**Table 2** QSRR models selected from the literature review

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| References | Type of contaminant | Number of contaminants | MDs  selected | Best machine learning algorithms used | RT max measured  (min) | R²  test set | RMSE  test set  (min) | Percentage of error |
| Aalizadeh et al., 2019 | Emerging contaminants | 1830 | LogDa, CIC1b, SeigZc, RDF020pd, AlogPe | SVM | 14.4 | 0.88 | 1.04 | 7% |
| McEachran et al., 2018 | Environmental contaminants | 97 | LogPf, LogD, molecular weight, molecular  volume, polar surface areag, molar refractivityh, H\_donorsi, H\_acceptorsj | ACD  /ChromGenius® | 40.8 | 0.92 | 2.66 | 6.5% |
| Bade et al., 2015 | Emerging contaminants | 544 | nDBk, nTBl, nCm, nOn, nR04-nR09o, UIp, Hyq, MlogPr, AlogP, logP, logD | MLP | 16.5 | 0.91 | 0.89 | 5.4% |
| Munro et al., 2015 | Pharmaceuticals | 166 | nDB or nTB, nC or nO, nR04-nR09, UI, Hy,  MlogP, AlogP , LogD, nBnzs, pKat | GRNN | 23.2 | 0.88 | 1.39 | 5.9% |
| Noreldeen et al., 2018 | Veterinary drugs | 95 | ACDlogPu, ALOGP, ALOGP2v, Hy, Ui, ibw, BEHp1x, BEHp2y,GATS1mz, GATS2ma2. | MLR | 9.3 | 0.95 | 0.62 | 6.6% |
| Bride et al., in press | Environmental contaminants | 274 | logD, DBEa3, nO, nC, nH, molecular weight, H\_donors, logSwa4 | MLR | 14.7 | 0.76 | 1.36 | 9.2% |
| Yang et al., 2020 | Pharmaceuticals | 133 | XlogPa5, BCUTp.1ha6, AATS1ia7, AATS3ia8, GATS1ea9, ALogP, AATSC0pa10,  ETA\_EtaP\_Ba11, AATS4ia12, AATS5ia13 | MLR | 15.0 | 0.63 | 1.42 | 9.4% |

* a: logD is the measure of hydrophobicity for the ionizable compounds
* b: CIC1 is the Complementary Information Content index (neighborhood symmetry)
* c: SeigZ is the eigenvalue sum from a Z weighted distance matrix of a Hydrogen-depleted Molecular Graph
* d: RDF020p is radial distribution function weighted by atomic polarizabilities,
* e: AlogP is logP estimated by the Ghose–Crippen method.
* f: LogP or LogKow, LogP is equal to the logarithm of the ratio of the concentrations of the test substance in octanol and water. This value allows apprehending the hydrophilic or hydrophobic (lipophilic) character of a molecule.
* g: defined as the surface sum over all [polar](https://en.wikipedia.org/wiki/Chemical_polarity) atoms or molecules, primarily [oxygen](https://en.wikipedia.org/wiki/Oxygen) and [nitrogen](https://en.wikipedia.org/wiki/Nitrogen), also including their attached hydrogen atoms.
* h: is a measure of the total [polarizability](https://en.wikipedia.org/wiki/Polarizability) of a [mole](https://en.wikipedia.org/wiki/Mole_(unit)) of a substance
* i: the number of H-bond donor as descriptors of the H-bonding property
* j: the number of H-bond acceptor groups as descriptors of the H-bonding property
* k: number of double bonds
* l: number of triple bonds
* m: number of Carbon
* n: number of Oxygen
* o: the number of 4–9 membered rings
* p: unsaturation index
* q: hydrophilic factor
* r: Moriguchi logP
* s: number of benzen groups
* t: equilibrium constant of the dissociation reaction of an acid species in acid-base reactions
* u: ACDlogPa molecular properties octanol-water partitioning coefficients
* v: ALOGP2 molecular properties Ghose-Crippen octanol water coefficient squared
* w: Ib information indices information bond index.
* x: BEHp1 burden eigenvalue descriptors highest eigenvalue n. 1 of burden matrix/weighted by atomic polarizabilities.
* y: BEHp2 burden eigenvalue descriptors highest eigenvalue n. 2 of burden matrix/weighted by atomic polarizabilities.
* z: GATS1mb 2D autocorrelation descriptors Geary autocorrelation-lag 1/weighted by atomic masses.
* a2: GATS2mb 2D autocorrelation descriptors Geary autocorrelation-lag 2/weighted by atomic masses.
* a3: the double-bond equivalent descriptor is the number of unsaturations present in a organic molecule
* a4: the water solubility described by the logarithm of water solubility in mg/L at 25°C.
* a5: XlogP is the constitutional descriptors-describe hydrophobic/hydrophilic properties
* a6: BCUTp.1h is the BCUT descriptor/nlow highest polarizability weighted BCUTS
* a7: AATS1i is the autocorrelation descriptor/average Broto-Moreau autocorrelation - lag 1 / weighted by first ionization potential
* a8: AATS3i is the autocorrelation descriptor/average Broto-Moreau autocorrelation - lag 3 / weighted by first ionization potential
* a9: GATS1e is the autocorrelation descriptor/Geary autocorrelation - lag 1 / weighted by Sanderson electronegativities
* a10: AATSC0p is the autocorrelation descriptor/ average centered Broto-Moreau autocorrelation - lag 0 / weighted by first ionization potential
* a11: ETA\_EtaP\_B is the extended topochemical atom descriptor/branching index EtaB relative to molecular size
* a12: AATS4i is the autocorrelation descriptor/average Broto-Moreau autocorrelation - lag 4 / weighted by first ionization potential,
* a13: AATS5i is the autocorrelation descriptor/average Broto-Moreau autocorrelation - lag 5 / weighted by first ionization potential