ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Quantitative Structure Activity Relationship And Artificial Neural Networks In Design Of Benzimidazoles As Antiproliferative Agents.

Layla Abdel-Ilah[1,*], Selma Zukic[2], Elma Veljovic[2], Almir Badnjevic[1,3], and Lejla Gurbeta[1,3].

[1]International Burch University, Faculty of Engineering and Natural Sciences, Genetics and Bioengineering Department, Francuskerevolucije bb, Ilidza, 71210 Sarajevo.
[2]University of Sarajevo, Faculty of Pharmacy, ZmajaodBosne 8, Sarajevo.
[3]Medical Devisce Inspection Laboratory Verlab, Ismeta Mujezinovica 30, 71000 Sarajevo.

ABSTRACT

Designing safe and effective medications is the aim of drug discovery and development process to improve drug's potency and to reduce their side effects. The aim of this study was to test the efficiency of Quantitative Structure Activity Relationship (QSAR) and Artificial Neural Networks (ANN) in prediction antiproliferative activity of benzimidazoles. Data set of 27 compounds with antiproliferative activity on MCF-7 breast cancer cell line was collected from literature. QSAR and ANN models were developed. Descriptors with a strong negative influence on antiproliferative activity were: Mor23p, Mor23m, X1A, Mor31v. Trial and error method was used by changing the parameters of ANN: number of layers, number of neurons in hidden layers and type of transfer functions. Networks with 6 hidden layers, 5 neurons, Back propagation type and LOGSIG activation function achieved the most accurate outputs. It was found that ANN study is more suitable for antiproliferative activities testing than QSAR study.

Keywords: Benzimidazoles, Quantitative Structure Activity Relationship (QSAR), Artificial Neural Networks (ANN), Antiproliferative activity, MCF-7.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Benzimidazole derivatives have many practical uses in laboratory and industry as a result of their interesting heterocyclic structure.[1] More importantly, they constitute an important class of biologically active compounds due to their broad spectrum of pharmacological activities such as antiproliferative.[2]antimicrobial,[3,4]anticonvulsant,[5] anti-inflammatory,[6] analgesic [7] properties. Some drugs with benzimidazole as structural motif, such as tiabendazole, parbendazole, mebendazole, albendazole and flubendazole have been widely used in therapy of gastrointestinal worm infections.[8]In particular, nowadays we face some major problems in medicine, including resistance to antibiotics and lacking of cure for some chronic diseases and cancer. Related to the problem of microbial resistance, a number of enzymes and receptors important for microbial growth as well as for the tumour cell cycle progression maybe considered as potential targets for new drugs. This is important as it is largely been recognized that chronic inflammation due to microbes may induce mutations due to microbial release of genotoxic agents which adds to tumour development.[9]

As such validated targets are recognized as major pillars in the drug discovery and drug development process, a number of *in silico* methods are used to identify potential targets and interactions of small molecules with target proteins in the cell. The application of these methods is a necessity in the drug discovery process to reduce the phase time, as well as to amplify the design of new molecules with better biological activities and minimal side effects for a disease specific target. Quantitative Structure Activity Relationship (QSAR) and Artificial Neural Networks (ANN) are examples on *in silico* methods and are widely used in this *in silico* prediction studies.

QSAR methods have been applied in several scientific studies including chemistry, biology and toxicology and drug discovery to predict and classify biological activities of virtual or newly-synthesized compounds. QSAR models can also be used in designing new chemical entities and are now regarded as essential tools in pharmaceutical industries to identify promising hits and generate high quality leads in the early stages of drug discovery.[10]

ANNs are integrate systems inspired by biological neuron networks. In machine learning and cognitive sciences, these networks estimate functions depending on inputs.[11]Artificial Neural Networks are considered as an important research area in health care and medicine because of their ability to increase the accuracy of diagnosis, reduce time and costs.

ANNs use nonlinear statistical data modeling tools where the complex relationships between inputs and outputs are modeled or patterns are found. In medicine, ANN can help doctors to diagnose diseases earlier and more efficiently based on diagnostic input values.[12-21]

In drug design ANN models applications include: compounds screening, quantitative structure activity relationship studies, receptor modeling, formulation development, pharmacokinetics, and processes involving complex mathematical modeling.[11,22] Training process is based on weights of the data set, and is adjusted until the errors are minimized. It allows finding predictive, robust, and accurate models.[23]

The aim of this study is testing the efficiency of Quantitative Structure Activity Relationship (QSAR) and Artificial Neural Networks (ANN) in predicting the antiproliferative activity of benzimidazole derivatives (against Michigan Cancer Foundation (MCF) -7 cell line).

EXPERIMENTAL SECTION

Data set

In order to investigate efficiency of QSAR and ANN in design benzimidazole derivatives we used alredy descrabed benzimidazole with antiproliferative activity. Data set of 27 compounds (Table. 1) with antiproliferative activity on MCF-7 breast cancer cell line were collected from three different articles published in 2000, 2007, and 2008 years. All collected compounds were tested on MCF-7 cell line using the same cytotoxicity assay with activities expressed as IC50 and they all show a noticeable activity against this line.

ISSN: 0975-8585

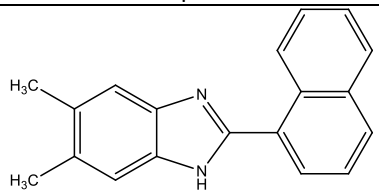
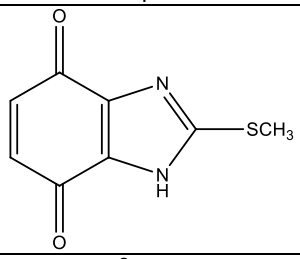
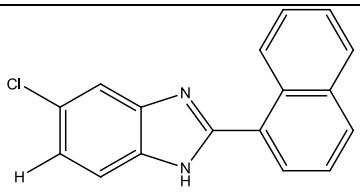
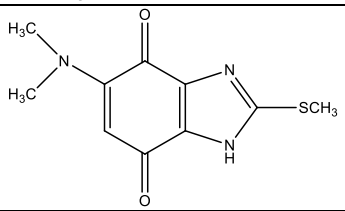
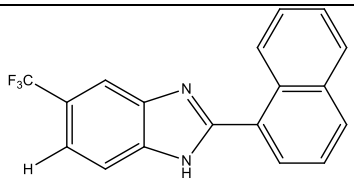
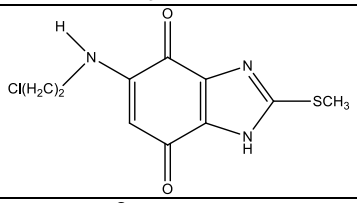
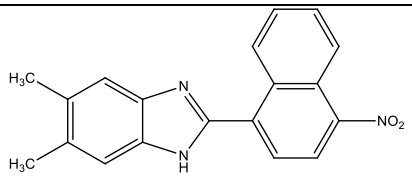
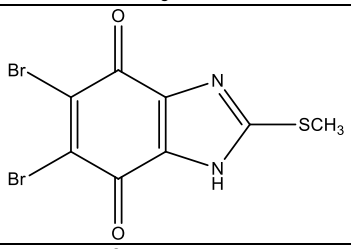
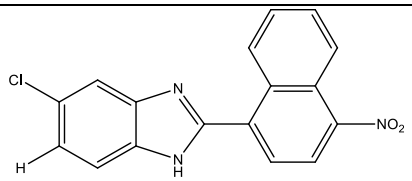
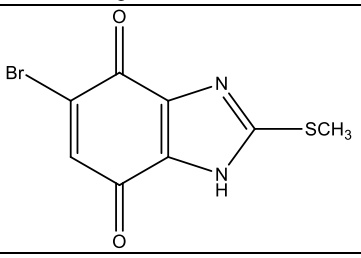
Compounds 1-12 are a series of 2-arylbenzimidazoles were synthesized and tested for antiproliferative activity against a panel of exponentially growing cell lines derived from human haematological and solid tumors. The effect of compounds was determined under experimental conditions allowing cells to undergo exponential growth for at least three cell cycles.[24]

Compound 13-21 are a series of benzimidazole-4,7-diones bearing at the 2-position the thiomethyl group or the 2-pyridyl moiety. They were synthesized and tested in vitro on the tumor cell line.[25]

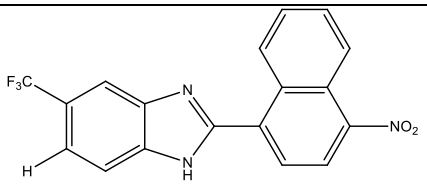
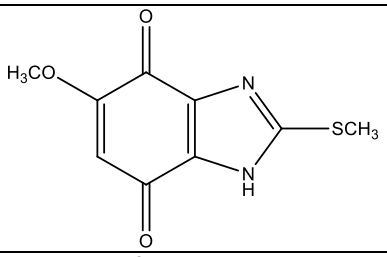
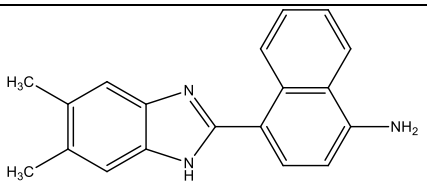
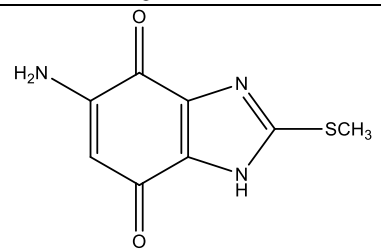
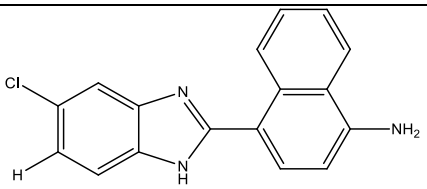
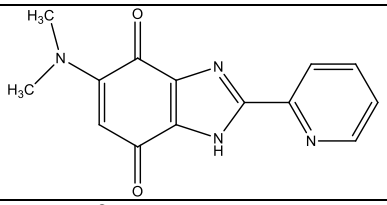
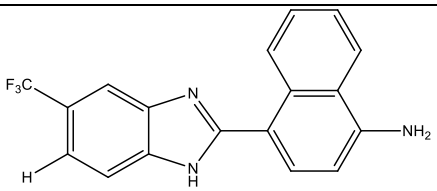
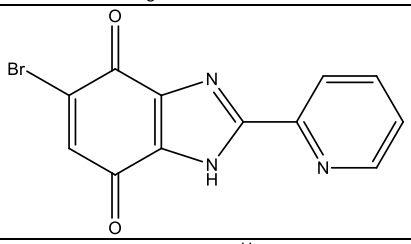
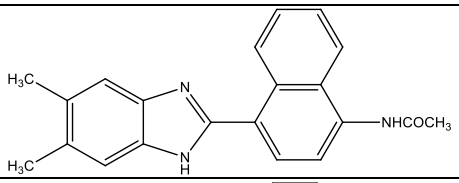
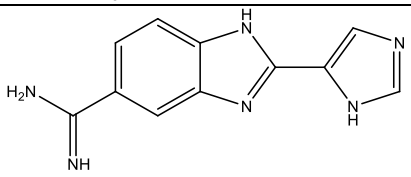
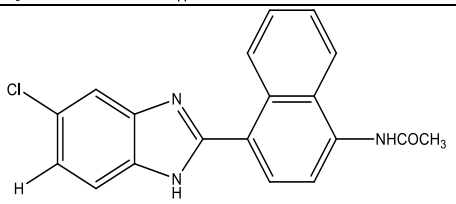
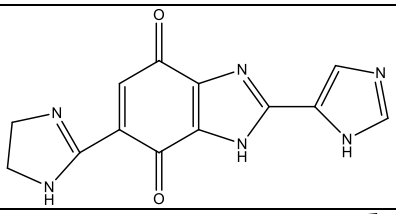
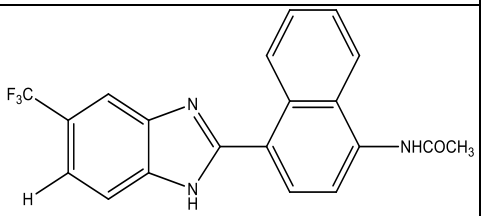
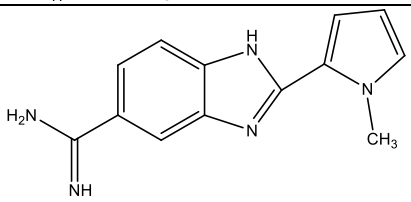
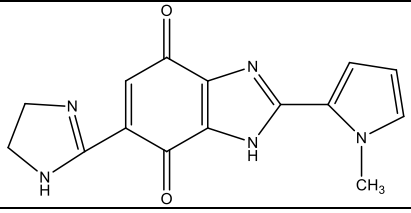
Compounds 22-27 are a group of heterocyclic benzimidazoles with amidino substituents at C-5 of benzimidazole ring. They were manufactures by applying various heterocyclic nuclei at C-2 position. They were tested against tumor cells.[26]

Initial data set (Table 1) was divided on training set consisting of 19 molecules (compounds 2,5,6,7,8,9,10,11,12,13,14,16,17,19,21,23,25,26,27) and test set consisting of 8 molecules (compounds 1,3,4,15,18,20,22,24). Compounds from both sets cover wide range of activity. Negative logarithm of IC50 (pIC50) was calculated to represent Y-variable.

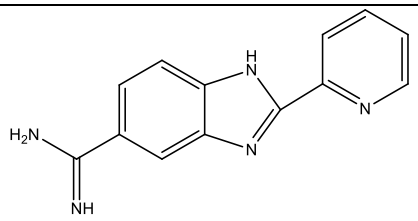
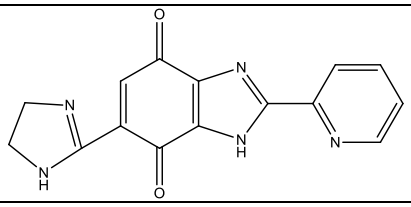
Table 1: Structure and antiproliferative activity of tested benzimidazole derivatives

	Compound	IC_{50}		Compound	IC_{50}
1		4.76955	13		4.20308
2		4.4318	14		4.04576
3		5.37675	15		4.04576
4		5.14267	16		4.39469
5		5.14267	17		4.20204

ISSN: 0975-8585

6		5.02228	18		4.4437
7		6.1549	19		4.04576
8		5.55284	20		4.04576
9		5.67778	21		4.04576
10		4.25181	22		5.14267
11		5	23		4
12		5.22185	24		4.92082
			25		5.30103

ISSN: 0975-8585

	26		4
	27		4.92082

**IC50: The concentration that causes a reduction in the cell growth by half.*

Quantitative Structure Activity Relationship study

Molecules selection for dominant microspecies at physiological pH 7,4 was performed using Marvin Sketch 6.1.0 software,[27] which helps in drawing compounds structures.

Structures of all dominant form were pre-optimized with semiempirical/PM3 Method (*Parameterized Model revision 3*),[28] using Gaussian software, [29] with Chem3D Ultra7 program.[30] Structures were than optimized using quantum chemical HartreeFock/3-21G method for energy minimization.[31]

Molecular descriptors were calculated with E-Dragon software,[32] with descriptors divided in several categories including constitutional descriptors, ring descriptors, geometrical descriptors, 2D-matrix-based descriptors, connectivity indices etc. Highly intercorrelated descriptors were excluded from the matrix.

2D-QSAR model was developed with *Partial least square regression,* PLS,[33] using SIMCA 14.1 software (*Soft Independent Modeling of Class Analogy)*.[34]

Selection of descriptor is based on calculation of variable of importance in the projection (VIP) that represent the influence of each x-variable on Y and X matrices. Descriptors with VIP score smaller than 0,5 were considered as unimportant and were excluded from PLS model. Next model was created and parameters of the model (R2, Q2(Y), RMSEE, F *ratio*, p vrijednost) were calculated.[35]

The best model with 4 descriptors and good statistical parameters was obtained.
Parameters Q2 (*cross-validated squared correlation coefficient*) and PRESS (*Predicted residuals sum of squares*) were used for internal validation.[36]

Predictive ability of the model was estimated with RMSEE (*root mean square error of estimation*) and the cross-validated R2 (Q2). For predictive model Q2 should be greater than 0,5.[37]

In addition, to assess the robustness of the model, response permutation test (*Y*-scrambling), as a measure of the model overfitting, was performed.[38]

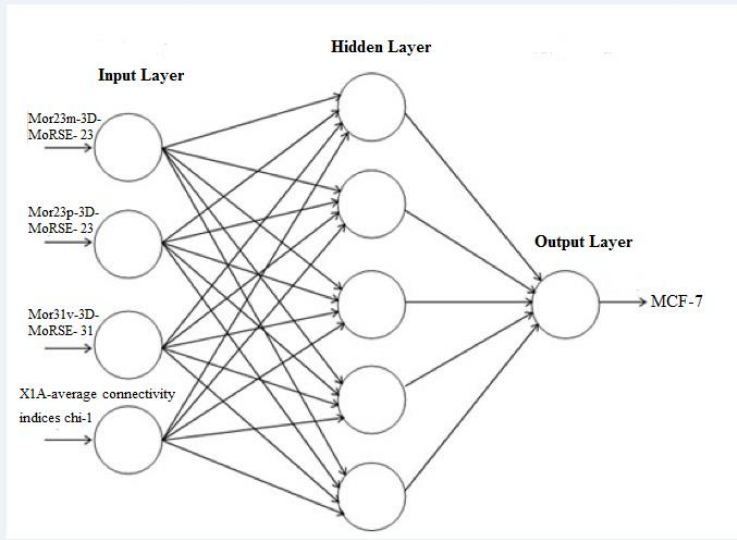
For external validation is used external test set with 8 molecules, RMSEP (*root mean square error of prediction)* and correlation coefficient for the test set R2 pred. For predictive QSAR model values of R2pred should be higher than 0,5.[37] For furter support to good predictivity, r2m parameters were calculated using „Applicability domain using standardization approach".[39]

Artificial Neural Networks study

In this research, the ANN method was applied to obtain a nonlinear models to test the anti-proliferative activity of molecules in the data set, to develop model's accuracy. Networks applied here have a special structures and units: (1) the number of neurons in the input layer was 4, like the number of chosen

descriptors, (2) the number of hidden neurons was optimized and changed, and (3) one neuron was put in the output layer to present MCF-7 value[40] (Figure. 1).

Figure 1: The structure of constructed networks.



When developing an artificial neural network to solve a specific problem, type of the network, number of hidden layers, number of neurons in each layer, dataset division percentages and type of the training algorithm can affect the performance of the model and the accuracy of the obtained results.[41]
In this study, not only impact of network's architecture type on the performance of the model and the accuracy of the obtained results are tested, but also the effect of changing hidden layers numbers, number of neurons in the hidden layer and type of activation function in neurons.

Trial and error method was used by changing three parameters: number of hidden layers, number of neurons in each layer and the type of transfer function, to compare the performance of different ANN architectures using Mean Square Error (MSE) and Absolute Error calculations (Equations. 1 &2).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(X_{predicted} - X_{actual}\right)^2$$ **Equation. 1**

$$Error_{absolute} = Output_{target} - ANN_{output}$$ **Equation. 2**

The number of neurons within the hidden layer is chosen to be from 2 to 25. The number of hidden layers is from 1 to 10. The transfer functions are: TANSIG, LOGSIG, and PURELINE. The performance received from these systems were compared. After changing parameters, the optimal classifiers were identified and employed for network architecture comparison. To investigate optimal ANN architecture for testing the antiproliferative activity, data was applied to three types of neural networks and their performances were compared. The trained networks were Feed Forward Back propagation network (BPA), Radial Basis Function (RBF), and Probabilistic neural network (PNN).
Levenberg- Marquardt algorithm was applied in this research because it shows a good performance and low output errors in many classification tasks.[42]

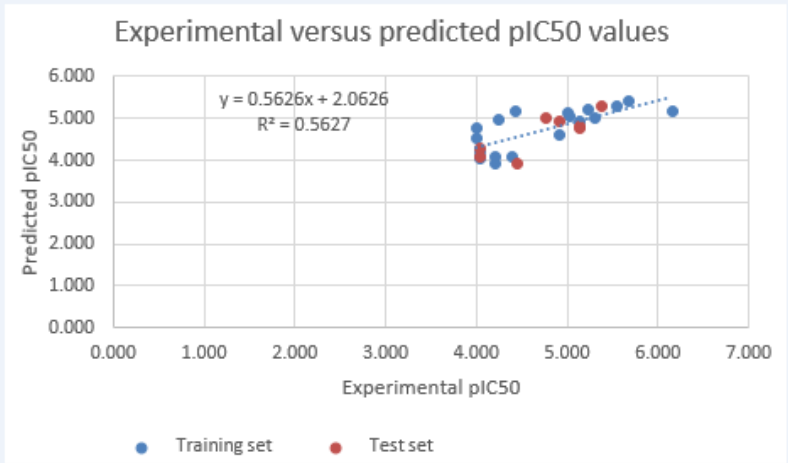
RESULTS AND DISCUSSION

QSAR study results

Once the quantitative and qualitative models were built, each molecule of the training and test sets has its antiproliferative activity predicted. Experimental and predicted values are presented in Table 2 and Figure 2.

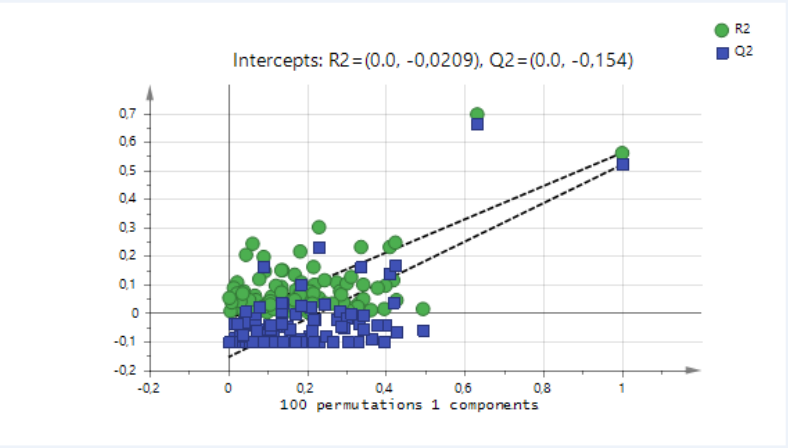
ISSN: 0975-8585

Figure 2: Plot of experimental versus predicted pIC50 values for QSAR model.



The Y-scrambling test gave the following intercept values of R2 and Q2 regression lines: R2=(0.0,-0.0209), Q2=(0.0,-0.154), which showed that model was not obtained by chance. However, it is obtained by linking the pIC50 values with structure's properties (Figure. 3)

Figure 3: *Y*-scrambling test with intercept values of R^2 and Q^2 regression lines: R^2=(0.0,-0.0209), Q^2=(0.0,-0.154).



Several partial Least Squares (PLS) models with different sets of variables were developed using Soft Independent Modeling of Class Analogy SIMCA 14,1. software. Variables were selected based on their Variable Importance in Projection VIP score. Those with smaller VIP values were removed from the model. Optimal model with four variables was selected.

Statistical parameters of the model are listed in Table 2.

Table 2: Results of predicted antiproliferative activity by using QSAR study.

Compounds	Experimental pIC_{50}	Predicted pIC_{50}	Absolute Error
1	4.76955	5.02291	0.25336
2	4.4318	5.16684	0.73504
3	5.37675	5.29702	0.07973
4	5.14267	4.8098	0.33287
5	5.14267	4.92109	0.22158
6	5.02228	5.05907	0.03679
7	6.1549	5.1709	0.984

ISSN: 0975-8585

8	5.55284	5.29489	0.25795
9	5.67778	5.39984	0.27794
10	4.25181	4.95787	0.70606
11	5	5.11221	0.11221
12	5.22185	5.20358	0.01827
13	4.20308	3.92251	0.28057
14	4.04576	4.07907	0.03331
15	4.04576	4.08902	0.04326
16	4.39469	4.07919	0.3155
17	4.20204	4.07924	0.1228
18	4.4437	3.92418	0.51952
19	4.04576	4.05845	0.01269
20	4.04576	4.29942	0.25366
21	4.04576	4.21133	0.16557
22	5.14267	4.76262	0.38005
23	4	4.77383	0.77383
24	4.92082	4.91442	0.0064
25	5.30103	5.00344	0.29759
26	4	4.51955	0.51955
27	4.92082	4.60197	0.31885

Scatter plot of experimental *versus* predicted values of training and test set is shown in Figure 2. Four most important descriptors that have influence on antiproliferative activity are:

Mor23m-3D-MoRSE-signal 23/weighted by atomic masses
Mor23p-3D-MoRSE-signal 23/weighted by atomic polarizability
Mor31v-3D-MoRSE-signal 31/weighted by atomic van der Waals volumes
X1A-average connectivity indices chi-1

VIP values of these descriptors (Figure 4). According to the coefficient plot all descriptors have negative influence on biological activity (Figure 5)

Figure 4: VIP values of most important descriptors used in 2D-QSAR.

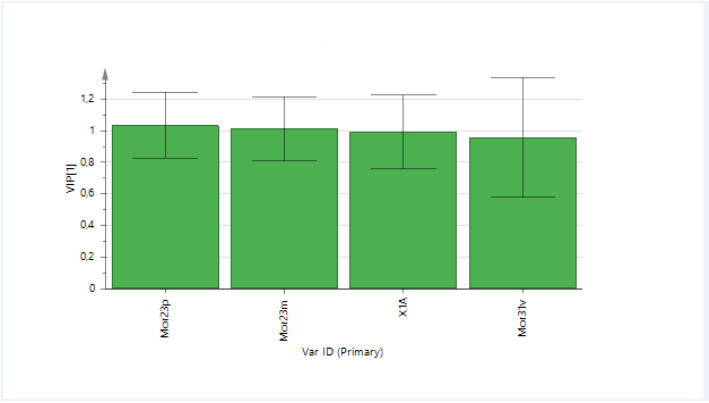
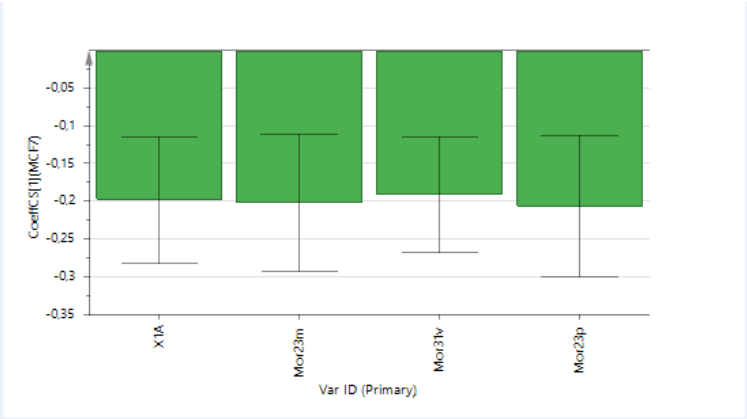


Figure 5: Coefficient plot of most important descriptors used in 2D-QSAR.

ISSN: 0975-8585



Values of the four most important descriptors that have influence on antiproliferative activity are presented in Table 3 and Table 4.

Table 3: Values of the four most important descriptors for training set.

Primary ID	Set	X1A	Mor23m	Mor31v	Mor23p	MCF7	YPredPS[1](MCF7)
2PM3-HF.MOL	WS	0.427	-0.796	-0.077	-0.76	4.4318	5.16684
5PM3-HF.MOL	WS	0.428	-0.564	-0.055	-0.658	5.14267	4.92109
6PM3-HF.MOL	WS	0.426	-0.809	-0.012	-0.687	5.02228	5.05907
7PM3-HF.MOL	WS	0.425	-0.834	-0.017	-0.8	6.1549	5.1709
8PM3-HF.MOL	WS	0.426	-0.844	-0.138	-0.761	5.55284	5.29489
9PM3-HF.MOL	WS	0.423	-0.967	-0.123	-0.76	5.67778	5.39984
10PM3-HF.MOL	WS	0.429	-0.714	0.021	-0.769	4.25181	4.95787
11PM3-HF.MOL	WS	0.43	-0.728	-0.121	-0.73	5	5.11221
12PM3.HF.MOL	WS	0.428	-0.863	-0.105	-0.728	5.22185	5.20358
13PM3-HF.MOL	WS	0.444	-0.229	0.13	-0.245	4.20308	3.92251
14PM3-HF.MOL	WS	0.444	-0.323	0.089	-0.32	4.04576	4.07907
16PM3-HF.MOL	WS	0.441	-0.304	0.104	-0.282	4.39469	4.07919
17PM3-HF.MOL	WS	0.442	-0.377	0.118	-0.275	4.20204	4.07924
19PM3-HF.MOL	WS	0.442	-0.229	0.054	-0.241	4.04576	4.05845
21PM3-HF.MOL	WS	0.433	-0.399	0.161	-0.292	4.04576	4.21133
23PM3-HF.MOL	WS	0.427	-0.525	-0.045	-0.455	4	4.77383
25PM3-HF.MOL	WS	0.427	-0.562	-0.119	-0.626	5.30103	5.00344
26PM3-HF.MOL	WS	0.437	-0.528	0.007	-0.452	4	4.51955
27PM3-HF.MOL	WS	0.43	-0.412	-0.021	-0.424	4.92082	4.60197

Table 4: Values of the four most important descriptors for test set.

Primary ID	Set	X1A	Mor23m	Mor31v	Mor23p	MCF7	YPredPS[1](MCF7)
1PM3-HF.MOL	TS	0.426	-0.792	0.061	-0.798	4.76955	5.02291
3PM3-HF.MOL	TS	0.424	-0.985	-0.045	-0.775	5.37675	5.29702
4PM3-HF.MOL	TS	0.427	-0.599	0.081	-0.708	5.14267	4.8098
15PM3-HF.MOL	TS	0.454	-0.413	-0.008	-0.341	4.04576	4.08902
18PM3-HF.MOL	TS	0.448	-0.296	0.129	-0.301	4.4437	3.92418
20PM3-HF.MOL	TS	0.435	-0.411	0.11	-0.37	4.04576	4.29942
22PM3-HF.MOL	TS	0.434	-0.697	-0.031	-0.515	5.14267	4.76262
24PM3-HF.MOL	TS	0.432	-0.638	-0.088	-0.628	4.92082	4.91442

According to results QSAR study, a negative coefficient values of this variables means that increasing of this properties leads to decreasing antiproliferative activity. For further support to biological activity,

decreasing these properties which are described by following descriptors will lead to increasing in activity. (Table 3, 4)

Increasing the antitumor activity of molecules in the data set requires decreasing the atomic mass, atomic polarizability, van der Waals volume and the average connectivity indices. The selected four descriptors are the most important with the highest influence for the development of QSAR model and testing the MCF7 values with accepted results of outputs.

ANN results

Networks with 1 hidden layer, 5 neurons in the hidden layer, and LOGSIG function achieve the smallest errors among all other networks (Figures 6, 7 &8), (Tables 5, 6 &7).

Figure 6: Absolute error values for different number of hidden layers.

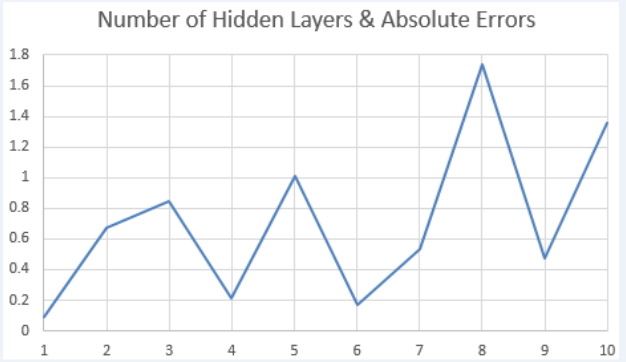


Figure 7: Absolute error values for different number of neurons in the hidden layer.

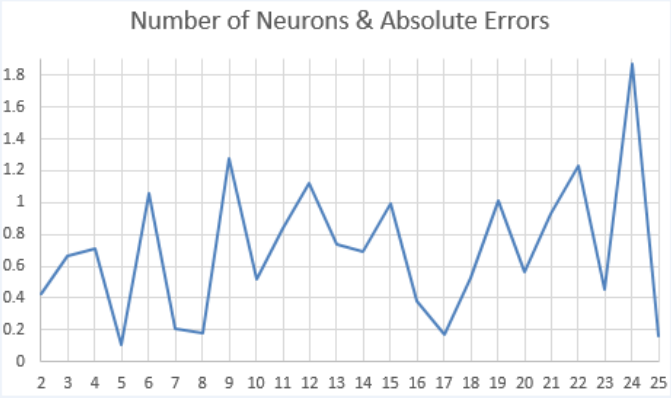


Figure 8: MSE values for three different networks: with TANSIG, PURELINE, and LOGSIG transfer functions. The MSE results were: 0.059, 0.0058, and 0.002 respectively.

ISSN: 0975-8585

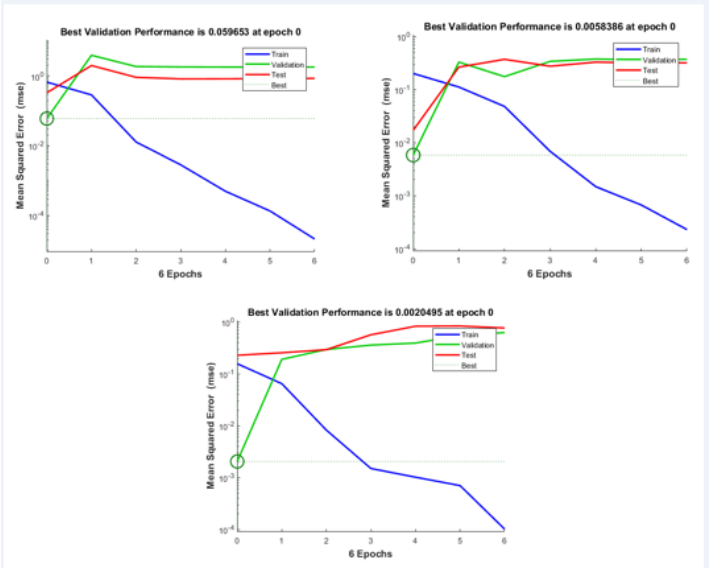


Table 5: Absolute error values for different number of neurons in the hidden layer.

Number of neurons	Absolute Errors
2	0.43
3	0.66
4	0.71
5	0.11
6	1.06
7	0.21
8	0.18
9	1.28
10	0.52
11	0.84
12	1.12
13	0.74
14	0.69
15	0.99
16	0.38
17	0.17
18	0.53
19	1.01
20	0.56
21	0.93
22	1.23
23	0.45
24	1.87
25	0.16

Table 6: Absolute error values for different number of hidden layers.

Number of hidden layers	Absolute errors
1	0.09
2	0.67
3	0.85
4	0.21

ISSN: 0975-8585

5	1.01
6	0.17
7	0.53
8	1.74
9	0.47
10	1.36

Table 7: MSE values for networks with TANSIG, PURELINE, and LOGSIG transfer functions.

Transfer function	MSE
TANSIG	0.63
LOGSIG	0.18
PURELINE	0.31

Feed forward back propagation network achieves the lowest average error = 0.263, and 0.183 in training and test respectively. This accuracy is due to its ability to minimize errors and make strong relations between inputs and outputs. Average absolute errors for RBF, BPA, and PNN networks are shown in Figure 9.

Figure 9: Average Absolute Errors for RBF, BPA, and PNN networks after training and testing.

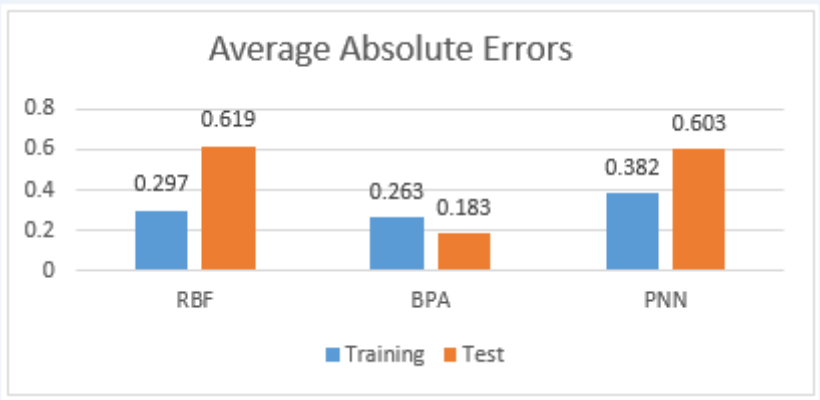


Figure 10. represents the regression plot of BPA network for training, test, and validation sets. R values are very near to or equals 1, and most of compounds are near or on the line. This explains that the predicted results are very close to the experimental results, which means a successful training process.

Figure 11 shows the calculated average absolute errors for ANN and QSAR studies for both training and test sets.

Figure 10: The regression plot of BPA network for training, test, and validation sets.

ISSN: 0975-8585

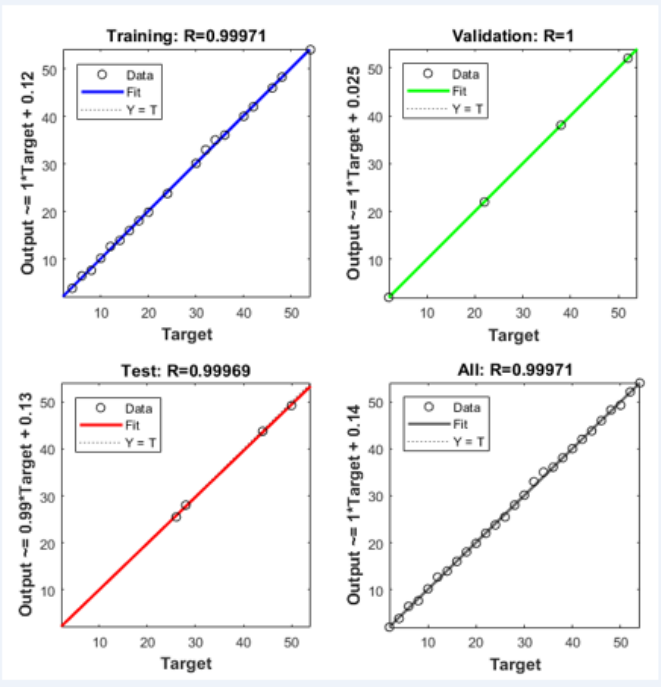
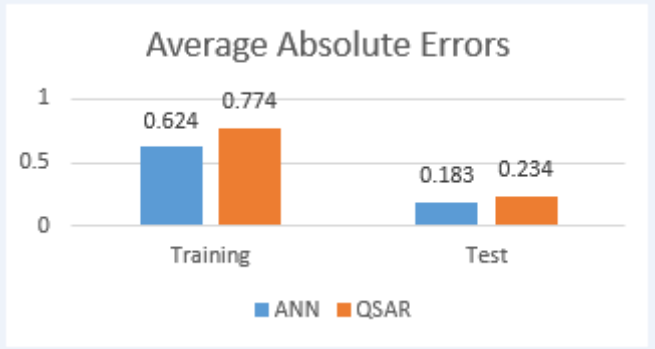


Figure 11: The calculated average absolute errors for ANN and QSAR studies for both training and test sets.



DISCUSSION

Quantitative Structure Activity Relationship

We can say that this QSAR study attempts to relate physical and chemical properties of a benzimidazole derivatives to their antiproliferative activities by simply using easily calculable descriptors and simple statistical methods to build a model which both describes the activity of the data set and can predict activities for further sets of untested compounds.

For a model to be predictive, it is recommended that Q2 and R2 values should be higher than 0.5 (Leach, 2001). It is clear from the table that both Q2 and R2 have values of 0,883 and 0,521 respectively. These results mean that the model has a high predictivity.

It was reported that for good predictability R2 – Q2 value should not be larger than 0.3 (Leach, 2001). R2 – Q2 values are calculated and it is equal 0.362. However, this value is not far from 0.3.

It can be concluded that the model has an acceptable values of parameters after validation process. Q2 and R2pred values larger than 0.5, which mean that the prediction ability of the model is high enough for this task.

A negative coefficient values of descriptors indicate that any increasing of these properties causes a decreasing in the antiproliferative activity. Supporting the biological activity requires decreasing these properties. Increasing the antitumor activity of molecules in the data set requires decreasing the atomic mass, atomic polarizability, van der Waals volume and the average connectivity indices.

It can understood that the molecular descriptors of drugs can strongly effect their antiproliferative activities. The selected four descriptors are the most important with the highest influence for the development of QSAR model and testing the MCF7 values with accepted results of outputs.

Types of descriptors analyzed in this research fails to take into account the three-dimensional nature of chemical structures which obviously play an important part in an antiproliferative activity. Steric, hydrophobic and electrostatic interactions are crucial to whether a molecule will interact optimally at its active site. It is logical to model these potential interactions to find the location in space around the molecule that are both acceptable and forbidden. This QSAR method does not take into account the 3-D structure of the molecules. So, further researches can be added to improve these results.

Artificial Neural Networks

It was found that increasing the number of hidden layers and neurons in the hidden layer does not correlate to a better accuracy each time. Our results match Bishop's (1995) study that found that one hidden layer is often enough.

Back propagation network achieved the best performance due to its ability to minimize errors and build a strong connections between inputs and outputs. Implementing different number of hidden layers, neurons in the hidden layers and types of transfer functions leads to different performance levels. Acceptable results cannot be achieved by testing one parameter only.

Network's performance can be developed by including many new parameters in the comparison. This step will generate more accurate and efficient networks to achieve medical, social, and economical benefits. A new parameters can be analyzed in the future and their impact on results can be tested. Testing all possible aspects will help in designing a better approach for neural networks to test antiproliferative activities for Benzimidazole derivatives.

ANN study is more suitable for antiproliferative activities testing than QSAR study. The results obtained here are very similar to previous researches. For example, both (Wesolowski, M. &Suchacz, B., 2012)[43] and (Bicciato, S., 2004)[44] concluded that ANN is a powerful tools for drug discovery and monitoring complex interactions between drugs, and the physiological system.

Sardari et al., 2014[45] applied ANN to correlate the chemical structures of different compounds with their pharmacological activities, and the result was an ANN model with high predictive aptitude that are helpful when studying new structures created by systematic modification of lead tuberculosis agents.
Ventura, et al., 2013[46] analyzed a series of hydrazide and isoniazid derivatives by generating Multiple Linear Regression models and Neural Network models. The neural network's ability to predict antitubercular behavior was better than that obtained with Multiple Linear Regression.

CONCLUSION

Designing safe and effective medications is the aim of drug discovery and development process to improve life's quality and to decrease toxicity. Developing drug design and synthesis procedures can be extensive and challenging, particularly if the association between input data and the biological activities is non-linear. Therefore, applying computational methods is a necessity in the drug discovery process to reduce the phase time, as well as to amplify the design of new molecules with better biological activities and minimal side effects for a disease specific target. These methods provides a possibility of replacing some old and less efficient drug designing procedures to save resources and reducing the economic and mental burden on health care institutes and governments as a result of long and expensive drug synthesis procedures .

This study aims to investigate ANN models and QSAR studies to test the antiproliferative activity on MCF-7 line of Benzimidazole derivatives.

The results of this research explain how ANN and QSAR are a promising tool for a nonlinear approximation with their abilities to solve a complicated technological and scientific problems and can be a suitable method to predict the biological activities of molecules. Moreover, they can aid in drug designing processes by rationalizing a large number of experimental observations and allow saving of both time and money in the drug design process.[47,48]

Residuals obtained by QSAR study and the best ANN model for both training and test sets were compared. ANN model had 0.624 and 0.183 absolute errors in training and test data respectively. However, QSAR study had 0.774 and 0.234 absolute errors in both training and test set respectively. The lower errors means higher accuracy and better performance. It can be concluded that ANN study is more suitable for antiproliferative activities testing than QSAR study.

Artificial Neural Networks (ANN) can be used for drug synthesis purposes. The idea behind these systems lies in parallel information analyzing in structures that mimic biological nervous system. ANNs have the ability to perform many functions: classification, fault detection, speech analysis, and processing of inaccurate inputs.[49]

In pharmaceutical research, ANN models applications include: compounds screening, quantitative structure activity relationship studies, receptor modeling, formulation development, pharmacokinetics, and processes involving complex mathematical modeling. Training process is based on weights of the data set, and is adjusted until the errors are minimized. It allows finding predictive, robust, and accurate models. Moreover, this quantitative approach helps rational drug planning generating more significant results.

REFERENCES

[1] M. Wang, X. Han, Z. Zhou. New substituted benzimidazole derivatives: A patent review (2013–2014). Expert Opin. Ther. Pat. 2015, 25, 595–612.

[2] K. J. Soderlind, B. Gorodetsky, A. K. Singh, N. R. Bachur, G. G. Miller, J. W. Lown. Bis-benzimidazole anticancer agents: Targeting human tumour helicases. Anticancer Drug Des. 1999, 14, 19–36.

[3] E. Mentese, H. Bektas, S. Ulker, O. Bekircan, B. Kahveci. Microwave-assisted synthesis of new benzimidazole derivatives with lipase inhibition activity. J. Enzyme Inhib. Med. Chem. 2014, 29, 64–68.

[4] Y. Ke, X. Zhi, X. Yu, G. Ding, C. Yang, H. Xu. Combinatorial synthesis of benzimidazole-azo-phenol derivatives as antifungal agents. Comb. Chem. High Throughput Screen 2014, 17, 89–95.

[5] R. V. Shingalapur, K. M. Hosamani, R. S. Keri, M. H. Hugar. Derivatives of benzimidazolepharmacophore: Synthesis, anticonvulsant, antidiabetic and DNA cleavage studies. Eur. J. Med. Chem. 2010, 45, 1753–1759.

[6] K. C. S. Achar, K. M. Hosamani, H. R. Seetharamareddy. In vivo analgesic and anti-inflammatory activities of newly synthesized benzimidazole derivatives. Eur. J. Med. Chem. 2010, 45, 2048–2054.

[7] P. A. Datar, S. A. Limaye. Design and Synthesis of Mannich bases as Benzimidazole Derivatives as Analgesic Agents. Anti-Inflamm. Anti-Allergy Agents Med. Chem. 2015, 14, 35–46.

[8] H. S. Cho, P. F. Lopes. Injectable Formulation of a Macrocyclic Lactone and Levamisole. (2011): United States Patent NO. 20130090296. Available online: http://patentscope.wipo.int/search/en/detail.jsf?docId=WO2011161209&recNum=34&maxRec=497&office=&prevFilter=&sortOption=&queryString=DORAMECTIN&tab=PCTDescription (accessed on 10 June 2015).

[9] F. Grasso, T. Frisan. Bacterial Genotoxins: Merging the DNA Damage Response into Infection Biology. Biomolecules, 2015, 5(3), 1762-1782.

[10] F. Soualmia, S. Belaidi, H. Belaidi, N. Tchouar, Z. Almi. Quantitative Structure Anti-Proliferative Activity Against HEPG2 and SW1116 Relationships in a Series of PyrazineDerivatives.JournalofBionanoscience Vol. 11, 1–8, 2017.

[11] C. Liang,& L. Peng. An automated diagnosis system of liver disease using artificial immune and genetic algorithms. Journal of medical systems, 2013, 37(2), 1-10.

[12] B. Alić, L. Gurbeta, A. Osmanovic and A. Badnjević, "Machine learning techniques for classification of diabetes and cardiovascular diseases," 2017 6th Mediterranean Conference on Embedded Computing (MECO), Bar, Montenegro, 2017, pp. 1-4. doi: 10.1109/MECO.2017.7977152
[13] A. Badnjevic, M. Cifrek, L. Gurbeta, E. Feric. Classification of Chronic Obstructive Pulmonary Diseases Based on Neuro-Fuzzy Software. in the Book: „Chronic Obstructive Pulmonary Disease (COPD): Clinical Symptoms, Emerging Treatment Strategies and Impact on Quality of Life“, 2016, Nova Science Publisher, ISBN: 978-1-63484-500-7, 2016.
[14] A. Catic, L. Gurbeta, A. Kurtovic-Kozaric, S. Mehmedbasic, A. Badnjevic. Application of Neural Networks for classification of Patau, Edwards, Down, Turner and Klinefelter Syndrome based on first trimester maternal serum screening data, ultrasonographic findings and patient demographics. BMC Medical Genomics (2018) 11:19, DOI: 10.1186/s12920-018-0333-2.
[15] A. Badnjevic, L. Gurbeta, M. Cifrek, D. Marjanovic. Diagnostic of Asthma Using Fuzzy Rules Implemented in Accordance with International Guidelines and Physicians Experience. IEEE 39th International convention on information and communication technology, electronics and microelectronics (MIPRO), 30. May to 03. June 2016. Opatija, Croatia.
[16] A. Badnjevic, L. Gurbeta, M. Cifrek, D. Marjanovic. Classification of Asthma Using Artificial Neural Network“, IEEE 39th International convention on information and communication technology, electronics and microelectronics (MIPRO), 30. May to 03. June 2016. Opatija, Croatia.
[17] A. Aljovic, A. Badnjevic, L. Gurbeta. Artificial Neural Networks in the Discrimination of Alzheimer's disease Using Biomarkers Data, IEEE 5th Mediterranean Conference on Embedded Computing (MECO), 12 – 16 June 2016, Bar, Montenegro.
[18] B. Alic, D. Sejdinovic, L. Gurbeta, A. Badnjevic. Classification of Stress Recognition using Artificial Neural Network. IEEE 5th Mediterranean Conference on Embedded Computing (MECO), 12 – 16 June 2016, Bar, Montenegro.
[19] S. Halilovic, H. Avdihodžić, L. Gurbeta. Micro cell culture analog Apparatus (µCCA) output prediction using Artificial Neural Network, 5th Mediterranean Conference on Embedded Computing (IEEE MECO 2016), Bar, Montenegro;
[20] B. Alic, L. Gurbeta, A. Badnjevic. Classification of Metabolic Syndrome patients using implemented Expert System, CMBEBIH 2017. IFMBE Proceedings, vol 62. pp 601-607, Springer, Singapore. DOI: 10.1007/978-981-10-4166-2_91
[21] D. Sejdinovic, L. Gurbeta, A. Badnjevic, M. Malenica, T. Dujic, A. Causevic, T. Bego, L. Divovic, Classification of prediabetes and Type 2 Diabetes using Artificial Neural Network, CMBEBIH 2017. IFMBE Proceedings, vol 62. pp 685-689, Springer, Singapore. DOI: 10.1007/978-981-10-4166-2_103.
[22] E. Veljovic, S. Spirtovic – Halilovic, S. Muratovic, A. Osmanovic, A. Badnjevic. Artificial Neural Network and Docking Study in Design and Synthesis of Xanthenes as Antimicrobial Agents, CMBEBIH 2017. IFMBE Proceedings, vol 62. pp 617-626, Springer, Singapore. DOI: 10.1007/978-981-10-4166-2_93.
[23] T. Mitchell, 1997, Machine Learning. New York: McGraw-Hill.
[24] G. Vitale, A. Carta, M. Loriga, G. Paglietti, P. La Colla, B. Busonera, D. Collu and R. Loddo, 2-Arylbenzimidazoles as Antiviral and Antiproliferative Agents-Part 1, Medicinal Chemistry, 2008, 4, 605-615.
[25] L. Garuti, M. Roberti, M. Malagoli, T. Rossi, M. Castelli, Synthesis and Antiproliferative Activity of Some Benzimidazole4,7-dione Derivatives, Bioorganic & Medicinal Chemistry Letters 10, 2000, 2193-2195.
[26] K. Starþevic, M. Kralj, K. Ester, I. S. M. Grce, K. Pavelic, G. Karminski – Zamola, Synthesis, Antiviral and Antitumor Activity of 2-substituted-5-amidinobenzimidazoles, Bioorg Med Chem. 2007 Jul 1;15(13):4419-26.
[27] ChemAxon Marvin 6.1.0 program, Budapest, Hungary, 2013. www.chemaxon.com.
[28] J. J. P. Stewart. Optimization of parameters for semiempirical methods I.Method. J ComputChem 1989; 10: 209–20. Stewart JJP. Optimization of parameters for semiempirical methods II. Applications. J ComputChem 1989; 10:221–64.
[29] Gaussian 98 (RevisionA.7) Frisch MJ etal. Gaussian, Inc, Pittsburgh PA, 1998.
[30] Chem3D Ultra7.0, Cambridge Soft Corporation, 100 Cambridge Park Dr, Cam-bridge, USA, 2001, http://www.cambridgesoft.com.
[31] C. C. J. Roothan. New developments in molecular orbital theory. Rev Mod Phys 1951, 23, 69–89.
[32] E-Dragon. http://www.vcclab.org/lab/edragon/.
[33] L. Eriksson, J. Jaworska, A. P. Worth, M. T. Cronin, R. M. Mcdowell, P. Gramatica. Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs. Environ Health Perspect 2003;111:1361–75.

[34] AB Umetrics. SIMCA Program,Version 14.1. Umea, Sweden, 2015, www.umetrics.com
[35] T. Mehmood, H. Martens, S. Sæbø, J. Warringer, L. Snipen. A partial leastsquares algorithm for parsimonious variable selection. Algorithms MolBiol 2011; 6:27–38.
[36] P. K. Ojha, K. Roy. Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection. ChemometrIntell Lab 2011, 109:146–61.
[37] A. Tropsha. Best practices for QSAR model development, validation and exploration. MolInf 2010; 29:476–88.
[38] L. Eriksson, E. Johansson, N. KettanehWold, J. Trygg, J. Wikstrom, S. Wold. Multi and mega variate data analysis. Basic principles and applications I. 2nd ed. Umeå: Umetrics Academy, 2006.
[39] K. Roy, S. Kar, P. Ambure. On a simple approach for determining applicability do-main of QSAR models. ChemometrIntell Lab 2015, 145, 22–9.
[40] J. Caballero, F. M. Zampini, S. collina, M. Fernandez. Chem. Biol. Drug. Des. 2007, 69, 48.
[41] C. Devika, S. M. Amita, Case Study on Classification of Glass using Neural Network Tool in MATLAB. International Journal of Computer Applications® (IJCA), 2014, (0975 – 8887), International Conference on Advances in Computer Engineering & Applications (ICACEA) at IMSEC,GZB.
[42] N. Khan, D. Gaurav, T. Kandl. Performance evaluation of LevenbergMarquardt technique in error reduction for diabetes condition classification. Procedia Computer Science, 2013, 18, 2629-2637.
[43] M. Wesolowski, B. Suchacz, Artificial Neural Networks: Theoretical Background and Pharmaceutical Applications: A Review. J. AOAC Int., 2012, 95(3), 652-668.
[44] S. Bicciato, Artificial neural network technologies to identify biomarkers for therapeutic intervention. Curr. Opin. Mol. Ther., 2004, 6(6), 616-623.
[45] S. Sardari, H. Kohanzad, G. Ghavami. Artificial neural network modeling of antimycobacterial chemical space to introduce efficient descriptors employed for drug design. Chemometr. Intell. Lab. Syst., 2014, 130, 151-158.
[46] C. Ventura, D. Latino, F. Martins. Comparison of Multiple Linear Regressions and Neural Networks based QSAR models for the design of new antitubercular compounds. Eur. J. Med. Chem., 2013, 70, 831-845.
[47] L. Eriksson, J. Jaworska, A. P. Worth, M. T. Cronin, R. M. Mcdowell, P. Gramatica. Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs. Environ Health Perspect 2003, 111, 1361–75.
[48] G. A. Carpenter, S.Grossberg. A self-organizing neural network for supervised learning, recognition and prediction. IEEE Communications Magazine, 1992, 38–49.
[49] S. Singh, T. V. Murthy. Neural network-based sensor fault accommodation in flight control system. Journal of Intelligent Systems, 2013, 22(3), 317-333.