ijpbs 2023;doi10.22376/ijpbs.2023.14.3.p6-20

International Journal of Pharma and Bio Sciences

ISSN 0975-6299

Original Research Article

QSAR of Pyrimidamine for Antibrestcancer

Quantitative Structure-Activity Relationship (Qsar) Modeling Studies of 4, 6-Diaryl-2- Pyrimidinamine Derivatives as Anti-Breast Cancer Agents

Neha Sharma[1*], Vipin Saini[2] Shivani Sharma[3], Sushma Gautam[4], Jyoti Sharma[5], Vipra Kohli[6], Munsaka Siankuku[7], Monika Sen[8], Sonali Bhanbhani[9], Surbhi Chopra[10], Aditi Dwivedi[11] and Satinder Singh Minhas[12]

[1,3-5,8-11]Department of Biotechnology, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana (Ambala)-133207 (Haryana)

[2] Maharishi Markandeshwar College of Pharmacy, Maharishi Markandeshwar (Deemed to be University), Mullana (Ambala)-133207 (Haryana)

[6,12]Maharishi Markandeshwar Medical College and Hospital, Solan-173229 (Himachal Pradesh)

[7]Department of Biomedical Sciences, School of Health Sciences, University of Zambia, Lusaka, Zambia-15101 (Africa)

Abstract: Comprehensive preponderance of breast cancer and its high recurrence attracts the attention of many research programs in drug discovery. Several statistics are used to evaluate its occurrence. The discovery of new and more efficient anti-cancer agents is a key area in chemotherapy. A Quantitative-Structure-Activity Relationship (QSAR) study was performed on 4, 6-diaryl-2-pyrimidinamine derivatives; several highly descriptive and predictive QSAR models for these compounds were obtained. In the present study, QSAR studies have been performed on 31 compounds of 4, 6-diaryl-2-pyrimidinamine derivatives as anti-breast cancer agents using CHEM sketch and NCSS Software. The software includes; ChemDraw version 12.0.2, Spartan'14 (version 1.1.2), Material Studio (V8) software, Pyrex software, PADEL V2.20, DTC data lab software version, and Auto Dock Visualizer version 4.2. QSAR models have been developed by using multiple linear regressions to identify descriptors, which focus on biological activity, Leave One out Method was employed in cross-validation analysis to validate the developed model and has indicated that activity can be best modeled in multi-parametric regression). Upon applying the Leave One out (LOO) method, one compound which exists as an outlier leads to qualitative findings with a high degree of statistical significance and good predictive ability.

Keywords: Breast cancer, Drug discovery, Chemotherapy, Regression analysis, Leave One Out, Cross-validation analysis

***Corresponding Author**

Neha Sharma , Department of Biotechnology, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana (Ambala)-133207 (Haryana)

Date of Receiving 23 February, 2023

Date of Revision 10 May, 2023

Date of Acceptance 31 May, 2023

Date of Publishing 3 July, 2023

Funding This research did not receive any specific grant from any funding agencies in the public, commercial or not for profit sectors

Citation Neha Sharma, Vipin Saini, Shivani Sharma, Sushma Gautam, Jyoti Sharma, Vipra Kohli, Munsaka Siankuku, Monika Sen, Sonali Bhanbhani, Surbhi Chopra, Aditi Dwivedi and Satinder Singh Minhas , Quantitative Structure-Activity Relationship (QSAR) Modeling Studies of 4, 6-Diaryl-2- Pyrimidinamine Derivatives as Anti-Breast Cancer Agents.(2023).Int J Pharm Sci.14(3), p6-20 http://dx.doi.org/10.22376/ijpbs.2023.14.3.p6-20

1. INTRODUCTION

One of the crucial public health concerns having captured global attention is cancer. Until recently, many corresponding incidences and mortality statistics have indicated an uptick in cancer in economically developing and developed nations [1,2]. Breast cancer is amongst a few of the many types which have witnessed an unprecedented rise. Although a progression in the treatment regime is upscaled, a ray of hope exists in identifying novel anticancer agents to ward off breast cancer [3]. Currently, anti-cancer agents obtained from natural sources have found immense usage [4]. Breast cancer originates from breast tissues [5] clinically manifested as a lump in the breast, change in breast shape, dimpling of the skin, red or scaly patch of skin, or inverted nipple[6]. These attributes can be accompanied by shortness of breath, bone pain, swollen lymph nodes, or yellowing of skin [7]. Reportedly, in 80% of the cases, it is a lump that leads to a preliminary diagnosis of breast cancer [8]. Other pathophysiological indicators include inflamed armpits and collarbone [9], which might not be accompanied by pain. Pain in the breast could be associated with other health issues, not necessarily malignancy [10]. Early detection of breast cancers was made possible through mammography [11]. Based on recent molecular techniques like immune- histology and gene expression profiling, breast cancer can be categorized into four major types [12] (Figure 1). Treatment of breast cancer is undertaken by different therapies [13-18] (Figure 2).

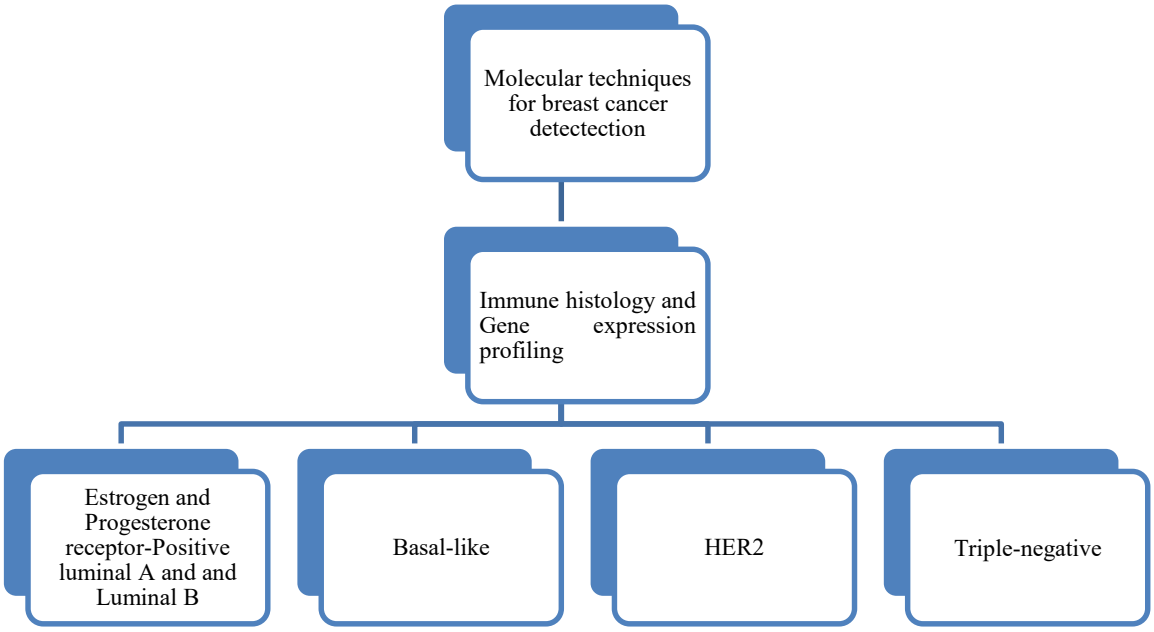


Fig. 1: Techniques for Breast-cancer detection

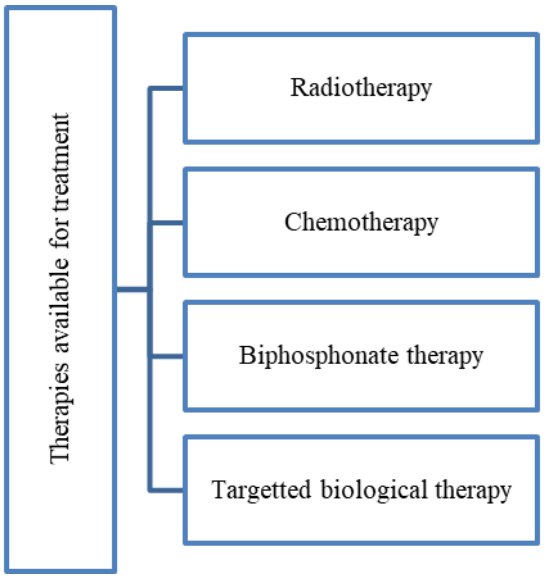


Fig. 2: Treatment regime of Breast cancer

India had an estimated 11.6 lakh new cancer cases in 2018, according to a report by the World Health Organization (WHO), which speculated one in 10 Indians will develop cancer during their lifetime and one in 15 will die of the disease. Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among females worldwide, according to Global Cancer (GLOBOCAN) 2018 [19]. Primary sources which trigger breast cancer include

environmental factors and dietary aberrations; secondary ones include virus-mediated genetic disturbances [20]. The deregulation mainly recognized in breast cancer is the inactivation of tumor suppressor genes and hyper-activation of proto-oncogenes [21]. It is described as a heterogeneous group of neoplasm having posed an immense challenge in modern medicine [22]. This fact has necessitated an urge to screen plant-derived novel bio-therapeutics as promiscuous anti-cancer tools (Figure 3) elucidates a generalized mechanism of breast cancer path physiology, risk factors, and sources.

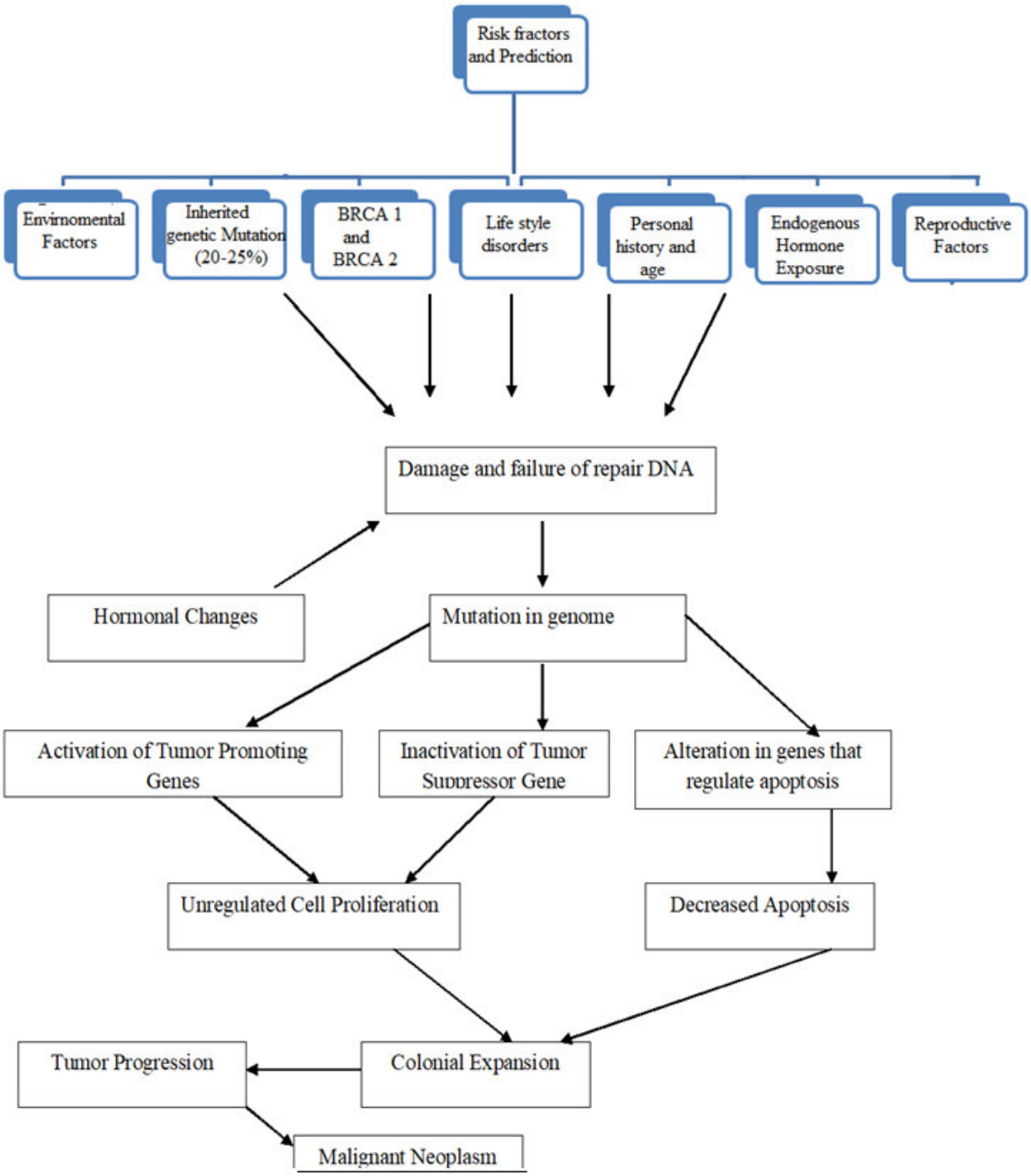


Fig. 3: Pathophysiology of Breast Cancer

4, 6-diaryl-2-Pyrimidine derivatives and their related compounds are known to be associated with biodynamic properties[23]. They possess potential pharmacological properties exhibiting anti-HIV, anticancer, analgesic, and anti-inflammatory effects. The 2-pyrimidine amine scaffold is the key skeleton of pazopanib, the Vascular Endothelial Growth Factor Receptor-2 (VEGFR-2) inhibitor [24]. It is known that VEGFR-2 plays a key role in the angiogenesis pathway, which participates in the proliferation, metastasis, and invasion of breast cancer cells. The Quantitative Structure-Activity Relationship (QSAR) method provides a promising approach for estimating the activity based on the descriptors derived from the molecular structures[25,26]. QSAR also revealed the positive and negative electrostatics regions of the active compounds. Structure prediction through QSAR plays a key tool in medicinal chemistry [27,28]. QSARs and Quantitative Structure-Property Relationships (QSPRs) are twin "computational chemistry" disciplines, which consist of searching for similarities between molecules in large databases of existing molecules with known properties [29]. Developing such a relationship helps us identify compounds' physical, chemical, and biological properties based on understanding the

phenomena observed or developing new theories. The current scope of the study was based on developing a novel model for studying the relationship between structure and anti-breast cancer activity of 4, 6-diaryl-2-pyrimidinamine and their derivatives. Mathematical methods can be used to establish the relationship between the structural characteristics of the molecule and its properties. Multiple Linear Regression (MLR), cross-validation analyses, Partial Least Squares (PLS), and Leave One out (LOO) were applied to a series of 4, 6-diaryl-2-pyrimidinamine to develop a QSAR model to predict anti-breast cancer activity [30, 31] reliably. The best model was validated by the regression coefficient (R^2).

1.1 Molecular optimization

The physico-chemical properties of the molecules can be analyzed using various electronic and molecular parameters through geometrical optimization [32]. (ChemDraw (V12.0.2) software was employed to sketch the compounds in 2D format. These structures were then converted into a 3D format, and additional optimization was included using Spartan 14 (V1.1.4) software. The complete optimization based on these parameters, Density functional theory (DFT), includes applying the B3LYP version and 6-311G* basis set [33, 34].

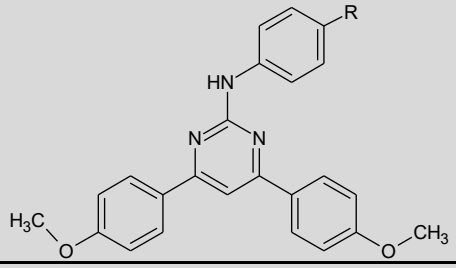
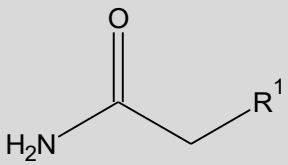
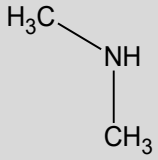
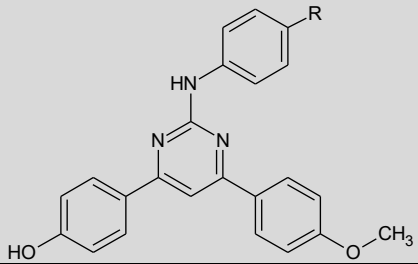
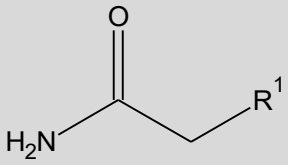
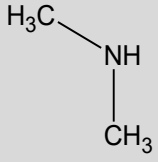
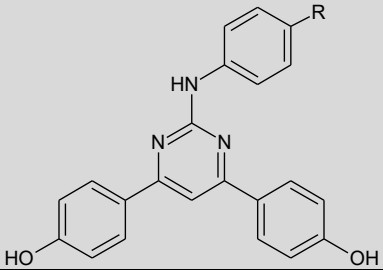
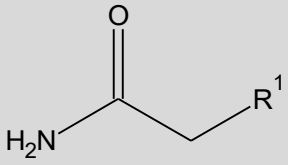
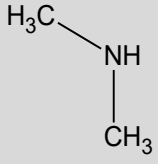
2. EXPERIMENTAL

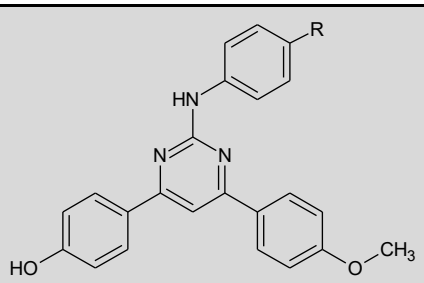
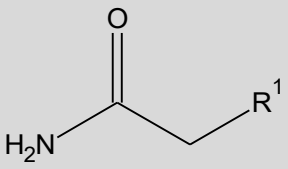
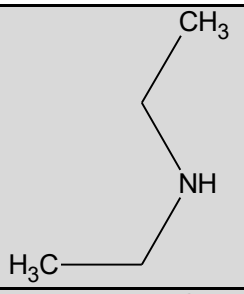
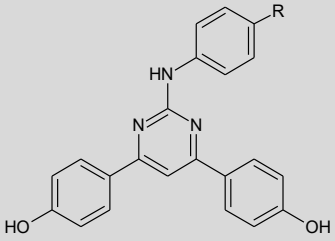
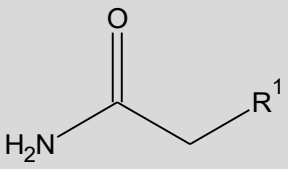
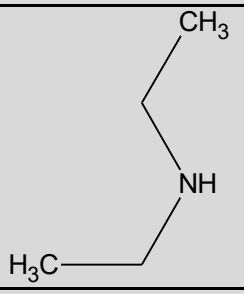
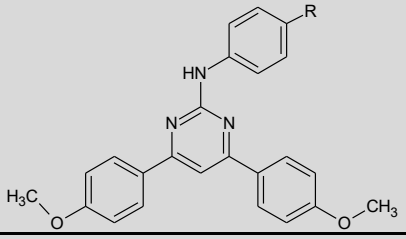
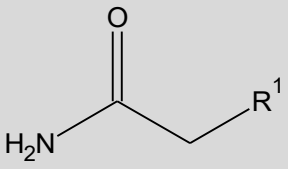
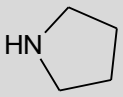
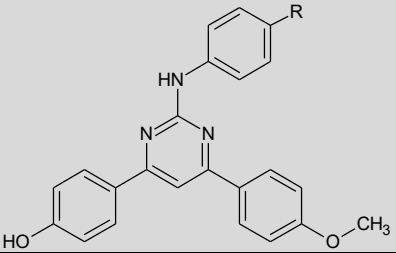
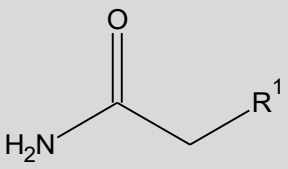
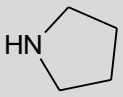
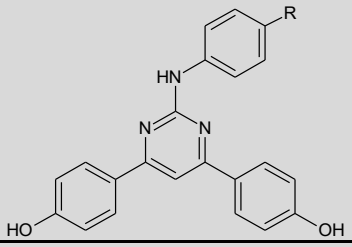
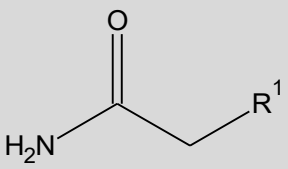
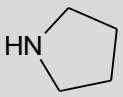
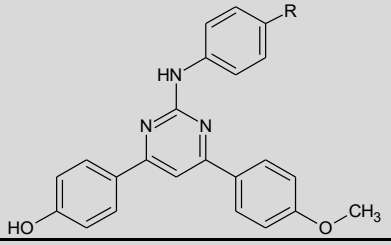
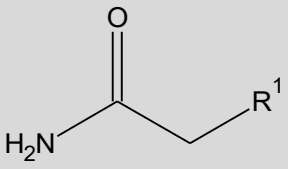
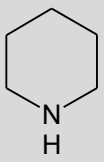
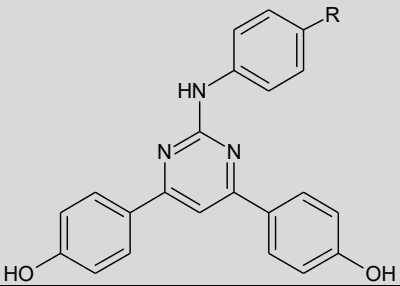
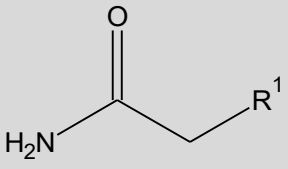
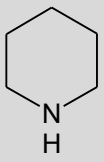
All computational studies were performed using Chem. Sketch software and NCSS-11 [35]. It provides sequential and stepwise multiple regression analysis with linear and parabolic relationships to generate the QSAR model. Table 1 corroborates the recommended values of QSAR equation [36]. Chemical structures of 31 (Table 2) analogs were drawn with the help of ACD labs software which helps in the calculation of physicochemical descriptors, and the log P has been drawn from the literature. NCSS-11 Software does multiple regressions. The regression coefficient (R^2) & the cross-regression coefficient (Q^2) validated the best model.

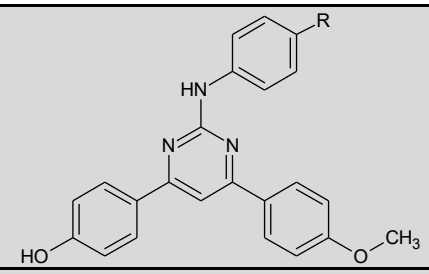
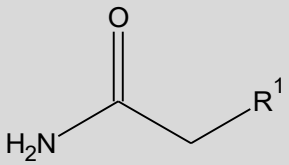
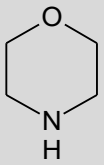
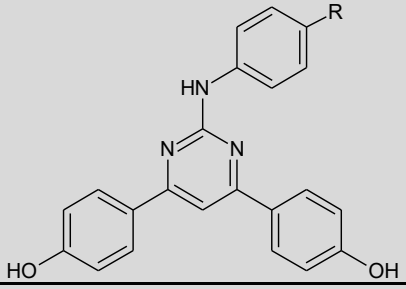
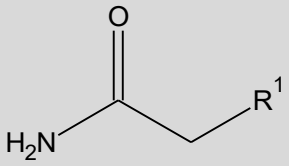
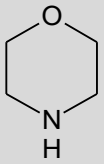
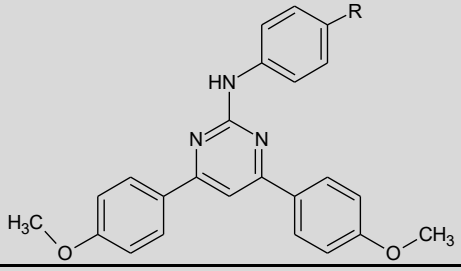
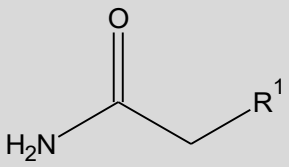
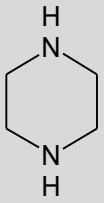
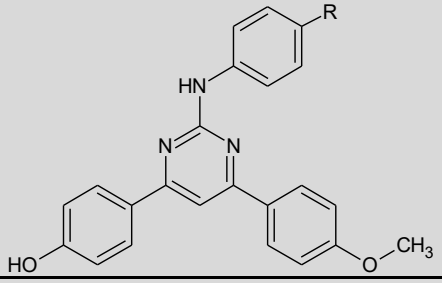
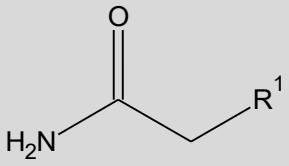
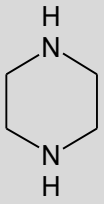
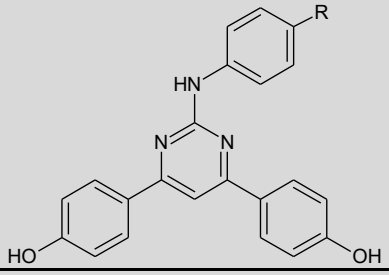
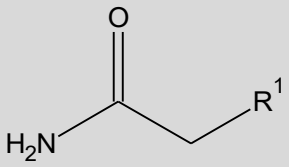
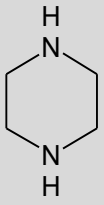
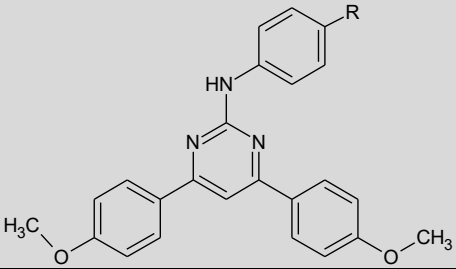
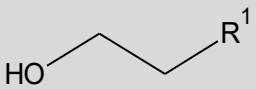
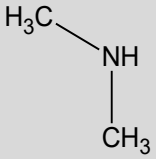
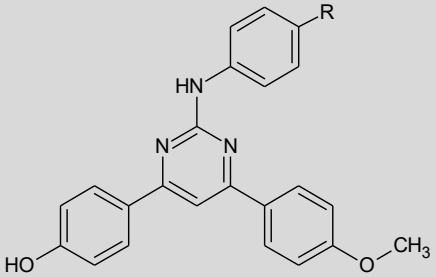
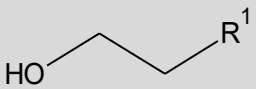
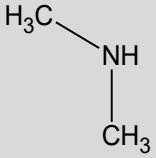
Table 1: Recommended values for evaluating the QSAR equation

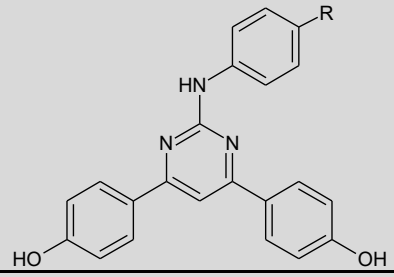
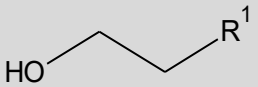
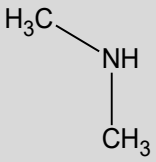
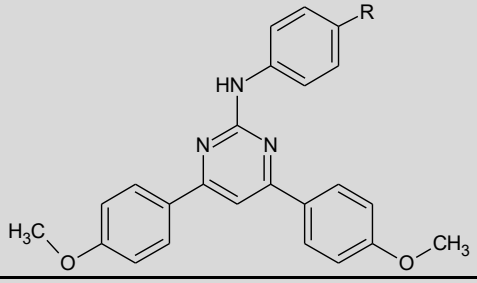
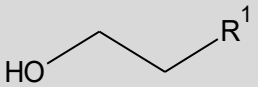
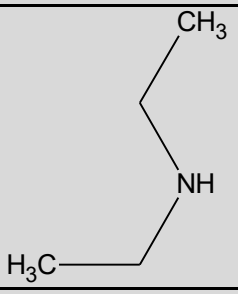
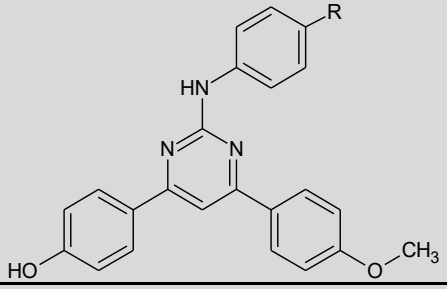
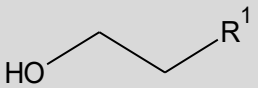
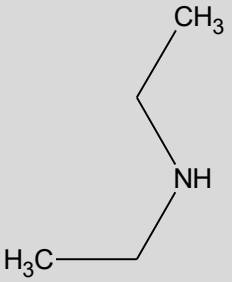
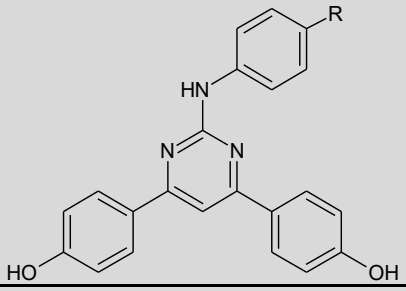
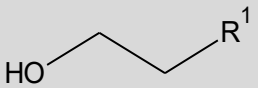
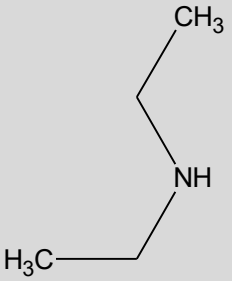
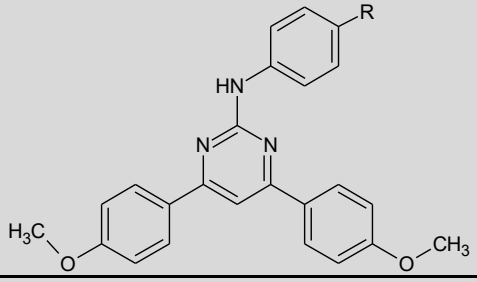
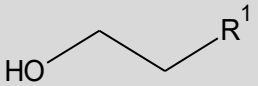
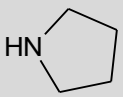
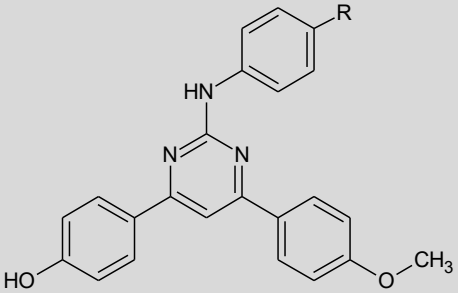
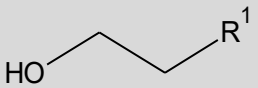
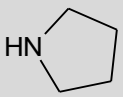
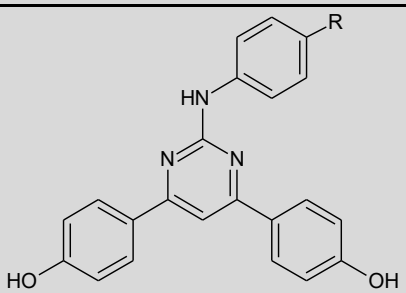
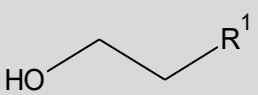
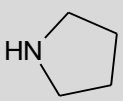
Character	Name	Value
R^2	Regression coefficient	>0.88
N	Number of compounds	31
Q^2	Cross-regression coefficient	0.7
R^2_A	Regression value	>0.86
F-ratio	Fishers ratio	>50.9
Se	Standard error	<0.10

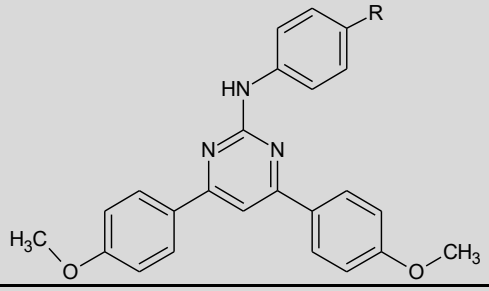
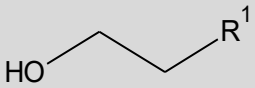
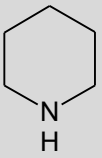
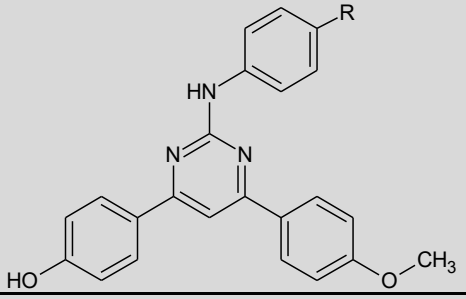
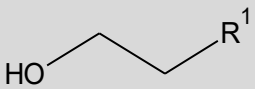
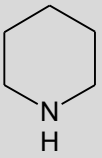
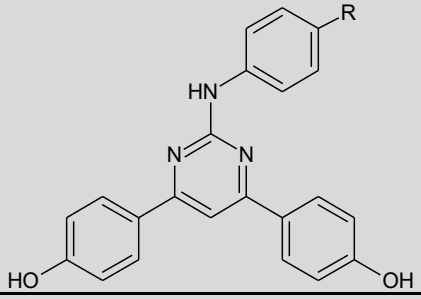
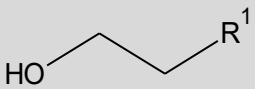
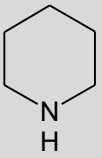
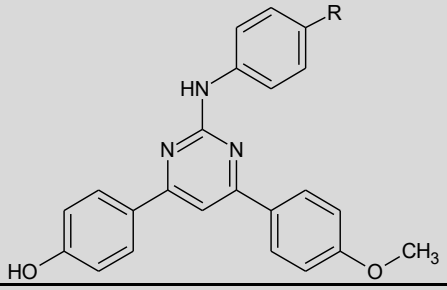
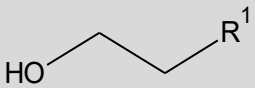
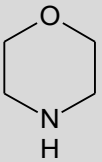
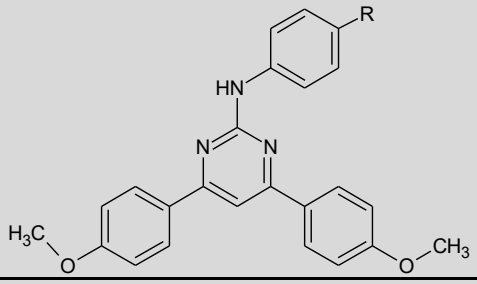
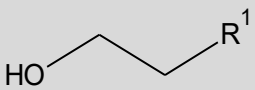
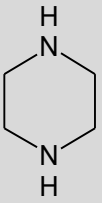
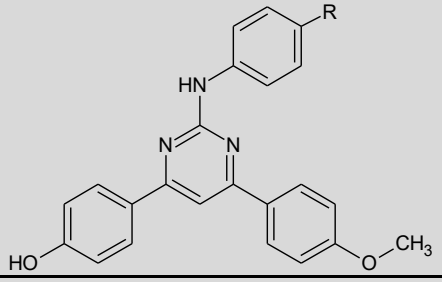
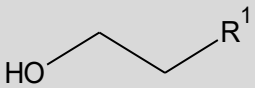
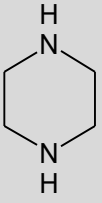
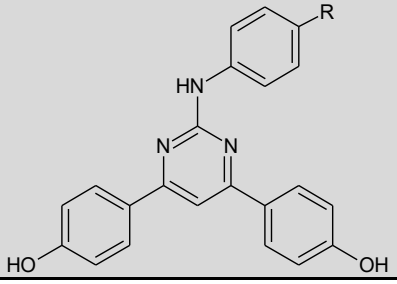
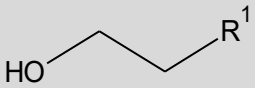
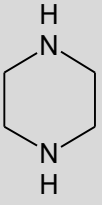
Table 2:- Structural details along with their LogP values.

Comp.No.	Parent Structure	R	R^1	LogP
1.				3.11
2.				2.52
3.				1.93

4.				3.58
				2.99
6.				2.68
7.				4.14
8.				3.55
9.				4.70
10.				4.11

11.				3.39
12.				2.79
13.				3.70
14.				3.11
15.				2.52
16.				5.16
17.				4.57

18.				3.98
19.				6.22
20.				5.63
21.				5.04
22.				5.74
23.				5.15
24.				4.56

25.				6.31
26.				5.72
27.				5.13
28.				4.06
29.				4.38
30.				3.78
31.				4.38

3. RESULTS AND DISCUSSION

The principle underlying QSAR computational methods are based on implementing and deriving a mathematical relationship that quantitatively links molecular descriptors with a macroscopic observable (physicochemical property or biological activity) for a series of matching chemical compounds using statistical analytics. A generalized mathematical form of QSAR is represented below.

Activity ¼ f (X) X

3.1 Physicochemical and structural properties [37]

Of vital significance is an analysis of structural data to detect the determining factors for the property or activity measured. In the concluding step, the developed models undergo various internal and external validation procedures to test their statistical significance, robustness, and predictive power [38]. The current challenge in the QSAR model development process is undeniably in developing a model that can accurately predict the therapeutic activities of new chemicals [39]. Pyrimidine, being indigenous to synthetic chemistry, acts as an enabler of structurally diverse derivatives, such as analogs originating from the substitution of the aryl ring, derivatization of the nitrogen in pyrimidine, and substitutions at the carbon at positions 2, 4, 5, and 6 [40]. Some novel pyrimidine derivatives have been studied previously for their anti-cancerous activity, including 2, 4-disubstituted [41-43] 2, 5-Disubstituted [44] and 4, 6-Disubstituted [45].4, 6-diaryl-2-pyrimidinamine derivatives exhibit appropriate activities in pharmacological studies [46-48]. Recently, QSAR modeling has been used to quantitatively validate the structure relationship of 2-anilinopyrimidine [49]. In the present study, the data set from Table 2 were subjected to regression analysis which subsequently gave a correlation matrix showing inter-correlation among the selected descriptors and the anti-breast cancer agent (log P) activity. The variable selection for multiple regression analysis has indicated the possibility of using only eight physicochemical parameters for modeling the anti-breast cancer activity (log P) (Table 3). Regression analysis is reported in Table 4, generated as a result of successive use of one to eight descriptors. The statistically significant models obtained from MLR analysis are reported below.

3.2 MLR

Multiple linear regressions is a statistical method used to investigate the relationship between one dependent variable and several independent variables, where the dependent variable is sometimes referred to as the response and the independent variables are referred to as the predictors. It is a mathematical method for reducing the difference between experimental and predicted values of biological activities [50, 51]. A regression diagnostics study was conducted on known chemical moieties to detect potential outliers by computing leverage values (hii) and studentized deleted residuals (r*i). In addition, the hat matrix's diagonal elements, known as leverage values, were used as identification tools for outlying X observations. The study output revealed leverage values, which were greater than 2(p1/n), were projected as outliers in terms of their X values [52]. In a similar study conducted, the magnitude of the studentized residuals, on the other hand, was used to identify outlying Y observations. The reliability and predictability of any model are indicated by a high correlation coefficient (R 14 0,768) and a low standard deviation error in prediction [53]. An *in silico* design study on pyridinyl imidazoles as c-jun N-terminal kinase-3 inhibitors were carried out, creating two-dimensional QSAR models. Briefly, the observed squared correlation coefficient (R2 > 0.6), a relative indicator of the fit quality, was one of the statistical criteria used to select the best model. Internal and external validations were utilized to assess the significance of created models and their precise prediction capacity [54].

3.3 *One-Variable Model*

Subsequent regression analysis indicated that IR as a correlating descriptor in the One-Variable Model is the best model for modeling the log P. This model is as follows:

$$\log P = -29.4590(\pm 5.3727)\ IR + 52.7663. \quad (1)$$

N=19, Se=0.1985, R^2=0.5090, R^2_A=0.4921, F-ratio=30.066.

Here and hereafter, N is the number of compounds, Se is the standard error, R^2 is the squared correlation coefficient, R^2A is adjusted R^2, and F-ratio is the Fishers ratio. The negative coefficient of IR indicates that the decrease in its magnitude will enhance the activity log P.

3.3.1 Two-Variable Model

When MW and PC were taken together, the model significantly improved all the parameters. For example, the R^2 value changes from 0.5090 to 0.7130. Similarly, improvement in adjusted R^2 also shows that the bi-parametric model has been obtained with improved statistics as follows:

$$\log P = -0.1428(\pm 0.0179)\ MW + 0.0601(\pm 0.0072)\ PC + 9.2142 \quad (2)$$

N =31, Se=0.1544, R^2 =0.7130, R^2_A=0.6925, F-ratio =34.778

3.3.2 Three-Variable Model

A Tri-parametric model was obtained when MW, MR, and PC were taken together. In this model, the R^2 shows good improvement.

$$\log P = -0.0971(\pm 0.0193)MW - 0.4951(\pm 0.1339)MR + 0.0994(\pm 0.0122)\,PC + 15.0878 \quad (3)$$

N=31, Se=0.1281, R^2 =0.8095, R^2A =0.7883, F=38.224

3.3.3 ***Four-Variable Model***

A tetra–parametric model was obtained when MW, PC, IR, and ST were taken together. In this model, the shows significant improvement.

$$\log P = -0.1536(\pm 0.0227\ MW + 0.0593(\pm 0.0100)\ PC - 74.3551(\pm 13.6973)IR + 0.3177\,(\pm 0.0504)ST + 119.8908 \quad (4)$$

N=31, Se =0.1006, R^2 =0.8869, R^2 A=0.8695, F= 50.991

To investigate the predictive ability of the proposed model, we have estimated various cross-validated parameters using the cross-validation method. These parameters demonstrated in Table 5 indicate that multi-parametric models are best for modeling.

3.4 MLR Results after deletion

One of the compounds is found to be an outlier; upon deletion of this compound no 6, a tetra metric model with the best modeling activity is noted mentioned as follows:

$$\log P = -0.1591368\ (\pm 0.01792478)\ MW + 0.06291306\ (\pm 0.007978716)\ PC - 67.71675(\pm 10.87715)\ IR + 0.2986334\ (\pm 0.0398771)\ ST + 108.8756 \quad (5)$$

N=30, Se =0.0781, R^2 =0.9290, R^2 A=0.9176, F= 81.763

This model is excellent for modeling the activity of the compounds of the present study, as reported in Table 5. To obtain further support in favor of our result, we have also estimated the activity of compounds and found that the estimated values are very close to the observed activity, as demonstrated in Table 6. The residue, i.e., the difference between the observed and estimated values, confirms these feelings. The final support in our favor is obtained by plotting the estimated activity against the observed activity, and such a plot shown is the best suitable model for modeling of 4, 6-diaryl-2-pyrimidinamine derivative used in the present study demonstrated in (Figures 4 and 5).

Table 3: -Values of the calculated physicochemical parameters along the objective used in the present study

Comp.No.	MW	MR	MV	PC	IR	ST	D	POL
1	483.56	141.60	393.80	1057.90	1.64	52.00	1.23	56.13
2	469.53	136.81	386.20	1016.30	1.67	57.90	1.27	54.23
3	455.51	132.01	342.70	974.60	1.70	65.40	1.33	52.33
4	497.59	146.07	401.20	1095.80	1.65	55.60	1.24	57.91
5	483.56	141.28	375.70	1054.20	1.68	61.90	1.29	56.00
6	509.60	148.64	409.30	1127.00	1.65	57.40	1.24	58.92
7	495.57	143.85	383.80	1083.60	1.67	63.50	1.29	57.02
8	481.54	139.05	358.20	1040.20	1.70	71.10	1.34	55.12
9	509.60	148.46	401.50	1123.70	1.66	61.30	1.27	58.85
10	495.57	143.66	375.90	1080.30	1.69	68.10	1.32	56.95
11	511.57	145.45	392.50	1103.90	1.66	62.50	1.30	57.66
12	497.54	140.65	367.00	1060.50	1.69	69.70	1.35	55.76
13	538.64	156.92	438.10	1193.40	1.64	55.00	1.23	62.20
14	524.61	152.12	412.60	1150.00	1.66	60.30	1.27	60.30
15	510.59	147.32	387.00	1106.60	1.69	66.80	1.32	58.40
16	470.56	138.64	401.20	1050.60	1.61	47.00	1.17	54.96
17	456.53	133.85	375.60	1009.00	1.63	52.00	1.21	53.06
18	442.51	129.05	350.10	967.30	1.66	58.20	1.26	51.16
19	498.51	147.91	434.20	1130.20	1.60	45.80	1.15	58.63
20	484.59	143.11	408.80	1088.50	1.62	50.30	1.18	56.73
21	470.56	138.32	383.10	1046.90	1.64	55.70	1.23	54.83
22	496.60	145.68	416.80	1113.90	1.62	51.00	1.19	57.75
23	482.57	140.89	391.20	1070.50	1.64	56.00	1.23	55.85
24	468.55	136.09	365.60	1027.10	1.67	62.20	1.28	53.95

25	510.63	150.29	434.50	1154.00	1.61	49.70	1.17	59.58
26	496.60	145.50	408.90	1110.60	1.63	54.30	1.21	57.68
27	482.57	140.70	383.30	1067.20	1.66	60.00	1.24	55.78
28	498.57	142.49	400.00	1090.90	1.63	53.30	1.25	56.48
29	525.64	153.96	445.50	1180.30	1.61	49.20	1.78	61.03
30	511.61	149.16	420.00	1136.90	1.63	53.60	1.22	59.13
31	497.59	144.36	394.40	1093.50	1.65	59.00	1.26	57.23

Table 4: -Regression parameters and quality of correlation when physicochemical parameters are used

Model No.	Parameters Used	A_i=(1-6)	B	Se	R^2	R^2A	F
1	IR	-29.4590±5.3727	52.7663	0.1985	0.5090	0.4921	30.064
2	ST	-0.1052±0.0257	10.2144	0.2258	0.3648	0.3429	16.657
3	MV	0.2014±0.0076	-3.8033	0.2544	0.1933	0.1654	6.947
4	D	-3.4341±0.0076	8.5052	0.2683	0.1030	0.0721	3.332
5	PC	0.0053±0.0038	-1.6675	0.2742	0.0630	0.0307	1.951
6	POL	0.0028±0.0854	0.5773	0.2807	0.0183	0.0079	0.541
7	MR	0.0248±0.0338	0.5843	0.2807	0.0183	0.0079	0.593
8	MW PC	-0.1428±0.0179 0.0601±0.0072	9.2142	0.1544	0.7130	0.6925	34.778
9	MR PC	-0.9248±0.1407 0.1115±0.0163	15.8026	0.1750	0.6315	0.6052	23.990
10	PC POL	0.1116±0.0163 -2.3356±0.3559	15.8467	0.1752	0.6308	0.6044	23.915
11	IR ST	-77.6576±18.4564 0.2109±0.7779	120.1582	0.1798	0.6109	0.5831	21.977
12	MW MR PC	-0.0971±0.0193 -0.4951±0.1339 0.0994±0.0122	15.0878	0.1281	0.8095	0.7883	38.224
13	MW PC POL	-0.0972±0.0193 0.0994±0.0122 -1.2476±0.3393	15.0955	0.1284	0.8007	0.7875	38.055
14	MW MV PC	-0.1923±0.0261 -0.0369±0.0151 0.0946±0.0155	10.8684	0.1423	0.7650	0.7389	29.300
15	MW PC ST	-0.1998±0.0302 0.0848±0.0128 0.0710±0.0313	6.4631	0.1442	0.7588	0.7320	28.313
16	MW PC IR ST	-0.1536±0.0227 0.0593±0.0100 -74.3551±13.6973 0.3177±0.0504	119.8908	0.1006	0.8869	0.8695	50.991
17	MW MR PC ST	-0.1563±0.0254 -0.5163±0.1166 0.1268±0.0138 0.0754±0.0242	12.3444	0.1145	0.8610	0.8397	40.277
18	MW PC ST POL	-0.1565±0.0254 0.1269±0.0138 0.0757±0.0243 -1.2879±0.2955	12.3548	0.1117	0.8606	0.8392	40.157

Table 5: -Regression and Modeling parameters and correlation quality after deleting one compound 6.

Model No.	Parameters Used	A_i=(1-6)	B	Se	R^2	R^2A	F
19	MW PC IR ST	-0.1591368±0.01792478 0.06291306±0.007978716 -67.71675±10.87715 0.2986334±0.0398771	108.8756	0.0781	0.9290	0.9176	81.763

Table 6:-Observed and estimated values logP using models 16 and 19 with residual reports.

Compd.No.	Using model no.16			Using model no.19	
	Obs.logP	Est.logP	Residual	Est.logP	Residual
1.	3.110	2.916	0.194	2.953	0.157

2.	2.520	2.249	0.271	2.299	0.221
3.	1.930	2.082	-0.152	2.114	-0.184
4.	3.580	3.408	0.172	3.502	0.078
5.	2.990	2.868	0.122	2.968	0.022
6.	2.680	3.985	-1.305	-	-
7.	4.140	4.017	0.123	4.061	0.079
8.	3.550	3.783	-0.233	3.801	-0.251
9.	4.700	4.285	0.415	4.371	0.329
10.	4.110	3.796	0.314	3.873	0.237
11.	3.390	3.188	0.202	3.171	0.219
12.	2.790	2.827	-0.037	2.791	-0.001
13.	3.700	3.442	0.258	3.608	0.092
14.	3.110	3.220	-0.110	3.339	-0.229
15.	2.520	2.636	-0.116	2.749	-0.229
16.	5.160	5.122	0.038	5.100	0.060
17.	4.570	4.911	-0.341	4.855	-0.285
18.	3.980	4.332	-0.352	4.282	-0.302
19.	6.220	5.910	0.310	5.979	0.241
20.	5.630	5.519	0.111	5.561	0.069
21.	5.040	5.436	-0.396	5.434	-0.394
22.	5.740	5.402	0.338	5.456	0.284
23.	5.150	5.085	0.065	5.097	0.053
24.	4.560	4.405	0.155	4.418	0.142
25.	6.310	5.955	0.355	6.035	0.275
26.	5.720	5.511	0.209	5.557	0.163
27.	5.130	4.674	0.456	4.730	0.400
28.	4.060	3.722	0.338	3.705	0.355
29.	4.380	5.050	-0.670	5.152	-0.772
30.	3.780	4.542	-0.762	4.614	-0.834
31.	4.380	4.351	0.029	4.373	0.007

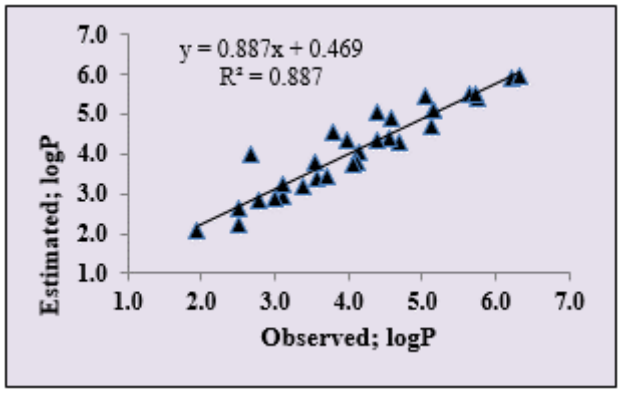


Fig. 4: Correlation between Observed and Estimated logP using model 16 (Table 6)

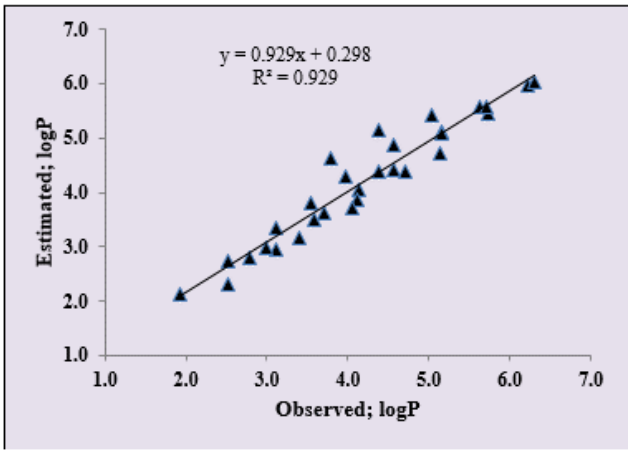


Fig. 5: Correlation between Observed and Estimated logP using model 19 (Table 5)

Table 6:-Cross – validated parameters for the best-obtained models of Tables 3 & 4.

Model no.	Parameters used	PRESS/SSY	R^2_{cv}	Spress	PSE
1	IR	0.9646	0.0354	0.1529	0.1430
8	MWPC	0.4023	0.5977	0.1210	0.1093
12	MWMRPC	0.2353	0.7647	0.1023	0.0891
16	MWPCIRST	0.1272	0.8728	0.0817	0.0685
19	MWPCIRST	0.1147	0.8853	0.1953	0.1692

4. CONCLUSION

A QSAR analysis using 31 compounds was successfully carried out to build statistically significant models with good correlative and predictive ability. The detailed structural investigation revealed that the anti-breast cancer activity is modeled best by using a multi-parametric model with physicochemical parameters (MW, PC, IR, and ST); upon applying the Leave One out (LOO) method, one compound which exists as an outlier leads to qualitative findings with a high degree of statistical significance and good predictive ability. It also provided important information about structural characteristics contributing to the inhibitory potency of 4, 6-diaryl-2-pyrimidinamine derivatives. Analysis of model parameters provided details of the structure-activity relationship, thus providing information regarding structural modification to design derivatives with better activity before synthesis. These models are also used to predict the anti-cancer activity of newly designed compounds.

5. ACKNOWLEDGEMENT

The authors are thankful to the Department of Biotechnology, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar University, Mullana, Ambala (Haryana), and Maharishi Markandeshwar Medical College and Hospital, Solan (Himachal Pradesh) for providing requisite resources to conduct the study. However, the current study is not a part of any financial assistance.

6. AUTHORS CONTRIBUTION STATEMENT

Suresh Verma, Neha Sharma, and Vipin Saini conceived and designed the study. Suresh Verma and Vipra Kohli conducted the study. Shivani, Sushma Gautam, Monika Sen, and Surbhi Chopra wrote the initial draft. Aditi Dwivedi, Jyoti Sharma, and Sonali Bhambhani validated the data. Neha Sharma wrote the final manuscript with minor amendments. Satinder Singh Minhas approved the final draft.

7. CONFLICT OF INTEREST

Conflict of interest declared none.

8. REFERENCES

1. Siegel P, Zou MJ, Jemal Z. A. Cancer statistics. J Clin. 2014;64(1):9-29.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61(2):69-90. doi: 10.3322/caac.20107, PMID 21296855.
3. Ndagi U, Mhlongo N, Soliman ME. Metal complexes in cancer therapy – an update from a drug design perspective. Drug Des Dev Ther. 2017;11:599-616. doi: 10.2147/DDDT.S119488, PMID 28424538.
4. Fridlender M, Kapulnik Y, Koltai H. Plant-derived substances with anticancer activity: from folklore to practice. Front Plant Sci. 2015;6:799. doi: 10.3389/fpls.2015.00799, PMID 26483815.
5. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global cancer in women: burden and trends. Cancer Epidemiol Biomarkers Prev. 2017;26(4):444-57. doi 10.1158/1055-9965.EPI-16-0858, PMID 28223433.
6. Akram M, Iqbal M, Daniyal M, Khan AU. Awareness and current knowledge of breast cancer. Biol Res. 2017;50(1):33. doi: 10.1186/s40659-017-0140-9, PMID 28969709.
7. Siu AL, US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2016;164(4):279-96. doi: 10.7326/M15-2886, PMID 26757170.
8. Watson M. Assessment of suspected cancer. InnovAiT. 2008;1(2):94-107. doi: 10.1093/innovait/inn001.
9. Hayes J, Richardson FC. Population attributable risks for modifiable lifestyle factors and breast cancer in New Zealand women. Int Med J. 2013;43(11):1198-204.
10. Costa WA, Monteiro MN, Queiroz JF, Gonçalves AK. Pain and quality of life in breast cancer patients. Clinics (Sao Paulo). 2017;72(12):758-63. doi 10.6061/clinics/2017(12)07, PMID 29319722.
11. Narod SA. Reflections on screening mammography and the early detection of breast cancer: a Countercurrents Series. Curr Oncol. 2014;21(5):210-4. doi 10.3747/co.21.2068, PMID 25301304.
12. Yersal O. Biological subtypes of breast cancer: prognostic and therapeutic implications. World J Clin Oncol. 2014;5(3):412. doi: 10.5306/wjco.v5.i3.412.
13. Heron DE, Shogan JE, Mucenski JW. Innovations in chemotherapy and radiation therapy: implications and opportunities for the Asia-Pacific Rim. Biomed Imaging Interv J. 2008;4(3):e40. doi: 10.2349/biij.4.3.e40, PMID 21611006.
14. Kharfan-Dabaja MA, Fahed R, Hussein M, Santos ES. The evolving role of monoclonal antibodies in the treatment of chronic lymphocytic leukemia. Expert Opin Investig Drugs. 2007;16(11):1799-815. doi: 10.1517/13543784.16.11.1799, PMID 17970639.
15. Smith RP, Heron DE, Huq MS, Yue NJ. Modern radiation treatment planning and delivery—from röntgen to real-time. Hematol Oncol Clin North Am. 2006;20(1):45-62. doi 10.1016/j.hoc.2006.01.006, PMID 16580556.
16. Eandi JA, Link BA, Nelson RA, Josephson DY, Lau C, Kawachi MH et al. Robotic-assisted laparoscopic salvage prostatectomy for radiation-resistant prostate cancer. J Urol. 2010;183(1):133-7. doi: 10.1016/j.juro.2009.08.134, PMID 19913249.

17. Camacho LH. Novel therapies targeting the immune system: CTLA4 blockade with tremelimumab (CP-675,206), a fully human monoclonal antibody. Expert Opin Investig Drugs. 2008;17(3):371-85. doi: 10.1517/13543784.17.3.371, PMID 18321236.
18. Arruebo M, Vilaboa N, Sáez-Gutierrez B, Lambea J, Tres A, Valladares M et al. Assessment of the evolution of cancer treatment therapies. Cancers. 2011;3(3):3279-330. doi: 10.3390/cancers3033279, PMID 24212956.
19. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424. doi: 10.3322/caac.21492, PMID 30207593.
20. O'Gorman CO, Lucas R, Taylor B. Environmental risk factors for multiple sclerosis: a review with a focus on molecular mechanisms. Int J Mol Sci. 2012;13(12):11718-52. doi: 10.3390/ijms130911718.
21. Sever R, Brugge JS. Signal transduction in cancer. Cold Spring Harb Perspect Med. 2015;5(4):1. doi: 10.1101/cshperspect.a006098, PMID 25833940.
22. Ogino S, Nowak JA, Hamada T, Phipps AI, Peters U, Milner DA et al. Integrative analysis of exogenous, endogenous, tumor, and immune factors for precision medicine. Gut. 2018;67(6):1168-80. doi: 10.1136/gutjnl-2017-315537, PMID 29437869.
23. Sharma V, Chitranshi N, Agarwal AK. Significance and biological importance of pyrimidine in the microbial World. Int J Med Chem. 2014;2014:202784. doi: 10.1155/2014/202784, PMID 25383216.
24. Cherukupalli S, Karpoormath R, Chandrasekaran B, Hampannavar GA, Thapliyal N, Palakollu VN. An insight into synthetic and medicinal aspects of pyrazole [1,5-a]pyrimidine scaffold. Eur J Med Chem. 2017;126:298-352. doi: 10.1016/j.ejmech.2016.11.019, PMID 27894044.
25. Suresh K, Rekha K. Anti-influenza activity of monoterpene-containing substituted coumarins using physicochemical parameters. Org Chem Curr Res. 2018;7(187).
26. Suresh K, Vishali G. Phosphodiesterase 4 (PDE4) inhibition of 4, 5, 6, 7- tetrahydro-1H – 1, 2 – doxepin – 7- one Derivatives Using Physicochemical Parameters. Curr. Trends in biotech. Chem Res. 2018;8(1):28.
27. Hadaji EG, Ouammou A, Bouachrine M. QSAR study of anthra. Adv Chemother. 2018; 1;1pyrazol-6(2H)-one Derivatives as Potential Anticancer Agents Using Statistical Methods.
28. Hadaji EG, Bourass M, Ouammou A, Bouachrine M. 3D-QSAR models to predict anti-cancer activity on a series of protein P38 MAP kinase inhibitors. J Taibah Univ Sci. 2017;11(3):392-407. doi: 10.1016/j.jtusci.2016.05.006.
29. Perkins R, Fang H, Tong W, Welsh WJ. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. Environ Toxicol Chem. 2003;22(8):1666-79. doi: 10.1897/01-171, PMID 12924569.
30. Sharma MC, Kohli DV, Sharma S. Molecular modeling studies of substituted 2,4,5-Trisubstituted triazolinones aryl and nonaryl derivatives as angiotensin II AT1Receptor antagonists. J Chem. 2013;2013:1-14. doi: 10.1155/2013/427181.
31. Alam S, Khan F. 3D-QSAR studies on Maslinic acid analogs for Anticancer activity against Breast Cancer cell line MCF-7. Sci Rep. 2017;7(1):6019. doi: 10.1038/s41598-017-06131-0, PMID 28729623.
32. Putri DEK, Pranowo HD, Haryadi W. Study on anti-tumor activity of novel 3-substituted 4 aniline-coumarin derivatives using quantitative structure-activity relationship (QSAR). Mater Sci Forum. 2019;948:101-8. doi: 10.4028/www.scientific.net/MSF.948.101.
33. Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A. In-silico studies of some oxadiazoles derivatives as anti-diabetic compounds. J King Saud Univ Sci. 2018;32(1):432.
34. Abdullahi M, Uzairu A, Shallangwa GA, Mamza P, Arthur DE, Ibrahim MT. In-silico modeling studies on some C14-urea-tetrandrine derivatives as potent anticancer agents against prostate (PC3) cell line. J King Saud Univ Sci. 2019;32(1), 770:2020.
35. Hussain KA, Radhi Wisam AH. Quantitative structure-activity relationships (QSAR) study and improving it of some Schiff-base ligands as anticancer for prostate cancer. J Chem Pharm Res. 2012;4(3):1702.
36. Lawal HA, Uzairu A, Uba S. QSAR, molecular docking, design, and pharmacokinetic analysis of 2-(4-fluorophenyl) imidazole-5-ones as anti-breast cancer drug compounds against MCF-7 cell line. J Bioenerg Biomembr. 2020;52(6):475-94. doi: 10.1007/s10863-020-09858-0, PMID 33247393.
37. Free SM, Wilson JW. A mathematical contribution to structure-activity studies. J Med Chem. 1964;7(4):395-9. doi: 10.1021/jm00334a001, PMID 14221113.
38. Acharya C, Coop A, Polli JE, MacKerell AD. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Curr Comput Aided Drug Des. 2011;7(1):10-22. doi: 10.2174/157340911793743547, PMID 20807187.
39. Yousefinejad S, Hemmateenejad B. Chemometrics tools in QSAR/QSPR studies: a historical perspective. Chemometrics Intell Lab Syst. 2015;149:177-204. doi: 10.1016/j.chemolab.2015.06.016.
40. Selvam TP, James CR, Dniandev PV, Valzita SK. A mini-review of pyrimidine and fused pyrimidine marketed drugs. Res Pharm. 2012;2(4):01.
41. Xu Y, Hao SY, Zhang XJ, Li WB, Qiao XP, Wang ZX et al. Discovery of novel 2, 4-disubstituted pyrimidines as aurora kinase inhibitors. Bioorg Med Chem Lett. 2020;30(3):126885. doi: 10.1016/j.bmcl.2019.126885, PMID 31862411.
42. Hu G, Wang C, Xin X, Li S, Li Z, Zhao Y et al. Design, synthesis, and biological evaluation of novel 2, 4-diaminopyrimidine derivatives as potent antitumor agents. New J Chem. 2019;43(25):10190-202. doi: 10.1039/C9NJ02154J.
43. Kamal A, Dastagiri D, Ramaiah MJ, Reddy JS, Bharathi EV, Reddy MK et al. Synthesis and apoptosis-inducing ability of new aniline substituted pyrimidine sulfonamides as potential anticancer agents. Eur J Med Chem. 2011;46(12):5817-24. doi: 10.1016/j.ejmech.2011.09.039, PMID 22000207.
44. Reddy OS, Suryanarayana CV, Narayana KJP, Anuradha V, Babu BH. Synthesis and cytotoxic evaluation for some new 2, 5-disubstituted pyrimidine derivatives for

anticancer activity. Med Chem Res. 2015;24(5):1777-88. doi: 10.1007/s00044-014-1276-6.
45. Mule SNR, Nurbhasha S, Kolla JN, Jadav SS, Jayaprakash V, Bhavanam LR et al. Synthesis, biological screening, and molecular docking studies of novel 4, 6-pyrimidine derivatives as EGFR-TK inhibitors. Med Chem Res. 2016;25(11):2534-46. doi: 10.1007/s00044-016-1668-x.
46. Guan XW, Wang HQ, Ban WW, Chang Z, Chen HZ, Jia L et al. Novel HDAC inhibitor Chidamide synergizes with rituximab to inhibit diffuse large B-cell lymphoma tumor growth by upregulating CD20. Cell Death Dis. 2020;11(1):20. doi: 10.1038/s41419-019-2210-0, PMID 31907371.
47. Han YT, Kim K, Son D, An H, Kim H, Lee J et al. Fine-tuning of 4,6-biphenyl-2-(3-alkoxy aniline)pyrimidine focusing on the activity-sensitive aminoalkoxy moiety for a therapeutically useful inhibitor of receptor for advanced glycation end products (RAGE). Bioorg Med Chem Lett. 2015;23(3):579-87.
48. Liu L, Tang Z, Wu C, Li X, Huang A, Lu X et al. Synthesis and biological evaluation of 4,6-diaryl-2-pyrimidinamine derivatives as anti-breast cancer agents. Bioorg Med Chem Lett. 2018;28(6):1138-42. doi: 10.1016/j.bmcl.2017.12.066, PMID 29482944.
49. Abdulrahman HL, Uzairu A, Uba S. *In silico* studies of some 2-anilinopyrimidine derivatives as anti-triple-negative breast cancer agents. J Basic Appl Sci. 2020;9:13.
50. Hadaji EG, Bourass M, Ouammou A, Bouachrine M. Organic compounds based on (E)-N-Aryl-2-ethene-sulfonamide as microtubule targeted agents in prostate cancer: QSAR study. Adv Phys Chem. 2017;2017:1-14. doi: 10.1155/2017/7629056.
51. Mata J. Interpretation of concrete dam behavior with artificial neural network and multiple linear regression models. Eng Struct. 2011;33(3):903-10. doi: 10.1016/j.engstruct.2010.12.011.
52. Rawlings JO, Pantula SG, Dickey DA. Applied regression analysis: a research tool. 2nd ed. Springer Science+Business Media; 2001.
53. Tropsha A, Gramatica P, Gombar VK. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. The Importance of Being Earnest. QSAR Comb Sci. 2003;22(1):69-77. doi: 10.1002/qsar.200390007.
54. El Aissouq A, Toufik H, Stitou M, Ouammou A, Lamchouri F. In silico design of novel tetra-substituted pyridinyl imidazole derivatives as c-Jun N-terminal kinase-3 inhibitors, using 2D/3D-QSAR studies, molecular docking, and ADMET prediction. Int J Pept Res Ther. 2020;26(3):1335-51. doi: 10.1007/s10989-019-09939-8.