**Machine learning-, rule- and pharmacophore-based classification on the inhibition of**

**P-glycoprotein and NorA**

Trieu-Du Ngo, Thanh-Dao Tran, Minh-Tri Le, and Khac-Minh Thai\*

Department of Medicinal Chemistry, Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, 41 Dinh Tien Hoang, Dist. 1, Ho Chi Minh City, Viet Nam

\*Corresponding author. Tel.: +84-909-680-385; Fax: +84-8-3822-5435; E-mail: thaikhacminh@gmail.com; [thaikhacminh@uphcm.edu.vn](mailto:thaikhacminh@uphcm.edu.vn)

**Supplementary material 1**

SM1-Table 1. List of 24 attributes serving as independent variables in the classification models.

SM1-Table 2. List of 21 attributes serving as variables in perceptual mapping.

SM1-Table 3. Average values of 18 descriptors computed for 4 activity classes in the development of the MDS map.

SM1-Table 4. Frequency of 3 fingerprints counted for 4 activity classes in the development of the CA map.

SM1-Table 1. List of 24 attributes serving as independent variables in the classification models.

|  |  |  |
| --- | --- | --- |
| **Code** | **Type** | **Description** |
| PEOE\_VSA\_FPPOS | 2D MOE descriptor | Fractional polar positive vdw surface area |
| AATSC7c | 2D PaDEL descriptor | Average centered Broto-Moreau autocorrelation - lag 7/weighted by charges |
| MATS1s | Moran autocorrelation - lag 1/weighted by I-state |
| SpMax3\_Bhv | Largest absolute eigenvalue of Burden modified matrix - n 3/weighted by relative van der Waals volumes |
| SpMin2\_Bhe | Smallest absolute eigenvalue of Burden modified matrix - n 2/weighted by relative Sanderson electronegativities |
| SpMin8\_Bhe | Smallest absolute eigenvalue of Burden modified matrix - n 8/weighted by relative Sanderson electronegativities |
| SpMin5\_Bhi | Smallest absolute eigenvalue of Burden modified matrix - n 5/weighted by relative first ionization potential |
| ETA\_Shape\_X | Shape index X |
| n8HeteroRing | Number of 8-membered rings containing heteroatoms (N, O, P, S, or halogens) |
| nF8HeteroRing | Number of 8-membered fused rings containing heteroatoms (N, O, P, S, or halogens) |
| JGI6 | Mean topological charge index of order 6 |
| JGI7 | Mean topological charge index of order 7 |
| VE1\_D | Coefficient sum of the last eigenvector from topological distance matrix |
| MACCSFP30 | MACCS  key | CQ(C)(C)A\*  (non-C Q4 bonded to >= 3 C) |
| PubchemFP122 | Pubchem fingerprint | >= 2 any ring size 3 |
| PubchemFP534 | S-C:C-O |
| PubchemFP741 | Oc1cc(S)ccc1 |
| PubchemFP749 | Nc1cc(N)ccc1 |
| PubchemFP772 | Nc1c(Br)cccc1 |
| SubFP41 | Substructure fingerprint | 1,2-Diol |
| SubFP48 | Aldehyde |
| SubFP62 | Halogen acetal like |
| SubFP84 | Carboxylic acid |
| SubFP157 | Semicarbazone |

\*A: Any valid periodic table element symbol; Q: Hetero atoms; any non-C or non-H atom

SM1-Table 2. List of 21 attributes serving as variables in perceptual mapping.

|  |  |  |
| --- | --- | --- |
| **Code** | **Type** | **Description** |
| diameter (dia) | 2D MOE desriptor | Largest vertex eccentricity in graph |
| BCUT\_PEOE\_2 (BP2) | PEOE Charge BCUT (2/3) |
| GCUT\_PEOE\_2 (GP2) | PEOE Charge GCUT (2/3) |
| balabanJ (bJ) | Balaban averaged distance sum connectivity |
| Q\_VSA\_FNEG (QVF) | Fractional negative vdw surface area |
| ATSC2m (A2m) | 2D PaDEL descriptor | Centered Broto-Moreau autocorrelation - lag 2/weighted by mass |
| ATSC4m (A4m) | Centered Broto-Moreau autocorrelation - lag 4/weighted by mass |
| ATSC1s (A1s) | Centered Broto-Moreau autocorrelation - lag 1/weighted by I-state |
| AATSC6v (AA6v) | Average centered Broto-Moreau autocorrelation - lag 6/weighted by van der Waals volumes |
| AATSC4s (AA4s) | Average centered Broto-Moreau autocorrelation - lag 4/weighted by I-state |
| MATS4s (M4s) | Moran autocorrelation - lag 4/weighted by I-state |
| SpMAD\_DzZ (SpM) | Spectral mean absolute deviation from Barysz matrix/weighted by atomic number |
| ASP-3 (ASP3) | Average simple path, order 3 |
| AVP-6 (AVP6) | Average valence path, order 6 |
| nHCsatu (nHCs) | Count of atom-type H E-State: H on C sp3 bonded to unsaturated C |
| minHCsatu (minHCs) | Minimum atom-type H E-State: H on C sp3 bonded to unsaturated C |
| ETA\_BetaP\_ns\_d (EBPnsd) | A measure of lone electrons entering into resonance relative to molecular size |
| MDEO-22 (MDEO22) | Molecular distance edge between all secondary oxygens |
| MACCSFP128 (MFP128) | MACCS  key | ACH2AAACH2A\*  (CH2s separated by 4 bonds) |
| MACCSFP144 (MFP144) | Anot%A%Anot%A\*  (atoms separated by (!:):(!:)) |
| PubchemFP2 (PFP2) | Pubchem fingerprint | >= 16 H |

\*A: Any valid periodic table element symbol

SM1-Table 3. Average values of 18 descriptors computed for 4 activity classes in the development of the MDS map.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **P** | **A** | **D** | **N** |
| diameter | 0.2074 | 0.3782 | 0.4985 | 0.3464 |
| BCUT\_PEOE\_2 | 0.4752 | 0.7795 | 0.5078 | 0.7652 |
| GCUT\_PEOE\_2 | 0.4359 | 0.6802 | 0.4840 | 0.7108 |
| balabanJ | 0.5112 | 0.3455 | 0.3231 | 0.3853 |
| Q\_VSA\_FNEG | 0.4937 | 0.7748 | 0.4523 | 0.6809 |
| ATSC2m | 0.5740 | 0.5056 | 0.3797 | 0.3006 |
| ATSC4m | 0.5791 | 0.5091 | 0.3934 | 0.5436 |
| ATSC1s | 0.6517 | 0.5384 | 0.8090 | 0.5947 |
| AATSC6v | 0.2655 | 0.3210 | 0.4712 | 0.5143 |
| AATSC4s | 0.5190 | 0.3865 | 0.3665 | 0.3659 |
| MATS4s | 0.5676 | 0.3789 | 0.3529 | 0.4004 |
| SpMAD\_DzZ | 0.1839 | 0.2917 | 0.5616 | 0.2783 |
| ASP-3 | 0.4111 | 0.5237 | 0.5766 | 0.5718 |
| AVP-6 | 0.3856 | 0.5031 | 0.5892 | 0.3944 |
| nHCsatu | 0.3684 | 0.0000 | 0.2105 | 0.0000 |
| minHCsatu | 0.3827 | 0.0000 | 0.4041 | 0.0000 |
| ETA\_BetaP\_ns\_d | 0.3364 | 0.1979 | 0.3596 | 0.2272 |
| MDEO-22 | 0.1783 | 0.0000 | 0.1449 | 0.0000 |

SM1-Table 4. Frequency of 3 fingerprints counted for 4 activity classes in the development of the CA map.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **P** | **A** | **D** | **N** |
| MACCSFP128 | 3 | 1 | 16 | 1 |
| MACCSFP144 | 16 | 1 | 17 | 4 |
| PubchemFP2 | 13 | 7 | 19 | 7 |