Supplemetary materials to the manuscript

"Comparative prediction of gas chromatographic retention indices for chemicals related to Chemical Weapons Convention by incremental and machine learning methods"

by

Albert Kireev[1*], *Sergey Osipenko*[2], *Gary Mallard*[3], *Eugene Nikolaev*[4] *and Yury Kostyukevich*[1*]

[1]Center for Molecular and Cellular Biology, Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, 121205 Moscow, Russia

[2]Center for Computational and Data-Intensive Science and Engineering, Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, 121205 Moscow, Russia

[3]Teal Consulting, Chevy Chase, Maryland, United States

[4]Laboratory of Mass-spectrometry, Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, 121205 Moscow, Russia

*Corresponding authors Albert Kireev, Email: A.Kireev@skoltech.ru

Yury Kostyukevich, Email: Y.Kostyukevich@skoltech.ru

Table S1. Set of hyperparameters for grid search cross-validation

Parameter	Values
Tree depth	2,4,6,8,10
Learning rate	0.01, 0.05, 0.075, 0.1, 0.2, 0.3
Gamma	0, 0.5, 1, 5, 10

Table S2. Examples of calculated difference of retention index (RI) using molecular pairs matched by RO radical. 1A1, 1A2, 1A3 and selected 2B4 with 1A1 data sets are considered

RO pairs	1A1 -2B4			Thio 2B4			1A2			1A3		
	RI diff	SD	N	RI diff	SD	N	RI diff	SD	N	RI diff	SD	N
methyl : ethyl	-70	9	11	-69	3	3	-57	4	10	-45	5	37
methyl : isopropyl	-101	8	11	-102	6	5	-75	11	10	-52	8	39
methyl : propyl	-160	4	12	-161	4	5	-147	4	10	-133	7	38

methyl : sec-butyl	-200	10	12	-205	3	4	-172	9	10	-145	9	39
1-Propylcyclohexyl :1,2-Dimethylbutyl	334	2	3	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table S3. Retention Indices predicted by R-P differences for 1A1 toxic chemicals with available combination of RO and RP, but missing RI data in OCAD library

RO_P	Propyl_P	isoPropyl_P	Ethyl_P	Methyl_P
1-Propylcyclohexyl	1600	1565	PRESENT IN OCAD	PRESENT IN OCAD
1-Propylheptyl	1637	1602	PRESENT IN OCAD	PRESENT IN OCAD
2,2-Dimethylcyclohexyl	1494	1460	1409	PRESENT IN OCAD
2,2-Dimethylhexyl	1436	1402	1351	PRESENT IN OCAD
2,3-Dimethylpentyl	1390	1356	1305	PRESENT IN OCAD
2,5-Dimethylcyclohexyl	1500	1465	PRESENT IN OCAD	PRESENT IN OCAD