotoxic**, **carcinogenic** and **estrogenic** effects of small molecules. The platform is freely available, without downloading any software, by using reliable machine learning models.

Non-Toxic – High confidence	Non-Toxic – Low confidence
Toxic – Low confidence	Toxic – High confidence

DEVELOPMENT OF MACHINE LEARNING MODELS
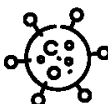
Modeling Datasets
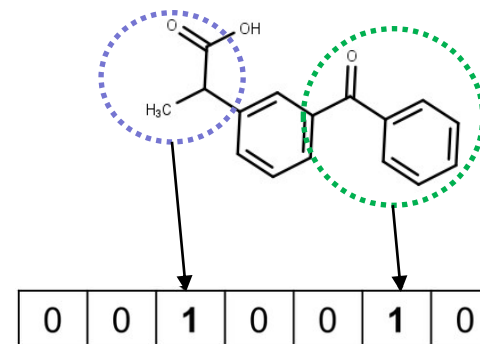
VEGA QSAR datasets

Mutagenicity
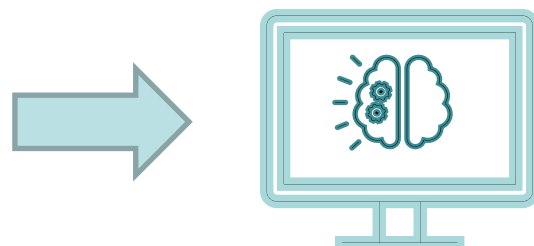
Hepatotoxicity

Carcinogenicity

Estrogenicity

Molecular representation

5 different types of molecular fingerprints (FPs)

RDKit
Morgan
Pharm2D
PubChem
LINGO

| 0 | 0 | 1 | 0 | 0 | 1 | 0 |

Classification Models

4 different classification algorithms

Random Forest
Support Vector Machine
k-Nearest Neighbor
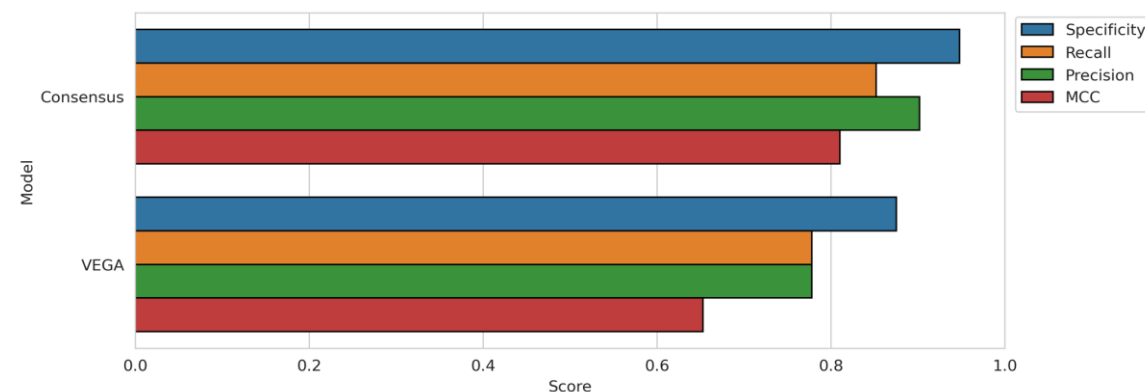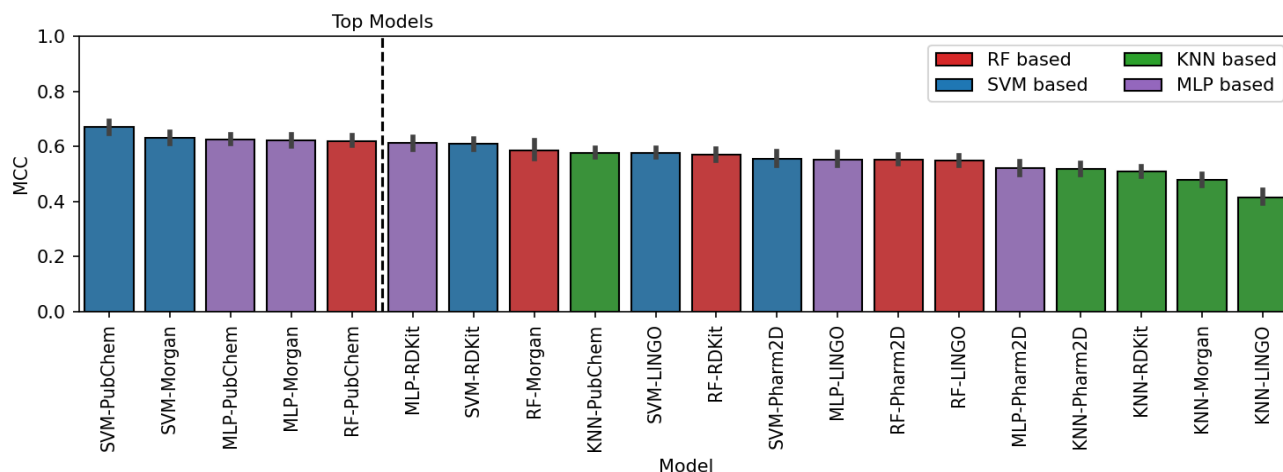**Multi-Layer Perceptron**

Model Building and Evaluation

Combination of FPs with each algorithm, setting the best hyperparameters and application of a 20-fold cross-validation (CV) in order to select the top 5 models for each endpoint for further analysis
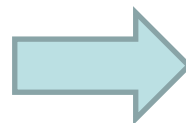
Consensus Strategy

Application of a consensus strategy combining the best-performing models in order to improve predictive performance. A consensus score (CS) for each tested molecule is computed by averaging the Probability Scores produced by the top-scored models.

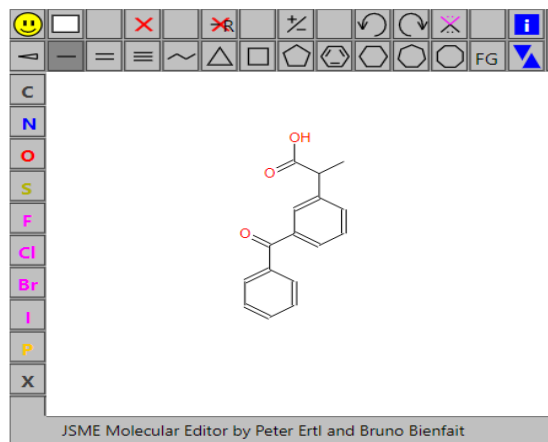**NON-TOXIC** : CS <0.5
TOXIC: CS ≥ 0.5

VenomPred Platform

1

http://www.mmvsl.it/wp/venompred/

2

JSME Molecular Editor by Peter Ertl and Bruno Bienfait

3

Name

Organization

Email

Endpoints
- ☐ Mutagenicity
- ☐ Carcinogenicity
- ☐ Hepatotoxicity
- ☐ Estrogenicity

4

SMILES*

CC(C1=CC(=CC=C1)C(=O)C2=CC=CC=C2)C(=O)O

Send Request

5
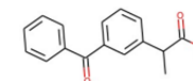
Name	Entry_1
SMILES	CC(C1=CC(=CC=C1)C(=O)C2=CC=CC=C2)C(=O)O

Endpoint	Probability
Mutagenicity	8 %
Carcinogenicity	23 %

Endpoint	Probability
Hepatotoxicity	53 %
Estrogenicity	45 %

Università di Pisa

Molecular Modeling & Virtual Screening Laboratory

Dipartimento di Farmacia

THANKS FOR YOUR ATTENTION

ACKNOWLEDGMENTS:
Prof. Dr. Marco Macchia, Prof. Dr. Tiziano Tuccinardi,
Dr. Giulio Poli, Salvatore Galati

mmvsl.pisa@virgilio.it
mmvsl.pisa