Journal of Computational Science &Engineering 60 (2023) 11-24

Available at http://www.asocse.org ISSN 1710-4068

Inhibitory prediction of thiazole derivatives based on machine learning for MCF-7

Tianci Huang[1], Jiajian Huang[1], Yili Wang[1], Peijian Zhang[1*]

[1]School of Data Science and Software Engineering, Qingdao University, 266071, China

[2]Institute for Computational Science Engineering, Qingdao University, Qingdao, 266071, China

Abstract: **The quantitative structure activity relationship (QSAR) study was carried out to predict the anti-breast cancer effect of 2-Aminothiazole derivatives. heuristic method (HM), random forest (RF), support vector machine (SVM), and genetic expression programing (GEP), four QSAR models were established based on four descriptors. And the model built by SVM showed the best prediction ability and strongest model robustness: In the training set, R^2 = 0.959, RMSE = 0.010. In the test set, R^2=0.941 RMSE=0.013. This experimental study showed the role of the four key descriptors of 2-Aminothiazole derivatives in the treatment of breast cancer, which can provide a new basis for the prediction and screening of new drugs.**

Keywords: Inhibitor; QSAR; Thiazole Derivatives; Heuristic method; Random Forest; Support Vector Machine; Gene expression programming

Received 3 December 2022, Revised 24 March 2023, Accepted 10 April 2023

*Corresponding Author: Peijian Zhang, peijianzh@126.com

1. Introduction

According to data released by WHO, nearly 10 million people died of cancer in 2020, accounting for about one sixth of the total human deaths, and breast cancer is one of the most common cancers[1]. Most patients need comprehensive treatment, such as chemotherapy, surgical resection, and radiation protocols and so on. Among them, drug treatment is the key link in cancer treatment[1].

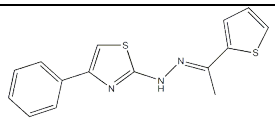
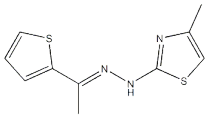
In the past, 2-Aminothiazole derivatives have played an important role in the inhibition of cancer cells[2]. Thiazole derivatives have excellent biological activity and high anti proliferative activity against different human cancer cells[3]. As a special pharmacophore, 2-aminothiazole is widely used in the development of anticancer or antibacterial drugs due to its unique properties[4].

Using the QSAR and machine learning method to carry out regression prediction analysis on the selected molecular descriptors can not only obtain the anticancer value of 2-Aminothiazole drugs more accurately, but also avoid repeated experiments on a large number of drugs, thus reducing human and material expenditure[5-7]. In this experiment, HM, RF, SVM, and GEP were used for regression analysis. The SVM algorithm obtained better prediction results, and the RF algorithm also analyzed the importance of the selection of molecular descriptors, further verifying the rationality of this experiment[8] [9] [10]. Generally speaking, this experimental conclusion can be used in the analysis of thiazole related compounds against breast cancer.

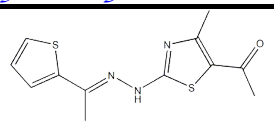
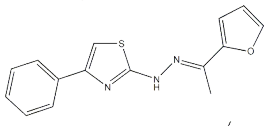
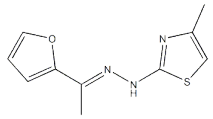
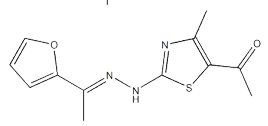
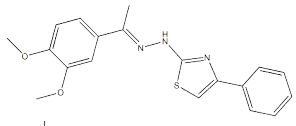
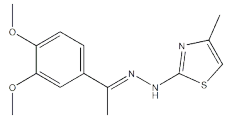
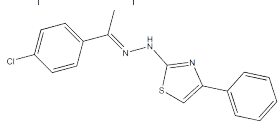
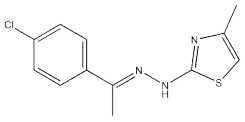
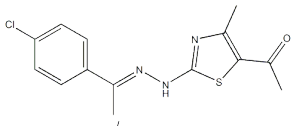
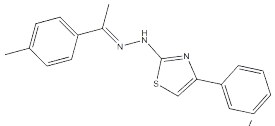
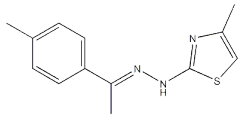
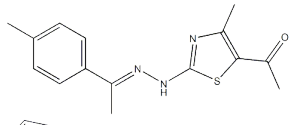
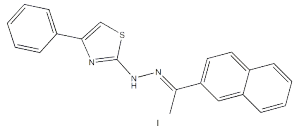
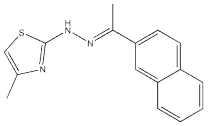
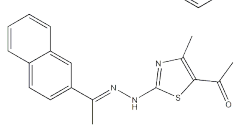
2. Material and method

2.1. Data set

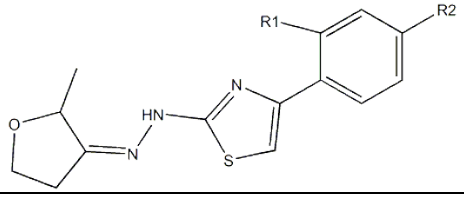
A total of 65 compounds were selected from literatures ([11-16]), and their MCF-7 IC_{50} values were recorded in this paper. All compounds were divided into training set and test set according to the ratio of 4:1, including 52 compounds as training set for building models, and the other 13 compounds as test set for validating models. All compounds and IC_{50} values can be found in Table 1.

Compound	Structure	Measured $IC_{50}(\mu m)$	Measured -lg(IC_{50})	Predicted -lg(IC_{50})			
				HM	RF	SVM	GEP
1*		18.48	-1.27	-1.35	-1.29	-1.24	-1.41
		53.76					
2*			-1.73	-1.24	-1.27	-1.53	-1.50

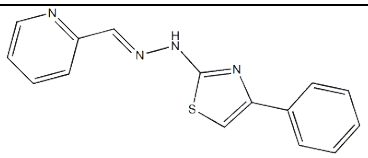
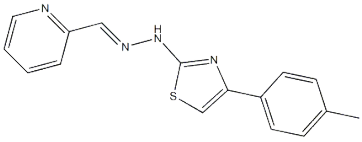
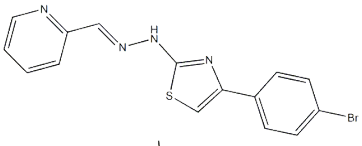
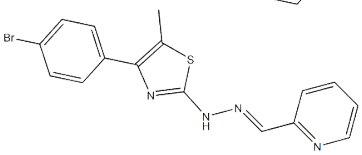
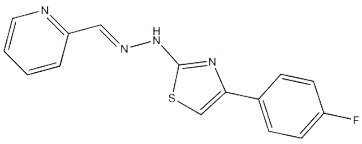
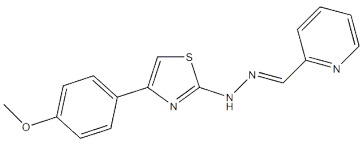
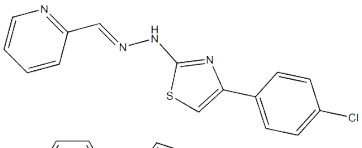
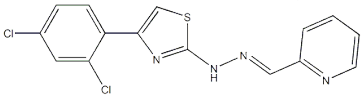
2023 The American Computational Science Society. All rights reserved.

3*		19.18	-1.28	-1.14	-1.05	-1.30	-1.25
4*		67.37	-1.83	-1.51	-1.34	-1.96	-1.40
5		17.09	-1.23	-1.47	-1.40	-1.13	-1.14
6		8.97	-0.95	-1.33	-0.98	-1.05	-1.24
7		82.32	-1.92	-1.47	-1.61	-1.82	-1.37
8		45.66	-1.66	-1.63	-1.55	-1.76	-1.41
9		16.28	-1.21	-1.46	-1.31	-1.31	-1.43
10		37.31	-1.57	-1.44	-1.49	-1.47	-1.39
11*		41.29	-1.62	-1.55	-1.35	-1.66	-1.55
12		15.17	-1.18	-1.50	-1.39	-1.11	-1.45
13		9.23	-0.97	-1.53	-1.18	-1.07	-1.20
14		32.45	-1.51	-1.43	-1.28	-1.61	-1.33
15		27.25	-1.44	-1.67	-1.36	-1.34	-1.48
16		23.94	-1.38	-1.39	-1.31	-1.28	-1.31
17		59.81	-1.78	-1.37	-1.49	-1.88	-1.37

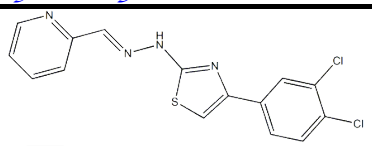
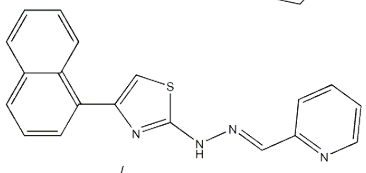
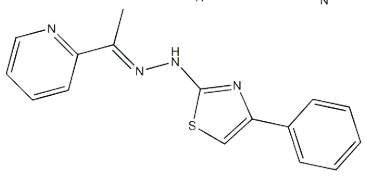
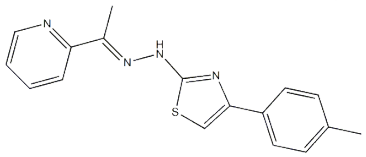
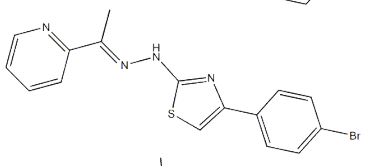
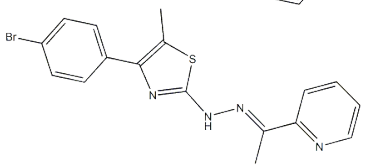
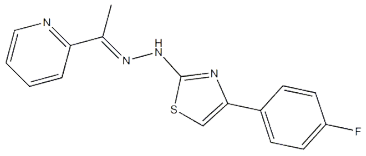
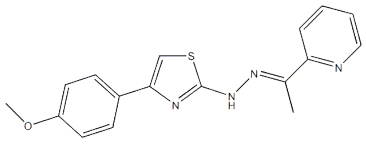
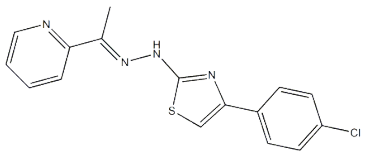
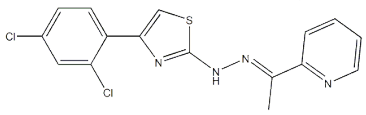
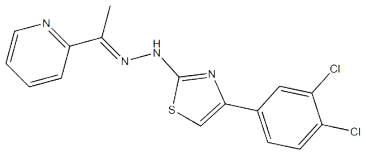
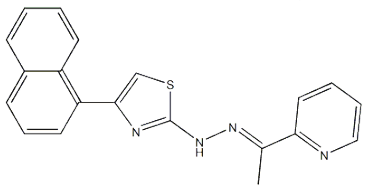
2023 The American Computational Science Society. All rights reserved.



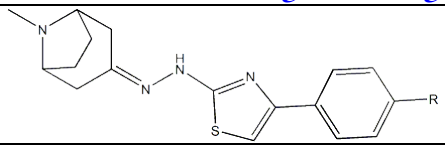
Compound	R1	R2	Measured IC_{50}(μm)	Measured -lg(IC_{50})	Predicted -lg(IC_{50})			
					HM	RF	SVM	GEP
18*	-H	-H	15.96	-1.20	-1.40	-1.32	-1.35	-1.23
19*	-H	$-CH_3$	36.27	-1.56	-1.40	-1.25	-1.45	-1.34
20	-H	$-OCH_3$	26.76	-1.43	-1.40	-1.38	-1.33	-1.31
21	-H	-CN	29.26	-1.47	-1.32	-1.48	-1.57	-1.41
22	-H	-F	37.62	-1.58	-1.35	-1.39	-1.68	-1.35
23	-H	-Cl	12.34	-1.09	-1.31	-1.14	-1.30	-1.34
24	-H	-Br	27.49	-1.44	-1.25	-1.36	-1.52	-1.36
25	$-CH_3$	$-CH_3$	30.40	-1.48	-1.42	-1.34	-1.57	-1.31
26	$-OCH_3$	$-OCH_3$	39.24	-1.59	-1.45	-1.59	-1.49	-1.42
27*	-F	-F	69.76	-1.84	-1.34	-1.35	-1.81	-1.41
28	-Cl	-Cl	8.36	-0.92	-1.26	-1.09	-1.02	-1.24

Compound	Structure	Measured IC_{50}(μm)	Measured -lg(IC_{50})	Predicted -lg(IC_{50})			
				HM	RF	SVM	GEP
29		11.00	-1.04	-0.76	-0.85	-0.94	-0.81
30		3.80	-0.58	-0.55	-0.53	-0.48	-0.56
31		4.40	-0.64	-0.61	-0.65	-0.74	-0.63
32		2.50	-0.40	-0.41	-0.38	-0.50	-0.38
33		4.60	-0.66	-0.72	-0.70	-0.76	-0.67
34		5.70	-0.76	-0.76	-0.75	-0.69	-0.81
35*		4.70	-0.67	-0.68	-0.59	-0.70	-0.62
36*		2.50	-0.40	-0.59	-0.53	-0.63	-0.60

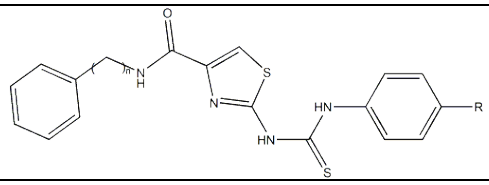
2023 The American Computational Science Society. All rights reserved.

37		2.90	-0.46	-0.55	-0.51	-0.56	-0.58
38		7.20	-0.86	-0.66	-0.66	-0.76	-0.59
39		1.01	0.004	-0.37	-0.07	-0.10	-0.39
40		2.10	-0.32	-0.37	-0.31	-0.27	-0.36
41		1.10	-0.04	-0.24	-0.20	-0.14	-0.31
42		1.61	-0.21	-0.21	-0.27	-0.31	-0.37
43		1.50	-0.18	-0.33	-0.29	-0.08	-0.37
44		1.30	-0.11	-0.37	-0.21	-0.31	-0.35
45		3.50	-0.54	-0.29	-0.38	-0.44	-0.36
46		1.50	-0.18	-0.24	-0.29	-0.28	-0.29
47		3.90	-0.59	-0.17	-0.48	-0.49	-0.34
48		2.70	-0.43	-0.25	-0.42	-0.35	-0.24

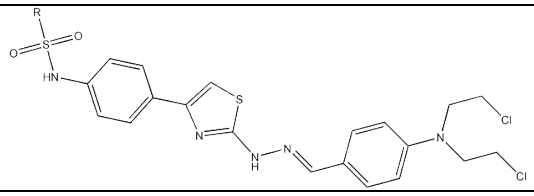
2023 The American Computational Science Society. All rights reserved.



Compound	R	Measured $IC_{50}(\mu m)$	Measured -lg(IC_{50})	Predicted -lg(IC_{50})			
				HM	RF	SVM	GEP
49	-F	19.16	-1.28	-1.14	-1.23	-1.18	-1.13
50	-Cl	9.08	-0.96	-1.10	-1.06	-0.95	-1.21
51	-Br	7.67	-0.88	-1.04	-0.88	-0.88	-0.97
52	$-CH_3$	11.06	-1.04	-1.17	-1.10	-0.99	-1.13
53	$-CF_3$	6.02	-0.78	-1.06	-0.82	-0.88	-1.02
54	-3,4-diCl	7.19	-0.86	-0.98	-0.84	-0.96	-0.89
55	-2,4-diCl	7.05	-0.85	-1.04	-0.90	-0.90	-0.93



Compound	R	n	Measured $IC_{50}(\mu m)$	Measured -lg(IC_{50})	Predicted -lg(IC_{50})			
					HM	RF	SVM	GEP
56	$-OCH_3$	1	24.61	-1.39	-1.72	-1.40	-1.54	-1.55
57*	$-OCH_3$	2	10.56	-1.02	-1.18	-1.09	-1.08	-1.22
58*	$-OCH_3$	3	63.82	-1.80	-1.55	-1.68	-1.70	-1.41
59	-Cl	1	56.52	-1.75	-1.76	-1.61	-1.65	-1.51
60	-Cl	2	17.23	-1.24	-1.21	-1.23	-1.14	-1.22
61	-Cl	3	45.06	-1.65	-1.43	-1.61	-1.75	-1.47



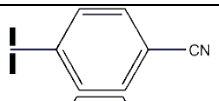
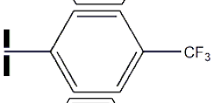
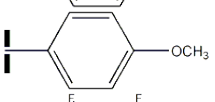
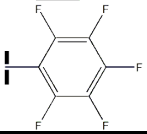
Compound	R	Measured $IC_{50}(\mu m)$	Measured -lg(IC_{50})	Predicted -lg(IC_{50})			
				HM	RF	SVM	GEP
62		8.30	-0.92	-0.66	-0.92	-0.82	-0.67
63		3.03	-0.48	-0.61	-0.86	-0.58	-0.60
64*		4.16	-0.62	-0.66	-0.68	-0.70	-0.71
65		3.42	-0.53	-0.46	-0.58	-0.63	-0.67

Table 1. Measured and Predicted -lg(IC_{50}) of 2-Aminothiazole derivatives

2.2. Calculation of the Descriptors

Before QSAR analysis, the molecular descriptors must be calculated first. Only by accurately calculating and expressing the molecular descriptors can the rationality and accuracy of the final model be guaranteed[17]. Now molecular descriptors were mainly calculated by some professional software including CODESSA, SYBYL, and Dragon. In this experiment, CODESSA software was used to calculate all possible descriptors. At the same time, HM can select several key descriptors, and RF algorithm was used to analyze the importance of these descriptors[9, 10]. Finally, four descriptors (1. Min e–n attraction for a H–N bond, 2. Kier&Hall index (order 3) 3. Min resonance energy for a C-C bond

2023 The American Computational Science Society. All rights reserved.

4. Avg electroph. react. index for a C atom) were selected to express the relationship between different compounds and IC_{50} values. The Steps of Calculation Descriptors are shown in Figure 1.

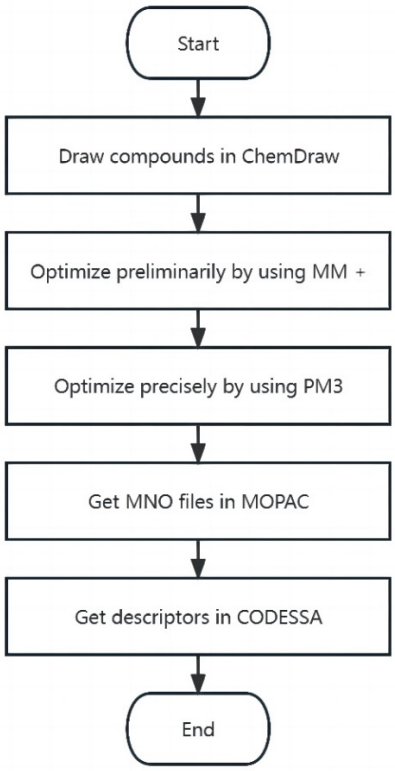


Figure 1. The Steps of Calculation Descriptors

To obtain molecular descriptors, it is necessary to draw and express each compound reasonably. From the experience of previous experiments, we know that compounds with lower molecular energy have more stable chemical structure, so many ways can be used to make all compounds have lower energy as possible. First of all, ChemDraw software was used to draw all compounds. In order to reduce its energy, the MM+ method in HyperChem software[18] can be pre-optimized, and its semi-empirical PM3 method can further optimize the compounds to obtain a set of lower and more stable compounds. After that, in order to obtain the molecular descriptors expressing the characteristics of compounds through CODESSA software, the above drawn compound files were imported into MOPAC to obtain the corresponding MNO files, which were used as the processing objects of CODESSA software[19, 20].

The files obtained above were calculated in CODESSA, combined with subsequent algorithms, four best descriptors were selected in the experiment. All descriptors are shown in Table 2. Obtaining more accurate molecular descriptors through the above methods is the basic condition for establishing QSAR models and conducting subsequent research.

Symbol	Physical-chemical meaning	X	DX	t-test
MEAHN	Min e–n attraction for a H–N bond	-5.36	0.41	-13.18
KHI	Kier&Hall index (order 3)	0.18	0.04	5.14
MRECC	Min resonance energy for a C-C bond	-0.23	0.07	-3.28
AERC	Avg electroph. react. index for a C atom	115.28	42.08	2.74

Table 2. The Descriptors Selected by HM and Their Physical-Chemical Meaning

2.3. Linear Model by HM [21] [22]

Using HM in CODESSA can quickly select the applicable descriptors, one of the reasons is that it can eliminate some unnecessary descriptors speedily, such as those that do not exist in the 65 compounds above, and collinear descriptors (correlation coefficient greater than 0.8)[23] [24]. At the same time, it can be used for weighted regression analysis, and a linear regression model is established based on it, which greatly improved the selection efficiency and accuracy of molecular descriptors[25, 26]. In this experiment, R^2 and mean square error (MSE) were used to test and evaluate the models established by HM, RF, SVM, and GEP. In addition, the robustness of the experimental model can be verified by cross validation.

2023 The American Computational Science Society. All rights reserved.

2.4. Nonlinear Model by RF

Because different compounds have their own characteristics, the above linear model may not be accurate in describing IC_{50}, so many nonlinear models were used in this study to predict the IC_{50} of the above compounds. RF can not only calculate R^2 and MSE of training set and test set, but also combine the GINI coefficients of different molecular descriptors to judge their importance[10, 27].

In general, several important descriptors derived from CODESSA can be used as RF feature selection. Each tree in the random forest can provide its most important descriptors, and then the descriptors of all trees will be averaged to generate the most relevant descriptors[28]. It is easy to conclude that the higher the GINI coefficient, the more important this descriptor is. So, the GINI coefficients of these descriptors were used as the basis for sorting and selection. At the same time, RF regression were also crucial. After determining the number of trees, the maximum depth of trees, and the number of related features, RF regression modeling can be carried out[29].

In this experiment, 618 descriptors were derived through CODESSA, and the RF model was built using this data. The GINI coefficients of each descriptor can be known, and the GINI coefficients were arranged in descending order. This sequence shown the importance of different descriptors. Finally, four descriptors can make the best performance of the experimental results, and can also prove the rationality and accuracy of the selection of descriptors by HM linear model.

2.5. Nonlinear Model by SVM

SVM is a common algorithm in machine learning, which was proposed by Vapnik in 1964[30]. If this algorithm is used for regression calculation, it is also called support vector regression (SVR). SVM algorithm has strong robustness to outliers, good generalization ability and high prediction accuracy, therefor it has shown advantages in solving small sample, nonlinear and high-dimensional pattern recognition[31, 32]. Some parameters are also introduced in SVR, such as ε-insensitive loss function, punish coefficient C, and the coefficient gamma of the kernel function. The value of C is a number greater than 0, and the greater the value of C, the higher the attention to the error. In other words, the greater the value of C, the higher the penalty to the model, and the weaker the generalization ability (over fitting).

Non-linear separable low-dimensional sample points are linearly separable in higher-dimensional space by formula(1)-(2), so we can use Gaussian kernel function to map the input space into the infinite-dimensional feature space, and apply linear SVM method in the infinite-dimensional space to obtain nonlinear SVM[33]. At the same time, in order to obtain a better separation method of hypersurface and decision function, it is better to try to compare different kernel functions for experiments. The Gaussian kernel function formula is shown in the following illustration. Attempt to minimize formula (3) when the constraints are formula (4)-(6)

$$\mathrm{y} = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \cdot \langle \varphi(x_i), \varphi(x) \rangle + b \tag{1}$$

$$\mathrm{y} = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b \tag{2}$$

$$\frac{1}{2}\|\mathrm{w}\|^2 + C\sum_{i=1}^{N} (\xi_i + \xi_i^*) \tag{3}$$

$$\mathrm{y}_i - wx_i - b \le \varepsilon + \xi_i \tag{4}$$

$$wx_i + b - y_i \le \varepsilon + \xi_i^* \tag{5}$$

$$\xi_i, \xi_i^* \ge 0 \tag{6}$$

When searching for the optimal hyperplane, the error of the predicted value of each training data is at most equal to ε. In order to allow outliers, data with prediction error greater than ε will be penalized by ε-insensitive loss function. The larger the value of ε, the less the number of support vectors. It can also be understood that the smaller the value of ε, the more over-fitting the model will be.

2.6. Nonlinear Model by GEP [34]

The establishment of a nonlinear model GEP can also be used to predict the IC_{50} values of compounds and to demonstrate the anticancer activity of such compounds. GEP algorithm was proposed by Ferreira according to the principles of genetic algorithm (GA) and genetic planning (GP), and had the characteristics of simple coding and easy operation[34, 35].

In the GEP algorithm, it is necessary to initialize the population first, that is, randomly generate a certain number of individual sets, and then calculate the fitness values of each individual. If the conditions are met, it exits, and if the conditions are not met, it starts to generate a new population. During the process of generating a new population, the optimal individual is saved, and then the probability of individual selection is determined based on the size of the fitness

2023 The American Computational Science Society. All rights reserved.

value. The larger the value, the easier it is to be selected. All selected individuals are either directly copied to the next generation, or undergo genetic operations with a certain probability, such as mutation, recombination, etc. After the above steps, a new species group is generated and checked to see if it meets the conditions. Repeat the process until the expected results are met[34, 36]. The Steps of GEP Algorithm are shown in Figure 2.

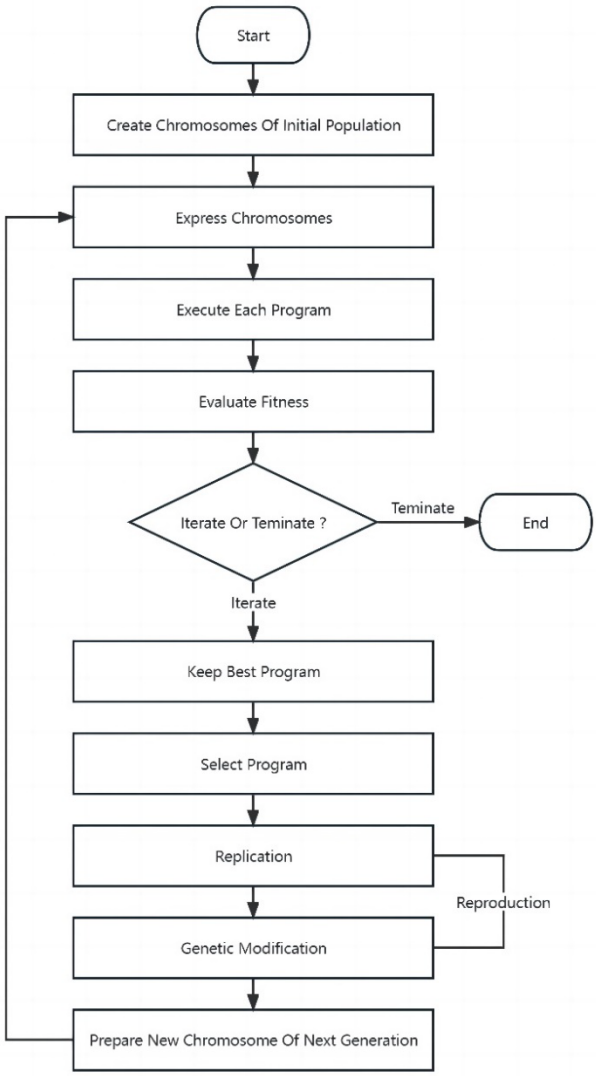


Figure 2. The Steps of GEP Algorithm

This experiment ran GEP through the software Automatic Problem Solver (APS)[37]. The software can combine the compound descriptors selected by the above HM and RF to establish a nonlinear model for predicting the antitumor activity of 2-Aminothiazole derivatives[37]. Appropriate functions can be selected for prediction and analysis when establishing the model. The GEP function used in this experiment is shown in the following Table 3.

Number	Name	Representation	Arity	Definition
1	Additiion	+	2	(x+y)
2	Subtraction	-	2	(x-y)
3	Multiplication	*	2	(x*y)
4	Division	/	2	(x/y)
5	Floating-point Remainder	Mod	2	mod(x,y)
6	Power	Pow	2	pow(x,y)
7	Square Root	Sqrt	1	sqrt(x)
8	Exponential	Exp	1	exp(x)
9	10^x	Pow10	1	pow(10,x)
10	Natural Logarithm	Ln	1	ln(x)

2023 The American Computational Science Society. All rights reserved.

11	Sine	Sin	1	sin(x)
12	Cosine	Cos	1	cos(x)
13	Tangent	Tan	1	tan(x)

Table 3. The function used in GEP

3. Results and Discussion

3.1. Results of HM

A total of 618 descriptors were matched in CODESSA software. In order to select the best descriptor to describe the anticancer properties of 2-Aminothiazole derivatives, the name of the descriptor, the number of descriptors, relevant information and R^2, R^2_{cv} were recorded, and a line graph was drawn based on the above data. Figure 3. shows that when the number of descriptors increases but R^2 does not increase significantly, the most suitable number and type of descriptors can be determined. The values of R^2, R^2_{cv} and S^2 are shown in Table 4.

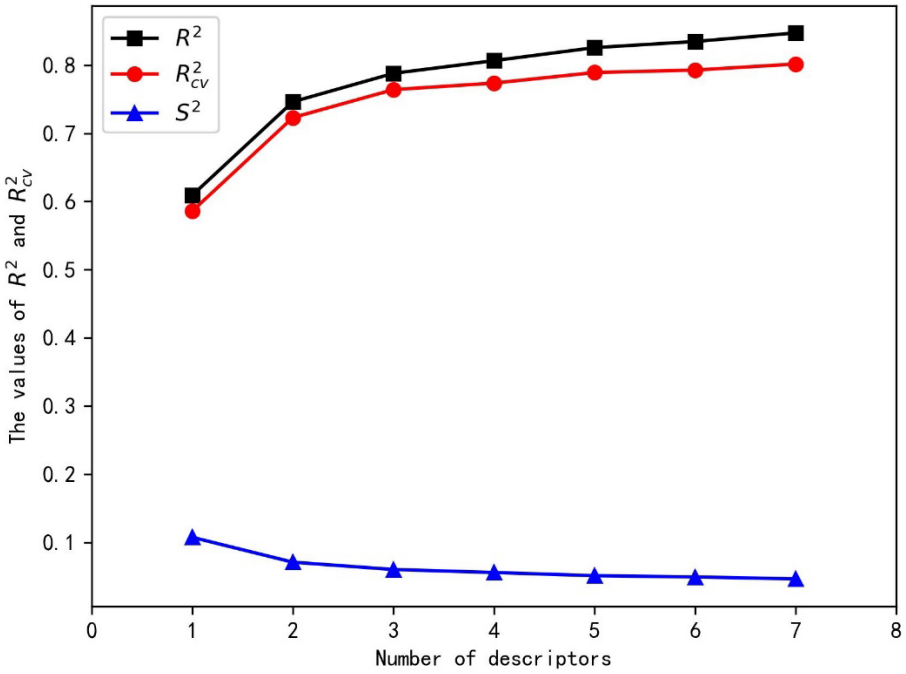


Figure 3. Influence of the Number of Descriptors on R^2, R^2_{cv} and S^2

Number of descriptors	1	2	3	4	5	6	7
R^2	0.60	0.74	0.78	0.80	0.82	0.83	0.84
R^2_{CV}	0.58	0.72	0.76	0.77	0.78	0.79	0.80
S^2	0.10	0.07	0.06	0.05	0.05	0.04	0.04

Table 4. The values of R^2, R^2_{cv} and S^2

All descriptors and R^2 data are presented in the table. It can be seen from the table that R^2 has reached 0.806 and R^2_{cv} has reached 0.773 when four descriptors are selected. At this time, if the number of descriptors continues to increase, R^2 is still improving, but it is not significant, so it can be determined that four descriptors were sufficient to describe the anticancer properties of this compound. The plot of HM is shown in Figure 4.

2023 The American Computational Science Society. All rights reserved.

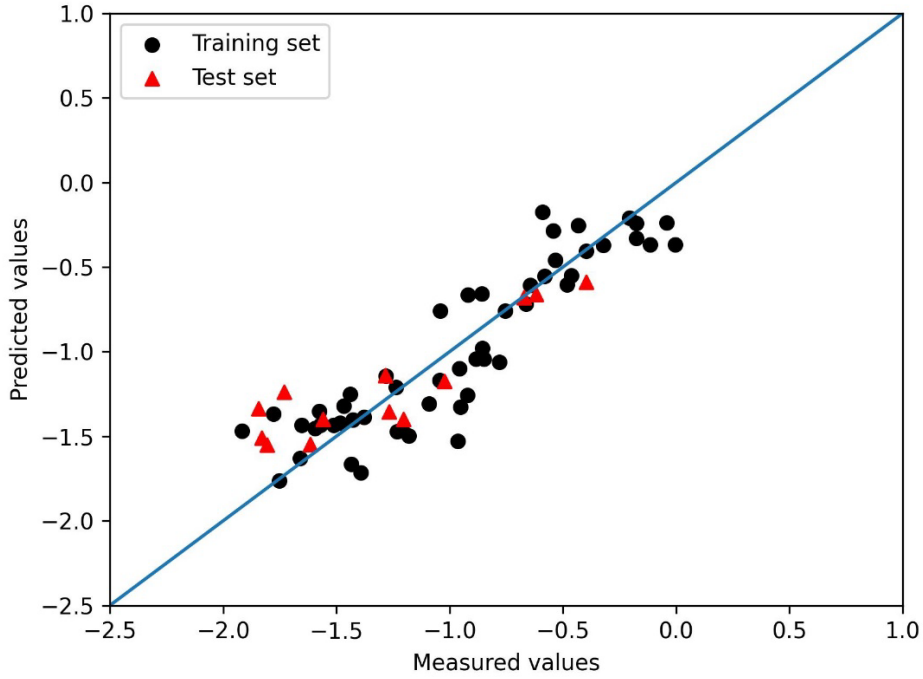


Figure 4. The Plot of Measured and Predicted −lg (IC_{50}) by HM

3.2. Results of RF

For the 618 descriptors obtained from CODESSA software, the RF is applied again to obtain the GINI coefficients of descriptors, the histogram of GINI coefficient is shown Figure 5. The GINI coefficients of Descriptors are shown in Table 5.

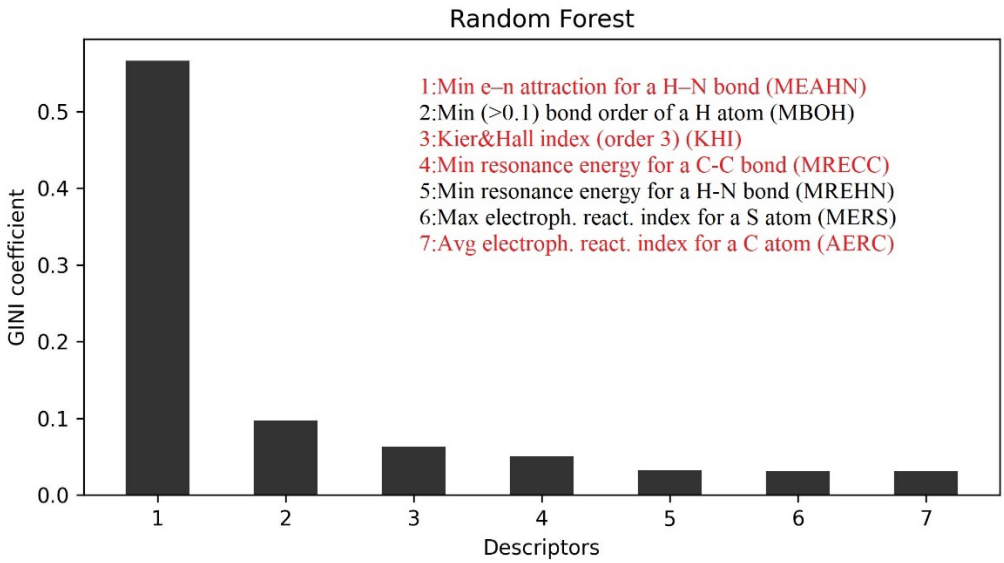


Figure 5. The Descriptors Selected by RF

Table 5. The GINI coefficient of Descriptors

Descriptors	MEAHN	MBOH	KHI	MRECC	MREHN	MERS	AERC
GINI coefficient	0.566	0.097	0.063	0.050	0.032	0.031	0.031

Combining HM and RF, it is found that three of the four most important descriptors for the selection of RF nonlinear model have been selected in the HM linear model, which can prove that the selection of descriptors for the HM is

2023 The American Computational Science Society. All rights reserved.

reasonable and correct. Figure 6. shows Measured and Predicted Results by RF.

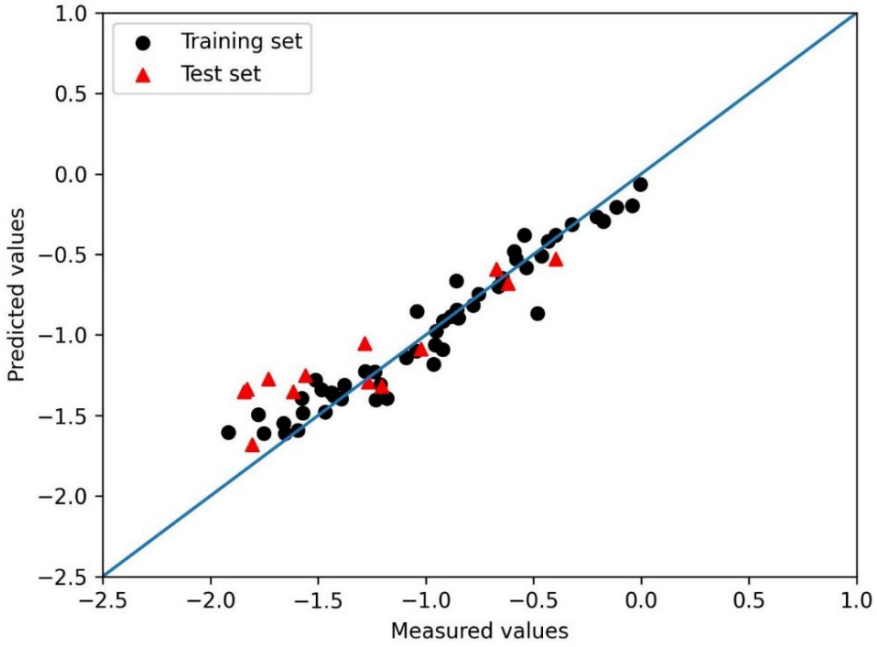


Figure 6. The Plot of Measured and Predicted −lg (IC_{50}) by RF

3.3. Results of SVM

The regression of SVM requires multiple parameters, such as push coefficient C, ε, and ε- insensitivity function, kernel function κ, where C= 291.505, ε=0.1, γ(gamma)= 10.235. In the final fitting result, R^2 of training set and test set were 0.959 and 0.941, and MSE of training set and test set were 0.010 and 0.013. The best prediction results are shown in the following Figure 7. It is easy to conclude that the fitting effect of SVM model for data is very good.

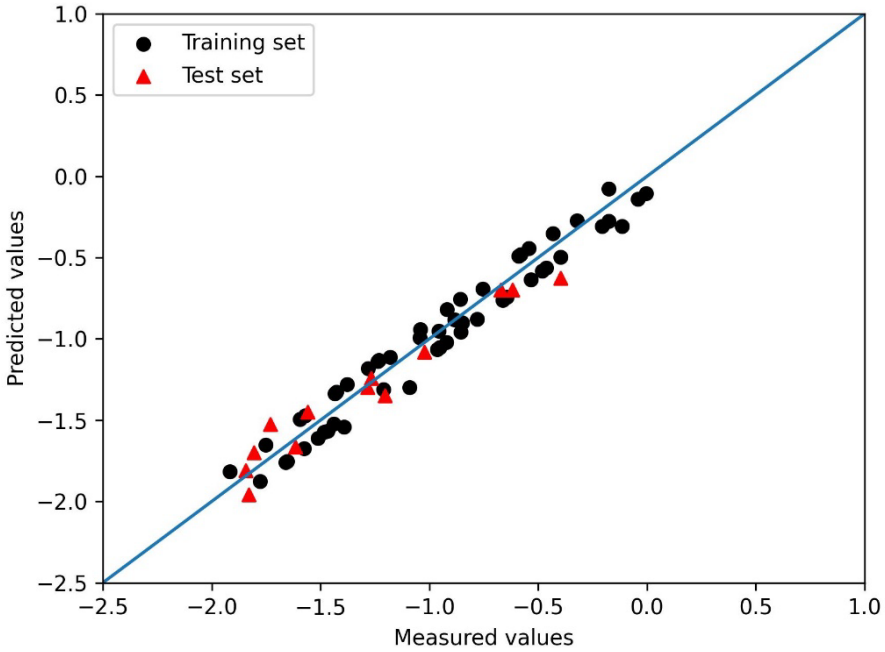


Figure 7. The Plot of Measured and Predicted −lg (IC_{50}) by SVM

3.4. Results of GEP

2023 The American Computational Science Society. All rights reserved.

After verifying the rationality of the four descriptors, the above four descriptors can also be used normally in other nonlinear models. Put these four descriptors and the corresponding IC_{50} values into the APS software, select functions such as addition, subtraction, multiplication and division, and run the GEP algorithm for iterative evolution. The final results are shown in the following Table 6. The formula Eq. (7) obtained by APS software is as follows. Figure 8. shows Measured and Predicted Results by GEP.

GEP	Best of Run - Training	Best of Run - Testing
Fitness	959.928	945.914
R^2	0.842	0.843
Max. Fitness	1000	1000
MSE	0.04	0.05
RMSE	0.20	0.23

Table 6. The Results of GEP

$$\begin{aligned} -\lg(IC_{50}) = {} & (\tan(f \bmod)((d_4 + d_3), \tan(d_4))) \cdot ((d_4 - d_3) - (d_3 + d_3))) \\ & + (\cos(sqrt(\exp(d1))) \cdot \cos(pow(\log(d_1), \sin(d_2)))) \\ & + ((d_2 - (\tan((d_3 + d_1)) \cdot (d_3 + d_3))) \cdot d_4) \\ & + ((d_2 - f \bmod(((d_1 + d_4) \cdot d_3), (d_4 + d_4))) \cdot d_4) \\ & + \sin(d_1) \end{aligned} \tag{7}$$

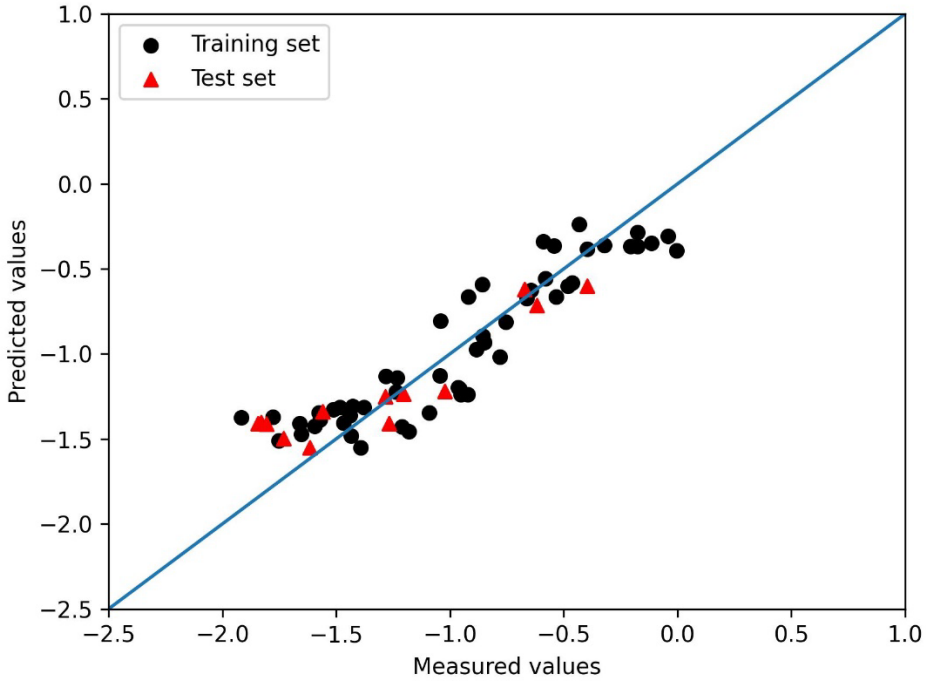


Figure 8. The Plot of Measured and Predicted −lg (IC_{50}) by GEP

4. Discussion

After the screening of HM and the verification of the GINI coefficient of RF, the rationality of the descriptor selection has been verified. Among the descriptors selected by the RF model, three of the first four descriptors have been selected by the HM model, and the first seven descriptors selected by the RF include all four descriptors (MEAHN, KHI, MRECC, and AERC) selected by the HM model, further verifying the accuracy of descriptor selection. MEAHN is the most important descriptor because it is the first selected one in both HM and RF. Several other descriptors also directly or indirectly affect the molecular anticancer activity.

Considering that the prediction effect of linear prediction model is relatively common, many nonlinear prediction models have been established. The prediction results of all the above prediction models are shown in the Table 7. It is easy to find that SVM has the best prediction ability and robustness, and its results are the most satisfactory, because SVM is

2023 The American Computational Science Society. All rights reserved.

more suitable for small data sets. GEP model is slightly inferior to SVM, and can also achieve good results in prediction and analysis. The combination of prediction results from multiple models collectively demonstrates the impact of different molecular structure on anticancer activity.

Compounds 39 and 41 in Table 1 have lower IC_{50} values. In combination with this experiment, other molecules with similar chemical structures are likely to have anticancer properties and have the potential to produce new anticancer drugs. In conclusion, the relationship between the four descriptors mentioned above and cancer resistance has high research value.

Table 7. Comparison of Prediction Results of Different Methods

Method	R^2		MSE	
	Training set	Test set	Training set	Test set
HM	0.806	0.773	0.048	0.063
RF	0.935	0.670	0.016	0.074
SVM	0.959	0.941	0.010	0.013
GEP	0.842	0.843	0.041	0.057

5. Conclusion

SVM model shows the best prediction ability. In the future, SVM model is believed to be widely used in the study of 2-Aminothiazole derivatives. Among the molecular descriptors of anticancer properties, Min exchange energy for a H-N bond, Kier&Hall index (order 3), Min response energy for a C-C bond, Avg electroph. react. index for a C atom and other important descriptors play a key role, and these descriptors are effective in the development and screening of new drugs.

Conflict and interest

The authors confirm that this article content has no conflict of interest.

Acknowledgements

The authors are immensely grateful for the APS software, Chemdraw software, HyperChem software, MOPAC software, and PyCharm software providers. The authors are also immensely grateful to Qingdao University for their tremendous support in this school level innovation and entrepreneurship project(X2021110650416).

References

[1] World Health Organization: Cancer – Key Facts， Treatment. 2022; http://www.who.int/news-room/fact-sheets/detail/cancer.
[2] W.D. Alrohily, M.E. Habib, S.M. El-Messery, A. Alqurshi, H. El-Subbagh, E.-S.E. Habib. Microbial Pathogenesis, 136 (2019) 103674.
[3] Z.K. Laczkowski, K. Misiura, M. Switalska, J. Wietrzyk, B.A. Laczkowska, B. Fernandez, A. Paneth, T. Plech. Anti-Cancer Agents in Medicinal Chemistry, 14 (2014) 1271-1281.
[4] P. Li, S.-F. Yang, Z.-L. Fang, H.-R. Cui, S. Liang, H.-Y. Tian, B.-G. Sun, C.-C. Zeng. Journal of Environmental Chemical Engineering, 10 (2022) 107487.
[5] A. Pogorzelska, J. Sławiński, K. Brożewicz, S. Ulenberg, T. Bączek. Molecules, 20 (2015) 21960-21970.
[6] S. Kamboj, A. Rajput, A. Rastogi, A. Thakur, M. Kumar. Computational and Structural Biotechnology Journal, 20 (2022) 3422-3438.
[7] Z. Cheng, Q. Chen, S. Liu, Y. Liu, Y. Ren, X. Zhang, Z. Shen. Journal of Hazardous Materials, 430 (2022) 128269.
[8] R. Vankdothu, M.A. Hameed. Measurement: Sensors, 24 (2022) 100440.
[9] R. Kasarda, N. Moravčíková, G. Mészáros, M. Simčič, D. Zaborski. Livestock Science, 267 (2023) 105143.
[10] W. Guo, J. Zhang, D. Cao, H. Yao. Construction and Building Materials, 330 (2022) 127219.
[11] A.M. El-Naggar, M.A. El-Hashash, E.B. Elkaeed. Bioorg Chem, 108 (2021) 104615.
[12] D. Osmaniye, S. Gorgulu, B.N. Saglik, S. Levent, Y. Ozkay, Z.A. Kaplancikli. Bioorg Chem, 114 (2021) 105123.
[13] T.D. Dos Santos Silva, L.M. Bomfim, A.C.B. da Cruz Rodrigues, R.B. Dias, C.B.S. Sales, C.A.G. Rocha, M.B.P. Soares, D.P. Bezerra, M.V. de Oliveira Cardoso, A.C.L. Leite, G.C.G. Militao. Toxicol Appl Pharmacol, 329 (2017) 212-223.
[14] K. Piechowska, M. Switalska, J. Cytarska, K. Jaroch, K. Luczykowski, J. Chalupka, J. Wietrzyk, K. Misiura, B. Bojko, S. Kruszewski, K.Z. Laczkowski. Eur J Med Chem, 175 (2019) 162-171.
[15] M.A. Sabry, M.A. Ghaly, A.R. Maarouf, H.I. El-Subbagh. Eur J Med Chem, 241 (2022) 114661.
[16] K.Z. Łączkowski, M. Świtalska, A. Baranowska-Łączkowska, T. Plech, A. Paneth, K. Misiura, J. Wietrzyk, B. Czaplińska, A. Mrozek-Wilczkiewicz, K. Malarz, R. Musioł, I. Grela. Journal of Molecular Structure, 1119 (2016) 139-150.

2023 The American Computational Science Society. All rights reserved.

[17] R.O. Oyewole, A.K. Oyebamiji, B. Semire. Heliyon, 6 (2020) e03926.
[18] HyperChem. 4.0, Hypercube, 1994.
[19] A.R. Katritzky, S. Perumal, R. Petrukhin, E. Kleinpeter. Journal of Chemical Information and Computer Sciences, 41 (2001) 569-574.
[20] W. Ren, D. Kong. Computers and Applied Chemistry, 26 (2009) 1455–1458.
[21] W. Tong, D.R. Lowis, R. Perkins, Y. Chen, W.J. Welsh, D.W. Goddette, T.W. Heritage, D.M. Sheehan. Journal of Chemical Information and Computer Sciences, 38 (1998) 669-677.
[22] A.R. Katritzky, L. Mu, V.S. Lobanov, M. Karelson. The Journal of Physical Chemistry, 100 (1996) 10400-10407.
[23] D. He, J. Ma, X. Shi, C. Zhao, M. Hou, Q. Guo, S. Ma, X. Li, P. Zhao, W. Liu, Z. Yang, J. Mou, P. Song, Y. Zhang, J. Li. Chemical and Pharmaceutical Bulletin, 62 (2014) 967-978.
[24] B. Yang, H. Si, H. Zhai. Letters in Drug Design & Discovery, 18 (2021) 406-415.
[25] W.-X. Wang, H. Si, Z. Zhang. Molecular Simulation, 38 (2012) 38-44.
[26] Y. Si, X. Xu, Y. Hu, H. Si, H. Zhai. Chemical Biology & Drug Design, 97 (2021) 978-983.
[27] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston. Journal of Chemical Information and Computer Sciences, 43 (2003) 1947-1958.
[28] K. Inthajak, N. Khamsemanan, C. Nattee, P. Toochinda, L. Lawtrakul. Monatshefte für Chemie - Chemical Monthly, 148 (2017) 1697-1709.
[29] H. Tyralis, G. Papacharalampous, A. Langousis. Water, 11 (2019).
[30] C.J.C. Burges. Data Mining and Knowledge Discovery, 2 (1998) 121-167.
[31] C.Z. Cai, L.Y. Han, Z.L. Ji, X. Chen, Y.Z. Chen. Nucleic Acids Research, 31 (2003) 3692-3697.
[32] C.Z. Cai, W.L. Wang, L.Z. Sun, Y.Z. Chen. Mathematical Biosciences, 185 (2003) 111-122.
[33] R. Khemchandani, Jayadeva, S. Chandra. Expert Systems with Applications, 36 (2009) 132-138.
[34] C. Ferreira. Complex Syst, 13 (2001) 87-129.
[35] C. Ferreira. Springer, (2002) 635-653.
[36] P. Dey, A.K. Das. Energy, 95 (2016) 447-458.
[37] K.J. Zhang. Zhejiang University, (2010) 1-13.

2023 The American Computational Science Society. All rights reserved.