Supplementary material

Equation 1

$$v' = (v - \min_A)(\text{new\_max}_A - \text{new\_min}_A) / (\max_A - \min_A) + \text{new\_min}_A$$

where v' is the normalized descriptor value, v is the original descriptor value, $\max_A$ and $\min_A$ represent original maxima and minima of concerned descriptor. New_max_A and new_min_A represent maxima and minima defined for normalized descriptor value **[16]**. Table 1 shows the values of descriptors before and after normalization.

	MW	LP	ADE	PDE	HBD	HBA	TPSA	Ch	RB
Original Descriptor Vectors	249.354	2.58	-1.44	-40.22	0	2	41	1	3
	162.192	1.06	-6.25	-5.6	0	2	47	0	3
	152.149	0.37	-1.17	-46.53	0	2	57	-1	3
Normalized Descriptor vectors	0.249	0.483	0.382	0.911	0	0.067	0.228	0.5	0.1
	0.162	0.419	0.114	0.988	0	0.067	0.261	0.333	0.1
	0.152	0.390	0.397	0.897	0	0.067	0.317	0.167	0.1

Table 1: Sample representation of original and normalized descriptor vectors is given.

	Test data for molecules	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
Nine descriptor data set	20000	9945	9280	720	55	99.45%	92.80%	96.12%
Lipinski's rule descriptor data set	20000	9045	7451	2549	955	90.45%	74.51%	82.48%

Table 2: Detailed results of both datasets are given. TP= number of true positives, TN= number of true negatives, FP= number of false positives, FN=number of false negatives