HIGHER EFFICIENCY IN PREDICTION OF TIBO ACTIVITY BY EVOLUTIONARY NEURAL NETWORK IN COMPARISON WITH MULTIPLE LINEAR REGRESSION AND MULTILAYERED PERCEPTRONS

Abhik Seal[1*]

Doeacc Society Kolkata(Jadavpur University Campus) Kolkata-700032 Department of Bioinformatics.
Email:abhik1368@gmail.com
*corresponding author

Abstract

The treatment of acquired immunodeficiency syndrome (AIDS) is a challenging medical problem. TIBO is a non-nucleoside reverse transcriptase inhibitor, which binds noncompetitively to the hydrophobic pocket on the p66 subunit of RT enzyme. We used a dataset consisting of physicochemical properties and reverse transcriptase inhibitor activities of 88 set of 4,5,6,7-tetrahydro-y-imidazo-[4,5,1-jk][1,4]-x-benzodiazepin-2-(1h)one derivatives that are variously substituted by halogens ,alkyl groups. The dataset was taken from the BIOBYTE database at (www.davidhoekman.com). The concentration of the compound leading to 50% effect has been measured and expressed as IC_{50}. The logarithm of the inverse of this parameter has been used as biological end points (log 1/C) in the QSAR studies. The evolutionary neural network (ENN) is a new system for modeling multivariate data. The strengths of ENN's are that they can extract insignificant predictors, choose the size of the hidden layers and nodes and fine tune the parameters needed in training the network. We have used an ENN to predict the biological activities of Reverse Transcriptase Inhibitors. We have found out that Evolutionary Neural networks are better predictor of activity values than Multiple linear regression and Multilayered Perceptrons. We have calculated the correlation coefficient of each of the methods where we have found ENNs are the best.

Keywords: Evolutionary Neural Networks, Multiple Linear Regression, QSAR, Genetic Algorithm ,Reverse Transcriptase Inhibitor.

INTRODUCTION

The Quantitative structure activity relationships (QSAR) are certainly a major factor in contemporary drug design. The term QSAR is applied to the methods which correlate molecular structure to some kind of invitro or invivo biological property. QSAR was first developed by Hansch and Fujita [1] which has been invaluable for understanding of drug structure –activity relationship for lead discovery and optimization.QSAR related biological activity for members of a congeneric series with substituted parameters. Many methods are have been used to map the molecular descriptor to properties. The majority are regression methods of which multiple linear regression was first used [1].Regression methods[11] attempt to fit the a specific function with parameters to a set of data .They usually do this using the gradient descent methods such as the method of least squares which finds the best set of parameters that minimize the sum of the squares of the errors between the measured values of dependent variables .Some QSAR problems have linear response surfaces while most QSAR problems involve at least some degree of nonlinearity. Initially this was tackled using the bilinear, exponential or polynomial regression together with the cross terms to find the relationship. This has created problems because the true and complex nature relationship was often not found.

Recently there has been a growing interest in neural networks in the field of QSAR. The main strength of neural networks is that they can be trained to perform non linear mapping of the physicochemical parameters on the corresponding biological activity [2-5].The class of neural networks often used for empirical structure property modeling ,the back propagation neural net which learns in a similar way the brain learns. However, neural networks left some structure activity mapping problem unsolved and introduced few extra problems[6].The most major problem is optimal construction of the network and the parameters needed in training them has to be determined by trial and error methods. The problem is solved by the use of Evolutionary Neural Networks which can build Neural Network models with right hidden units and number of nodes in the hidden units.In the paper we have used two different methods of predicting the activity the traditional regression analysis and Evolutionary Neural Networks[7].We have mainly used correlation coefficient (R^2) to compare the ability of the models to predict biological activities.

Materials and Methods

1. Dataset

The dataset for the project is usually obtained from the http://davidhoekman.com with the help of BioByte's BIOLOOM software[8].BioByte's new Bio-Loom program weaves several different threads of data into a cohesive whole, Bio-Loom still calculates hydrophobic and molecular refractivity parameters via CLOGP & CMR, calculations which have been the world standard for decades, but it now has the ability to access Bio Byte's entire Thor Masterfile database, which includes over 60,000 measured log P and log D values (in many solvent systems), as well as 14,000 pKas, including associated references. The biological activity searches are now much more powerful than before. The entire QSAR Database is now available for easy and flexible searching online, accessible only through the Bio-Loom program. The dataset used in the paper is set no **5250** of 4,5,6,7-tetrahydro-y-imidazo-[4,5,1-jk][1,4] -x-benzodiazepin-2-(1h)one(1) derivatives with 88 molecules and among them 81 molecules are studied and rest were outliers .A QSAR equation is developed which is given below as Eq(1) with R^2 of 0.87 and σ 0.532 .
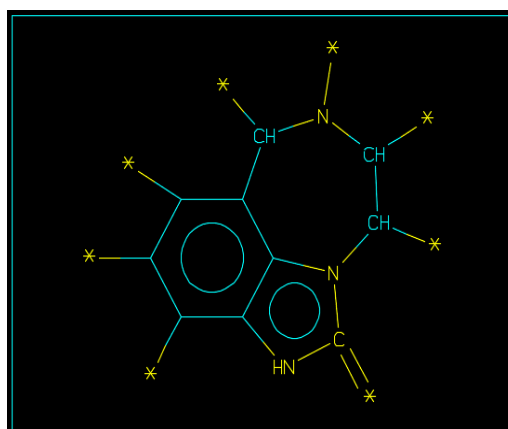
Fig1. 4,5,6,7-tetrahydro-y-imidazo-[4,5,1-jk][1,4]-x-benzodiazepin-2-(1h)one(1) derivatives

The dataset is given in table 1 in appendix along with the outliers row coloured in blue :

The Qsar Equation obtained from the database is as follows:

$$\text{Log } 1/C = 0.4 \text{ Log P} + 1.1\beta\ 1\text{-}8 + 0.9 \text{ I-DMA} + 1.5 \text{ I}\sigma + 1.9 \qquad (1)$$

The equation (1) was generated using mainly four descriptors i.e Clog P (the calculated octanol/water partition coefficient of the molecule), B1 (8-x) (Verloop's sterimol parameter (width parameter of the X substituent at the position 8) and I-S and I-DMA which are indicator variables.The Eq(1) is based on typical regression equation .It is a simple way of determining though it can only model linear relationships. The regression equation is usually estimated by using the least squares fit method.

2. Evolutionary Neural Networks.

Neural Networks are systems which model the estimation in a different approach. Many types of neural network exist but here we focus on the feed forward neural networks. In neural networks at first step a data is taken which is to be trained and then a test set is taken to validate the data whether the neural net performed well or not.Here the dependent variables i.e the biological activity of the molecules are set as targets for the target layer and independent variables are set as inputs in the input layer. A layer with appropriate number of nodes is taken and is placed between the target and the input layer this layer is called the hidden layer. These layers are fully connected to each other with weights at each connection. The networks are then trained using the back propagation algorithm.

$$E = \sum_j (O_j - t_j)^2 \qquad (2)$$

The algorithm is runned until the Error (E) becomes small enough. The user can choice the value of E as desired. The use of test set is always important when using neural network by which we can monitor the performance of the networks. If the network performs well in unseen data then we can conclude that the network has learned the training data well. But, if we continue training the network then it would learn the training data to well which means it will be not able to predict the test data well. The network is nonlinear because the neurons squeezes the summations through an activation function. An activation function for a backpropagation net should have several important characteristics: i.e should be continuous, differentiable, and monotonically nondecreasing. One of the most typical activation functions is the binary sigmoid function, which has range of (0,1) and is defined as :

$$f_1(x) = \frac{1}{1 + \exp(-x)}$$

Another common activation function is bipolar sigmoid, which has range of (- 1, 1) and is defined as

$$f_2(x) = \frac{1}{1 + \exp(x)} - 1$$

The problem in using the neural network is that the optimal structure of the network is not known. The structure is usually determined by the trial and error method which consumes a lot of time and is unreliable. The evolutionary neural network (ENN) solves the problem and optimizes the size of the hidden layer. However there are other methods of optimizing networks[9-10] but here we are concerned about ENN.

In the paper we used Pythia (evolutionary optimizer) which is obtained from http://www.runtime.org/pythia.html . Pythia features Back propagation Networks. The network parameters are initially set to random value. During the "Training Phase", the actual output of the network is compared with the desired output and the error propagated back toward the input of the network. A special feature of Pythia is the Evolutionary Optimizer that automatically generates suitable networks for a given training data set. It uses evolutionary algorithms for the selection and generation of neural networks. In this paper a initial neural network is assigned with a total of six neurons i.e 4-5-1 , 4 is the number of inputs(indicates number of descriptors used) 5 is the number of neurons in the hidden layer and 1 is the output. Starting with the evolutionary optimizer an initial population of 100 was chosen and was runned for 500 epochs with mutation rate of 0.05 and crossover rate of 0.20. A Neural Network expects any input and output value to be between 0 and 1.The learning rate was set to 0.5. Therefore the pattern sets must be normalized before being processed by the network.

The Evolutionary Optimizer initially creates a generation containing 50 randomly created networks. Each Network within this generation will be trained shortly and its fitness determined according to the parameters in Goals to achieve. Then a new generation of networks will be created from the old one according to the procedure two "parent" networks will be chosen out of the old generation. The selection algorithm will choose networks with a high fitness by a higher probability. Two "children" networks will be created from the two "parent" networks. With the probability of 0.2, the two "children" networks will be crossed over. This means they will swap level with each other. The "children" now will be mutated with a probability of 0.05. Mutation means insertion or deletion of a level, insertion or deletion of a neuron into a level or change of weights.The two "parent" networks will now be checked if they belong to the 10 "fittest" of the old generation.   If they do they will be mutated and rolled over into the new generation. The selection continues until the new generation has 50 members too. After completion the new generation will be evaluated.

For Prediction of the datasets we splited the datasets into three parts with 35,44 and 66 datasets. We have used 35 as training set and tested the Evolutionary neural network model with 44 and 66 datasets. Again we selected 44 as train set and tested the Evolutionary Neural Model with 35 and 66 dataset also train 66 set and test with 35 and 44 as test sets. We also fitted three regression models on datasets 35,66 and 66 and one neural network model based on the complete data.

Results

When applying regression analysis to a data set a number of important factors must be considered. One of the most important is called multicollinearity. This means that some of the independent variables correlate strongly with one another. Regression provides information about the linear relationship between the actual value and the prediction. The correlation coefficient must be sufficiently high if we are to assert that a strong linear relationship between the measured variables exists. The Dataset we have found in the biobyte database contains values of variable descriptors but four variables correlate very strongly and the QSAR equation(1) formed with R^2 of 0.87 and σ 0.532 with 81 number of molecules. Regression models on datasets 35,44 and 66 are generated using Molegro data modeler[12] given below.

Dataset 35:

$$\text{Log } 1/C = 0.65678 \text{ Log P} + 0.448448 \text{ β 1-8} + 1.75995 \text{ I-DMA} + 0.814527 \text{ Iσ} + 1.629$$

$$R^2 = 0.867, \text{ Mean Square Error (MSE)} = 0.308141$$

Dataset 44 :

$$\text{Log } 1/C = 0.596728 \text{ Log } P + 1.24356 \text{ } \beta \text{ } 1\text{-}8 + 1.33031 \text{ I-DMA} + 0.4846 \text{ I}\sigma + 1.16972$$

$$R^2 = 0.7268 \text{ ,MSE} = 0.4488$$

Dataset 66:

$$\text{Log } 1/C = 0.4790 \text{Log } P + 1.3707 \text{ } \beta \text{ } 1\text{-}8 + 1.39151 \text{ I-DMA} + 0.610441 \text{ I}\sigma + 1.3867$$

$$R^2 = 0.8139, \text{ MSE} = 0.3623$$

The optimized evolutionary Neural network models generated by pythia for datasets 35,44,66 are given below at first 4 represent number of inputs followed by neurons in the hidden layer .1 represents the output of the network. Initially a 4-5-1 network architecture was choosen as the seed network, weights was choosen randomly and from it 100 generations where produced with more than 2 hidden layered models. No bias was used in the models that were predicted the learning rate was assumed 0.5 , no momentum was used. Also each network containing the number of neurons is also shown in the table below .

Dataset 35 Architecture:

Neural network Architecture	Number of neurons
4-5-6-7-1	18
4-5-6-1	11
4-4-5-5-1	14
4-5-4-6-1	15
4-5-7-1	12
4-5-1	5

Dataset 44 Architecture:

Neural Network Architecture	Number of Neurons
4-7-2-6-1	15
4-4-4-1	8
4-7-2-7-6-1	22
4-7-6-1	13

Dataset 66 Architecture:

Neural Network Architecture	Number of Neurons
4-5-4-6-1	15
4-5-4-5-6-1	20
4-5-4-4-6-1	19
4-5-4-1	9
4-4-4-1	8

The 81 molecule dataset is divided into 3 groups i.e 35, 44, 66 groups. The 35 train set is tested 44 and 66 test set and the 0Squared Correlation coefficient(R^2)with 0.87 and 0.88 are predicted with 44 test set and R^2 0.912and 0.906 with 66 test set . Train set 44 is tested with 35 and 66 test sets and the predicted R^2 0.967 and 0.963 with 35 test set and R^2 0.897 and 0.894 with 66 test set. Train set 66 is tested with 35 and 44 test sets and predicted R^2 with 0.967 and 0.953 with 35 test set and R^2 0.9 and 0.88 with 44 test set respectively. The details of all architectures obtained for each of the three sets is given table (2a, 2b, 2c, 2d, 2e, 2f) in the appendix section.

We have also made a ENN models with the whole dataset i.e number of molecules is 81 and it predicted R^2 0.946 and 0.944 . The table 3 in appendix gives the details of the architectures of models.

Conclusion

In this paper we have suggested that evolutionary neural networks have advantages over classical QSAR methods, and that they will find their own place in the arsenal of medical and computational chemistry in the future. With the advancement of technology newer drug designing methods have been and also more drugs are being developed evolved .The generation of Qsar equations are very important in predicting the biological activity of a molecule .But when regression methods are performed on the dataset then the results usually differs from the neural network approach since neural network regression is a nonlinear method of predicting and better predictions are performed . But since Neural network hidden layers are very difficult to predict so Evolutionary strategy have been implemented to find the most suitable layers and the number of neurons .

So while designing drugs in predicting biological activity if ENN methods are performed then better results are expected than normal linear regression methods. These methods could be applied to newer groups of drugs like Zidovudine, Didanosine,Stavudine,Lamivudine,Nevirapine,Delavirdine ,Etravirine .Also now more newer drugs are being developed and much computer power are provided so it is possible now to design a drug and in predicting its activity correctly, it would take much less time . Other than Neural networks and multilayered perceptrons radial bais functions could be used which could predict much better than neural networks . Both Evolutionary strategy and RBF technique could be applied in traning the datasets and good results good be obtained.

In future these studies could be performed and one can expect much better results is proper parameters are provided.

References

[1] Hansch,C. and Fujita,T ; J.Am.Chem.Soc.1964,vol 86,pp1616-1626.

[2] Salt, D.W, Yildiz, N, Livingstone, D.J and Tinsley, C.J; The use of artificial neural networks in QSAR, Pestic.sci,vol 36,(1992) pp161-170.

[3] Tetko, I.V, Luik, A.I and Poda, G.I .Application of neural networks in structure activity relationships of a small number of molecules. J. Med. chem.(1993) Vol 36 pp 811-814.

[4] Andrea, T.A, Kalayeh, H. Application of Neural networks in quantitative structure activity relationships of dihydrofolate reductase inhibitors. J.Med Chem,(1991) vol34 pp 2834-2836.

[5] Gasteiger, J and Zupan, J. Neural networks in chemistry. Angew .Chem. Int. Ed, Engl (1993) vol 32 pp 503-527.

[6] Manallack, D.T and Livingstone,D.J.Neural networks in drug discovery:Have they lived up to their promise ?, Eur .J.Med.Chem,(1999) , vol 34 pp 195-228.

[7] Weekes, D and Fogel, G.B.Evolutionary optimization, backpropagation, and data preparation issues in QSAR modeling of HIV inhibition by HEPT derivatives. Biosystems(Nov 2003), Volume 72(1-2), pp 149-158.

[8] Biobyte's Bioloom. http://www.biobyte.com/bb/prod/bioloom.html

[9] Hagiwara,M., A Simple and Effective method for Removal of Hidden Units and Weights,1994 Neurocomputing,Vol 6(2), pp207-218.

[10] Weigend,A.S., Huberman,B.A. and Rumelhart,D.E.,Predicting the Future: A Connectionist Approach, Internation Journal of Neural Systems,1990, Vol 1(3),pp 193-209.

[11] Montgomery,D.C. and Peck,E.A., Introduction to linear regression analysis, John Wiley & Sons, New York 1982, pp. 34-38.

[12] Molegro bioinformatics solutions.Molegro data modeller. http://www.molegro.com/mdm-product.php October 04,2007

Appendix

OBSERVED Log(1/IC50)	PREDICTED Log(1/IC50)	DEVIATION	CLOGP	B1-8	I-S	I-DMA	CMR	MGVOL	PI-X	MR-X	L-8	I-5	PI-Y	MR-Y	I-8	L-6	B1-6	B5-6	MR-6	PI-6
7.36	6.78	0.58	3.53	1	1	1	9	2.29	0	0.31	2.06	1	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
7.47	7.06	0.41	4.24	1	1	1	9.5	2.41	0.71	0.81	2.06	1	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
8.37	7.94	0.43	4.24	1.8	1	1	9.5	2.41	0.71	0.81	3.52	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
8.24	7.22	1.02	3.67	1.35	1	1	9.02	2.3	0.14	0.3	2.65	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
8.3	7.77	0.53	4.09	1.7	1	1	10.27	2.59	0.61	1.59	4.3	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
7.47	7.14	0.33	3.45	1.35	1	1	9.62	2.48	-0.02	0.99	3.98	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
7.02	7.34	-0.32	3.98	1.35	1	1	10.09	2.63	0.38	1.45	4.8	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
5.94	6	-0.06	3.76	1.6	0	1	8.63	2.33	-0.57	0.84	4.23	1	-0.11	1.06	1	6.39	1.52	4.82	2.25	2.01
7.25	7.22	0.03	2.96	1.6	1	1	9.48	2.44	-0.57	0.84	4.23	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
6.73	7.19	-0.46	2.88	1.6	1	1	9.5	2.44	-0.65	0.89	3.53	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
5.2	5.44	-0.24	2.59	1.5	0	1	9.02	2.44	-1.49	1.19	4.06	1	-0.11	1.06	1	6.39	1.52	4.82	2.25	2.01
7.33	6.83	0.5	4.91	1.95	0	1	8.93	2.36	0.86	1.09	3.82	1	-0.11	1.06	1	6.39	1.52	4.82	2.25	2.01
8.52	8.16	0.36	4.39	1.95	1	1	9.78	2.46	0.86	1.09	3.82	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
7.06	7.15	-0.09	5.17	2.15	0	1	9.45	2.44	1.12	1.6	4.23	1	-0.11	1.06	1	6.39	1.52	4.82	2.25	2.01
7.32	8.48	-1.16	4.65	2.15	1	1	10.31	2.54	1.12	1.6	4.23	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
6.36	6.14	0.22	4.11	1.6	0	1	8.95	2.38	0.4	1.16	4.66	1	-0.11	1.06	1	6.39	1.52	4.82	2.25	2.01
7.53	7.55	-0.02	3.8	1.6	1	1	9.81	2.48	0.4	1.16	4.66	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
6	6.14	-0.14	4.34	1.52	0	1	8.61	2.32	0.56	0.77	2.87	1	-0.11	1.06	1	6.39	1.52	4.82	2.25	2.01
7.87	7.55	0.32	4.03	1.52	1	1	9.47	2.43	0.56	0.77	2.87	1	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
4.48	4.24	0.24	3.3	1	0	0	8.18	2.15	-0.28	0.94	2.06	1	-0.11	1.06	0	5.14	1.52	4.36	1.82	1.48
3.07	4.07	-1	1.88	1.35	0	0	7.94	2.07	-1.23	0.75	2.78	1	-0.11	1.06	1	5.14	1.52	4.36	1.82	1.48
5.18	4.61	0.57	3.27	1.35	0	0	8.87	2.36	0.18	1.76	3.53	1	-0.11	1.06	1	5.14	1.52	4.36	1.82	1.48
4.22	3.69	0.53	1.88	1	0	0	7.94	2.07	-1.23	0.75	2.06	1	-0.11	1.06	0	5.14	1.52	4.36	1.82	1.48
5.18	4.23	0.95	3.27	1	0	0	8.87	2.36	0.18	1.76	2.06	1	-0.11	1.06	0	5.14	1.52	4.36	1.82	1.48
3.8	3.51	0.29	1.42	1	0	0	8.44	2.23	-0.97	1.7	2.06	1	-0.11	1.06	0	4.14	1.55	3.24	1.35	1.14
5.61	5.47	0.14	2.55	1	1	0	9.04	2.25	-0.28	0.94	2.06	1	0.95	1.69	0	5.14	1.52	4.36	1.82	1.48
7.6	6.84	0.76	3.67	1	1	1	9.02	2.3	0.14	0.3	2.06	1	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
5.23	5.86	-0.63	5.09	1	0	1	8.66	2.37	0.88	0.71	2.06	1	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01
6.31	7.12	-0.81	4.41	1	1	1	9.51	2.48	0.88	0.71	2.06	1	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
6.5	4.97	1.53	5.21	1	0	0	9.54	2.6	0.56	0.77	2.06	1	-0.11	1.06	0	.	.	.	.	.
5.18	5.32	-0.14	3.7	1	0	1	8.77	2.38	-0.02	0.99	2.06	1	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01

OBSERVED Log(1/IC50)	PREDICTED Log(1/IC50)	DEVIATION	CLOGP	B1-8	I-S	I-DMA	CMR	MGVOL	PI-X	MR-X	L-8	I-5	PI-Y	MR-Y	I-8	L-6	B1-6	B5-6	MR-6	PI-6
5.33	6.68	-1.35	3.26	1	1	1	9.62	2.48	-0.02	0.99	2.06	1	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
7.6	7.29	0.31	4.84	1	1	1	9.99	2.53	1.42	1.31	2.06	1	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
5.97	7.12	-1.15	4.39	1	1	1	9.78	2.46	0.86	1.09	2.06	1	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
4.15	4.08	0.07	2.91	1	0	0	7.22	1.9	0	0.31	2.06	1	-0.11	1.06	0	5.11	1.52	3.78	1.45	1.1
4.33	4.24	0.09	3.31	1	0	0	7.68	2.04	0	0.31	2.06	1	-0.11	1.06	0	.	.	.	.	.
3.07	3.76	-0.69	2.08	1	0	0	7.43	2.02	0	0.31	2.06	1	-0.11	1.06	0	5.98	1.52	4.4	1.65	-0.3
3.24	4.02	-0.78	2.74	1	0	0	7.04	1.86	0	0.31	2.06	1	-0.11	1.06	0	3.99	1.52	4.49	1.29	0.21
3.97	4.14	-0.17	3.05	1	0	0	8.04	2.09	0	0.31	2.06	1	-0.11	1.06	0	.	.	.	2.28	1.25
4.18	4.08	0.1	2.91	1	0	0	7.22	1.9	0	0.31	2.06	1	-0.11	1.06	0	5.11	1.52	3.78	1.45	1.1
4.3	4.21	0.09	3.24	1	0	0	7.68	2.04	0	0.31	2.06	1	-0.11	1.06	0	6.35	1.52	4.55	1.91	1.61
4.05	4.12	-0.07	3.01	1	0	0	7.25	1.94	0	0.31	2.06	1	-0.11	1.06	0	4.92	1.52	3.49	1.5	1.55
4.72	4.24	0.48	3.31	1	0	0	7.68	2.04	0	0.31	2.06	1	-0.11	1.06	0	.	.	.	.	.
4.36	4.16	0.2	3.11	1	0	0	7.57	1.97	0	0.31	2.06	1	-0.11	1.06	0	5.14	1.52	4.36	1.82	1.48
4.24	4.22	0.02	3.25	1	0	0	7.68	2.04	0	0.31	2.06	1	-0.11	1.06	0	5.89	1.52	4.82	1.91	1.61
4.46	4.22	0.24	3.25	1	0	0	7.68	2.04	0	0.31	2.06	1	-0.11	1.06	0	5.89	1.52	4.82	1.91	1.61
4	4.33	-0.33	3.54	1	0	0	7.71	2.08	0	0.31	2.06	1	-0.11	1.06	0	6.17	1.52	4.54	1.96	2.13
4.9	5.37	-0.47	3.84	1	0	1	8.15	2.18	0	0.31	2.06	1	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01
4.21	4.35	-0.14	3.6	1	0	0	8	2.07	0	0.31	2.06	1	-0.11	1.06	0	4.16	1.52	4.62	.	.
4.54	4.37	0.17	3.65	1	0	0	8.15	2.18	0	0.31	2.06	1	-0.11	1.06	0	.	.	.	.	.
4.66	5.37	-0.71	3.84	1	0	1	8.15	2.18	0	0.31	2.06	1	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01
5.4	5.37	0.03	3.84	1	0	1	8.15	2.18	0	0.31	2.06	1	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01
4.43	4.37	0.06	3.65	1	0	0	8.15	2.18	0	0.31	2.06	1	-0.11	1.06	0	.	.	.	.	.
3.91	4.7	-0.79	4.51	1	0	0	9.81	2.51	0	0.31	2.06	1	-0.11	1.06	0	.	.	.	.	.
4.15	4.28	-0.13	3.41	1	0	0	8.28	2.14	0	0.31	2.06	1	-0.11	1.06	0	.	.	.	.	.
7.34	7.67	-0.33	3.54	1.8	1	1	9.03	2.27	0.71	0.81	3.52	0	0.39	1.23	1	6.39	1.52	4.82	2.25	2.01
6.8	6.79	0.01	3.54	1	1	1	9.03	2.27	0.71	0.81	2.06	0	0.39	1.23	0	6.39	1.52	4.82	2.25	2.01
4.64	4.57	0.07	4.17	1	0	0	8.61	2.18	0	0.31	2.06	0	.	.	0	.	.	.	.	.
4.5	4.12	0.38	3	1	0	0	8.54	2.14	0	0.31	2.06	0	0.95	1.69	0	.	.	.	.	.
6.17	5.93	0.24	3.72	1	1	0	9.03	2.27	0.71	0.81	2.06	0	0.95	1.69	0	.	.	.	.	.
5.66	5.85	-0.19	3.52	1	1	0	8.92	2.2	0.71	0.81	2.06	0	0.95	1.69	0	5.14	1.52	4.36	1.82	1.48
4.13	4.48	-0.35	3.93	1	0	0	8.17	2.22	0	0.31	2.06	0	0.86	1.99	0	4.92	1.52	3.49	1.5	1.55

OBSERVED Log(1/IC50)	PREDICTED Log(1/IC50)	DEVIATION	CLOGP	B1-8	I-S	I-DMA	CMR	MGVOL	PI-X	MR-X	L-8	I-5	PI-Y	MR-Y	I-8	L-6	B1-6	B5-6	MR-6	PI-6
4.9	4.6	0.3	4.24	1	0	0	8.61	2.32	0	0.31	2.06	0	0.86	1.99	0	.	.	.	.	.
4.32	4.65	-0.33	4.37	1	0	0	8.61	2.32	0	0.31	2.06	0	0.88	1.99	0	.	.	.	.	.
4.92	5.37	-0.45	3.84	1	0	1	8.15	2.18	0	0.31	2.06	0	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01
6.84	6.61	0.23	4.76	1.8	0	1	8.64	2.3	0.71	0.81	3.52	0	-0.11	1.06	1	6.39	1.52	4.82	2.25	2.01
6.8	5.73	1.07	4.76	1	0	1	8.64	2.3	0.71	0.81	2.06	0	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01
5.61	5.53	0.08	2.7	1	1	0	8.1	2.05	0	0.31	2.06	0	0.95	1.69	0	4.92	1.52	3.49	1.5	1.55
7.11	6.78	0.33	3.53	1	1	1	9	2.29	0	0.31	2.06	0	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
7.92	7.94	-0.02	4.24	1.8	1	1	9.5	2.41	0.71	0.81	3.52	0	0.95	1.69	1	6.39	1.52	4.82	2.25	2.01
7.64	7.06	0.58	4.24	1	1	1	9.5	2.41	0.71	0.81	2.06	0	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
4.25	5.57	-1.32	4.36	1	0	1	8.61	2.32	0	0.31	2.06	0	0.45	1.52	0	6.39	1.52	4.82	2.25	2.01
5.65	6.98	-1.33	4.05	1	1	1	9.47	2.43	0	0.31	2.06	0	1.51	2.15	0	6.39	1.52	4.82	2.25	2.01
4.87	5.77	-0.9	3.32	1	1	0	8.89	2.22	0	0.31	2.06	0	1.51	2.15	0	5.14	1.52	4.36	1.82	1.48
4.84	6.98	-2.14	4.05	1	1	1	9.47	2.43	0	0.31	2.06	0	1.51	2.15	0	6.39	1.52	4.82	2.25	2.01
7.38	6.98	0.4	4.05	1	1	1	9.47	2.43	0	0.31	2.06	0	1.51	2.15	0	6.39	1.52	4.82	2.25	2.01
5.94	6.98	-1.04	4.05	1	1	1	9.47	2.43	0	0.31	2.06	0	1.51	2.15	0	6.39	1.52	4.82	2.25	2.01
6.64	5.93	0.71	5.28	1	0	1	9.1	2.44	0.71	0.81	2.06	0	0.45	1.52	0	6.39	1.52	4.82	2.25	2.01
6.32	7.26	-0.94	4.76	1	1	1	9.96	2.55	0.71	0.81	2.06	0	1.51	2.15	0	6.39	1.52	4.82	2.25	2.01
4.59	6.98	-2.39	4.05	1	1	1	9.47	2.43	0	0.31	2.06	0	1.51	2.15	0	6.39	1.52	4.82	2.25	2.01
6.74	5.73	1.01	4.76	1	0	1	8.64	2.3	0.71	0.81	2.06	1	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01
7.47	5.85	1.62	3.52	1	1	0	8.92	2.2	0.71	0.81	2.06	1	0.95	1.69	0	5.14	1.52	4.36	1.82	1.48
7.22	5.57	1.65	2.8	1	1	0	8.43	2.08	0	0.31	2.06	1	0.95	1.69	0	5.14	1.52	4.36	1.82	1.48
4.22	4.12	0.1	3.01	1	0	0	7.25	1.94	0	0.31	2.06	0	-0.11	1.06	0	4.92	1.52	3.49	1.5	1.55
5.78	5.53	0.25	2.7	1	1	0	8.1	2.05	0	0.31	2.06	0	0.95	1.69	0	4.92	1.52	3.49	1.5	1.55
4.46	4.24	0.22	3.31	1	0	0	7.68	2.04	0	0.31	2.06	0	-0.11	1.06	0	.	.	.	.	.
7.01	6.78	0.23	3.53	1	1	1	9	2.29	0	0.31	2.06	0	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
5.48	5.37	0.11	3.84	1	0	1	8.15	2.18	0	0.31	2.06	1	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01
4.46	4.24	0.22	3.31	1	0	0	7.68	2.04	0	0.31	2.06	0	-0.11	1.06	0	.	.	.	.	.
7.01	6.78	0.23	3.53	1	1	1	9	2.29	0	0.31	2.06	0	0.95	1.69	0	6.39	1.52	4.82	2.25	2.01
5.48	5.37	0.11	3.84	1	0	1	8.15	2.18	0	0.31	2.06	1	-0.11	1.06	0	6.39	1.52	4.82	2.25	2.01

Table 1 showing QSAR dataset obtained from Bioloom software from the BIOBYTE DATABSE In the table above marked with blue rows indicates the rejected or the outliers in the dataset

Table 2a **The table below shows the Neural network models predicted the Evolutionary Neural network in which 35 train set is given and 44 test set is used. The results of the test set are as follows:**

	4-5-6-7-1	4-5-6-1	4-4-5-5-1	4-5-4-6-1	4-5-7-1	4--5--1
Correlation(r)	0.932	0.933	0.879	0.824	0.900	0.889
Squared Correlation(R^2)	0.8688	0.8815	0.773	0.679	0.8115	0.776
Speraman Rank correlation(Rho)	0.918	0.9211	0.864	0.796	0.9012	0.852
RMSD	0.4871	0.4418	0.6151	0.747	0.572	0.616

Table 2b: **The table below shows the Neural network models predicted the Evolutionary Neural network in which 35 train set is given and 66 test set is used. The results of the test set are as follows:**

	4-5-6-7-1	4-5-6-1	4-4-5-5-1	4-5-4-6-1	4-5-7-1	4-5--1
Correlation (r)	0.9770	0.951	0.9	0.914	0.871	0.952
Squared Correlation (R^2)	0.912	0.906	0.903	0.836	0.76	0.824
Spearman Rank correlation(Rho)	0.944	0.945	0.939	0.907	0.858	0.899
RMSD	0.415	0.432	0.435	0.571	0.72	0.606

Table 2c: **The table below shows the Neural network models predicted the Evolutionary Neural network in which 44 train set is given and 35 test set is used. The results of the test set is as follows:**

	4-7-2-6-1	4-4-4-1	4-7-2-7-6-1	4-7-6-1
Correlation(r)	0.974	0.953	0.983	0.981
Squared correlation (R^2)	0.949	0.909	0.9677	0.9637
Spearman Rank correlation(Rho)	0.9544	0.9238	0.9781	0.966
RMSD	0.347	0.5	0.277	0.295

Table 2d: **The table below shows the Neural network models predicted the Evolutionary Neural network in which 44 train set is given and 66 test set is used. The results of the test set are as follows:**

	4-4-4-1	4-7-2-6-1	4-7-2-7-6-1	4-7-6-1
Correlation(r)	0.902	0.947	0.930	0.945
Squared Correlation(R^2)	0.814	0.897	0.866	0.894
Spearman Rank correlation(Rho)	0.903	0.947	0.915	0.942
RMSD	0.630	0.468	0.511	0.466

Table 2e : **The table below shows the Neural network models predicted the Evolutionary Neural network in which 66 train set is given and 35test set is used. The results of the test set are as follows:**

	4-5-4-6-1	4-5-4-5-6-1	4-5-4-4-6-1	4-5-4-1	4-4-4-1
Correlation(r)	0.983	0.976	0.970	0.966	0.949
Squared Correlation(R^2)	0.967	0.9539	0.941	0.934	0.9021
Spearman Rank correlation(Rho)	0.975	0.9605	0.9563	0.957	0.918
RMSD	0.278	0.327	0.378	0.392	0.476

Table 2f: **The table below shows the Neural network models predicted the Evolutionary Neural network in which 66 train set is given and 44 test set is used. The results of the test set are as follows:**

	4-4-4-1	4-5-4-6-1	4-5-4-5-6-1	4-5-4-4-6-1	4-5-4-1
Correlation(r)	0.898	.940	0.942	0.948	0.911
Squared Correlation(R^2)	0.807	0.884	0.888	0.9	0.831
Spearman Rank correlation(Rho)	0..883	0.916	0.91	0.915	0.898
RMSD	0.569	0.438	0.429	0.412	0.536

Table 3 **The table below shows the Neural Network models predicted by Evolutionary Neural network in which the whole dataset is used the results of the models are as follows:**

	4-4-6-6-2-1-1	4-5-5-2-1-1	4-4-4-5-2-1-1	4-5-6-5-2-1-1	4-5-5-5-2-1-1	4-4-3-6-1-1
Correlation(r)	0.967	0.960	0.963	0.972	0.971	0.963
Correlation squared(R^2)	0.936	0.9223	0.929	0.946	0.944	0.929
Spearman Rank correlation(Rho)	0.942	0.937	0.943	0.945	0.95	0.936
RMSD	0.371	0.414	0.394	0.342	0.347	0.386