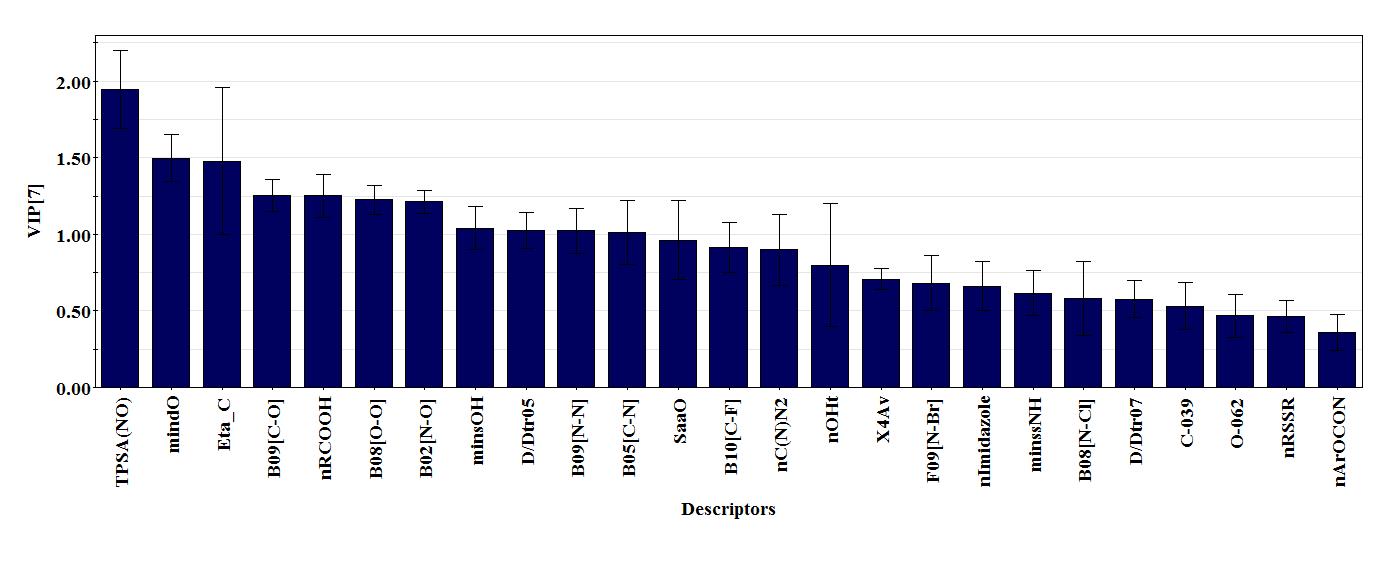
**Supplementary Information**

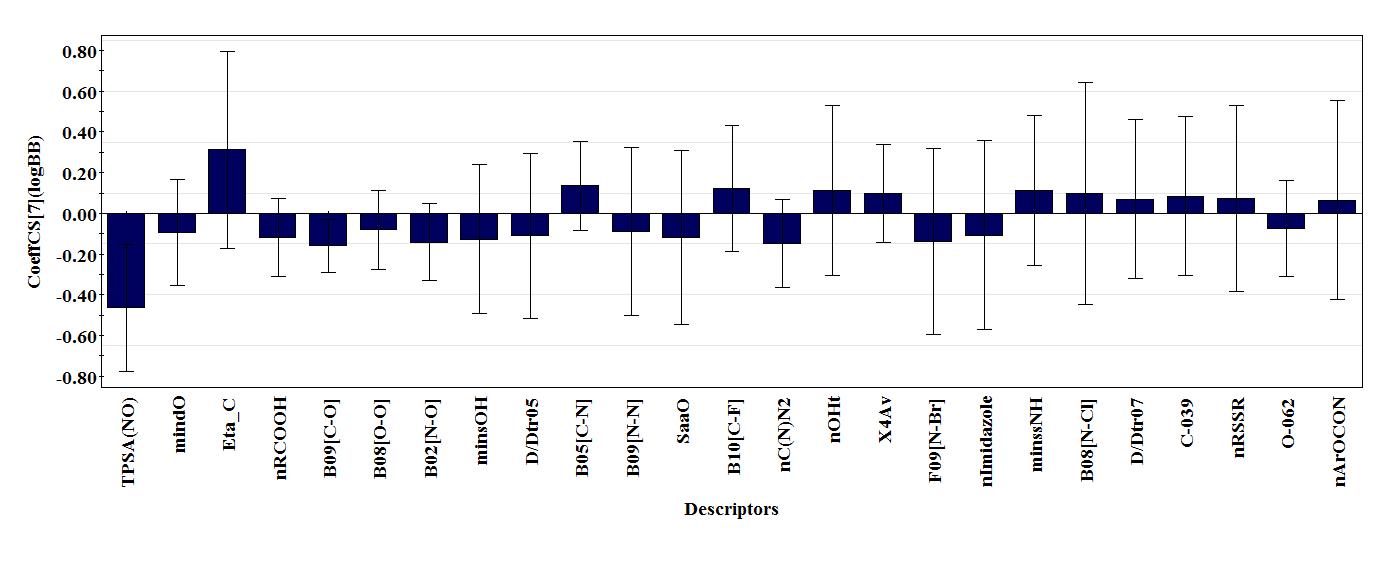
**Innovative Strategies for the Quantitative Modeling of Blood-Brain Barrier (BBB) Permeability: Harnessing the Power of Machine Learning-based q-RASAR Approach**

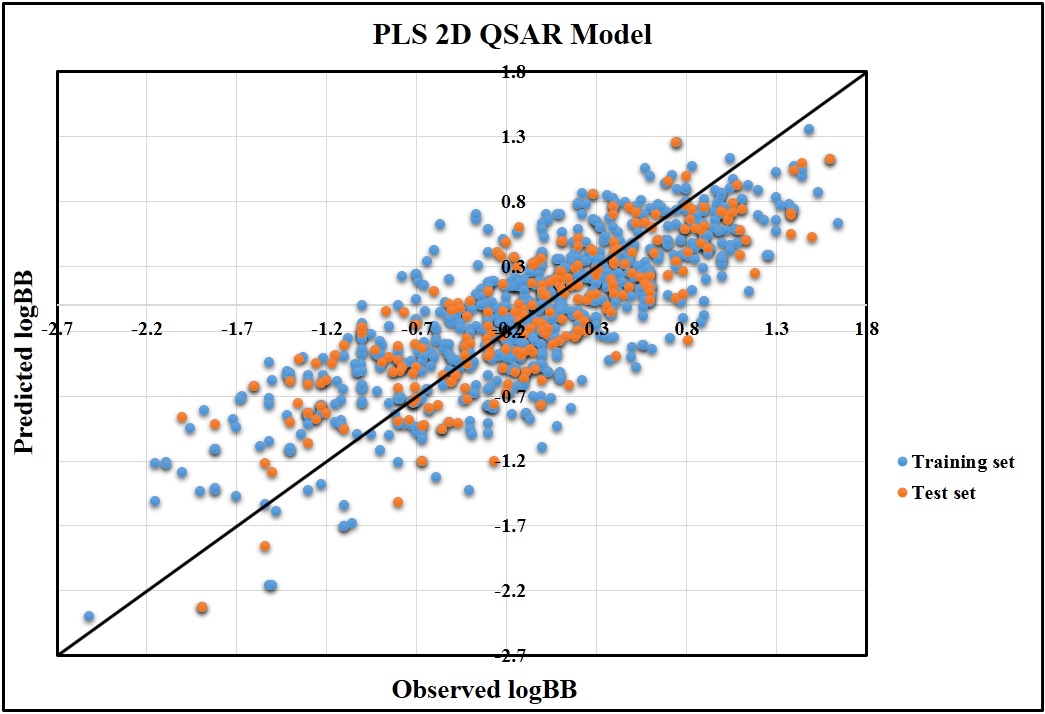
[**Vinay Kumar**](https://orcid.org/0000-0002-6809-7633)**,** [**Arkaprava Banerjee**](https://orcid.org/0000-0001-8468-0784)**,** [**Kunal Roy**](https://orcid.org/0000-0003-4486-8074)**\***

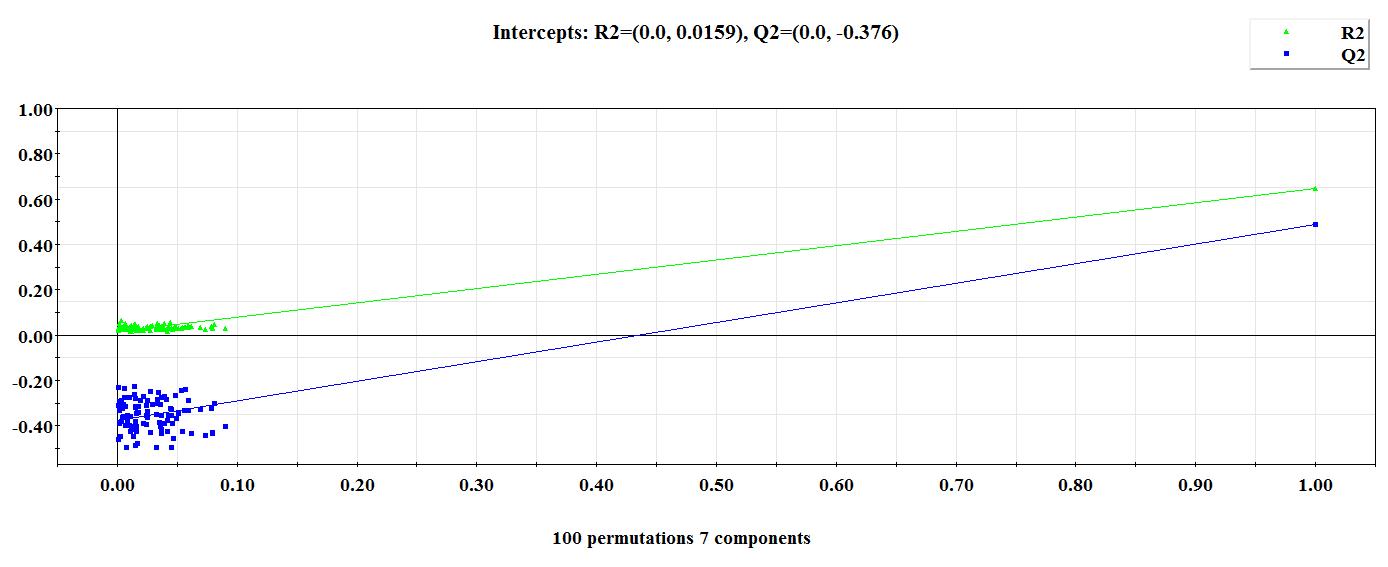
*Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India*

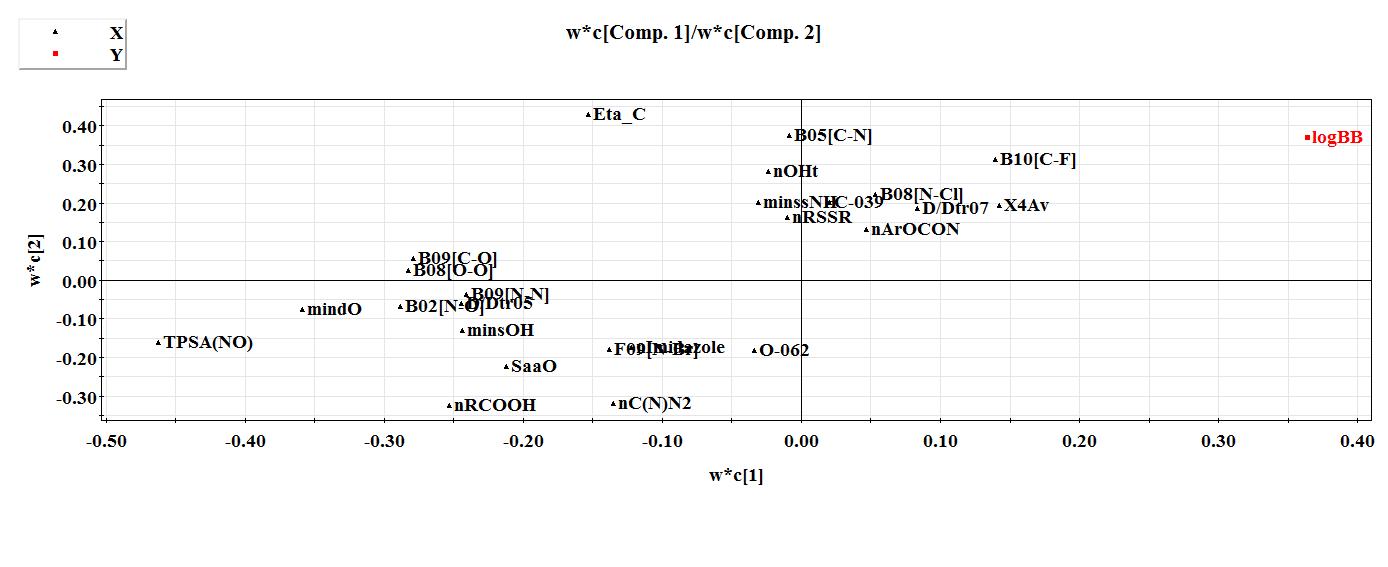
*E-mail:*[*kunalroy\_in@yahoo.com*](mailto:kunalroy_in@yahoo.com)*;*[*kunal.roy@jadavpuruniversity.in*](mailto:kunal.roy@jadavpuruniversity.in)

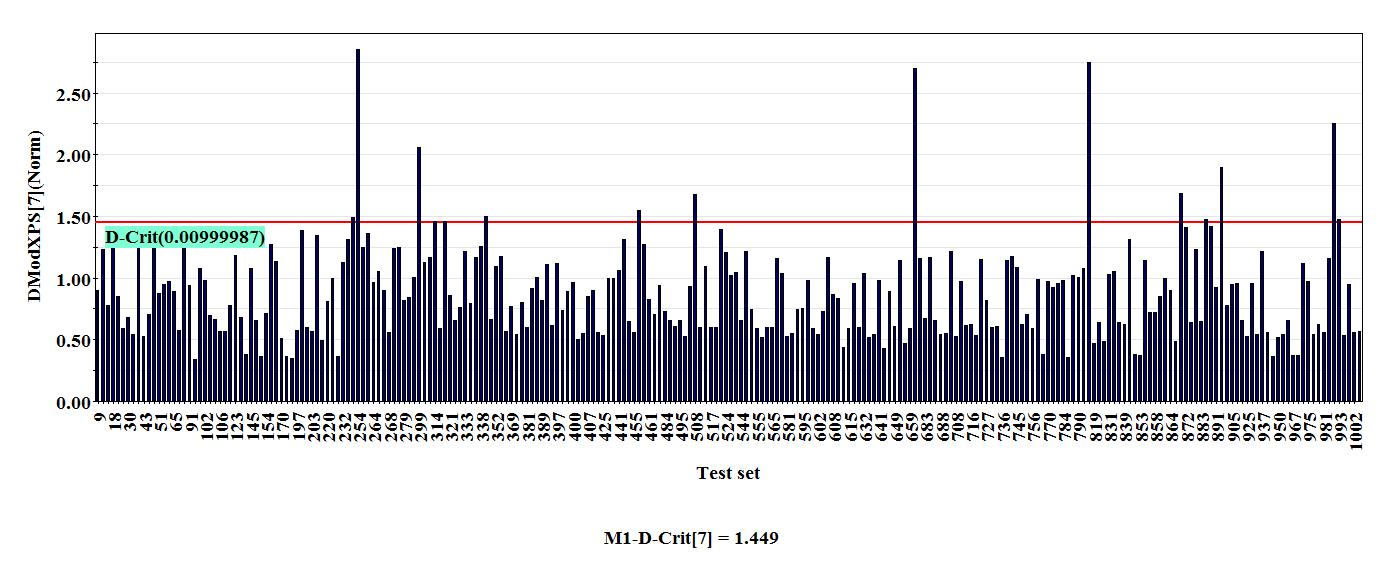
**Fig. S1** Variable importance plot of 2D QSAR model.

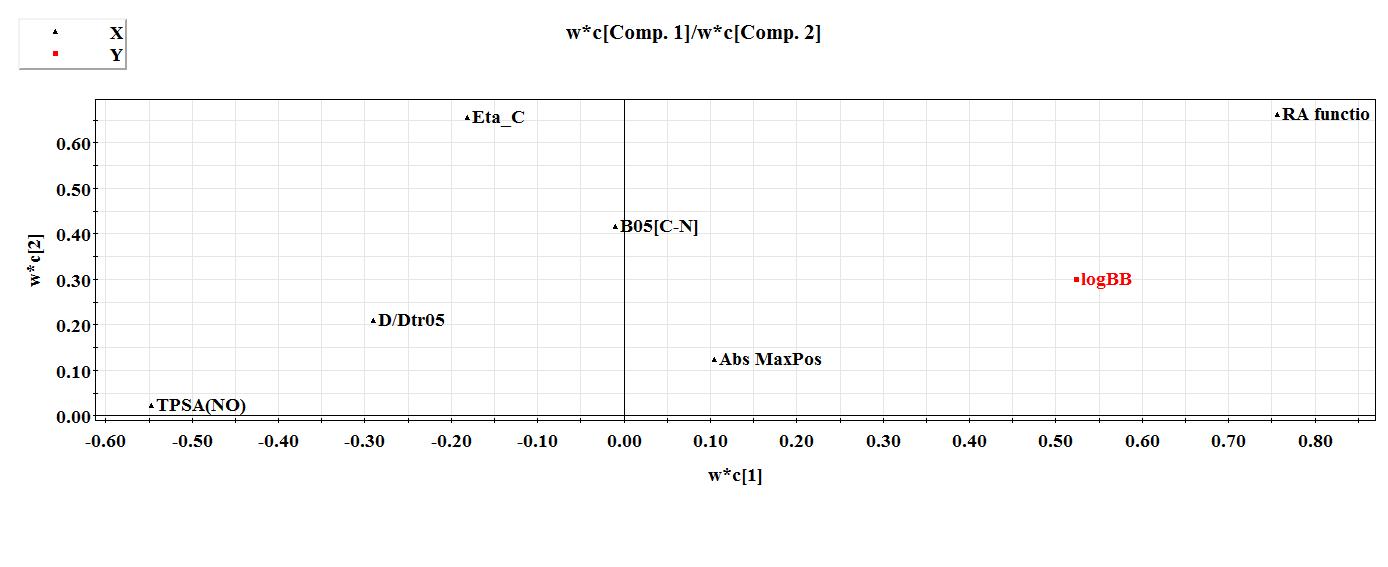
**Fig. S2** Regression coefficient plot of 2D QSAR model.

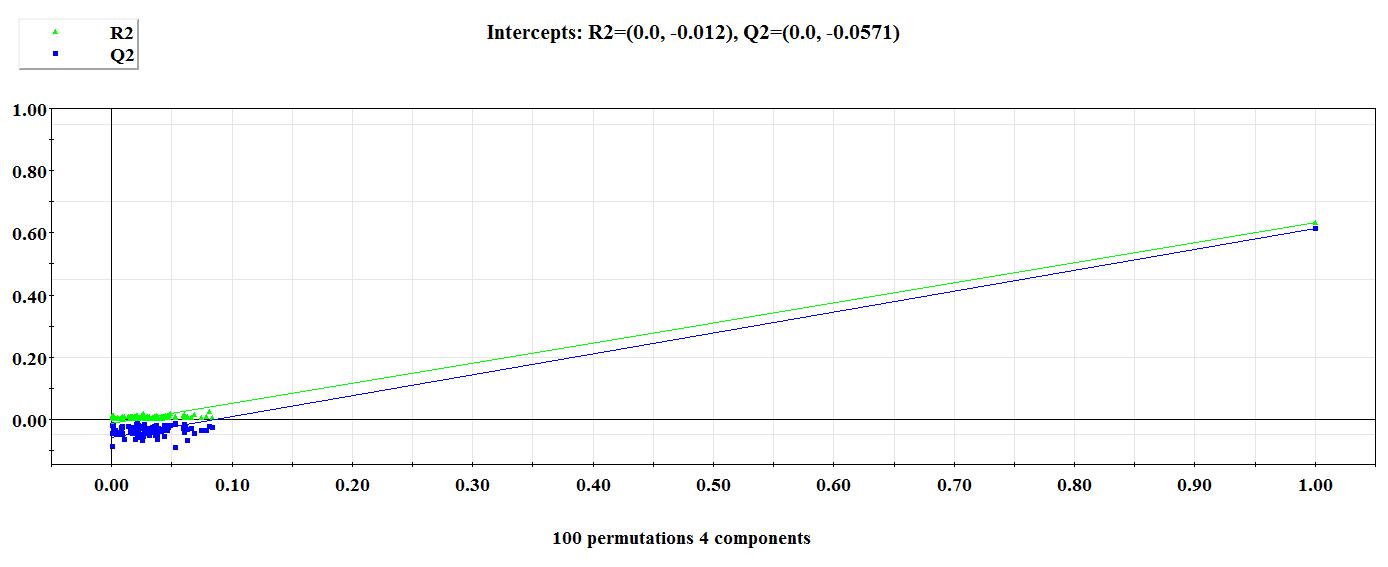
**Fig. S3** Scatter plot of 2D QSAR model.

**Fig. S4** Y-randomization plot of 2D QSAR model.

**Fig. S5** Loading plot of 2D QSAR model.

**Fig. S6** Applicability domain plot of 2D QSAR model for test set compounds.

**Fig. S7** Loading plot of q-RASAR model.

**Fig. S8** Y-randomization plot of q-RASAR model.

**Table S1** List of structural and physicochemical descriptors used to develop the 2D QSAR and q-RASAR models.

|  |  |  |  |
| --- | --- | --- | --- |
| **Descriptors** | **Class** | **Description** | **Contribution** |
| Eta\_C | ETA indices | eta composite index | +Ve |
| B05[C-N] | 2D Atom Pairs | Presence/absence of C – N at topological distance 5 | +Ve |
| B10[C-F] | 2D Atom Pairs | Presence/absence of C – F at topological distance 10 | +Ve |
| minssNH | Atom-type E-state indices | Mimimum ssNH | +Ve |
| nOHt | Functional group counts | number of tertiary alcohols | +Ve |
| X4Av | Connectivity indices | average valence connectivity index of order 4 | +Ve |
| B08[N-Cl] | 2D Atom Pairs | Presence/absence of N – Cl at topological distance 8 | +Ve |
| C-039 | Atom-centred fragments | Ar-C(=X)-R | +Ve |
| nRSSR | Functional group counts | number of disulfides | +Ve |
| D/Dtr07 | Ring descriptors | distance/detour ring index of order 7 | +Ve |
| nArOCON | Functional group counts | number of (thio-) carbamates (aromatic) | +Ve |
| O-062 | Atom-centred fragments | O- (negatively charged) | -Ve |
| B08[O-O] | 2D Atom Pairs | Presence/absence of O – O at topological distance 8 | -Ve |
| B09[N-N] | 2D Atom Pairs | Presence/absence of N – N at topological distance 9 | -Ve |
| mindO | Atom-type E-state indices | Mimimum dO | -Ve |
| nImidazoles | Functional group counts | number of Imidazoles | -Ve |
| D/Dtr05 | Ring descriptors | distance/detour ring index of order 5 | -Ve |
| nRCOOH | Functional group counts | number of carboxylic acids (aliphatic) | -Ve |
| SaaO | Atom-type E-state indices | Sum of aaO E-states | -Ve |
| minsOH | Atom-type E-state indices | Mimimum sOH | -Ve |
| F09[N-Br] | 2D Atom Pairs | Frequency of N – Br at topological distance 9 | -Ve |
| B02[N-O] | 2D Atom Pairs | Presence/absence of N – O at topological distance 2 | -Ve |
| nC(=N)N2 | Functional group counts | number of guanidine derivatives | -Ve |
| B09[C-O] | 2D Atom Pairs | Presence/absence of C – O at topological distance 9 | -Ve |
| TPSA(NO) | Molecular properties | topological polar surface area using N,O polar contributions | -Ve |
| *RA function(LK)* | RASAR | Read-Across derived composite function | **+Ve** |
| Abs MaxPos-MaxNeg | RASAR | Absolute difference between the MaxPos and MaxNeg values | **+Ve** |

**Table S2** Comparison of the different models (2D QSAR, and machine learning (ML) based q-RASAR) in terms of their different validation metrics. (**Note:** Descriptors in PLS and q-RASAR models are arranged according to their importance (VIP)). (**Note** = **RF**: Random Forest, **GB**: Gradiant Boost, **EGB**: Extreme Gradiant Boost, **SVM**: Support Vector Machine, **LSVM**: Linear Support Vector Machine, **RR**: Ridge Regression, **LV**: Latent variable).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Internal validation parameters** | | | | **External validation parameters** | | | |
| **nTrain** | **R2** | **Q2**Loo | **MAETrain** | **nTest** | **Q2F1** | **Q2F2** | **MAETest** |
| **2D QSAR** **model** | | | | | | | | |
| PLS (LV: 7) | 764 | 0.645 | 0.612 | 0.356 | 248 | 0.661 | 0.661 | 0.338 |
|  | | | | | | | | |
| **Chemical Read-Across prediction** | | | | | | | | |
| Similarity based method: Laplacian kernel, Sigma (σ): 1, No. of similar train comp.: 10, Gamma (γ): 1 | | | | | | 0.638 | 0.638 | 0.310 |
| **ML based q-RASAR model** | | | | | | | | |
| **PLS (LV: 4)** | **764** | **0.634** | **0.627** | **0.334** | **248** | **0.697** | **0.697** | **0.305** |
|  | | | | | | | | |
| **Other machine learning (ML) algorithms** | | | | | | | | |
| RF | 764 | 0.950 | 0.629 | 0.158 | 248 | 0.679 | 0.679 | 0.423 |
| AdaBoost | 0.670 | 0.590 | 0.461 | 0.628 | 0.628 | 0.476 |
| GB | 0.817 | 0.601 | 0.328 | 0.679 | 0.679 | 0.427 |
| EGB | 0.853 | 0.555 | 0.293 | 0.653 | 0.653 | 0.447 |
| SVM | 0.697 | 0.635 | 0.401 | 0.671 | 0.671 | 0.425 |
| LSVM | 0.620 | 0.616 | 0.456 | 0.681 | 0.681 | 0.424 |
| **RR** | **0.635** | **0.627** | **0.462** | **0.698** | **0.698** | **0.426** |

**Table S3.** Predictive outcomes of structurally modified compounds using the developed q-RASAR model, juxtaposed with experimental data of the original compound for comparison.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Structure** | **AD status** | **logBB (Pred.)** |
| **Chlorpromazine**   **logBB (Exp.): 1.36** | | | |
| CPZ-M1 |  | Inside | 0.0816 |
| CPZ-M2 |  | Inside | 0.344 |
| CPZ-M3 |  | Inside | 0.0820 |
| CPZ-M4 |  | Inside | 0.033 |
| **Fluphenazine** **logBB (Exp.): 1.51** | | | |
| FPZ-M1 |  | Inside | 0.366 |
| FPZ-M2 |  | Inside | 0.261 |
| FPZ-M3 |  | Inside | 0.761 |
| FPZ-M4 |  | Inside | 0.294 |
| **Promazine** **logBB (Exp.): 1.23** | | | |
| PZ-M1 |  | Inside | 0.504 |
| PZ-M2 |  | Inside | 0.795 |
| PZ-M3 |  | Inside | 0.694 |
| PZ-M4 |  | Inside | 0.405 |
| **Promethazine**    **logBB (Exp.): 1.3** | | | |
| PMZ-M1 |  | Inside | -0.011 |
| PMZ-M2 |  | Inside | 0.070 |
| PMZ-M3 |  | Inside | 0.819 |
| PMZ-M4 |  | Inside | 0.427 |