Guest Editorial Preface

Special Issue on Application of Machine Learning theories in QSAR/QSPR

Giuseppina Gini, DEIB, Politecnico di Milano, Milan, Italy

Modern QSAR/QSPR models are more and more developed to help in creating new molecules and to assess the risks of chemicals, according to various advanced norms. The rapidly evolving fields of Artificial Intelligence (AI) together with their subfields of Data Mining (DM) and Machine Learning (ML) continue providing new methods and tools to learn from data and to build models. QSAR and QSPR embraced the new ML paradigm for such different uses as computing and selecting descriptors, developing non-parametric models able to deal with large varieties of chemicals, integrating symbolic and statistics knowledge. Even though ML methods are largely used in modeling physical and biological properties of chemicals, there is a lack of clearly assessing their advantages or disadvantages in building QSAR/QSPR models.

In this special issue a variety of models for drug discovery, industrial applications, agriculture, and toxicity prediction are presented, all of them using new approaches based on ML and DM principles.

We start with a drug design paper by Vinicius Gonçalves Maltarollo, which deals with antibiotic resistance, and develops a family of QSAR classifiers of FabI inhibitors. Despite the diversity of the chemical classes present in the dataset, a family of models, based on physicochemical properties and fingerprints, and trained with the most popular ML methods (decision trees, random forests, multilayer perceptron, k-nearest neighbors, naïve Bayes, and support vector machines) are effective.

Khalid Bouhedjar and collaborators develop a QSPR model for the critical temperature of organic compounds; their use of ML ideas is about exploiting optimization, which is one of the pillars of ML, to compute the optimal descriptor DCW; they also use a dataset of very diverse chemical classes, and show the stability of the results applying the method to 3 different splits of training and test set.

Artificial neural networks, a tool of resurgent interest in ML, are used by Jillella Gopala Krishna and collaborators in the evaluation of the odor threshold of black tea and coffee. Again there are many chemical classes responsible for producing the odors that characterize those drinks. The new results outline the importance of some descriptors and confirm previous findings.

The toxicity of deep eutetic solvents is the target of the paper by Amit Kumar Halder and collaborators, who develop 2D QSTR models against marine bacteria Aliivibrio fischeri. Their dataset contains pure compounds and mixtures, so requiring special computation of the descriptors. Random forest, neural nets, and support vector machines are the ML methods used to create models that are extensively compared and discussed.

Finally, the paper by Zhibin Liang and coworkers tackles the problem of defining the fitness function for forward feature selection; it uses the BindingDB database to predict IC_{50} or K_i for seven datasets, balancing the attention both at improving the model fitting and at reducing model complexity.

Giuseppina Gini
Guest Editor
IJQSPR