A new open-source python script for QSAR studyand its validation of cytotoxicity activity of thiazole analogues on MCF-7 cell line

Jayaprakash Venkatesan
BITS: Birla Institute of Technology
Prabha Thangavelu (drtpappa@yahoo.com)
Nandha College of Pharmacy https://orcid.org/0000-0001-6511-0428
Selvaraj Jubie
JSSCP Ooty: JSS College of Pharmacy Ooty
Sudeepan Jayapalan
BITS: Birla Institute of Technology
MVNL Chaitanya
Chitkara College of Pharmacy
Sivakumar Thangavel
Nandha College of Pharmacy

Research Article

Keywords: QSAR, Python, Supervised machine learning, Thiazole derivatives, MCF-7, Breast cancer, Cytotoxicity

Posted Date: February 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1306420/v1

Abstract

Our present study aimed to working on trending machine learning approach with anew open-source data analysis python script for the discovery of anticancer lead via building the QSAR model by using 53 compounds of thiazole derivatives.Total of 82 CDK molecular descriptor were downloaded from “chemdes” web server and used for our study. After training the model, we checked the model performance via cross-validation of external test set. The generated QSAR model afforded the ordinary least squares (OLS) regression as R^2 = 0.542, F=8.773, and adjusted R^2 (Q2) =0.481, std. error = 0.061, reg.coef_ developed were of, -0.00064 (PC1), -0.07753 (PC2), -0.09078 (PC3), -0.08986 (PC4), 0.05044 (PC5), and reg.intercept_ of 4.79279 developed through *statsmodels.formula* module. The performance of test set prediction was done by multiple linear regression, support vector machine, and partial least square regression classifiers of *sklearn*module, whichgenerated the model score of 0.5424, 0.6422, and 0.6422, respectively. Hence, we conclude that the R^2values(i.e. the model score) obtained using this script via three diverse algorithms were correlated well and there is not much difference between themand may be useful in the design of a similar group of thiazolederivatives as anticancer agents.

1. Introduction

Cancer is the world's second largest cause of mortality, following cardiovascular disease, and its prevalence is increasing at an alarming rate. As per the International Agency for Research on Cancer's (IARC) predictions, the rate of new cancer cases grew to 18.1 million in 2018, with 9.6 million cancer-related fatalities. More likely, the lung and breast cancer (11.6%) those are greatest prevalent kind of cancer types that kills people [1]. As a result, many researchers are presently focused on developing new medications with the fewest potential side effects and strong efficacy over the most often diagnosed cancers, as well as those that are most likely to be fatal.

Drug discovery requires the use of hybrid technologies for the discovery of new chemical substances that might be promising candidates. One of those interesting strategies is QSAR, or quantitative structure-property relationships (QSPR), and artificial intelligence systems that effectively predict how chemical alterations can impact biological activity via in-silico. QSAR approaches have been used to efficiently mimic several physiochemical aspects of compounds, including toxicity, metabolism, drug-drug interactions, and carcinogenesis [2]. Early QSAR models employed basic multivariate regression models to connect potency ($\log IC_{50}$) with substructure motifs and chemical parameters including, solubility (log P), hydrophobicity, substituent pattern, and electronic variables [3, 4]. However, these days, the thriving research on machine learning approaches use pattern recognition algorithms to differentiate mathematical relationships between empirical observations of organic compounds and extrapolate them to predict chemical, biological, and physical properties of novel compounds. Machine learning approaches are much more efficient than physical models and can readily expand to large datasets without requiring a lot of computer power [5, 6].

Thiazole analogues are one of the most recognized molecules in biomedical sciences, owing to the ease of structural optimization and the possibility to synthesis a large number of derivatives [7–10]. Thiazole is a heterocyclic compound that may be found in a variety of synthetic bioactive compounds and has gotten a lot of interest in drug development over the last decade. Thiazole compounds exhibited antitumor, anti-inflammatory, antioxidant, antibacterial, and anti-HIV properties. Sulphathiazole, ravuconazole, ritonavir, and meloxicam are examples of thiazole containing medicines that have been authorized for clinical usage. It's worth noting that thiazole has the potential to be a viable skeleton for the production of anticancer drugs [11–16].

This study aimed to develop a robust QSAR model to predict biological activity via a machine learning approach. The usage of expensive software packages often restricts many academicians from assessing and adopting available models. Although the current software that calculate descriptors could provide the lack of information for doing cross-validation and/or robust procedure for doing QSAR analysis. If the computational methods are created using fully accessible programs, they may be distributed more freely among scientists. We provide a novel python script for data analysis and validation of this script applying data concerning thiazole derivatives with cytotoxicity activity against MCF-7 cell line. CDK descriptors were calculated and used after appropriate preprocessing for building QSAR models. Machine learning algorithms like multiple Linear regression (MLR) and support vector machine (SVM) were employed to identify the correlation between the structures of thiazole (i.e., as described by the molecular descriptors) and their respective bioactivity (i.e., the IC_{50}values).

2. Materials And Methods

2.1. Data collection

In the present study, the cytotoxicity activity of various thiazole derivatives on MCF-7 breast cancer cell line were collected from the reported literature elsewhere [8–10, 17]. Such data are derived from different laboratories, have been generated at different times, most likely with different reagents and laboratory equipment.

2.2. Descriptor selection

We opted for CDK descriptors from the “Chemdes” webserver (http://www.scbdd.com/chemdes/) to calculate the descriptors of our thiazole derivatives. The webserver calculated around 138 various CDK descriptors for our compounds. However, after doing manual preprocessing such as removing missing values, zero,NAN (not a number) values, and removing highly correlated values, we have finally come up with 82 descriptors for our unsupervised machine learning QSAR model development workflow.

2.3. Interactive computing platform used

In our study, we used Google Collaboratory notebook (https://colab.research.google.com/) to execute our machine language. It allows us to write and execute Python in our browser, with zero configuration, free access to GPUs, and easy sharing options.

2.4. Machine Languages used

We have imported the necessary packagesof Python modules viz.

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import sklearn

import seaborn as sns

2.5. Input Data

Then we have imported the data for the program as a comma-separated values (.CSV) file, which contains 53 compound's code, their pIC_{50}value, and 82chosen CDK descriptors. The CSV file displays the calculated descriptors (X=independent variables) as the column and biological activity data (Y=dependent variable) in the first column.

2.6. Dataset preparation

A dataset of 53 compounds with their cytotoxicity activity values (IC_{50} in μM/ml) on MCF-7 cell line were collected from the diverse reported literature [8–10, 17]. Then, we converted the biological activity (IC_{50}) data into logarithmic scale [-log(IC_{50})] since the original IC_{50} value has an uneven distribution of the data points. To make the distribution more even we transformed it to a logarithmic value. Then we arranged all data as per their increasing order of activity. After that, we have picked 10 compounds based on "Rule of thumb" for the test set, and the remaining 43 compounds were kept as the training set. The structures of Thiazole derivatives are listed in Table 1.

2.7. Training the Model

The first step involves training the model by using 43 training set compounds and followed by cross-validation of the built model via 10 test set compounds. The built model was predicted the biological activity of test compounds and showed a similar value to that of the training set biological activity, which further confirms the enhanced model performance. The workflow of Built QSAR model has been shown in Fig. 1.

2.8. Data preprocessing and feature selection

Then we used three various regression algorithms to build the QSAR models with 82 descriptors. Before using these algorithms, we preprocessed the data using diverse modules available in Python language [18, 19]. In the preprocessing step, the independent and dependent variables are clearly stated by defining X and Y variables, respectively.

The first step in the preprocessing of data involves the dropping of missing values and strings if any presented in that loaded CSV file followed by applying dimensionality reduction using principal component analysis (PCA). Suppose, if we have more independent variables for our selected compound's biological activity, then we have to go for doing the dimensionality reduction via PCA, which could help us to select the best variables related to our biological activity. Subsequently, from *sklearn.preprocessing* module, we imported the *StandardScaler* to scale and transform the given values then we set the explained variance ratio, and the set variance between the descriptors are explained by 20 principal components. The reason behind checking the explained variance ratio is to know how well the PCA components are represented without any loss between the variables (selected descriptors). In our case, it captured 99.78%. i.e., the variables (descriptor information) are not missing their information and presented 99.78% within these 20 selected principal components. Besides, we have chosen the PCA component with 90% variability captured via the data preprocessing step, and then it is viewed through drawing the 2D and 3D plot (Fig. 2). Later, the module picked the adequate numbers of principal components to preserve 90% variability without any further loss of the features information, i.e., 90% of the descriptors information are captured in these five PC. However, in our case, it picked up of five principal components, the transformed data by principal components (90% variability) and the respective snippet was shown in Figs. 3 and 4. The next step involves creating the Pandas data frame for the selected five principal components so that it could be used for further regression analysis.

2.9. Algorithm or Classifier used

We have used three machine-learning algorithms, Multiple Linear Regression (MLR), Support Vector Machine (SVM), and Partial Least Square (PLS), to build the QSAR models. These three classifiers have predicted the accuracy of the effect of thiazole derivatives on MCF-7 breast cancer cell line.

2.9.1. Multiple linear regression model

From *sklearn*, we imported *linear_model* for performing MLR regression of our preprocessed data, from which we calculated reg.coef_, reg.intercept_, and derived the mathematical formula from the built QSAR model. Based on the summary of the regression equation, the researcher could be able to select the right model for predicting the activity of compounds whose activity is unknown. The number of nodes used in the input layer was equal to the number of the descriptors presented in the data set (i.e., principal component descriptors), while one node (test set molecules) was used in the output layer corresponding to the IC_{50} value. The overall neural network of our QSAR model showed in Fig. 5.

2.9.2. Support vector machine

From *sklearn.svm*, we imported an SVR classifier to train the model and the performance of the model was observed by its model score prediction. Further, the model was validated by test set and the corresponding predicted pIC_{50} value calculated and the respective snippet was shown in Fig. 6. In addition to this, we also viewed the linear plot for experimental versus predicted pIC_{50} values via *seaborn*

(*import matplotlib.pyplot as plt*)module, which displayed a narrower variance of the data points (Fig. 7). Residual and regression plots are saved as image files for quick analysis.

2.9.3. Partial Least Square (PLS)

Partial Least Squares (PLS) regression model that combines dimensionality reduction with multiple regressions to turn variables into uncorrelated variables that are optimally associated with the activity or feature of interest. From *sklearn* library, we imported the necessary packages, 'cross_decompositionimport PLSRegression' to fit the PLS regression model. Then we calculated the model score and compared it with the other two regression models.

2.10. Ordinary Least Square (OLS) regression

Subsequently, we imported *statsmodels.formula.api*as*sm*for statistical calculation of our data. A predictive mathematical model is built using selected descriptors, all statistical terms associated with the model like R-squared (R^2), Adj. R-squared (Q^2), 'F' statistics, 't-value', and 'p-value' were calculated and presented.

2.11. Saving the model using "Pickle"

Finally, the built model was saved by using "*Pickle*" module of Python, later could be retrieved further for the prediction of newly designed molecules with similar biological activity.

3. Results And Discussion

Developing machine learning algorithms have become an important tool in the drug discovery process. Nowadays, a variety of machine learning tools are used to establish QSAR models. From our study result, the generated QSAR model via an open-source python program was predicted well with external test set compounds. The generated statistical model afforded the ordinary least squares (OLS) regression as R^2 = 0.542, F=8.773, and adjusted R^2 (Q^2) =0.481, std. error = 0.061 (Table 2), reg.coef_ developed were of, -0.00064 (PC1), -0.07753 (PC2), -0.09078 (PC3), -0.08986 (PC4), 0.05044 (PC5), and reg.intercept_ of 4.79279 developed through *statsmodels.formula* module. The performance of test set prediction was done by MLR, SVM, and PLS classifiers of *sklearn*module, which threw the model score of 0.5424, 0.6422, and 0.6422, respectively. The model performance was validated through the test set, and the model predicted similar better values when compared to that of the training set. The linear curve has been plotted between the predicted and actual pIC_{50} value, which showed all the data fall over the middle linear line (Fig. 6). We have found that the model score obtained using these three algorithms were correlated well and there is not much variance between them and may be useful in the design of a similar group of thiazoleanalogs as anticancer agents.

Generated Multiple linear regression model:

pIC_{50} = 4.79279 -0.00064 (PC1) -0.07753 (PC2) -0.09078 (PC3) -0.08986 (PC4) + 0.05044 (PC5)

Table 2
OLS Regression Results

Dep. Variable:	**target**	**R-squared:**	**0.542**
Model:	OLS	**Adj. R-squared**:	0.481
Method:	Least Squares	**F-statistic**:	8.773
Date:	Mon, 27 Dec 2021	**Prob (F-statistic)**:	1.47e-05
Time:	13:43:07	**Log-Likelihood**:	-18.552
No. Observations:	43	**AIC**:	49.10
Df Residuals:	37	**BIC**:	59.67
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	4.7928	0.061	78.262	0.000	4.669	4.917
PC1	-0.0006	0.009	-0.074	0.941	-0.018	0.017
PC2	-0.0775	0.019	-4.072	0.000	-0.116	-0.039
PC3	-0.0908	0.022	-4.174	0.000	-0.135	-0.047
PC4	-0.0899	0.033	-2.761	0.009	-0.156	-0.024
PC5	0.0504	0.034	1.495	0.143	-0.018	0.119

Omnibus:	**2.933**	**Durbin-Watson:**	**1.185**
Prob(Omnibus):	0.231	**Jarque-Bera (JB)**:	2.283
Skew:	0.564	**Prob(JB)**:	0.319
Kurtosis:	3.054	**Cond. No.**	7.08

The purpose of linear regression analysis is to discover a linear function of a collection of predictor variables that minimizes the distances between data points along the dimensions of a set of outcome measures consider a dataset of training data points. Multivariate linear regression is used extensively in initial QSAR approaches for eg.Hansch and Free–Wilson analysis. Moreover, the use of linear regression models in QSAR is complicated by feature correlations and high-dimensional feature spaces, which could lead to model overfitting. To tackle these two issues that could complicate the accuracy and applicability, there are several approaches available like regularization, dimension reduction, and genetic algorithms [20]. Dimensionality reduction approaches, such as principal components analysis (PCA), on the other

hand, reduce the huge sets of correlated variables into smaller groups of unrelated variables [21]. Gao et al. 1999 [22], employed to reduce the correlation between the variables for estrogen receptor interaction prediction.

The supervised machine learning algorithm SVM, solve the classification problem by mapping data into a high-dimensional space with nonlinear kernel functions and determining the best-separated hyperplane [23]. The hyperplane is a linear data layout that maximizes the margin across support vectors, which are the locations nearest to the decision boundary. Nekoei et al. 2015 [24], recently employed a genetic variable selection technique in conjunction with SVMs to identify several structural features of aminopyrimidine-5-carbaldehyde oxime analogs that are essential for their strong VEGF-2 inhibition effect. In our study, the performance of the test set has been done through the SVM of *the sklearn* module, which generated the model score of 0.6422 and the linear curve between the predicted and actual pIC_{50} value plotted as shown in Fig. 7.

On the other hand, the PLS regression algorithm combines dimensionality reduction with multiple regressions to turn variables into uncorrelated variables that are optimally associated with the activity or feature of interest. PLS is widely utilized in 3D-QSAR [25], and Erikkson et al. 2003 [26], suggest it as a prime-line technique to QSAR model because of its improved efficiency and accuracy compared to deliberately merging unsupervised dimensionality with multiple regressions. Though the linear regression analysis has been effective in many drugs optimization applications. However, the underlying linearity and vector space constraints are not applicable for most QSAR applications. As a result, despite careful selection of features and the analyzed system, it is sometimes not quite enough to assure the success of linear regression models. Our study result on PLS showed a fit score of 0.6422.

4. Conclusions

Through 'data merging' approaches, which combine structural, genetic, and pharmacological data from the molecular to organism level, will be critical for the identification of safe and effective medications. Hence, modern machine learning algorithms are suitable for processing large amounts of data with high speed, accuracy, and flexibility is also required. Our present study handles the diverse machine learning algorithms that were used to build the QSAR model to know about the residual difference between observed and predicted cytotoxicity efficacy of the 53 selected molecules of thiazole derivatives. A total of 82 molecular descriptors were used for constructing the QSAR models, and their performances were comparatively evaluated. The QSAR studies on these training set compounds showed the predicted pIC_{50} values of the compounds have an acceptable correlation with the experimental values from the MLR, SVM, and PLS algorithms. Hence, the QSAR model generated via open-source Python script could be useful for designing a similar group with promising anticancer activity. It is anticipated that the knowledge gained from this study could be used as general guidelines for the design of novel anticancer drugs.

Declarations

Funding

Not applicable

Conflicts of interest/Competing interests

The authors declare no competing interests.

Availability of data and materials

The authors confirm that the data supporting the findings of this study are available within the article and as well the online version is available as a supplementary material at https://gist.github.com/Dr-T-Prabha/c89c573afc07612ad83908d8dd6f3773.

Code availability

https://gist.github.com/Dr-T-Prabha/c89c573afc07612ad83908d8dd6f3773

Authors' contributions

JV framed the concept and collected the data of our research, SJ drafted the manuscript, optimization was done by JS and MVNLC, TP executed and interpreted the python script and TS finalized and checked the manuscript for correction and plagiarism if any. All authors read and approved the manuscript.

Acknowledgements

The authors are thankful to the Nandha College of Pharmacy, Erode, Tamilnadu, the Department of Pharmaceutical Sciences & Technology, Birla Institute of Technology, Mesra, JSS College of Pharmacy, Ooty, and College of Pharmacy, Chitkara University, Punjab, Indiafor providing necessary support to carry out our research work.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, GlobalCancer Statistics 2018 GLOBOCAN estimates of incidence and mortalityworldwide for 36 cancers in 185 countries.Ca - Cancer J Clin68:394e424
2. Artem C, Eugene NM, Denis F, Alexandre V, Igor IB, Mark C, John D, Paola G, Yvonne CM, Roberto T, Viviana C, Victor EK, Richard C, Romualdo B, Chihae Y, James R, Lothar T, Johann G, Ann R Alexander T 2014 QSAR modeling: where have you been? Where are you going to?J Med Chem57:4977–5010
3. Prabha T, Sivakumar T 2018 Design, Synthesis, and Docking of Sulfadiazine Schiff Base Scaffold for their Potential Claim as INHA Enoyl-(Acyl-Carrier-Protein) Reductase Inhibitors.Asian J Pharm Clin Res11:233–237

4. Kubinyi H (1988) Free Wilson analysis. Theory, applications and its relationship to Hansch analysis. Quant Struct Act Relat 7:121–133
5. Prabha T, Selvinthanuja C, Hemalatha S, Sengottuvelu S, Senthil J (2021) Machine learning algorithm used to build a QSAR model for pyrazoline scaffold as anti-tubercular agent. J Med Pharma Alli Sci 10(6):4024–4030
6. Varnek A, Baskin I (2021) Machine learning methods for property prediction in chemoinformatics: Quo Vadis? J Chem Inf Model 52:1413–1437
7. de Siqueiraa LRP, de Moraes Gomes PAT, de Lima Ferreira LP, deMeloRego MJB Leite ACL 2019 Multi-target compounds acting in cancer progression: focus on thiosemicarbazone, thiazole and thiazolidinone analogues.Eur J Med Chem170:237e260
8. Mansour SA, Mahmoud SB, Saleh IA, Mostafa MG (2011) Anti-breast cancer activity of some novel 1,2-dihydropyridine, thiophene and thiazole derivatives. Eur J Med Chem 46:137–141
9. Saulo FPB, Nayara CF, Jonas PR, Elaine MSF, Renata BO 2016 Synthesis and cytotoxicity evaluation of thiosemicarbazones and their thiazole derivatives.Braz J Pharma Sci52:299–307
10. Wang G, Liu W, Fan M, He M, Li Y, Peng Z 20121 Design, synthesis and biological evaluation of novel thiazole-naphthalene derivatives as potential anticancer agents and tubulin polymerisation inhibitors.J Enzyme Inhib Med Chem36:1694–1702
11. Gümüş M, Yakan M, Koca İ (2019) Recent advances of thiazole hybrids in biological applications. Future Med Chem 11:1979–1998
12. Chhabria MT, Patel S, Modi P et al 2016 Thiazole: a review on chemistry, synthesis and therapeutic importance of its derivatives.Curr Top Med Chem16:2841–2862
13. Ayati A, Emami S, Moghimi S et al 2019 Thiazole in the targeted anticancer drug discovery.Future Med Chem11:1929–52
14. Jain S, Pattnaik S, Pathak K et al 2018 Anticancer potential of thiazole derivatives: a retrospective review.Mini Rev Med Chem18:640–55
15. Mishra R, Sharma PK, Verma PK et al 2017 Biological potential of thiazole derivatives of synthetic origin.J Heterocycl Chem54:2103–16
16. Sharma PC, Bansal KK, Sharma A et al 2020 Thiazole-containing compounds as therapeutic targets for cancer therapy.Eur J Med Chem188:112016
17. Alaa MA, Abrar AB (2021) Synthesis and antiproliferative activity studies of new functionalized pyridine linked thiazole derivatives. Arab J Chem 14:102914
18. Fabian P, Gaël V, Alexandre G, Vincent M, Bertrand T, Olivier G, Mathieu B, Peter P, Ron W, Vincent D, Jake V, Alexandre P, David C, Matthieu B, Matthieu P Édouard D 2011 Scikit-learn: Machine Learning in Python.J Mach Learn Res12:2825–2830
19. Kim S, Cho KH 2019 PyQSAR: A Fast QSAR Modeling Platform Using Machine Learning and Jupyter Notebook.Bull Korean Chem Soc40:39–44

20. Kubinyi H (1996) Evolutionary variable selection in regression and PLS analyses. J Chemom 10:119–133
21. John R, Owen IT, Nabney JL, Medina F, Fabian LV 2011 Visualization of Molecular Fingerprints.J Chem Inform Model51:1552–1563
22. Gao H, Williams C, Labute P, Bajorath J (1999) Binary quantitative structure-activity relationship (QSAR) analysis of estrogen receptor ligands. J Chem Inf Comput Sci 39:164–168
23. Noble WS (2006) What is a support vector machine? Nat Biotechnol 24:1565–1567
24. Nekoei M, Majid M, Eslam P (2015) QSAR study of VEGFR-2 inhibitors by using genetic algorithm-multiple linear regressions (GA-MLR) and genetic algorithm-support vector machine (GA-SVM): a comparative approach. Med Chem Res 24:3037–3046
25. Gasteiger J (2008) Handbook of Chemoinformatics: from Data to Knowledge. Wiley-VCH
26. Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P 2003 Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs.Environ Health Perspect111:1361–1375

Tables

Table 1 is available in the Supplementary Files section.

Figures

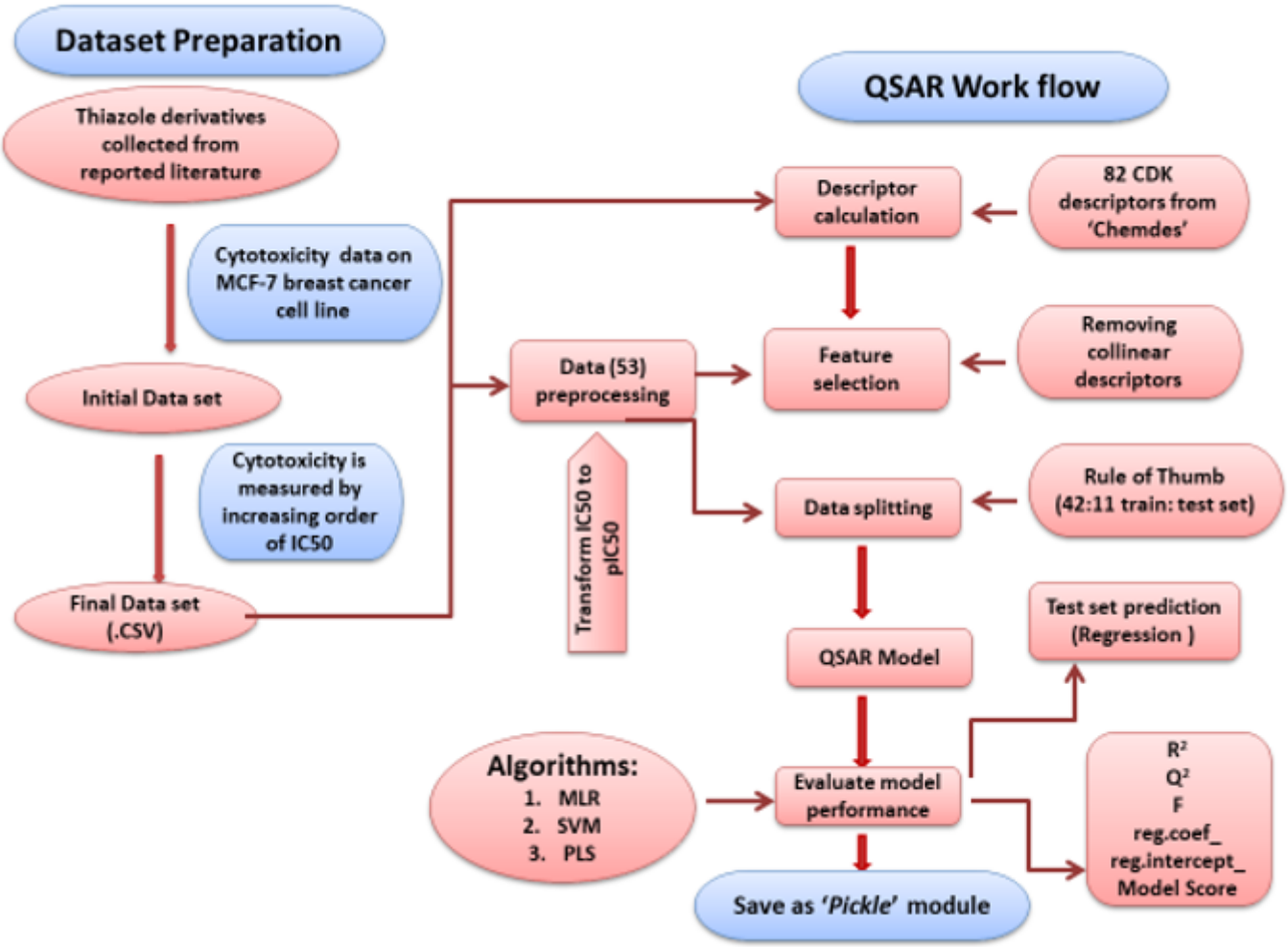
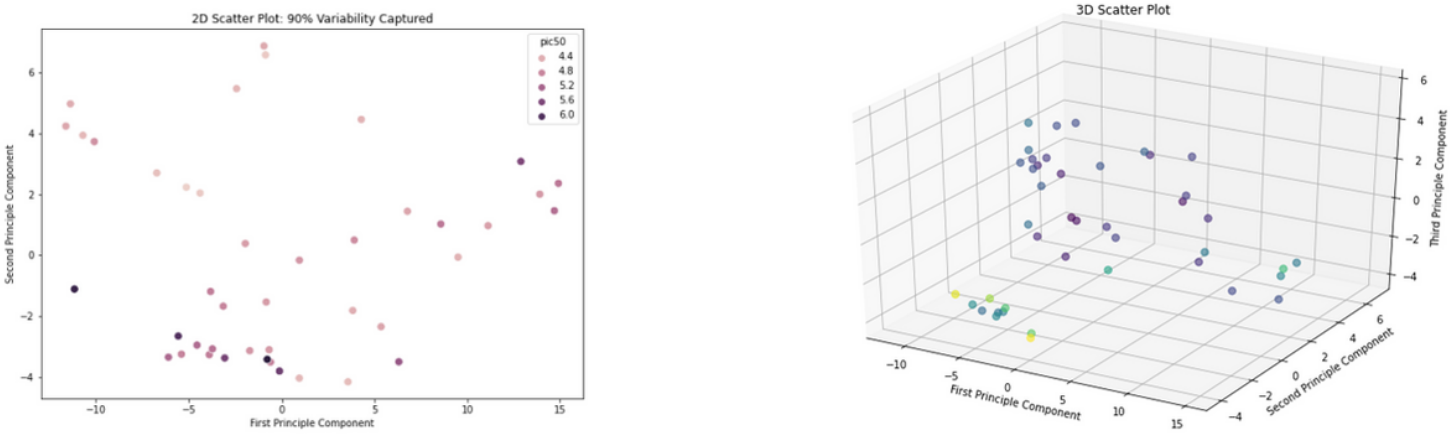


Figure 1

The workflow of Built QSAR model



(1a) (1b)

Figure 2

2D and 3D Scatter plot of principal component analysis that captured 90% variability

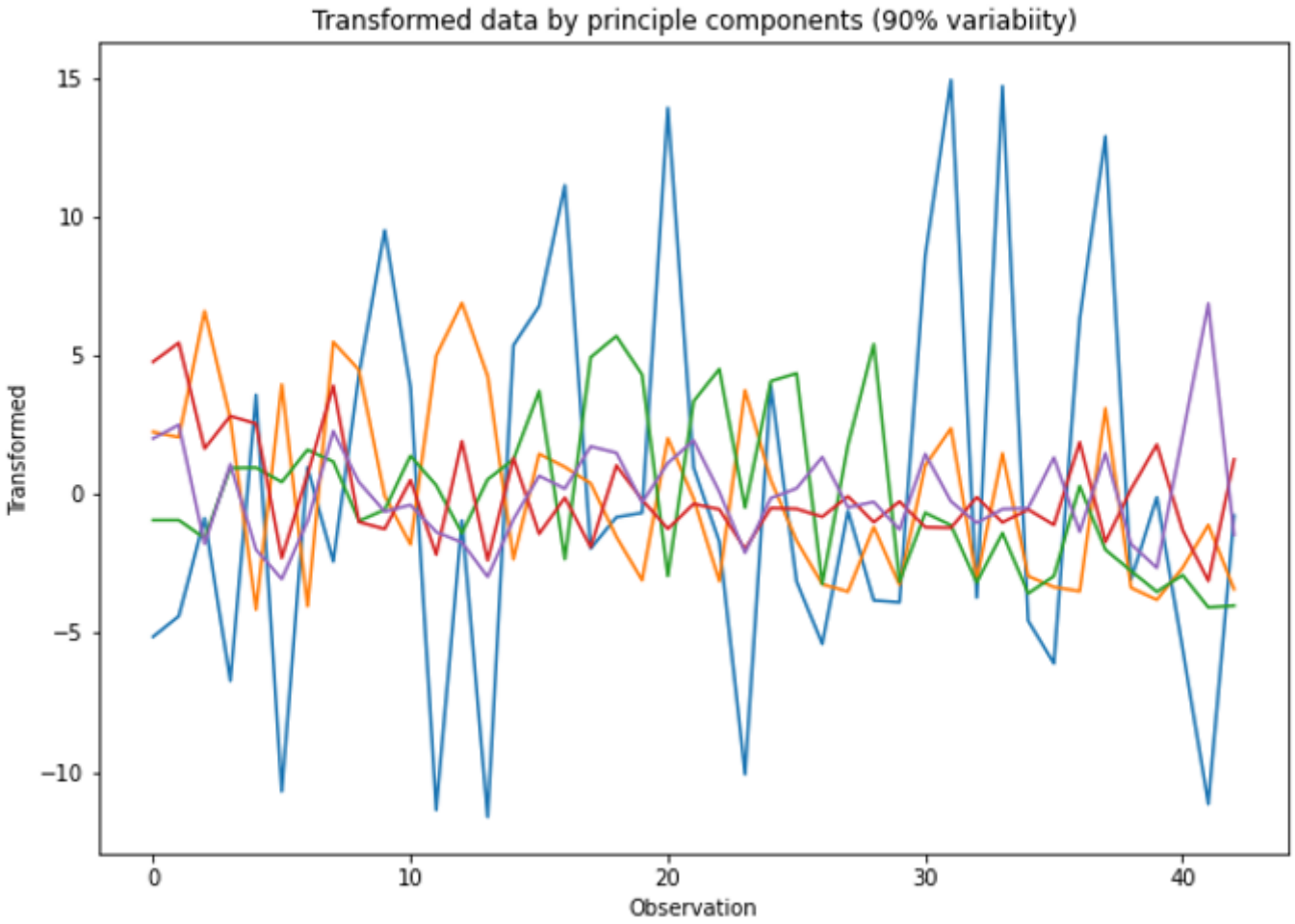


Figure 3

Transformed data by principal components (90% variability)

```
[ ]  df_new = pd.DataFrame(X_pca_90,columns=['PC1','PC2','PC3','PC4','PC5'])
     df_new['target'] = Y
     df_new.head()
```

	PC1	PC2	PC3	PC4	PC5	target
0	-5.137548	2.235943	-0.941772	4.759852	2.001487	4.056505
1	-4.396323	2.043850	-0.942504	5.444976	2.489944	4.123205
2	-0.870976	6.584207	-1.585581	1.637149	-1.793500	4.127261
3	-6.727808	2.705131	0.932989	2.808784	1.069419	4.221126
4	3.569939	-4.159506	0.951070	2.530539	-2.001315	4.251037

Figure 4

Snippet of principal components to preserve 90% variability

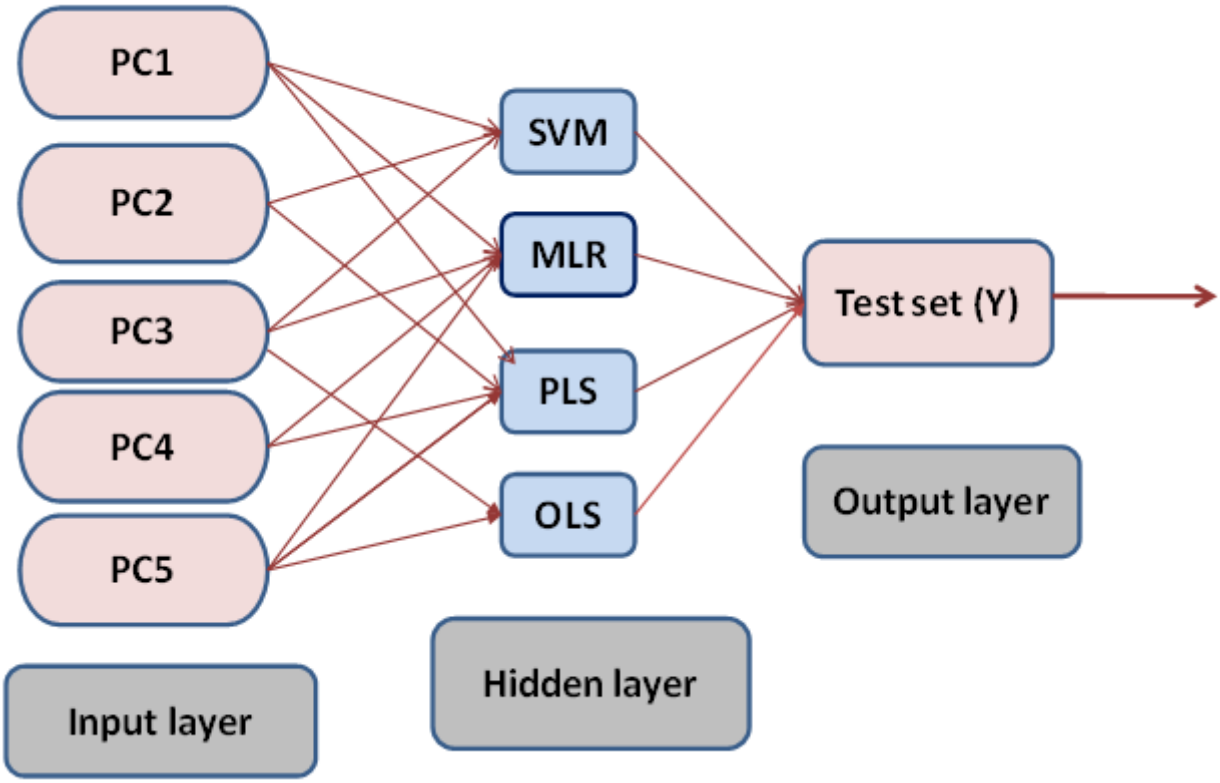


Figure 5

The neural network architecture of QSAR model

```
df1['target'] = df.target
df1
```

	Y_pred	target
0	4.190614	4.056505
1	4.192118	4.123205
2	4.345405	4.127261
3	4.267431	4.221126
4	4.653496	4.251037
5	4.485372	4.290730
6	4.714169	4.295849
7	4.207571	4.307153
8	4.454419	4.354578
9	4.540810	4.367037

Figure 6

Snippet shows the predicted biological activity

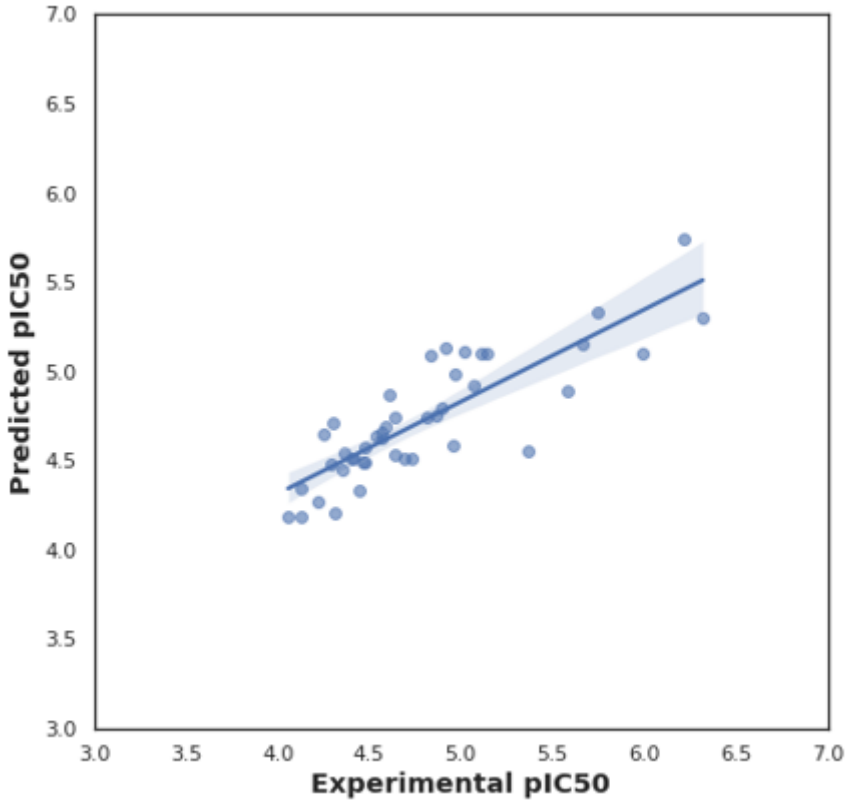


Figure 7

Linear regression plot of selected dataset compounds

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- SupplementaryMaterial28.01.2022.ipynb