Electronic Supplementary Material (ESI) for Nanoscale.
This journal is © The Royal Society of Chemistry 2019

A Chemoinformatic Approach for Characterization of Hybrid Nanomaterials: Safer and Efficient Design Perspective

Alicja Mikolajczyk[a*], Natalia Sizochenko[a,c], Ewa Mulkiewicz[c], Anna Malankowska[d], Bakhtiyor Rasulev[e] and Tomasz Puzyn[a*]

[a] *Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland*

[b] *Department of Environmental Analytics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland*

[c] *Department of Computer Science, Dartmouth College, 6211 Hinman, Hanover, New Hampshire 03755, United States of America*

[d] *Department of Environmental Technology, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland*

[e] *Department of Coatings and Polymeric Materials, North Dakota State University, NDSU Dept 2760, P.O. Box 6050, Fargo, ND 58108, United States of America*

† Footnotes relating to the title and/or authors should appear here.

*corresponding author e-mail: tomasz.puzyn@ug.edu.pl, mikolajczyk@qsar.eu.org

Table S1. Descriptors selected for the nano-QSPR$_{mix}$ model of photocatalysis, their values and the endpoint predictions

Sample label	Dataset splitting	Descriptor %mol$_{Pd}$	Descriptor BET$_{area}$	Descriptor φ_{Pt}	Descriptor φ_{mix}	Observed $UVVis_{degr}$	Predicted $UVVis_{degr}$
0.1Ag	Training	0.0	61.0	0.000	0.016	61.8	61.11
0.1Ag_0.1Au	Validation	0.0	106.5	0.000	0.182	63.0	61.11
0.1Ag_0.1Au_0.1Pd	Training	0.1	85.0	0.000	0.567	92.8	85.85
0.1Ag_0.1Pd	Training	0.1	79.0	0.000	0.135	77.4	85.85
0.1Ag_0.1Pt	Training	0.0	92.0	0.034	0.197	85.4	74.45
0.1Ag_0.1Pt_0.1Au	Training	0.0	94.0	0.034	0.705	84.4	74.45
0.1Ag_0.1Pt_0.1Au_0.1Pd	Training	0.1	94.0	0.034	1.556	100.0	99.18
0.1Ag_0.1Pt_0.1Pd	Validation	0.1	89.0	0.034	0.599	100.0	99.18
0.1Ag_0.5Au	Training	0.0	103.0	0.000	5.966	51.4	61.11
0.1Ag_0.5Pd	Training	0.5	115.0	0.000	3.817	99.2	99.65
0.1Ag_0.5Pt	Training	0.0	112.5	4.195	6.601	96.8	96.35
0.1Pt_0.1Au_0.1Pd	Training	0.1	92.0	0.034	0.727	100.0	99.18
0.5Ag	Training	0.0	91.0	0.000	1.949	64.4	61.11
0.5Ag_0.1Au	Training	0.0	112.0	0.000	3.889	52.9	61.11
0.5Ag_0.1Pd	Validation	0.1	77.5	0.000	3.458	72.1	85.85
0.5Ag_0.1Pt	Training	0.0	94.0	0.034	4.016	68.4	74.45
1.5Ag	Training	0.0	89.0	0.000	52.620	70.3	69.20
1.5Ag_0.1Au	Training	0.0	90.2	0.000	67.559	65.6	69.20
1.5Ag_0.1Pd	Training	0.1	98.0	0.000	64.500	96.0	93.94
1.5Ag_0.1Pt	Validation	0.0	145.6	0.034	68.462	85.6	96.35
2.5Ag	Training	0.0	91.0	0.000	243.612	73.5	69.20
2.5Ag_0.1Au	Validation	0.0	79.0	0.000	284.184	60.0	69.20
2.5Ag_0.1Pd	Training	0.1	98.0	0.000	275.784	94.9	93.94
2.5Ag_0.1Pt	Training	0.0	86.0	0.034	286.664	71.1	82.54
2.5Ag_0.5Au	Training	0.0	90.4	0.000	486.081	64.6	69.20
2.5Ag_0.5Pd	Training	0.5	107	0.000	432.209	99.5	93.94
2.5Ag_0.5Pt	Validation	0.0	119	4.195	501.992	96.1	96.35
4.5Ag	Training	0.0	67.5	0.000	1420.743	73.5	69.20
4.5Ag_0.5Au	Training	0.0	111.4	0.000	2129.422	69.5	69.20

Table S2. DT classifications for photocatalysis

Decision Stump 1 (amount of Pd)

For %mol$_{Pd}$ ≤ 0.05: -8.60

For %mol$_{Pd}$ > 0.05: 16.13

For Pd is missing, then set value to -8.65*10^{-15}

Decision Stump 2 (sedimentation potential of Pt nanoparticles)

φ_{Pt} ≤ 0.0168: -4.06

φ_{Pt} > 0.0168: 9.28

φ_{Pt} is missing, then set value to 7.723*10^{-16}

Decision Stump 3 (additive sedimentation potential for mixture)

φ_{mix} ≤ 6.28: -3.87

φ_{mix} > 6.28: 4.22

φ_{mix} is missing, then set value to 1.54: *10^{-16}

Decision Stump 4 (BET surface area)

$BET_{area} \leq 112.25$: -1.20
$BET_{area} > 112.25$: 12.61
BET_{area} is missing, then set value to $-7.723*10^{-17}$

Table S3. DT classifications for cytotoxicity
Decision Stump 1 (electron affinity)
A ≤ 198.67: -0.423
A > 198.67: 0.634

Decision Stump 2 (electron affinity)
A ≤ 39.81: -0.327
A > 39.81: 0.058

Decision Stump 3 (electron affinity)
A ≤ 490.92: -0.038
A > 490.92: 0.343

Decision Stump 4 (electron affinity)
A ≤ 198.67: 0.077
A > 198.67: -0.115

Table S4. Descriptors selected for the nano-QSAR$_{mix}$ model of cytotoxicity, their values and the endpoint predictions

Sample Label	Data Splitting	Descriptor χ_{mix}	Descriptor A	Observed pEC_{50}	Predicted pEC_{50} by MLR-GA model	Predicted pEC_{50} by DT model
0.1Ag_0.1Pd	Training	0.41	17.93	5.56	5.9	5.66
0.1Ag	Training	0.19	12.56	5.60	5.8	5.66
0.1Ag_0.1Au	Validation	0.45	34.84	5.67	5.9	5.66
0.1Ag_0.5Pd	Training	1.29	39.41	5.70	6.1	5.66
0.5Ag	Training	0.97	62.80	5.83	6.0	6.05
0.1Ag_0.1Pt_0.1Pd	Validation	0.64	38.46	5.86	5.9	5.66
0.1Ag_0.1Pt_0.1Au	Training	0.68	55.37	5.88	5.9	6.05
0.1Ag_0.1Au_0.1Pd	Training	0.67	40.21	5.88	5.9	6.05
0.1Ag_0.1Pt	Validation	0.42	33.09	5.92	5.9	5.66
0.1Pt_0.1Au_0.1Pd	Training	0.70	48.18	5.92	5.9	6.05
0.1Ag_0.5Pt	Training	1.33	115.21	5.97	6.1	6.05
0.1Ag_0.1Pt_0.1Au_0.1Pd	Validation	0.90	60.74	6.02	6.0	6.05
0.1Ag_0.5Au	Training	1.46	123.96	6.05	6.1	6.05
1.5Ag	Training	2.90	188.40	6.23	6.4	6.05
0.5Ag_0.1Au	Validation	1.22	85.08	6.25	6.0	6.05
0.5Ag_0.1Pt	Training	1.19	83.33	6.29	6.0	6.05
0.5Ag_0.1Pd	Training	1.19	68.17	6.46	6.0	6.05
1.5Ag_0.1Au	Validation	3.15	210.68	6.58	6.4	6.91
1.5Ag_0.1Pt	Training	3.12	208.93	6.63	6.4	6.91
2.5Ag	Training	4.83	314.00	6.76	6.8	6.91
1.5Ag_0.1Pd	Validation	3.12	193.77	6.77	6.4	6.05
2.5Ag_0.1Au	Training	5.08	336.28	6.80	6.8	6.91
2.5Ag_0.5Pd	Training	5.93	340.85	6.88	7.0	6.91
2.5Ag_0.5Au	Validation	6.10	425.4	7.01	70	6.91
2.5Ag_0.1Pt	Training	5.05	334.53	7.07	6.8	6.91
2.5Ag_0.5Pt	Training	5.97	416.65	7.10	7.0	6.91
2.5Ag_0.1Pd	Validation	5.05	319.37	7.20	6.8	6.91
4.5Ag	Training	8.69	565.20	7.40	7.5	7.29
4.5Ag_0.5Au	Training	9.96	676.60	7.41	7.7	7.29

Table S5. Description of basic quantum chemical descriptors

Number	Descriptor	Formula	Description
(1)	ionization potential (*IP*)	$IP = E_C^{N-1} - E_0^N$	E_C^{N-1} – the energy after losing one electron (cation);
(2)	electron affinity (*A*)	$A = E_0^N - E_A^{N+1}$	E_0^N – the basal state energy (neutral);
(3)	absolute electronegativity (μ)	$\mu = \frac{(I + A)}{2}$	E_A^{N+1} – the energy after gaining one electron (anion);
(4)	adsorption energy of metal (ΔE_{ads})	$\Delta E_{ads} = E_{tot} - (E_{TiO_2} + E_M$	E_{tot} – the total energy of the interacting metal – TiO_2 surface; E_{TiO2} is the total energy of the clean most stable TiO_2; E_{Me} – the total energy of the single metal atom

Table S6. List of nano-descriptors for obtained TiO_2–based hybrid nanomaterials derived from: (i) additive scheme and QM calculations; (ii) experimental measurement and (iii) Smoluchowski equation.

Sample label	**%mol$_{Ag}$**	**%mol$_{Pt}$**	**%mol$_{Au}$**	**%mol$_{Pd}$**	**BET$_{area}$**	A	χ_{mix}	IP	IP_1	φ_{Pt}	φ_{mix}
	[% mol]	[% mol]	[% mol]	[% mol]	[m^2/g]	[kJ/mol]	[kJ/mol]	[kJ/mol]	[kJ/mol]	-	-
0.1Ag_0.5Au	0.1	0	0.5	0	103.00	123.96	1.46	518.15	1197.00	0	5.966
0.5Ag_0.1Au	0.5	0	0.1	0	112.00	85.08	1.22	454.51	1233.00	0	3.889
2.5Ag_0.1Au	2.5	0	0.1	0	79.00	336.28	5.08	1916.51	5373.00	0	284.184
0.1Ag	0.1	0	0	0	61.00	12.56	0.19	73.10	207.00	0	0.016
0.1Ag_0.1Au	0.1	0	0.1	0	106.50	34.84	0.45	162.11	405.00	0	0.182
0.5Ag	0.5	0	0	0	91.00	62.80	0.97	365.50	1035.00	0	1.949
2.5Ag_0.5Au	2.5	0	0.5	0	90.40	425.40	6.10	2272.55	6165.00	0	486.081
1.5Ag_0.1Au	1.5	0	0.1	0	90.20	210.68	3.15	1185.51	3303.00	0	67.559
0.5Ag_0.1Pt	0.5	0.1	0	0	94.00	83.33	1.19	452.50	1214.10	0.034	4.016
4.5Ag_0.5Au	4.5	0	0.5	0	111.40	676.60	9.96	3734.55	10305.00	0	2129.422
1.5Ag	1.5	0	0	0	89.00	188.40	2.90	1096.50	3105.00	0	52.62
2.5Ag_0.1Pt	2.5	0.1	0	0	86.00	334.53	5.05	1914.50	5354.10	0.034	286.664
0.5Ag_0.1Pd	0.5	0	0	0.1	77.50	68.17	1.19	445.94	1222.00	0	3.458
2.5Ag	2.5	0	0	0	91.00	314.00	4.83	1827.50	5175.00	0	243.612
4.5Ag	4.5	0	0	0	67.50	565.20	8.69	3289.50	9315.00	0	1420.743
0.1Ag_0.1Pd	0.1	0	0	0.1	79.00	17.93	0.41	153.54	394.00	0	0.135
0.1Ag_0.1Pt_0.1Au	0.1	0.1	0.1	0	94.00	55.37	0.68	249.11	584.10	0.034	0.705
0.1Ag_0.1Pt	0.1	0.1	0	0	92.00	33.09	0.42	160.10	386.10	0.034	0.197
1.5Ag_0.1Pt	1.5	0.1	0	0	145.60	208.93	3.12	1183.50	3284.10	0.034	68.462
0.1Ag_0.1Au_0.1Pd	0.1	0	0.1	0.1	85.00	40.21	0.67	242.55	592.00	0	0.567
2.5Ag_0.1Pd	2.5	0	0	0.1	98.00	319.37	5.05	1907.94	5362.00	0	275.784
1.5Ag_0.1Pd	1.5	0	0	0.1	98.00	193.77	3.12	1176.94	3292.00	0	64.5
2.5Ag_0.5Pt	2.5	0.5	0	0	119.00	416.65	5.97	2262.50	6070.50	4.195	501.992
0.1Ag_0.5Pt	0.1	0.5	0	0	112.50	115.21	1.33	508.10	1102.50	4.195	6.601
0.1Ag_0.5Pd	0.1	0	0	0.5	115.00	39.41	1.29	475.30	1142.00	0	3.817
2.5Ag_0.5Pd	2.5	0	0	0.5	107.00	340.85	5.93	2229.70	6110.00	0	432.209
0.1Ag_0.1Pt_0.1Pd	0.1	0.1	0	0.1	89.00	38.46	0.64	240.54	573.10	0.034	0.599
0.1Pt_0.1Au_0.1Pd	0	0.1	0.1	0.1	92.00	48.18	0.70	256.45	564.10	0.034	0.727
0.1Ag_0.1Pt_0.1Au_0.1Pd	0.1	0.1	0.1	0.1	94.00	60.74	0.90	329.55	771.10	0.034	1.556

Table S7. Statistic parameters towards development of Nano-QSPR/Nano-QSAR model.*

No	Name	Suitable to	Coefficient	Description of Coefficient
1	The correlation coefficient The root mean square error of calibration	Measure of goodness-of-fit	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2}{\sum_{i=1}^{n}(y_i^{obs} - \tilde{y}^{obs})^2}$ $RMSEC = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2}{n}}$	where: y_j^{obs} – experimental (observed) value of the property for the i^{th} compound from the training set; y_i^{pred} – predicted value for i^{th} compound from the training set; y^{obs} – the mean experimental value of the property in the training set; n – the number of compounds in the training set.
	The cross-validated coefficient The root mean square error of calibration	Measure stability of the model	$Q^2_{LOO} = 1 - \frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{predcv})^2}{\sum_{i=1}^{n}(y_i^{obs} - \tilde{y}^{obs})^2}$ $RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{predcv})^2}{n}}$	where: y_j^{obs} – experimental (observed) value of the property for the i^{th} compound; y_i^{predcv} – cross-validated predicted valuef or i^{th} compound; y – the mean experimental value of the property in the training set; n – the number of compounds in the training set.
3	The external-validation coefficient The root mean square error of prediction	Measure external predictivity	$Q^2_{EXT} = 1 - \frac{\sum_{j=1}^{k}(y_j^{obs} - y_j^{pred})^2}{\sum_{j=1}^{k}(y_j^{obs} - \hat{y}^{obs})^2}$ $RMSEP = \sqrt{\frac{\sum_{j=1}^{k}(y_j^{obs} - y_j^{pred})^2}{k}}$	where: y_j^{obs} – experimental (observed) value of the property for the j^{th} compound from the validation set; y_j^{pred} – predicted valuef or j^{th}compound from the validation set; $\hat{y}^{obs}$ – the mean experimental value of the property in the validation set; k – the number of compounds in the validation set. Squared correlation coefficient values between the observed and predicted values of the compounds with/without intercept (r^2/ r_0^2).
4	Concordance Correlation Coefficient	The accuracy and different variants of r^2_m Restrictive parameter for expressing external predictivity	$r_m^2 = r^2\left(1 - \sqrt{r^2 - r_0^2}\right)$ $CCC = \frac{2\sum_{j=1}^{k_{EXT}}(y_j^{obs} - \hat{y}^{obs})(y_j^{pred} - \hat{y}^{pred})}{\sum_{j=1}^{k_{EXT}}(y_j^{obs} - \hat{y}^{obs})^2 + \sum_{j=1}^{k_{EXT}}(y_j^{pred} - \hat{y}^{pred})^2 + k_{EXT}(\hat{y}^{obs} - \hat{y}^{pred})^2}$	
		Mean absolute error	$MEA = \frac{1}{k}\sum y_j^{obs} - y_j^{pred}$	

*Please note that are details according to statistics calculations are described in presented study:

A. Golbraikh, A. Tropsha, *Journal of Molecular Graphics and Modelling*, 2002, **20**, 269-276. (Please see, MS, References section, position 42)

P. Gramatica, *QSAR & Combinatorial Science,* 2007, **26**, 694-701 (Please see, MS, References section, position 60)

K. Roy, S. Kar and P. Ambure, *Chemometrics and Intelligent Laboratory Systems,* 2015, ***145***, 22-29. (Please see, MS, References section, position 61)

K. Roy, N. R. Das, P. Ambure and N. B. Aher, *Chemometrics and Intelligent Laboratory Systems,* 2016, **152**, 18-33 (Please see, MS, References section, position 62)

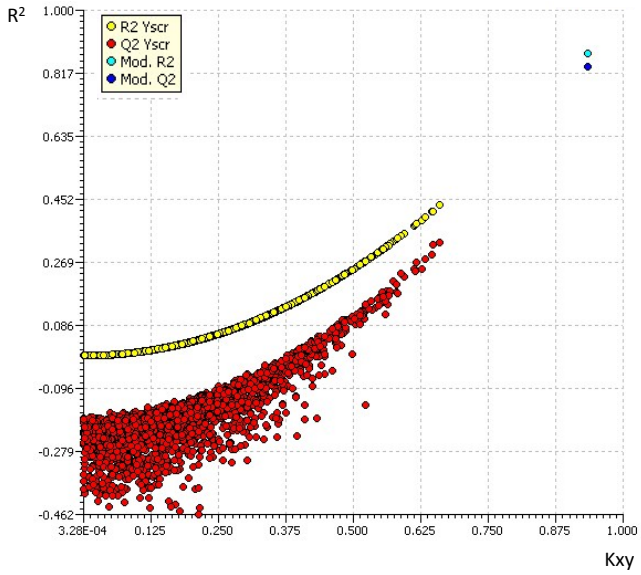

Figure S1. The results of the Y-scrambling test for Nano-QSAR$_{mix}$ model described in eq.4

Average R^2Y_{scr} = average 'R^2' of random models.
Average Q^2Y_{scr} = average 'Q^2' of random models

Table S8. The statistics of the Y-scrambling test for Nano-QSAR$_{mix}$ model described in eq.4

Obtained random Models Parameters for eq. 6:	**Value of selected parameters**
Average R^2Y_{scr} :	0.05
Average Q^2Y_{scr} :	-0.18
The y-rand R^2 range	-0.46 – 0.45

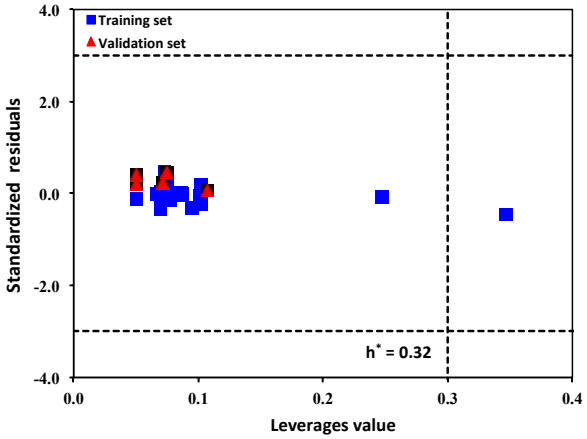

Figure S2. Williams plot describing applicability domains of MLR-GA model presented in eq. 4.

Table S9. Similarity matrix between the elements of sets A and B of the first case study; in the last column and row, the Hausdorff-like similarity contributions of all the individual elements are reported. (Normalized distance matrix based on Euclidean distance)*

	Set B																		
	B_0.1 Ag_0.5 Pd	**B_0.1 Ag_0. 1Pt**	**B_0.1 Ag_0. 5Pt**	**B_0.1 Ag_0.1 Au**	**B_0.1 Ag_0.5 Au**	**B_0.1 Ag_0.1 Pd**	**B_0.5 Ag_0.1 Au**	**B_0.5 Ag_0. 1Pt**	**B_0.5 Ag_0.1 Pd**	**B_1.5 Ag_0.1 Au**	**B_1.5 Ag_0. 1Pt**	**B_4.5 Ag_0.5 Au**	**B_1.5 Ag_0.1 Pd**	**B_2.5 Ag_0.1 Au**	**B_2.5 Ag_0.5 Pd**	**B_2.5 Ag_0.5 Au**	**B_2.5 Ag_0. 1Pt**	**B_2.5 Ag_0. 5Pt**	**B_2.5 Ag_0.1 Pd**	**max [Sab]**
Set A																			
A_0.1Ag_0.1 Pt_0.1Au_0.1 Pd	0.94	0.98	0.94	0.98	0.93	0.98	0.95	0.94	0.94	0.82	0.82	0.44	0.82	0.70	0.69	0.69	0.70	0.69	0.70	0.98
A_0.1Ag_0.1 Pt_0.1Pd	0.94	0.99	0.94	0.97	0.92	0.99	0.94	0.95	0.95	0.82	0.82	0.44	0.82	0.70	0.69	0.69	0.70	0.69	0.70	0.99
A_0.1Ag_0.1 Au_0.1Pd	0.94	0.97	0.92	0.99	0.93	0.98	0.95	0.94	0.95	0.82	0.82	0.44	0.82	0.70	0.69	0.69	0.70	0.69	0.70	0.99
A_0.1Ag_0.1 Pt_0.1Au	0.93	0.98	0.94	0.99	0.93	0.97	0.95	0.95	0.94	0.82	0.82	0.44	0.82	0.70	0.69	0.69	0.70	0.69	0.70	0.99
A_0.1Pt_0.1 Au_0.1Pd	0.94	0.98	0.94	0.98	0.93	0.98	0.94	0.94	0.94	0.82	0.82	0.45	0.82	0.70	0.69	0.69	0.70	0.69	0.70	0.98
Max [Sba]	0.94	0.99	0.94	0.99	0.93	0.99	0.95	0.95	0.95	0.82	0.82	0.45	0.82	0.70	0.69	0.69	0.70	0.69	0.70	

* Mauri, A.; Ballabio, D.; Todeschini, R.; Consonni, V. *Journal of Chemoinformatics*, 2016, **8,** 49.

SUM[s_{ba}] 15.69

SUM[s_{ab}] 4.91

$$Hs_{AB} = \frac{\sum_{a\epsilon A} max_{b\epsilon B}[S_{ab}] + \sum_{b\epsilon B} max_{a\epsilon A}[S_{ba}]}{n_A + n_B} = 0.86$$