DOI: https://doi.org/10.62985/j.huit_ojs.vol25.noS1_ICA.262

APPLICATION OF QSPR APPROACH FOR DEVELOPMENT OF NOVEL METAL-THIOSEMICARBAZONE COMPLEXES

Nguyen Minh Quang[1,*], Huynh Ngoc Chau[1], Pham Van Tat[2]

[1]*Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam*

[2]*Institute of Pharmaceutical Education and Research, Binh Duong University, Ho Chi Minh City, Vietnam*

*Email: *nguyenminhquang@iuh.edu.vn*

Received: 2 May 2025; Revised: 20 May 2025; Accepted: 31 May 2025

ABSTRACT

Twenty novel metal-thiosemicarbazone complexes (ML_2) were calculated the stability constants ($\log\beta_{12}$) based on the quantitative structure-property relationship (QSPR) models. The QSPR models were developed using multivariate linear regression (MLR), support vector regression (SVR), and artificial neural network (ANN) methods. Descriptors of the models were calculated from the PM7 and PM7/sparkle semi-empirical quantum mechanisms. The quality of the QSPR models was tightly controlled by the statistical values of OECD instructions and Tropsha's standards. As a result, the best $QSPR_{MLR}$ model includes five variables: *Dipole, xv2, xch5, SHBa, and* ^5C, with statistical values such as R^2_{train} = 0.922, Q^2_{LOO} = 0.861, and RMSE = 0.759. Besides, the best $QSPR_{SVR}$ model consists of capacity C = 10.0, gamma γ = 0.10, and epsilon ε = 0.1 with the number of support vectors equal to 42 and suitable regression parameters: R^2 = 0.925, and $RMSE_{CV}$ = 0.536. The $QSPR_{ANN}$ model with network architecture I(5)-HL(6)-O(1) and exponential transfer function was trained from descriptors of the MLR model and showed impressive results as R^2_{train} = 0.986; Q^2_{test} = 0.876 and $Q^2_{validation}$ = 0.921. In addition, this study used an external validation (EV) dataset of 25 $\log\beta_{12}$ experimental values to build complete QSPR models with $Q^2_{EV\text{-}MLR}$, $Q^2_{EV\text{-}SVR}$, and $Q^2_{EV\text{-}ANN}$ values of 0.834, 0.865, and 0.881, respectively. The positive results of the models can be used to find other new thiosemicarbazone and their complexes for applications in chemical, analytical, and environmental fields.

Keywords: ANN, MLR, QSPR model, stability constants $\log\beta_{12}$, SVR, metal-thiosemicarbazone complexes.

1. INTRODUCTION

In organic chemistry, thiosemicarbazone is a common compound utilized extensively in related domains. For this reason, they have drawn the interest of scientists and are frequently researched and combined in studies. Because thiosemicarbazone's structure specifically binds nitrogen and sulphur donors, it is also particularly easy to form complexes with metal ions. Their compounds have also been used in numerous related studies, particularly in analytical chemistry, where the stability constant is a defining feature of the capacity to form complexes with the appropriate ligands. Thus, researchers are constantly interested in discovering novel thiosemicarbazone compounds with exceptional benefits and various uses [1].

Metal-thiosemicarbazone complexes have a wide range of biological actions, including anticancer and antibacterial qualities, which makes their study a rapidly expanding field in medicinal chemistry [2]. A strong substitute for forecasting the effectiveness of novel compounds before their synthesis is the application of quantitative structure-property relationship (QSPR) techniques, since the synthesis and characterization of these complexes require a substantial investment of time and resources. To expedite the medication discovery process, researchers can evaluate different molecular characteristics and correlate them with biological activities using computational models. Recent developments in cheminformatics have significantly improved our ability to examine molecular interactions, which is essential for forming complexes. Thus, the use of QSPR techniques promotes a more effective synthetic pathway, allowing for the discovery of promising therapeutic intervention candidates for disease treatment and offering information on the chemical makeup of metal-thiosemicarbazone interactions [3].

A key tool in medicinal chemistry is QSPR modelling, which enables scientists to forecast a compound's biological activity based on its chemical structure. Drug development is streamlined by this computational strategy, which lessens the need for expensive and time-consuming experimental techniques. QSPR's ability to match chemical descriptors with pharmacological effects through statistical approaches makes finding lead compounds with desired qualities easier [3]. The importance of QSPR is obvious when creating new agents. In such metal-thiosemicarbazone complexes, knowledge of the delicate relationship between structure and activity can spur the development of innovative antibacterial treatments. Recent research that synthesized pyrazolone derivatives and evaluated their biological activities using computational techniques has shown that QSPR successfully directs the development of compounds with enhanced efficacy against infections resistant to several drugs [4]. In the end, this combined strategy of synthesis and modelling increases the possibility of finding novel, efficient treatments.

In order to determine the best models for creating new, significant compounds, this study builds QSPR models on complexes of thiosemicarbazone and metal ions using a comprehensive approach that combines quantum chemistry, new semi-empirical techniques (PM7 and PM7/sparkle) [5], and statistical computing/mathematical tools. The OECD's [6] and Tropsha's guidelines [7] are rigorously adhered to when creating QSPR models.

The multivariable linear regression method (MLR) was used in this work to successfully construct the QSPR model on a data set that included 99 experimental values of complexes between thiosemicarbazone and the metal ions Ni^{2+}, Cd^{2+}, Fe^{3+}, Co^{2+}, Cu^{2+}, Zn^{2+}, Fe^{2+}, and Mn^{2+} in aqueous solution (Table 1). In order to identify more accurate predictive models, we next created machine-learning models using the $QSPR_{MLR}$ model's descriptors, including the support vector regression ($QSPR_{SVR}$) and the artificial neural network model ($QSPR_{ANN}$). Using these models, the study added synthetic functional groups to possible places in the original structural framework to anticipate the stability constant values for 20 unique complexes on some newly created thiosemicarbazone derivatives (Figure 6 and Figure 7).

2. METHODOLOGY

The first step in this examination is to find experimental data since this study needs to begin with an empirical dataset [3].

2.1. Data mining

The mononuclear complex (ML_2) of two thiosemicarbazone ligands (L) and a metal ion (M) is the subject of the investigation; Fig. 1 depicts the ligand and complex structures.

Consequently, the following formula is used to get the stability constant (β_{12}):

$$\beta_{12} = \frac{[ML_2]}{[M][L]^2} \quad (1)$$

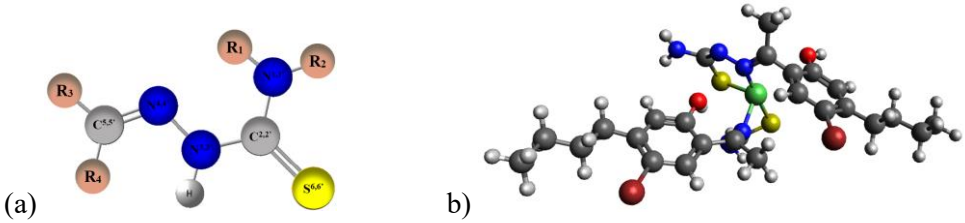

Figure 1. (a) Structure of thiosemicarbazone; (b) Structure of its complex [8]

Table 1. 99 stability constants of complexes with average ($\log\beta_{12,avg}$) values are included in the experimental dataset (n).

No.	Thiosemicarbazone				Metal ions	Number of complexes, n	$\log\beta_{12,avg}$	Ref.
	R_1	R_2	R_3	R_4				
1	H	H	$-CH_3$	$-C_{10}H_{13}OBr$	Ni^{2}+	1	6.549	[8]
2	H	H	$-C_2H_5$	$-C_{10}H_{13}O_2Br$	Ni^{2+}	1	8.807	[9]
3	H	H	$-C_2H_5$	$-C_{10}H_{13}O_2Br$	Cd^{2+}	1	6.927	[10]
4	H	H	H	$-C_{13}H_{16}NO_3$	Fe^{3+}	1	33.320	[11]
5	H	H	$-C_3H_7$	$-C_6H_4O_2Br$	Co^{2+}	1	10.4362	[12]
6	H	$-C_6H_5$	$-C_6H_5$	$-C_5H_4N$	Cu^{2+}	16	11.161	[13]
7	H	$-C_6H_5$	$-C_6H_5$	$-C_5H_4N$	Co^{2+}	16	10.399	[13]
8	H	$-C_6H_5$	$-C_6H_5$	$-C_5H_4N$	Zn^{2+}	16	11.161	[13]
9	H	$-C_6H_5$	$-C_6H_5$	$-C_5H_4N$	Ni^{2+}	16	10.186	[13]
10	H	H	$-C_2H_5$	$-C_{10}H_{13}O_2Br$	Fe^{2+}	1	7.086	[14]
11	$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Fe^{2+}	1	10.250	[15]
12	$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Co^{2+}	1	12.470	[15]
13	$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	2	11.445	[15,16]
14	$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	2	7.880	[15,16]
15	H	H	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	1	7.360	[16]
16	H	H	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	1	11.290	[16]
17	H	H	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	1	12.160	[16]
18	H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	1	7.000	[16]
19	H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	1	11.110	[16]
20	H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	1	12.430	[16]
21	H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	1	10.460	[16]
22	H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	1	7.2000	[16]
23	H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	1	11.130	[16]

No.	Thiosemicarbazone				Metal ions	Number of complexes, *n*	$\log\beta_{12,avg}$	Ref.
	R_1	R_2	R_3	R_4				
24	H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	1	12.580	[16]
25	H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	1	10.270	[16]
26	$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	1	7.7600	[16]
27	$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	1	12.490	[16]
28	H	$-C_6H_5$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	1	11.320	[16]
29	H	$-C_6H_5$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	1	12.410	[16]
30	H	$-C_6H_5$	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	1	10.210	[16]
31	H	$-C_3H_5$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	1	7.3300	[16]
32	H	$-C_3H_5$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	1	11.140	[16]
33	H	$-C_3H_5$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	1	12.530	[16]
34	H	$-C_3H_5$	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	1	10.220	[16]
35	H	H	$-C_6H_4OH$	$-C_6H_4OH$	Cu^{2+}	1	11.357	[17]
36	H	H	$-C_6H_4OH$	$-C_6H_4OH$	Co^{2+}	1	12.215	[17]
37	H	$-C_6H_5$	$-CH_3$	$-C_9H_{10}O_2Br$	Ni^{2+}	1	10.085	[18]

A big data set is first gathered from empirical investigations, and then smaller data sets are separated using the clustering method; in this case, the Agglomerative Hierarchical Clustering (AHC) method is applied sequentially. The results of a database including 99 metal-thiosemicarbazone complex stability constant values were chosen to build the models in this study. Table 1 displays the comprehensive statistics.

Table 2. Descriptive statistics for the training and test datasets for building models

Statistical values	Training data set	Test data set
mean	10.738	10.740
standard error	0.266	0.527
SD	2.641	2.635
sample variance	6.979	6.947
range	27.76	9.739
minimum	5.560	6.900
maximum	33.320	16.639
observations	99	25

Table 2 and Figure 2 show a high similarity between the training and validation data sets in terms of both descriptive statistics and distribution shape. This is favorable for training machine learning models, ensuring good representativeness and generalization ability when applying the model in practice. Because the training and validation data have similar distributions in terms of median, variance, distribution shape, and data point density, the data is reasonably divided, without "data leakage". The training model has a good generalization ability to unseen data (validation), reducing the risk of overfitting.

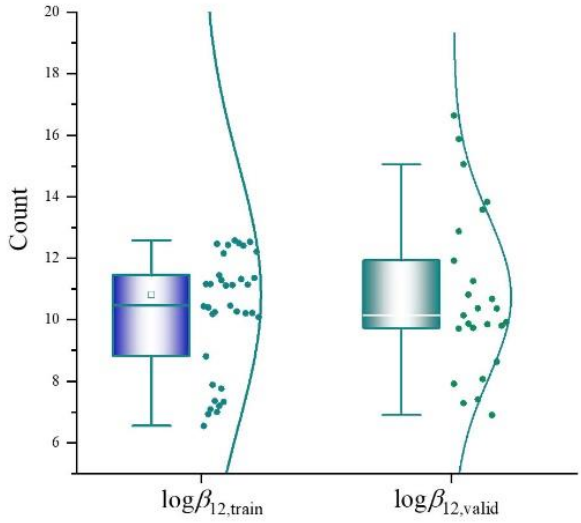

Figure 2. The descriptive statistics and distribution of the training and validation data sets

2.2. Descriptors

The variables of the built-in model equation are known as the descriptors of the model utilized in QSPR modelling research. They consist of molecular and quantum descriptors. While the molecular descriptors describe the structure of the study molecules in QSPR models and are computed from specialized tools for complicated structures that have been optimized, the quantum description is released from the structure optimization process [3].

Once the calculated descriptors have been summarized, the variables must be filtered; those with the same value of greater than 70% must be eliminated from the data set. The stability constant values ($\log\beta_{12,exp}$) of experimental complexes and structural parameters are thus included in the data collection. Lastly, we utilize this dataset to create QSPR models, including support vector regression, artificial neural networks, and multivariable linear regression models.

2.3. QSPR modeling

2.3.1. Multiple linear regression models

The QSPR model was constructed in the study using multivariable linear regression. One of the statistical techniques frequently employed in research projects is this one. The degree of the linear relationship between the dependent variables (the expected values) and several independent factors is ascertained using this method. Here, the independent variables are the chemical descriptors, and the dependent variable is the stability constant value ($\log\beta_{12}$) of the complexes between thiosemicarbazone and metal ions. Equation (2) represents the model of this approach: [3]

$$Y = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + ... + \beta_k.X_k + \varepsilon \qquad (2)$$

where Y is the dependent variable; β_0, β_1, β_2…, β_k are regression parameters of the model; X_i corresponds to the i^{th} explanatory variable (with i = 1 to k) and ε is the random error.

2.3.2. Support vector regression models

Support vector regression (SVR) is a machine learning (ML) technique that is commonly used in statistical mathematics and computer science to develop predictive models and data analysis algorithms such as classification, regression analysis, and prediction [3]. The support vector algorithm is a nonlinear algorithm, and the algorithm was first discovered in 1963 in Russia by Vapnik and Lerner [19]. In 1992, Vapnik et al. [20] proposed setting up a regression model by applying the kernel trick. The algorithm replaces vector scalar products with a nonlinear kernel function. This allows the algorithm to fit the maximization of the margins of the hyperplane in the transformation space, and the results show that the algorithm works well.

The kernel parameters, including capacity (C), gamma (γ), and the number of support vectors (n), as well as the kernel function selection, all affect SVR efficiency. Cross-evaluation is typically used to verify each parameter selection, and the kernel parameters with the highest accuracy are selected. The model is trained on the complete training dataset using an appropriate kernel function and optimizing the parameters. A variety of kernel functions are used when training a regression model. In order to find the best parameters, the study employed the Radial Basis Function (RBF) as follows [20]:

$$K(\vec{x}_i,\vec{x}_j)=\exp\{-\gamma\|\vec{x}_i.\vec{x}_j+r\|^2\},\ \gamma>0 \qquad (3)$$

here γ is the kernel parameter, r is a constant.

2.3.3. Artificial neural network models

An artificial neural network (ANN), which approximates neurons in the biological brain, functions on a collection of interconnected units or nodes known as artificial neurons. In traditional ANN implementations, each artificial neuron's output is determined by a non-linear function of the sum of its inputs, and the signal at the artificial neuron's link is genuine. Artificial neurons are typically put together in layers. Various layers can transform their input in different ways. The transfer function expresses this relationship, and the two transfer functions, hyperbolic tangent and exponential, were employed in this investigation [21].

Additionally, this study employs a back-propagation technique and a multi-layer perceptron (MLP) network type. An input layer, an output layer, and one or more hidden layers make up the structure of this MLP-type network. The least squares (LMS) algorithm is a generalised version of the back-propagation method. An approximation approach for determining the places at which network performance is optimal was proposed by Rumelhart et al. in 1986 [22]. The following procedures are used to accomplish the algorithm: back-propagation, direct propagation of calculations throughout the network, and appropriately updating the weights and offsets. The method terminates when the objective function's value is sufficiently tiny.

2.4. Model evaluation

An essential first step in verifying the accuracy of the constructed QSPR model is model evaluation. Two separate data sets are evaluated internally and externally as part of the model evaluation process. The initial training dataset with 99 experimental values (Table 1) was subjected to internal evaluation using the evaluation statistics index Q^2_{LOO} (>0.6) in conjunction with cross-validation (CV) based on the technique of eliminating leave-one-out (LOO) and evaluation index R^2 (>0.6). Furthermore, models with an evaluation index of Q^2_{EV} (>0.5) are evaluated externally (EV) on the independent data set [7]. Using various data sets, these amounts are computed using the same formula as follows [3]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (6)$$

where Y_i, $\hat{Y}_i$, and $\bar{Y}$ are the experimental, predicted, and average values.

The standard error (SE) is a crucial tool for detecting data and assessing how error models have changed. Since SE and sample size are theoretically inversely related, a higher number of observations results in a lower value. The formula is used to determine it [3]:

$$SE = \sqrt{\frac{\sum_{i=1}^{N}(\mathrm{Y_i} - \hat{\mathrm{Y}}_\mathrm{i})^2}{N - k - 1}} \quad (7)$$

here, N and k are the number of variables of the training set and the model, respectively.

Using an appropriate number of support vectors, the SVR model search method optimizes parameters like capacity (C), gamma (γ), and epsilon (ε). Development occurs until a slight variation exists between cross-validation ($RMSE_{CV}$) and calibration ($RMSE_C$). The $RMSE_{CV}$ values, which are determined using equation [20], should ideally be around the $RMSE_C$ values in the $QSPR_{SVR}$ modelling technique.

$$RMSE_{CV} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \quad (8)$$

here, Y_i and $\hat{Y}_i$ are the experimental and predicted values.

Additionally, this study compares the model's quality using mean absolute relative error (MARE,%) values. It is displayed as follows: [23]

$$MARE, \% = \frac{1}{n}\frac{\left|\log\beta_{12,\mathrm{exp}} - \log\beta_{12,\mathrm{pred}}\right|}{\log\beta_{12,\mathrm{exp}}}.100 \quad (9)$$

here, n is the number of samples; $\beta_{12,\mathrm{exp}}$ and $\beta_{12,\mathrm{pred}}$ are experimental and calculated stability constants, respectively.

3. RESULTS AND DISCUSSION

3.1. Development of $QSPR_{MLR}$ models

The multivariate regression models were constructed based on the initially collected dataset. The construction process was divided into a training subset of 79 values (about 80%). A CV subset of 20 values (approximately 20%) was extracted from the 99 sample values of experimental complexes that comprised the data set used to construct $QSPR_{MLR}$ models.

The $QSPR_{MLR}$ models were constructed using the backward elimination and forward regression techniques on the Regress system. The predictive ability of the $QSPR_{MLR}$ models was cross-validated using the leave-one-out (LOO) method with Q^2_{LOO} statistics. The models' quality is determined by statistical standards such as R^2_{train}, R^2_{adj}, Q^2_{LOO}, SE, and F_{stat}, which are used to identify a suitable model. Table 3 shows the statistical values of the entirely constructed $QSPR_{MLR}$ models. Therefore, the change in R^2_{train}, Q^2_{LOO}, and SE values determines which variables (k) will yield the best $QSPR_{MLR}$ model. A model is considered acceptable if its R^2_{train} and Q^2_{LOO} values are as close to 1.0 as possible and meet the statistical requirements (>0.6). In addition, the SE value should be as low as feasible. The results obtained seven $QSPR_{MLR}$ models that meet the statistical requirements presented in Table 3.

Table 3. Seven built $QSPR_{MLR}$ models (k = 5) with statistical values

Notation	$QSPR_{MLR}$ models
MLR1	$\log\beta_{12} = -27.70 + 63.47\times xch5 - 2.977\times S^6 + 0.206\times SHBa + 0.514\times vx1 + 0.362\times Dipole$; $R^2_{train} = 0.921$, $R^2_{adj} = 0.917$, $Q^2_{LOO} = 0.868$, $SE = 0.760$; $PRESS = 90.16$, $F_{stat} = 218.33$
MLR2	$\log\beta_{12} = -27.34 + 0.13\times LUMO + 0.48\times Dipole + 0.583\times xv1 + 61.61\times xch5 + 0.192\times SHBa$; $R^2_{train} = 0.906$, $R^2_{adj} = 0.901$, $Q^2_{LOO} = 0.877$, $SE = 0.832$, $PRESS = 83.86$, $F_{stat} = 178.88$
MLR3	$\log\beta_{12} = -22.32 + 0.365\times Dipole + 0.228\times xv2 + 78.88\times xch5 + 0.141\times SHBa + 10.60\times C^5$; $R^2_{train} = 0.922$, $R^2_{adj} = 0.918$, $Q^2_{LOO} = 0.861$, $SE = 0.759$, $PRESS = 95.08$, $F_{stat} = 218.98$
MLR4	$\log\beta_{12} = -13.76 + 0.38\times Dipole + 122.27\times xch5 + 45.68\times C^5 + 12.14\times N^4 - 0.831\times Saasc$; $R^2_{train} = 0.917$, $R^2_{adj} = 0.913$, $Q^2_{LOO} = 0.881$, $SE = 0.780$, $PRESS = 81.18$, $F_{stat} = 206.25$
MLR5	$\log\beta_{12} = -25.26 + 0.000053\times Core\text{-}core\ Repulsion + 0.422\times Dipole + 0.194\times SHBa + 56.74\times xch5 + 0.28\times xv2$; $R^2_{train} = 0910$, $R^2_{adj} = 0.905$, $Q^2_{LOO} = 0.827$, $SE = 0.814$, $PRESS = 118.63$, $F_{stat} = 187.83$
MLR6	$\log\beta_{12} = -24.79 + 0.95\times Ovality + 85.97\times xch5 + 10.83\times C^5 + 0.177\times SHBa + 0.192\times xv2$; $R^2_{train} = 0.902$, $R^2_{adj} = 0.897$, $Q^2_{LOO} = 0.840$, $SE = 0.847$, $PRESS = 109.57$, $F_{stat} = 172.13$
MLR7	$\log\beta_{12} = -21.46 + 0.538\times Dipole + 0.352\times xv2 + 65.63\times xch5 - 0.012\times \Delta H_f + 0.768\times Gmax$; $R^2_{train} = 0.917$, $R^2_{adj} = 0.913$, $Q^2_{LOO} = 0.839$, $SE = 0.781$, $PRESS = 109.79$, $F_{stat} = 205.70$

All seven models in Table 3 meet the statistical requirements; however, external validation with an independent dataset is necessary to select a complete model. The study uses a dataset of twenty-five experimental complexes and their stability constant values ($\log\beta_{12,exp}$) presented in Table 4.

Table 4. The 25 experimental $\log\beta_{12,exp}$ values of complexes in the EV dataset

1	Ligands				Metal ions	$\log\beta_{12.exp}$	ref.
	R1	R2	R3	R4			
1	H	H	- C6H5	-C2H4NO	Cu^{2+}	7.916	[24]
2	H	H	- C_6H_5	-C_2H_4NO	Ni^{2+}	9.711	[24]
3	H	-CH_3	-CH_3	-C_5H_4N	Ni^{2+}	11.919	[25]
4	H	-C_6H_5	-CH_3	-C_2H_4NO	Cu^{2+}	7.287	[26]
5	H	H	-CH_3	-C_5H_4N	Hg^{2+}	12.875	[27]
6	H	-C_6H_5	-CH_3	-$C_{10}H_{12}NO$	Co^{2+}	9.871	[28]
7	H	-C_6H_5	-CH_3	-$C_{10}H_{12}NO$	Ni^{2+}	10.140	[28]
8	H	-C_6H_5	-CH_3	-$C_{10}H_{12}NO$	Ag^{+}	9.736	[28]
9	H	-C_6H_5	-CH_3	-$C_{10}H_{12}NO$	Cd^{2+}	10.818	[28]
10	H	-C_6H_5	-CH_3	-$C_{10}H_{12}NO$	Hg^{2+}	11.256	[28]
11	H	H	-C_5H_4N	-C_5H_4N	Zn^{2+}	10.370	[17]
12	H	-C_6H_5	-C_5H_4N	-C_5H_4N	Mn^{2+}	7.410	[17]

1	Ligands				Metal ions	$\log\beta_{12.exp}$	ref.
	R1	R2	R3	R4			
13	H	H	$-CH_3$	$-C_5H_4N$	La^{3+}	13.580	[29]
14	H	H	$-CH_3$	$-C_5H_4N$	Pr^{3+}	13.830	[29]
15	H	H	$-CH_3$	$-C_{10}H_{12}NO$	Cu^{2+}	8.069	[30]
16	H	H	$-CH_3$	$-C_{10}H_{12}NO$	Hg^{2+}	10.675	[30]
17	H	H	$-CH_3$	$-C_{10}H_{12}NO$	Ag^{+}	9.854	[30]
18	H	H	$-CH_3$	$-C_{10}H_{12}NO$	Ni^{2+}	6.900	[30]
19	H	H	$-CH_3$	$-C_6H_4OH$	Ni^{2+}	8.630	[31]
20	H	H	$-CH_3$	$-C_6H_4OH$	Cu^{2+}	9.810	[31]
21	H	H	H	$-C_{10}H_6OH$	Cd^{2+}	9.920	[32]
22	H	H	-	$-C_9H_7NO$	Cu^{2+}	16.639	[33]
23	H	H	-	$-C_9H_7NO$	Ni^{2+}	15.876	[33]
24	H	H	-	$-C_9H_7NO$	Zn^{2+}	15.056	[33]
25	H	$-C_6H_5$	$-CH_3$	$-C_9H_{10}O_2Br$	Cu^{2+}	10.363	[34]

External validation (EV) methods are needed to create a predictive model. Although authors sometimes skip this phase, it is deemed crucial for creating a strong prediction model and needs to be carried out on a separate dataset. Accordingly, the study conducted the EV in conjunction with the search for the optimal ANN model using the external validation dataset, which consisted of 25 samples. Table 4 provides detailed data for 25 experimental observations.

Q^2_{EV} and MARE (%) are used to select the most suitable model. The value of Q^2_{EV} should be greater than 0.5 and as close to 1.0 as possible, and the smaller the value of MARE of the model, the closer the model can predict the actual value. According to the results in Figure 3a, the MLR3 model verified the correlation between the experimental and predicted values, with the highest Q^2_{EV} value.

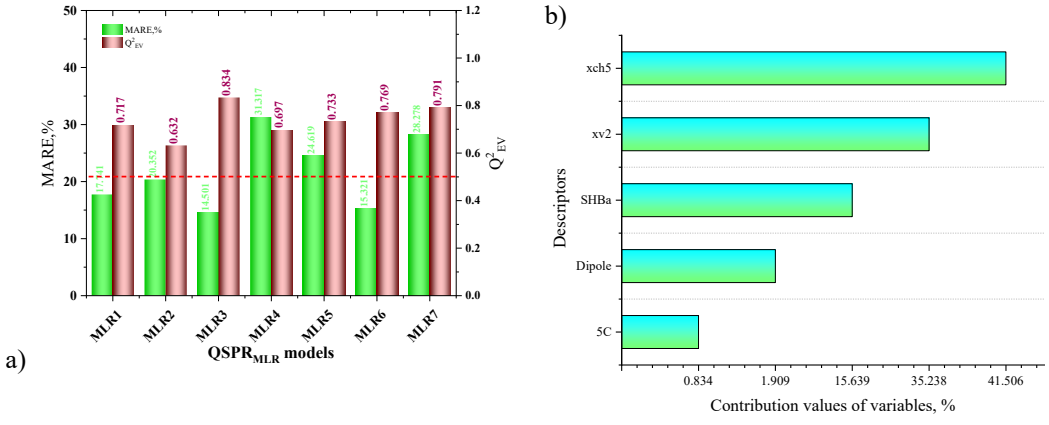

Figure 3. (a) Q^2_{EV} and MARE(%) values for selecting MLR (b) Contribution of variables in MLR3 model (k of 5)

Furthermore, the average percentage, MP_mx_k, is the contribution of each independent variable in the selected QSPR models (with k of 4 to 6) and is determined according to formula (10) [23]

$$MPx_k,\% = \frac{1}{n}\sum_{j=1}^{n}\left(100.\left|b_{m,i}x_{m,i}\right| \Big/ \sum_{i=1}^{k}\left|b_{m,k}x_{m,k}\right|\right) = \frac{1}{n}\sum_{j=1}^{n}\left(100.\left|b_{m,i}x_{m,i}\right| \Big/ C_{total}\right) \quad (10)$$

here n = 99 is the number of compounds; m is the number of compounds used to determine the P_mx_k value.

Table 6. Contribution of variables in MLR3 models with k of 4 to 6.

b_i	MLR3 model			MP_mx_k, %			GMP_mx_k, %
	k = 4	k = 5	k = 6	k = 4	k = 5	k = 6	
b_0	-23.85	-22.32	-23.47	-	-	-	-
b_1	0.389	0.365	0.408	2.100	2.033	1.596	2.100
b_2	0.326	0.228	0.457	38.850	28.247	38.618	38.850
b_3	59.34	78.88	76.69	37.762	52.072	34.685	37.762
b_4	0.196	0.141	0.126	21.288	15.816	9.815	21.288
b_5	-	10.60	5.393	-	1.832	0.670	0.834
b_6	-	-	0.825	-	-	14.616	4.872

As a result, the stability constant values ($\log\beta_{12,pred}$) are predicted based on the Q^2_{EV} coefficient and the MARE(%) values between the experimental and predicted values among the MLR models in Figure 3. As a result, the MLR3 model was chosen to develop the SVR, ANN models, and new ligands and complexes because it achieved the highest $Q^2_{EV\text{-}MLR3}$ value of 0.834 and the lowest MARE value (14.501%).

To prevent overfitting, the study performed variable selection for the selected MLR3 model by controlling the inclusion of variables in this model. The results are presented in Table 5. The changes in regression values and the increase in the number of variables k, as shown in Figure 4b, show that the choice of the 5-variable model of the MLR3 model is appropriate.

Table 5. The results of the built MLR3 models (k of 1 to 6).

k	Descriptors	SE	R^2_{train}	R^2_{adj}	Q^2_{LOO}	F_{stat}
1	x_1	2.238	0.290	0.283	0.094	39.60
2	x_1/x_2	1.909	0.489	0.478	0.108	45.87
3	$x_1/x_2/x_3$	1.591	0.648	0.637	0.251	58.40
4	$x_1/x_2/x_3/x_4$	0.904	0.888	0.883	0.722	185.91
5	$\boldsymbol{x_1/x_2/x_3/x_4/x_5}$	**0.759**	**0.922**	**0.918**	**0.861**	**218.98**
6	$x_1/x_2/x_3/x_4/x_5/x_6$	0.732	0.934	0.929	0.916	216.28

Notation of molecular descriptors			
Dipole	x_1	SHBa	x_4
xv2	x_2	^{5}C	x_5
xch5	x_3	knotpv	x_6

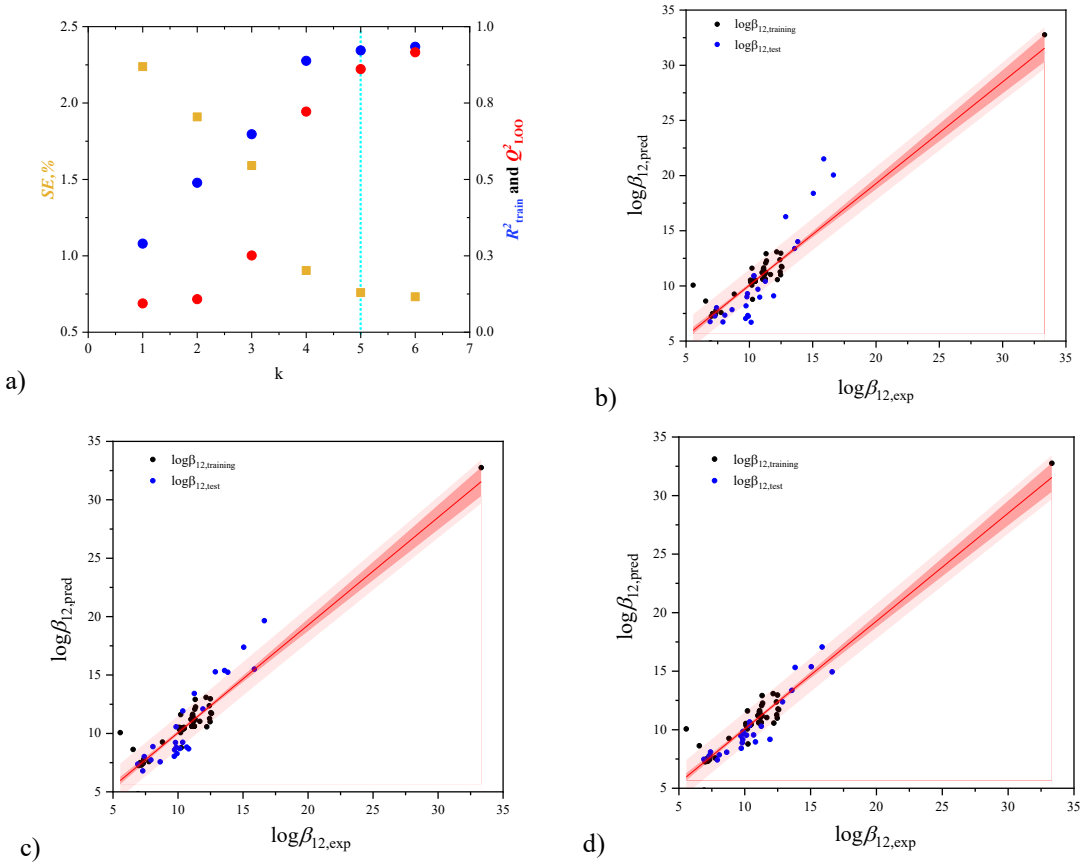

Figure 4. (a) Change of regression coefficient according to variable k; (b,c,d) Correlation of experimental ($\log\beta_{12,exp}$) versus predicted ($\log\beta_{12,pred}$) values of MLR3, SVR (a), and ANN (b) models on the training and test sets

The regression coefficients and the contributions of $MP_m x_k$ and $GMP_m x_k$ in the MLR3 models for k values ranging from 4 to 6 are presented in Table 6. The significant contributions of the variables are organized according to their $GMP_m x_k$ values in the following order: ^5C < *Dipole* < *SHBa* < *xv2* < *xch5*.

Figure 4 shows that the correlation between experimental stability constants and anticipated values derived from QSPR models proves their predictive capability, as evidenced by the high R^2_{train}, Q^2_{LOO}/Q^2_{CV}, and $SE/RMSE_{CV}$ statistical values. The results indicate that the calculated values are consistent with the experimental data, even though the complexes utilized in the validation set were not included in developing the QSPR models. Three QSPR models were developed, and all of them demonstrated strong predictability with low errors as evaluated by SE and MARE. This suggests that these models consistently apply to determining the stability constants $\log\beta_{12}$. The ANN model had the highest predictability among the models. The MLR3 model had the lowest predictability and the highest error rates. This disparity is also apparent when comparing the statistical performance of the various QSPR models, as shown in Figure 4. Thus, the correlation between the observed and predicted values shows that the performance, reliability, and validity of the QSPR prediction model are consistent with statistical standards.

3.2. Development of $QSPR_{SVR}$ models

As previously stated, the study uses machine-learning models to improve the predictive ability of MLR models. Therefore, these models are developed based on the descriptors selected from the MLR model results. In this case, SVR and ANN are two machine learning methods chosen in this work.

The $QSPR_{SVR}$ model is constructed using the best possible selection of kernel function types and parameters, such as gamma (γ) and capacity (C). This study chose the RBF radial function for vector machine regression. The predictive power of the SVR model was verified using the indices of the $RMSE_C$ (root mean squared error from calibration) and $RMSE_{CV}$ (root mean squared error from CV) values.

According to the results of the $QSPR_{MLR}$ model, this study can set up the 5-level parameters of the gamma (γ) in the range 0.01-0.05-0.1-0.5-1.0 and the capacity (C) in 1.0-5-10-20-50 using input factors such as *Dipole, xv2, xch5, SHBa, and C*[5]. The following is the display of the optimized parameters. Fundamentally, the $RMSE_{CV}$ value is the least, and the correlation coefficient (R^2) needs to be higher than 0.6 [5]. Hence, Table 7's results highlight the optimal values, such as the capacity (C) value of 10 and the gamma (γ) value of 0.1; in addition, the results include the epsilon (ε) of 0.1, the number support vector of 42, the R^2 value of 0.925, and the $RMSE_{CV}$ of 0.536.

Table 7. Optimal parameters of the SVR model with the R^2_{train} and $RMSE_{CV}$ values

$R^2/RMSE_{CV}$	C = 1.0	C = 5	C = 10	C = 20	C = 50
Gamma	1	2	3	4	5
0.01	0.252/3.152	0.487/1.825	0.636/1.024	0.562/1.512	0.436/2.021
0.05	0.461/1.831	0.672/0.987	0.736/0.786	0.626/1.125	0.556/1.668
0.1	0.535/1.756	0.762/0.689	**0.925/0.536**	0.806/0.608	0.642/1.002
0.5	0.621/1.622	0.659/0.982	0.769/0.723	0.765/0.774	0.545/1.625
1.0	0.539/1.888	0.526/1.699	0.669/0.976	0.698/0.865	0.469/1.873

3.3. Development of $QSPR_{ANN}$ models

Typically, a whole ANN model is constructed in two steps. In the initial phase, the models were trained using a dataset of 99 experimental values and surveyed using five descriptors of the aforementioned results (Table 1). A 70% training subset, a 15% test subset, and a 15% CV subset of approximation were used in the training procedure. The MLP-type ANN is I(5)-HL(*m*)-O(1), with the output layer O(1) being the $\log\beta_{12}$ values and the input layer I(5) being five neurons: *Dipole, xv2, xch5, SHBa, and C*[5]. The buried layer's neuron count (m) is rapidly determined. Table 8 displays the initial *m* values results.

An external dataset comprising 25 samples was used to assess the MLP-ANN models in the following stage (Table 5). In a similar approach to the choice of MLR models, the two indices used to identify the optimal model are Q^2_{EV} and MARE(%). Acceptable Q^2_{EV} values (>0.5) are then obtained via the best $QSPR_{ANN}$ model, with the MARE(%) value being the least value.

Table 8. The results of MLP-ANN models I(5)-HL(m)-O(1) in the first step

ANN	MLP-ANN models	R^2_{train}	Q^2_{test}	Q^2_{CV}	Train. error	Test Error	Valid. Error	Transfer Function
ANN1	I(5)-HL(3)-O(1)	0.978	0.713	0.903	0.201	0.103	0.153	Tangent
ANN2	I(5)-HL(10)-O(1)	0.989	0.958	0.956	0.100	0.017	0.163	Tangent
ANN3	I(5)-HL(4)-O(1)	0.988	0.972	0.923	0.111	0.011	0.149	Exponential
ANN4	I(5)-HL(5)-O(1)	0.980	0.761	0.894	0.180	0.088	0.151	Exponential
ANN5	**I(5)-HL(6)-O(1)**	**0.986**	**0.877**	**0.922**	**0.129**	**0.046**	**0.168**	**Exponential**

According to Figure 5a results, the ANN5 model is the ultimate choice, with a MARE value (%) of 7.665 and Q^2_{EV} statistical values of 0.881. Therefore, the ANN5 model with the network architecture I(5)-HL(6)-O(1) was selected to identify new ligands and complexes (Figure 5b).

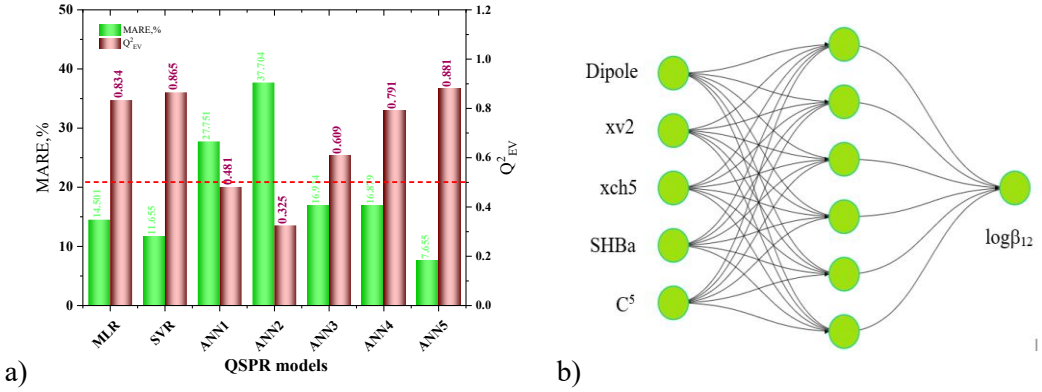

Figure 5. a) Q^2_{EV} and MARE(%) values for selecting ANN models; b) $QSPR_{ANN}$ model of network architecture MLP-ANN I(5)-HL(6)-O(1)

In addition, the values demonstrate the models' acceptable predictability (>0.5). Additionally, the MLR3, SVR, and ANN5 I(5)-HL(6)-O(1) models had MARE(%) values of 14.501, 11.655, and 7.655, respectively. The findings show that the SVR models outperform the MLR3 model regarding predictive power, while the ANN5 models have the highest predictive power. This indicates that the models developed using clever machine-learning methods are suitable for practical use. On the EV dataset of all three models, the difference between the estimated values ($\log\beta_{12,pred}$) and the initial values ($\log\beta_{12,exp}$) is also examined using the single-factor ANOVA approach. $F = 0.3688 < F\text{-}crit_{(0.05)} = 3.1787$; the results show that the difference between these results is not significant.

3.5. Design and prediction of new ligand and complexes

In order to create novel complexes, this effort looks for possible thiosemicarbazone compounds that correspond to the substance group of the building. By adding possible functional groups to the basic structure, these novel compounds are created (Figure 1a). To guide future research, we chose groups for the project that have been effectively synthesized in practice and have a high level of biological activity; these properties are the same as thiosemicarbazone and related complexes [35,36]. Five model descriptors were chosen to investigate new thiosemicarbazone-ligand complexes: *Dipole, xv2, xch5, SHBa, and* C^5.

Figure 6. The structure of 10H-phenothiazine (a) and 9H-carbazole (b)

Derivatives of 10H-phenothiazine and 9H-carbazole (Figure 6) are utilized to create novel thiosemicarbazone and complex them with common metal ions such as Zn^{2+}, Ni^{2+}, Ag^{+}, Cd^{2+}, and Cu^{2+}. Groups are attached to the ligand structure at the R_4 site to create new ligands; the hydrogen atoms are present at the original structure's R_1, R_2, and R_3 sites.

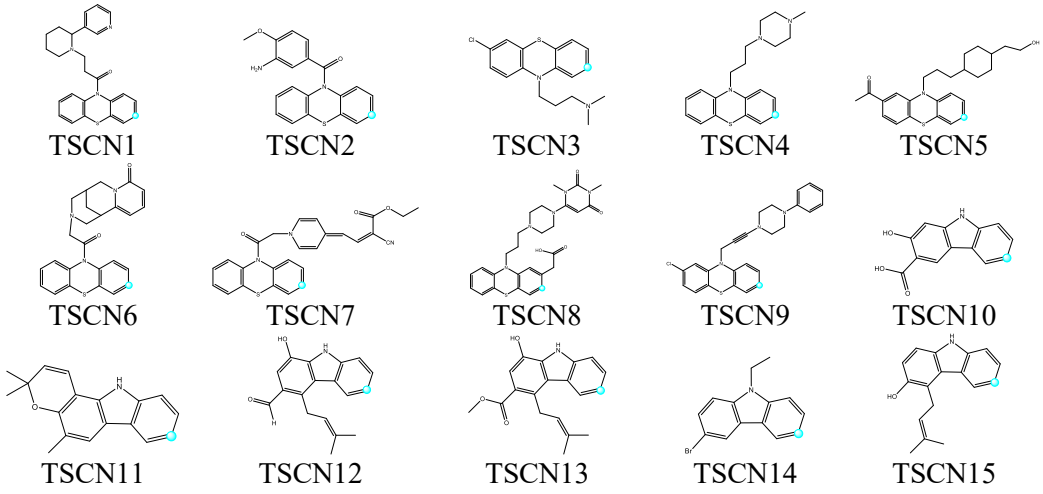

Figure 7. Structures at R_4 site of 15 new thiosemicarbazone derivatives

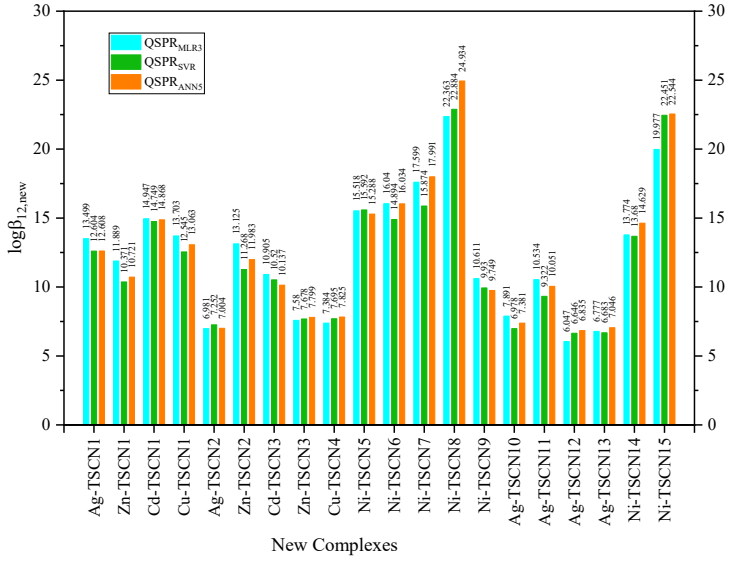

Figure 8. The $\log\beta_{12,new}$ calculated values of 20 new complexes from the models

Similar characteristics of newly optimized complexes were added to the training dataset to precisely construct and screen several novel thiosemicarbazone-ligand compounds. The

stability constants that have to surpass the AD and Outliers requirements were then predicted by new complexes using the absolute values D-Cook (|D-Cook| < 1.0) [6] and the results are presented in Figure 9. This led to the creation of 15 new thiosemicarbazone ligands and 20 new metal-ligand complexes, and three developed models were used to calculate the stability constants of the complexes ($\log\beta_{12,new}$). Figures 7 and 8 present the findings. The $QSPR_{MLR3}$, $QSPR_{SVR}$, and $QSPR_{ANN5}$ models are among the three whose predicted outcomes were compared using the one-way ANOVA technique. It indicates that the difference is not significant (F = 0.0489 < F-crit(0.05) = 3.1588).

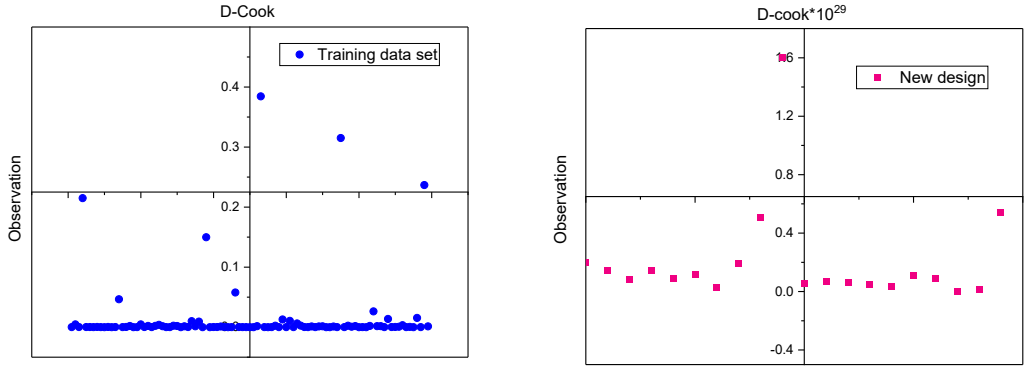

Figure 9. The D-Cook values of the training set and new design complexes

4. CONCLUSION

An initial success has been achieved in this study by building predictive models (QSPR) for the creation of novel thiosemicarbazone ligands and their complexes. Three models were constructed using support vector regression, multivariable linear regression, and artificial neural networks. They have been constructed using the external validation dataset of 25 samples of thiosemicarbazone-based complex values and the training dataset of 99 stability constant values. Furthermore, this work optimized complex structures using the semi-empirical QM/PM7-PM7/sparkle version approach. Using good statistical criteria (R^2_{train} = 0.922, Q^2_{LOO} = 0.861, SE = 0.759, and $Q^2_{EX\text{-}MLR}$ = 0.834), the chosen MLR model with five variables was constructed. The best $QSPR_{SVR}$ model findings included good regression parameters (R^2 = 0.925, $RMSE_{CV}$ = 0.536, and $Q^2_{EX\text{-}SVR}$ = 0.865), support vectors number equal to 42, and ideal model parameters such as capacity C = 10.0, gamma γ = 0.10, and epsilon ε = 0.1. Excellent results (R^2_{train} = 0.986, Q^2_{test} = 0.876, $Q^2_{validation}$ = 0.921, and $Q^2_{EX\text{-}ANN}$ = 0.881) were extracted by the MLP-ANN of the I(5)-HL(6)-O(1) model. Lastly, with anticipated stability constants ($\log\beta_{12,new}$), the derived models guide the design of 15 new thiosemicarbazone ligands and 20 new thiosemicarbazone-based complexes. These new compounds can be applied in a variety of disciplines. Furthermore, many ligands and their complexes may be found using the outcomes of QSPR-based models.

REFERENCES

1. Kumar, S., Dhar, D.N., & Saxena, P.N. - Applications of metal complexes of Schiff bases: A review, J. Sci. Ind. Res. **68** (3) (2009) 181-87.
2. Khan, T., Ahmad, R., Joshi S., & Khan, A.R. - Anticancer potential of metal thiosemicarbazone complexes: A review, Chem. Sin. **6** (12) (2015) 1-11.

3. Kunal R., Supratik K., & Rudra N. D. - A Primer on QSAR/QSPR Modeling, Fundamental Concepts, Springer, New York, USA (2015).
4. Muneeba, M., Sarfraz, A., Zahid, R. -Eccentric indices based QSPR evaluation of drugs for schizophrenia treatment, Heliyon **11** (2) (2025) e42222. https://doi.org/10.1016/j.heliyon.2025.e42222.
5. Stewart J. J. P. - Optimization of parameters for semi-empirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters, J. Mol. Model **19** (2013) 1-32.
6. Organization for Economic Co-operation and Development - Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships Models, France (2007).
7. Golbraikh A., & Tropsha A. - Beware of q2!, J. Mol. Graph Model. **20** (4) (2002) 269-276.
8. Patel, K.N., Parikh, K.S., & Patel, R.M. - 2-Hydroxy-4-n-butoxy-5-bromoacetophenone thiosemicarbazone as an extractive spectrophotometric reagent for nickel, Orbital. Elec. J. Chem. **2** (4) (2010) 341-346.
9. Patel, R.M., Parikh, K.S., & Patel, K.N. - 2-hydroxy-4-n-butoxy-5-bromopropiophenone thiosemicarbazone as an extraction spectrophotometric reagent for Nickel (II)., Int. J. ChemTech. Res. **2** (2) (2010) 1090-1093.
10. Parikh, K.S., Patel, R.M., & Patel, K.N. - New Spectrophotometric Method for Determination of Cadmium, E-J. Chem. **6** (S1) (2009) S496-S500
11. Milunovic, N.M., Enyedy, E.A., Nagy, N.V., Kiss, T., Trondl, R., Jakupec, M.A., Keppler, B.K., Novitchi, G., & Arion, V.B., - L and D Proline Thiosemicarbazone Conjugates: Coordination Behavior in Solution and the Effect of Copper(II) Coordination on Their Antiproliferative Activity, Inorg. Chem. **51** (2012) 9309-9321.
12. Patel, N.B., & Solanki, Y.J. - 2,4-Dihydroxy-5-Bromo [2'Methyl] Propiophenone Thiosemicarbazone [DHBMPT] as an Analytical Reagent: Studies on Co(II) Chelate, J. App. Chem. **5** (3) (2016) 654-660.
13. Atalay, T., & Akgemci, E.G. - Thermodynamic Studies of Some Complexes of 2-benzoylpyridine 4-phenyl-3-thiosemicarbazone., Tr. J. Chem. **22** (1998) 123-127.
14. Parikh, K., Patel, S.R.M., & Patel, K.N. - 2-Hydroxy-4-n-butoxy-5-bromo propiophenone Thiosemicarbazone as Spectrophotometric Reagent for Iron, Asian J. Chem. **22** (4) (2010) 2805-2810.
15. Gaál A., Orgován, G., Polgári, Z., Mihucz, V.G., Bősze, S., Szoboszlai, N., & Streli, C. - Complex forming competition and in-vitro toxicity studies on the applicability of di-2-pyridylketone-4,4,-dimethyl-3-thiosemicarbazone (Dp44mT) as a metal chelator, J. Inorg. Biochem. **130** (2014) 52-58.
16. Bernhardt, P.V., Sharpe, P.C., Islam, M., Lovejoy, D.B., Kalinowski, D.S., & Richardson, D.R. - Iron Chelators of the Dipyridylketone Thiosemicarbazone Class: Precomplexation and Transmetalation Effects on Anticancer Activity, J. Med. Chem. **52** (2) (2009) 407-415.
17. Toribio, F., Fernandez, J.M.L., Bendito, D.P., & Valcárcel, M. - 2,2'-dihydroxybenzophenone thiosemicarbazone as a spectrophotometric Reagent for the determination of copper, nickel, and iron trace amounts in mixtures without previous separations, Microchemical Journal **25** (1980) 338-347.

18. Nair, A.P., & Jeyaseelan, C. - Spectrophotometric determination of Cu(II) and Ni(II) using 4 phenyl-3-thiosemicarbazone of 2-hydroxy--5-bromoacetophenone as analytical reagent, Amer. Int. J. Res. in Form. Appl. Natur. Sci. **3** (1) (2013) 46-50.
19. Vapnik, V., & Lerner, A.Y. - Recognition of patterns with help of generalized portraits, Avtomat. i Telemekh **24** (6) (1963) 774-780.
20. Cortes, C., & Vapnik, V. - Support-vectornetworks, Machinelearning **20** (3) (1995) 273-297.
21. Gasteiger, J., & Zupan, J. - Neural Networks in Chemistry, Chiw. Inr. Ed. EngI. **32** (1993) 503-521.
22. Rumelhart, D.E., Hinton, G.E., & Williams, R.J. - Learning representations by back-propagating errors, Nature **323** (1986) 533-536.
23. Van Tat Pham, Development of QSAR and QSPR, Publisher of Natural sciences and Technique, Ha Noi (2009).
24. Reddy, K.H., Prasad, B.L., & Reddy, T.S. - Analytical properties of 1-phenyl-1,2-propanedione-2-oxime thiosemicarbazone: spectrophotometric determination of copper(II) and nickel(II) in edible oils and seeds, Talanta **59** (2003) 425-433.
25. Reddy, D.N., Reddy, K.V. - Tegegne B. M., & Reddy V. K.., Development of a Highly Sensitive Extractive Spectrophotometric Method the Determination of Nickel(II) from Environmental Matrices Using 2-Acetylpyridine-4-methyl-3-thiosemicarbazone, American J. Anal. Chem. **3** (2012) 719-726.
26. Gomaa, E.A., Ibrahim K. M ., & Hassan N. M. - Evaluation of thermodynamic parameters for the interaction of Cu(II) ion with 4-phenyl -1-diacetyl monoxime -3-thiosemicarbazone (BMPTS) in (60%V) ethanol (EtOH-H_2O) at different temperatures, Int J. Eng. Sci. **3** (1) (2014) 44-51.
27. Babu, S.V., & Reddy, K.H. - Direct Spectrophotometric Determination of Mercury (II) using 2-Acetylpyridine Thiosemicarbazone in Environmental Samples, Indian J. Adv. Chem. Sci. **1** (2012) 65-72.
28. Atalay T., & Ozkan E. - Thermodynamic stabilities. Thermodynamic parameters of some complexes of 4'-morpholinoacetophenone 4-phenyl-3-thiosemicarbazone, Thermochimica Acta. **246** (1994) 193-197.
29. Garg, B.S., Singh, S.R., Basnet, S.B. & Singh,R. P. - Potentiometric studies on the complexation equilibria between La(III), Pr(III), Sm(III), Tb(III), Dy(III), Ho(III) and 2-acetylpyridinethiosemicarbazone (2-APT)., Polyhedron. **7** (2) (1988) 147-150.
30. Atalay T., & Ozkan, E. - Thermodynamic studies of some complexes of 4'-acetophenone thiosemicarbazone, Thermochimica Acta. **237** (1994) 369-374.
31. Garg B.S., Ghosh S., Jain V. K., & Singh P. K. - Evaluation of thermodynamic parameters of bivalent metal complexes of 2-hydroxyacetophenonethiosemicarbazone (2-HATS), Thermochimica Acta. **157** (1990) 365-368.
32. Sahadev, Sharma, R. K., & Sindhwani, S. K. - Thermal studies on the chelation behaviour of biologically active 2-hydroxy-1-naphthaldehyde thiosemicarbazone (HNATS) towards bivalent metal ions: a potentiometric study, Thermochimica Acta **202** (1992) 291-299.
33. Sarkar, K., Garg & B.S. - Determination of thermodynamic parameters and stability constants of the complexes of p-MITSC with transition metal ions, Thermochimica Acta **113** (1987) 7-14.

34. Nair, A.P., & Jeyaseelan, C. - Spectrophotometric determination of Cu(II) and Ni(II) using 4 phenyl-3-thiosemicarbazone of 2-hydroxy-4-n-propoxy-5-bromoacetophenone as analytical reagent., Amer. Int. J. Res. in Form. Appl. Natur. Sci. **3** (1) (2013) 46-50.

35. Benjamin, A. B., Monika, M., Lekhnath, S., Olanike, O., & Oreoluwa, F. - Exploring the therapeutic potential of phenothiazine derivatives in medicinal chemistry, Results in Chemistry **8** (2024) 101565. https://doi.org/10.1016/j.rechem.2024.101565

36. Priyanka, S.W., Anuruddha, R.C., Kunal, G.R., & Pooja, T.G. - A Review on Carbazole and Its Derivatives as Anticancer Agents From 2013 to 2024, Chirality **37 (**2) (2025), e70021. https://doi.org/10.1002/chir.70021.

TÓM TẮT

ỨNG DỤNG MÔ HÌNH QSPR ĐỂ PHÁT TRIỂN CÁC PHỨC CHẤT MỚI GIỮA ION KIM LOẠI VÀ DẪN XUẤT THIOSEMICARBAZONE

Nguyễn Minh Quang[1,]*, Huỳnh Ngọc Châu[1], Phạm Văn Tất[2]

[1]Trường Đại học Công nghiệp Thành phố Hồ Chí Minh, Việt Nam

[2]Viện Đào tạo và Nghiên cứu Dược học, trường Đại học Bình Dương, Thành Phố Hồ Chí Minh, Việt Nam

*Email: *nguyenminhquang@iuh.edu.vn*

Hai mươi phức chất giữa ion kim loại và phối tử thiosemicarbazone mới (ML_2) đã được tính toán hằng số bền ($\log\beta_{12}$) dựa trên các mô hình mối quan hệ định lượng cấu trúc-tính chất (QSPR). Các mô hình QSPR được phát triển bằng các phương pháp hồi quy tuyến tính đa biến (MLR), hồi quy hỗ trợ vectơ (SVR) và mạng thần kinh nhân tạo (ANN). Các mô tả của mô hình được tính toán bằng phương pháp cơ học lượng tử (QM) bán thực nghiệm PM7 và PM7/sparkle. Chất lượng của các mô hình QSPR được kiểm soát chặt chẽ bởi các giá trị thống kê theo hướng dẫn của tổ chức OECD và các tiêu chuẩn của Tropsha. Kết quả mô hình $QSPR_{MLR}$ tốt nhất gồm năm biến: *Dipole, xv2, xch5, SHBa và* ^5C, với các giá trị thống kê như R^2_{train} = 0,922; Q^2_{LOO} = 0,861 và RMSE = 0,759. Bên cạnh đó, mô hình $QSPR_{SVR}$ tốt nhất bao gồm C = 10, gamma γ = 0,10 và epsilon ε = 0,1 với số vectơ hỗ trợ bằng 42 và các tham số hồi quy tốt: R^2 = 0,925 và $RMSE_{CV}$ = 0,536. Mô hình $QSPR_{ANN}$ với kiến trúc mạng I(5)-HL(6)-O(1) và hàm truyền *exponential* đã được tìm thấy qua quá trình luyện mạng từ các mô tả của mô hình $QSPR_{MLR}$ và cho kết quả khá ấn tượng: R^2_{train} = 0,986; Q^2_{test} = 0,876 và $Q^2_{validation}$ = 0,921. Ngoài ra, nghiên cứu này đã sử dụng tập dữ liệu đánh giá ngoại (EV) gồm 25 giá trị thực nghiệm $\log\beta_{12}$ để xây dựng các mô hình QSPR hoàn chỉnh với các giá trị $Q^2_{EV\text{-}MLR}$, $Q2EV\text{-}SVR$ và $Q^2_{EV\text{-}ANN}$ lần lượt là 0,834; 0,865 và 0,881. Kết quả nhận được từ các mô hình có thể được sử dụng để phát triển các dẫn xuất thiosemicacbazone và các phức chất mới nhằm tăng khả năng ứng dụng trong các lĩnh vực hóa học, phân tích và môi trường.

Từ khóa: ANN, hằng số bền $\log\beta_{12}$, MLR, mô hình QSPR, phức chất giữa ion kim loại và thiosemicarbazone, SVR.