Ali Niazi, Davood Nori-Shargh, Bahareh Yassar, and Ateesa Yazdanipour
Department of Chemistry, Faculty of Sciences, Islamic Azad University, Arak Branch, Iran, ali.niazi@gmail.com

A quantitative structure-property relationship (QSPR) study is suggested for the prediction of retention time ($R_t$) of central nervous drugs [1]. Ab initio geometry optimization was performed at the B3LYP level, with a known basis set, 6–31+G** [2, 3]. Local charges, electrostatic potential, dipole moment, polarizability, HOMO–LUMO energies, hardness, softness, electronegativity, electrophilicity and also structural descriptors (GETAWAY) were calculated for each compound. Among the investigation of QSAR, one of the most important factors affecting the quality of the model is the method to build the model. Many multivariate data analysis methods such as multiple linear regression (MLR), partial least squares (PLS) and artificial neural network (ANN) have been used in QSAR studies. MLR, as most commonly used chemometrics method, has been extensively applied to QSAR investigations. However, the practical usefulness of MLR in QSAR studies is rather limited, as it provides relatively poor accuracy. Modeling of the $R_t$ of central nervous drugs as a function of molecular structures was established by means of the least squares support vector machines (LS-SVM). The support vector machine (SVM) is a popular algorithm developed from the machine learning community. Due to its advantages and remarkable generalization performance over other methods, SVM has attracted attention and gained extensive applications. As a simplification of traditional of SVM, Suykens and Vandewalle [4] have proposed the use of least-squares SVM (LS-SVM). LS-SVM encompasses similar advantages as SVM, but its additional advantage is that it requires solving a set of only linear equations (linear programming), which is much easier and computationally more simple. In principle, LS-SVM always fits a linear relation ($y = w^T x + b$) between the regression ($x$) and the dependent variable ($y$) [3, 5-7].

This model was applied for the prediction of the $R_t$ of these drugs, which were not in the modeling procedure. The resulted model showed high prediction ability with root mean square error of prediction of 0.0034 for LS-SVM. Results have shown that the introduction of LS-SVM for quantum chemical and structural descriptors drastically enhances the ability of prediction in QSAR studies superior to multiple linear regression and partial least squares.

References