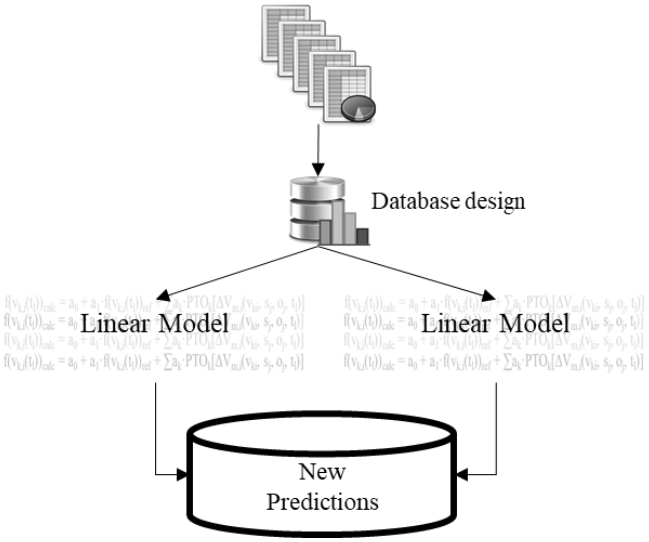
MOL2NET, International Conference Series on Multidisciplinary Sciences
Insert the title of the workshop here

Quantitative Structure-Activity Relationship (QSAR) Model Review

Harbil Bediaga Bañeres [a]

[a] *Department of Physical Chemistry, University of Basque Country UPV/EHU, 48940, Leioa, Spain.*

Graphical Abstract



Abstract.

The Quantitative Structure-Activity Relationship (QSAR) models are a very useful tool in the design of new chemical compounds. The QSAR methods are based on the assumption that the activity of a certain chemical compound is related to its structure.

Two types of QSAR analysis are summarized in this review: Linear Regression model (LR) and Linear Discriminant Analysis model (LDA).

Introduction

The Quantitative Structure-Activity Relationship (QSAR) models are a very useful tool in the design of new chemical compounds. The QSAR methods are based on the assumption that the activity of a certain chemical compound is related to its structure[1]. More precisely, this approach says that the activity, or the property, is related to the chemical structure through a certain mathematical algorithm, or rule. QSAR models are also called in silico methods, which actually refer to a somehow broader set of methods.

The number of published works in the last years has increased, this can be seen in **Figure 1**.

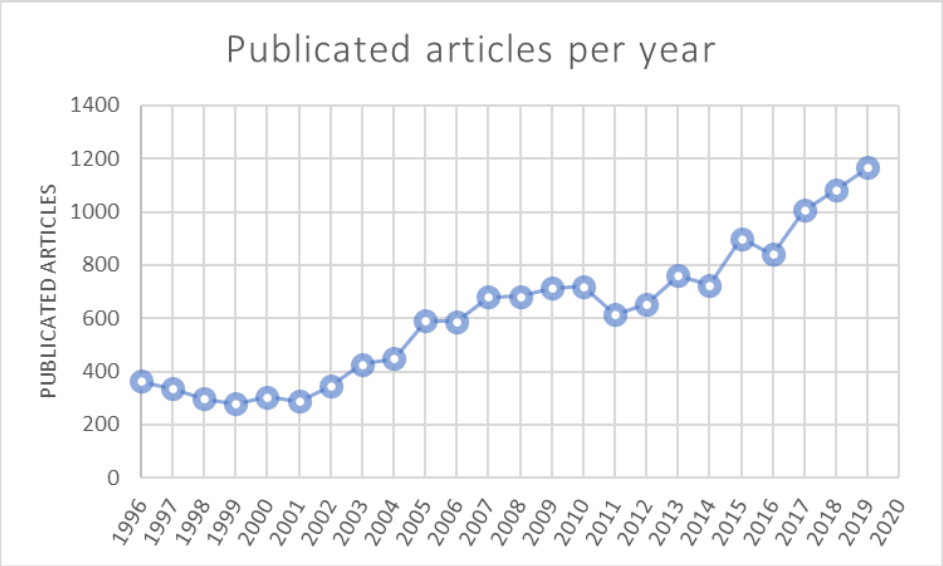


Figure 1. Review of published papers since 1996 with "QSAR" as keyword.

A QSAR method tries to find out a relationship between the response of activity or property, the dependent variable $f(v_{ij})_{calc}$ and the alterations in value of physical-chemical or structural attributes as change descriptors, **Equation 1**.

$$f(v_{ij})_{calc} = a_0 + a_1 D_1 + a_2 D_2 + \cdots + a_n D_n \quad (1)$$

Two types of QSAR analysis are summarized in this review: Linear Regression model[2] (LR) and Linear Discriminant Analysis model (LDA).

Materials and Methods

The main difference between these two techniques is that regression analysis deals with a continuous dependent variable, while discriminant analysis must have a discrete dependent variable.

The model generated with LDA method is a generalization of Fisher's linear discriminant analysis. It is a method used in statistics, pattern recognition, and machine learning to find a linear combination of features that characterize or separate two or more kinds of objects or events. The resulting combination can be used as a linear classifier. This type of model is used to predict bankruptcy[3], image recognition[4], marketing[5], medicinal studies[1], earth science[6]…

LR model is one of the most frequently used statistical methods and its used to predict a numeric response variable based on one or more numeric explanatory variables. Prediction is most straightforward when there is a straight-line relationship between a single explanatory variable and numeric response variable. This method is more appropriated when both variables are numeric. This model is a good option when the database has a homogeneous distribution, so it can be said that it has a homoscedastic distribution. The most common locations for these models are trend lines, medicine, informatics…

References

www.sciencedirect.com (12/06/2020)
1. Bediaga, H.; Arrasate, S.; González-Díaz, H., PTML Combinatorial Model of ChEMBL Compounds Assays for Multiple Types of Cancer. *ACS Combinatorial Science* **2018,** *20* (11), 621-632.
2. Pinder, J. P., Chapter 10 - Regression. In *Introduction to Business Analytics using Simulation*, Pinder, J. P., Ed. Academic Press: Boston, 2017; pp 313-369.
3. Lee, S.; Choi, W. S., A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications* **2013,** *40* (8), 2941-2946.
4. Moulin, C.; Largeron, C.; Ducottet, C.; Géry, M.; Barat, C., Fisher Linear Discriminant Analysis for text-image combination in multimedia information retrieval. *Pattern Recognition* **2014,** *47* (1), 260-269.
5. Haryani, S.; Motwani, B., Discriminant model for online viral marketing influencing consumers behavioural intention. *Pacific Science Review B: Humanities and Social Sciences* **2015,** *1* (1), 49-56.
6. Santos, V. S.; Nardini, V.; Cunha, L. C.; Barbosa, F.; De Almeida Teixeira, G. H., Identification of species of the Euterpe genus by rare earth elements using inductively coupled plasma mass spectrometry and linear discriminant analysis. *Food Chemistry* **2014,** *153*, 334-339.