Research Article

www.molinf.com

molecular informatics

doi.org/10.1002/minf.202200026

QSPR Modelling of the Solubility of Drug and Drug-like Compounds in Supercritical Carbon Dioxide

Imane Euldji,*[a] Cherif Si-Moussa,[a] Mabrouk Hamadache,[a] and Othmane Benkortbi[a]

Abstract: Quantitative structure–property relationship (QSPR) modeling was investigated to predict drug and drug-like compounds solubility in supercritical carbon dioxide. A dataset of 148 drug\drug-like compounds, accounting for 3971 experimental data points (EDPs), was collected and used for modelling the relationship between selected molecular descriptors and solubility fraction data achieved by a nonlinear approach (Artificial neural network, ANN) based on molecular descriptors. Experimental solubility data for a given drug were published as a function of temperature and pressure. In the present study, 11 significant PaDEL descriptors (AATS3v, MATS2e, GATS4c, GATS3v, GATS4e, GATS3 s, nBondsM, AVP-0, SHBd, MLogP, and MLFER_S), the temperature and the pressure were statistically proved to be sufficient inputs. The architecture of the optimized model was found to be {13,10,1}. Several statistical metrics, including average absolute relative deviation (AARD = 3.7748 %), root mean square error (RMSE = 0.5162), coefficient of correlation (r = 0.9761), coefficient of determination (R^2 = 0.9528), and robustise (Q^2 = 0.9528) were used to validate the obtained model. The model was also subjected to an external test by using 143 EDPs. Sensitivity analysis and domain of application were examined. The overall results confirmed that the optimized ANN-QSPR model is suitable for the correlation and prediction of this property.

Keywords: ANN · Drug-like Compounds · PaDEL Descriptors · QSPR · Solubility · Supercritical Carbon Dioxide

1 Introduction

Supercritical Fluid (SCF) technologies are nowadays well-known processing techniques in many industrial fields including chemical, biochemical, food, and pharmaceutical. SCF techniques are used in extraction, separation, and crystallisation procedure that are the main steps in active pharmaceutical ingredients (APIs) synthesis and particle engineering. Particle design of APIs is very critical in manufacture solid and liquid dosage forms. They are used to modify and to control numerous particle properties, including various sizes, shapes, crystallinity structures, polymorphism, processability, and solubilities.[1–5] Among the advantages of the SCF used are their potential regulating properties with temperature and pressure. Also, SCF are known to be non-toxic, non-flammable, and can achieve high mass and heat transfer rates.[1–6]

Current literature on applications of SCF technology applications in particle engineering, including crystal and polymorph modification, complex formation, solid dispersion, and other methods, used to improve the solubility of API, the bioavailability and processing capacity.[7]

Carbon dioxide is considered the most common solvent for SFT applications in the pharmaceutical industries (98 % of applications).[2,6,7] The most interesting characteristics of SC–CO_2 are their low critical point (31.3 °C, 7.4 MPa), that are easy to reach and to manipulate in an industrial device,[2,6,7] their appeal for heat-sensitive materials; it is inert, leaves no traces after the process, inexpensive, non-inflammable, GRAS (generally regarded as safe) status, and easy to recycle.[2–10] Another distinctive property of SC–CO_2 is its small linear structure, which increases its permeability.[11]

The solubilities of solid solutes in a SCF are a complex phase equilibrium at high pressure. Subsequently, it was demonstrated for the first time that solubility can be improved by using an SCF. The SCF community has actively measured and published solubility data.

A compilation of most of the experimental data published before 2004 was provided by Gupta and Shim[12] .A review series[13–18] on high pressure phase equilibria also present a compilation of experimental investigations on the solubility of solid solutes in SCF. Two other review papers on solubility of solids in sub- and SCFs gave a compilation of published experimental data.[19,20] A distinct advantage of SCFs is that the solubility of a solute is strongly a function of temperature and pressure. This advantage leads to a large demand for data.[21] Measuring the solubility of pharmaceutical compounds in SC–CO_2 at various temperatures and pressures became a fundamental necessity to

[a] *I. Euldji, C. Si-Moussa, M. Hamadache, O. Benkortbi*
Biomaterials and Transport Phenomena Laboratory (LBMPT), University of Yahia Fares, Faculty of Technology, Department of Process and Environmental Engineering, Medea 26000, Algeria
phone/fax: +213 78705031
E-mail: euldji.imane@univ-medea.dz
imene.eull@gmail.com

Supporting information for this article is available on the WWW under https://doi.org/10.1002/minf.202200026

© 2022 Wiley-VCH GmbH

assess whether the drug is suitable for treatment by this technology. In the supercritical conditions, the solute can create a metastable equilibrium with the solvent.

The addition of co-solvent or crystal grains allows to increase the solubility. In general, static and dynamic methods are applied for the solubility measurement of a solid in the supercritical solvent.[7] A reasonably good solubility measurement for a single solute in a single SCF requires at least 15 and possibly up to 40 experimental conditions (exp. 5–10 pressures at 3 or 4 different temperatures).[21] However, for low solubilities these measurements are not easy because they are time-consuming and expensive. This background explained why data on drugs solubility in SC–CO_2 is still not available for most pharmaceutical compounds.[1,22] Despite this, a large amount of experimental data on the solubility of solids (including solid drugs) in SC–CO_2 is now available, which allowed further attempts to develop models analogously to the measurement of the solubility. The objective behind the development of these models is to minimize the use of experiments, to offer the possibility of interpolating between the points measured experimentally, and to provide a tool to support process design and analysis. The ideal target for modelling is purely theoretical predictive models. However, up to now, these models are far from being effective. In the absence of predictive models, three approaches are used. The solubility of a solid solute in SCF is a phase equilibrium problem. Therefore, the obvious and straightforward approach of modelling solubility of a solid solute in a SCF is the use of a thermodynamic approach based on equations of state (EoS).[23–30] This can be divided into three main classes: cubic EoS, complex non-cubic EoS and conductor-like screening model (COSMO) based on quantum mechanical calculation results to determine the interaction between molecules in a mixture, which can be used to predict the solubility of a solid solute in SC–CO_2. The second approach is empirical or semi-empirical correlations[31–36] and the third is intelligent computer techniques.[37–42] All these modelling approaches rely on experimental data.

In addition to large and complicated computational methods, the cubic equation of state-based models requires the knowledge of critical and physicochemical properties of solutes such as critical temperature, pressure, acentric factor, sublimation pressure, molar volumes and vapor pressure of the solids unavailable in many cases, especially for pharmaceutical compounds and biomolecules. Obviously, the estimation of these properties with one of the several group-contribution methods leads to add a degree of uncertainty to the predicted results.[1,2,22] A review of the use of cubic EoSs in this context has been published by Yazdizadeh et al..[7,43]

Furthermore, the use of non-cubic complex EoS based on statistical associating fluids theory (SAFT) has also been investigated.[44–47] However, compared to cubic EoS, attempts to model the solubility of complex solid molecules in SC–CO_2 using complex non-cubic EoS are very few.

This is due to the lack of sufficient experimental data as this family of EoS use solute-specific parameters that usually are fitted to pure fluid such as vapor pressure or liquid density or few solubilities experimental data.[7] COSMO-based predictive models cannot be used for high-pressure phase equilibrium calculations without being coupled with an EoS. The experimental data required for solutes the melting temperature and the melting enthalpy of the solid drug. Both were used effectively to predict solubility of solids in SC–CO_2.[48–52]

Empirical and semi-empirical models are considered as a simple alternative to EoS because they are based only on density of the SCF and none of the solute parameters is required.[1,2,6] Typically,, the logarithm of solid drug solubility is assumed to be related to the SCF density for a given temperature and pressure. Several reviews related to density-based models for the correlation of solid solutes in SC–CO_2 were published.[7,53,54] The main conclusion of the various comparative studies[7,21,53–57] is that there is no model able to correlate solubility with accepted precision for all solids. The two key requirements for using the empirical correlation, as highlighted by Stiver in a recent paper,[21] are: a high-quality density source for SC–CO_2 and a robust, non-linear regression algorithm fitted to the AARD as objective function with a special effort to identify the global rather than a local minimum.

These last years. the application of artificial intelligent techniques in various fields of chemistry has been developed significantly. A review of the current trends and future direction of these applications has been reported recently by Baum et al..[58] Artificial neural network (ANN), as powerful soft computing tool, has been widely applied in various fields for modelling complex nonlinear problems. Applications of ANN modeling to correlate solubility of various solid solutes in SC–CO_2 have been reported.[59–66] In all these works, the physical properties such as critical temperature and pressure, acentric factor, normal boiling point, molecular weights of the solid solutes together with the temperature and pressure have been used as inputs to the ANN model. Another interesting approach still to be applied for modelling the solubility of drug compounds in SC–CO_2 is the quantitative structure-property relationship (QSPR).

The field of QSAR, QSPR, and QSTR modelling is the leading robust cheminformatics approach that links Activity Property and Toxicity to the molecular structure, respectively, to explore and exploit the dependence of chemical, biological, toxicological, or other types of activities or properties on their molecular characteristics.[67] QSAR modelling has been traditionally used as an advanced optimization technique in drug discovery research. However, in recent years QSAR modelling has grown, diversified and evolved from application to small series of homogeneous compounds using comparatively simple regression methods to the analysis of very large datasets including

 © 2022 Wiley-VCH GmbH

thousands of diverse molecular structures using a wide variety of machine learning and statistical techniques.

Additionally, the approach focuses on the field of drug-like property prediction and chemical hazard assessment.[67,68]

This paper deals for the first time the development of a combined ANN-QSPR model for prediction and correlation of drug and drug-like compounds solubility in supercritical carbon dioxide. The approach adopted in this work, aims first to find the molecular descriptors that are most correlated to solubility. to determine the ANN-QSPR architecture that best correlates the experimental data and to validate the model using recommended statistical metrics. Sensitivity analysis of the inputs was performed to verify the existence of outliers in the experimental data.

2 Materials and Methods

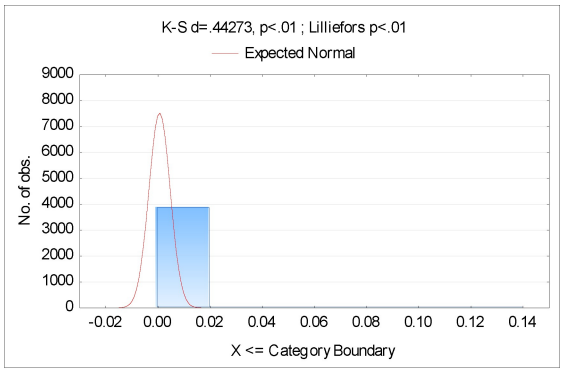
2.1 Experimental Dataset

A total of 3,971 experimental data points (EDP) of drug (148 drug-like compounds) solubility (mole fraction 0.000000001–0.131) measured in supercritical carbon dioxide from 298 K to 373.15 K and from 80 bar up to 500 bars were collected. The dataset of a previous work performed by Si-Moussa et al.[53] was updated by increasing the number of drug compounds from 99 (except for Azadirachtin) to 148 (Table 1), accounting from 2905 to 3971 experimental data points and increasing the number of inputs by taking the two inputs of the previous study (the temperature and the pressure) and adding the molecule descriptors (The detailed drug solubility data are tabulated in the Supporting Information). Then, the output property was converted to the corresponding pLn property "-Ln(y2)" to ensure a normality distribution, Figure 1.

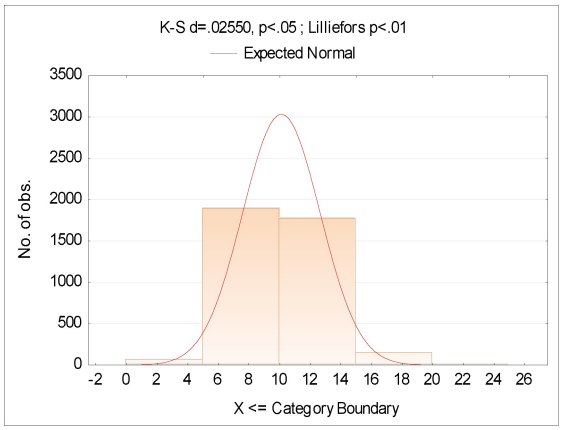
Chembioa drawing software was used to determine the SMILES notation of the 148 drug-like compounds. Their molecular descriptors (1544) were calculated using an online computer software program called PaDEL.[69] The 1544 descriptors were reduced to 124 using several variable reduction methods, starting by removing features with zero values. Then, the invariant descriptors, namely those with absent values or those with almost the same values "descriptor with all values ~ constant", were manually removed. Any descriptor that have identical values for 75% of the samples and all descriptors with a relative standard deviation <0.05 were removed. Finally, half of the descriptors showing an absolute value of the Pearson correlation coefficient >0.95 were also removed.

2.2 Selection of Relevant Descriptors

The genetic algorithm method (GA-MLR) was applied to reduce more the inputs to the smallest significant combination able to predict the target property with an acceptable error. The algorithm took as inputs the training set file (80% of database) with an inter-correlation of 0.9. The selection of the appropriate number of descriptors was carried out over an interval of 5 to 19 features, forming 14 combinations where each combination contained the data of the temperature, pressure, and from 3 to 17 descriptors. In order to select the best combination, the 14 models obtained were statistically examined. The results proved the over-performance of the model that has a combination of 13 features, best-fit combination. Multi-collinearity between the descriptors of the selected model was detected by calculating their variation inflation factors (VIF) using XLSTAT® (Addinsoft, New York, NY, USA). A model is judged acceptable if the VIF fell into the range of 1–5, Table 2. All VIF values of the descriptors were lower than 3.265, indicating that the obtained model has statistical significance, and the descriptors were found to be reasonably orthogonal. Therefore, the selected descriptors were as follows: AATS3v, MATS2e, GATS4c, GATS3v, GATS4e, GATS3 s, nBondsM, AVP-0, SHBd, MLogP, and MLFER_S, with the addition of T (K) and P (MPa). More details on the



(a) log-normal.



(b) log-transformed.

Figure 1. Normality distribution plot of mole fraction of solubility data.

 © 2022 Wiley-VCH GmbH

Table 1. Drug names and references.

No.	Names	References
1	4-aminoantipyrine	[91]
2	4-Aminosalicylic Acid	[92]
3	5-fluorouracil	[93]
4	10-undecenoic acid	[94]
5	all-trans retinoic acid	[95]
6	amiodarone hydrochloride	[96]
7	Amlodipine Besylate	[97]
8	Anastrozole	[98]
9	Aprepitant	[99]
10	Aspirin	[100]
11	Atorvastatine	[101]
12	Atropine	[102]
13	Azathioprine	[103]
14	Azelaic Acid	[104]
15	Benzocaine	[105]
16	Bisacodyl	[106]
17	Budesonide	[107]
18	Caffeine	[108]
19	Capecitabine	[109]
20	Carbamazepine	[102]
21	Cefixime Trihydrate	[110]
22	cefuroxime axetil	[111]
23	Celecoxib	[112]
24	Cetirizine	[113]
25	Chloramphenicol	[114]
26	Chlormezanone	[115]
27	Chlorpheniramine Maleate	[116]
28	Chlorpropamide	[117]
29	Chrysin	[118]
30	Cinildipine	[119]
31	Cinnarizine	[120]
32	Ciprofloxacin	[121]
33	clemastine fumarate	[122]
34	Climbazole	[123]
35	Clobetasole Propionate	[124]
36	Clofenamic acid	[125]
37	Clotrimazole	[126]
38	Clozapine	[127]
39	Codeine	[102]
40	Cortisone Acetate	[128]
41	Coumarin-7	[129]
42	Cyproheptadine	[130]
43	Cyproterone Acetate	[131]
44	Dasatinib monohydrate	[132]
45	Dexamethasone	[133,134]
46	Diazepam	[102]
47	Diclofenac Acid	[135]
48	Diflunisal	[136]
49	Docetaxel	[109]
50	Dutasteride	[137]
51	Enrofloxacin	[121]
52	Esomeprazole	[138]
53	Ethosuximide	[139]
54	Exemestane	[98]
55	Finasteride	[137]
56	Fluoxetine hydrochloride	[140]
57	Flurbiprofen	[141]
58	Flutamide	[137]
59	Fluvastatin	[100]
60	Gabapentin	[142]
61	Galantamine	[143]
62	gambogic acid	[144]
63	Gatifloxacin	[121]
64	Gemfibrozil	[145]
65	Ibuprofen	[93]
66	Imatinib mesylate	[146]
67	Imipramine HCl	[147]
68	Ipriflavone	[148]
69	Isoniazid	[149]
70	Ketoconazole	[126,150]
71	Ketoprofen	[151]
72	Ketotifen fumarate (KTF)	[152]
73	Lacidipine	[153]
74	Lamotrigine	[127]
75	Lansoprazole	[154]
76	Letrozole	[98]
77	Levonorgestrel	[155]
78	Loratadine	[156]
79	losartan potassium	[157]
80	Lovastatine	[100]
81	Mandelic Acid	[158]
82	Medroxyprogesterone Acetate	[131]
83	Mefenamic acid	[159]
84	Megestrol Acetate	[155]
85	Meloxicam sodium salt	[160]
86	Metaxalone	[115]
87	Methimazole	[161]
88	Methocarbamol	[115]
89	Methylparaben	[106]
90	Metronidazole Benzoate	[105]
91	Minoxidil	[162]
92	Nabumetone	[163]
93	Naproxen	[105]
94	Nicotinic Acid	[164]
95	Nifedipine	[153]
96	Niflumic Acid	[112]
97	Nimodipine	[165]
98	Nitrendipine	[166]
99	Norfloxacin	[167]
100	Ofloxacin	[167]
101	O-Nitrobenzoic Acid	[168]
102	Oxcarbazepine	[169]
103	Oxymetholone	[110]
104	Paclitaxel	[93]
105	Paeonol	[170]
106	Penicillin G	[171]
107	Penicillin V	[172]
108	Pentoxifylline	[120]
109	Phenazopyridine	[161]
110	Phenylephrine Hydrochloride	[173]
111	Pholcodine	[174]
112	Pindolol	[175]
113	Piracetam	[120]
114	Piroxicam	[176]
115	Prednisolone	[128]
116	Prednisone	[128]
117	Propranolol	[161]
118	protocatechuic acid	[118]
119	pyridin-4-amine	[177]
120	Pyrocatechol	[178]

© 2022 Wiley-VCH GmbH

Table 1. continued

No.	Names	References
121	Quetiapine hemifumarate	[179]
122	Quinine	[180]
123	Repaglinide	[181]
124	Rosuvastatin	[100]
125	Salicylamide	[160]
126	Salicylic Acid	[182]
127	sertraline hydrochloride	[183]
128	Simvastatine	[100]
129	sinapic acid	[118]
130	Sodium Valproate	[184]
131	Sorafenib tosylate	[185]
132	Spironolactone	[186]
133	Sulfabenzamide	[187]
134	Sulindac	[188]
135	Sunitinib malate	[189]
136	Tamsulosin	[190]
137	Teriflunomide	[191]
138	Tetramethylpyrazine	[192]
139	Theobromine	[108]
140	Theophylline	[108]
141	Thymidine	[93]
142	Tolbutamide	[117]
143	Tolfenamic acid	[125]
144	Triamterene	[193]
145	Triclocarban	[123]
146	Vitamin E Acetate	[194]
147	Warfarin	[195]
148	Zopiclone	[165]

characterization of the selected descriptors can be found in Table 2.

2.3 Artificial Neural Network

Artificial neural networks (ANNs) are computing systems consisting of a number of simple, highly interconnected processing units, which were developed by McCulloch and Pitts.[70–73] Such a system processes information through its dynamic state response to external inputs. These information processing patterns were inspired by mimicking the neural network of the human brain.[71] The ANN algorithm provides efficient performances for dynamic and nonlinear regression and classification problems.[72] The algorithm consists of input, hidden, and output layers. Each layer contains a set of interconnected processor components (neurons or nodes) whose output is the input layer of the next layer.

The output signal from one layer is connected to the next layer by means of weight factors through an intermediate that amplifies or weakens the signals.[71,74] A transfer function (activation function) like a linear or sigmoid function was used to calculate the outputs of nodes in the hidden and output layers. The number of neurons in the input and output layers is the number of input and output variables, respectively. There is no specific way to give the number of neurons in the hidden layer; thus, the number of hidden layers and the nodes in each hidden layer is usually related to the complexity of the problem and the trial-and-error method.[72,75] There are various neural networks with different training algorithms. A multilayer perceptron (MLP) trained with Bayesian regularisation backpropagation (trainbr) was considered in this study. MLP are feed-forward multilayer networks that supply flexible frameworks for non-linear function prediction.[76] The ANN modelling process can be summarised in two main phases: (1) assigning network structure and (2) adjusting the weight of connections between

Table 2. Characteristics of the descriptors selected in the optimal model and their VIF.

Descriptors Name	Type	Class	Description "Physical-chemical meaning"	VIF
AATS3v	2D	Auto correlation	Average Broto-Moreau autocorrelation – lag 3/weighted by van der Waals volumes	1.782
MATS2e	2D	Auto correlations	Moran autocorrelation of lag 2 weighted by Sanderson electro-negativity	1.226
GATS4c	2D	Auto correlations	Geary autocorrelation – lag 4/weighted by charges	2.875
GATS3v	2D	Auto correlations	Geary autocorrelation of lag 3 weighted by van der Waals volume	1.292
GATS4e	2D	Auto correlations	Geary autocorrelation of lag 4 weighted by Sanderson electro-negativity	3.265
GATS3 s	2D	Auto correlations	Geary autocorrelation of lag 3 weighted by I-state	1.247
nBondsM	2D	Bond Count	Total number of bonds that have bond order greater than one (aromatic bonds have bond order 1.5).	2.472
AVP-0	2D	Chi Path	Average valence path, order 0	2.090
SHBd	2D	Electro-topological State Atom	Sum of E-States for (strong) hydrogen bond donors	1.823
MLogP	2D	MannholdLogP	Moriguchi octanol-water partition coeff. (logP)	2.207
MLFER_S	2D	Molecular linear free energy relation "MLFER"	Combined dipolarity/polarizability	1.947
T(K)	–	–	Temperature	1.024
P(MPa)	–	–	Pressure	1.030

 © 2022 Wiley-VCH GmbH

neurons, as explained in [71]. More details can be found at [71, 75–77].

2.4 ANN-QSPR Modeling

For this study, the dataset of 3971 EDP were randomly divided into two main sets. The first set contained 3828 EDP that were used to develop the model, while the second set of 143 was hidden from the model and kept as an external test (validation set). Using the selected 13th early described features, the first set of 3828 EDP was taken to build the feed-forward MLP model, which contained a single input, hidden and output layer. The implementation of the algorithm was carried out under MATLAB® R2019b environment. The dataset was divided using MATLAB® divider and function into two random sets as follows: 70% (2695 EDP) for the training set and 30% (1155 EDP) for the test set. The net was trained using the training function trainbr "Bayesian regularisation back-propagation". The logsig "Log-sigmoid" function was chosen as a transfer function for hidden layer and purelin "Linear transfer" function for the output layer. The number of hidden neurons was selected to 10 "13 nodes of the input × 0.75 ~ 10". One output neuron was used to represent the observed solubility fraction (y_2). Several 25–30 iterations were carried out. The model was optimized based on their average absolute relative error (AARD%) as an objective function. Therefore, the MLP-ANN has the following architecture {13-10-1}. Figure 2 illustrates the architecture of MLP-ANN used in this study.

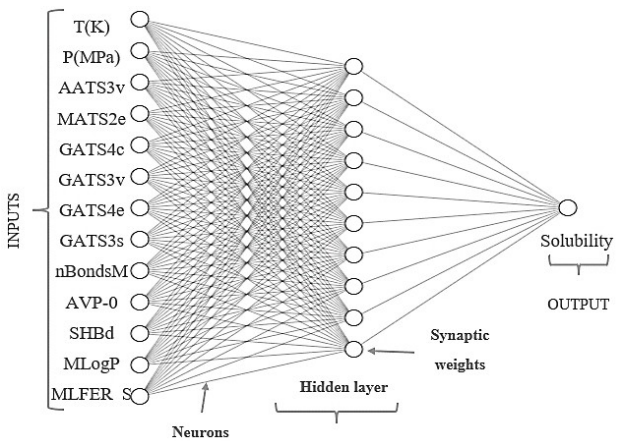


Figure 2. MLP-ANN structure.

2.5 Evaluation of the Model's Performance

Various statistics metrics were adopted to evaluate the performance of the QSAR model, as defined[78–81] by:

$$Q^2 = 1 - \frac{\sum_{i=1}^{N}\left(y^{obs}_{i\,(training)} - y^{pred}_{i\,(training)}\right)^2}{\sum_{i=1}^{N}\left(y^{obs}_{i\,(training)} - \bar{y}^{obs}_{(training)}\right)^2} \quad (1)$$

$$R^2 = r^2 = 1 - \frac{\sum_{i=1}^{N}\left(y^{obs}_i - y^{pred}_i\right)^2}{\sum_{i=1}^{N}\left(y^{obs}_i - \bar{y}^{obs}\right)^2} \quad (2)$$

$$R = r = \frac{\sum_{i=1}^{N}\left(y^{obs}_i - \bar{y}^{obs}\right)\cdot\left(y^{pred}_i - \bar{y}^{pred}\right)}{\sqrt{\sum_{i=1}^{N}\left(y^{obs}_i - \bar{y}^{obs}\right)^2}\cdot\sqrt{\sum_{i=1}^{N}\left(y^{pred}_i - \bar{y}^{pred}\right)^2}} \quad (3)$$

$$k = \frac{\sum_{i=1}^{N}\left(y^{obs}_i \cdot y^{pred}_i\right)}{\sum_{i=1}^{N}\left(y^{pred}_i\right)^2} \quad (4)$$

$$k' = \frac{\sum_{i=1}^{N}\left(y^{obs}_i \cdot y^{pred}_i\right)}{\sum_{i=1}^{N}\left(y^{obs}_i\right)^2} \quad (5)$$

$$r_0^2 = 1 - \frac{\sum_{i=1}^{N}\left(y^{obs}_i - k y^{pred}_i\right)^2}{\sum_{i=1}^{N}\left(y^{obs}_i - \bar{y}^{obs}\right)^2} \quad (6)$$

$$r'^2_0 = 1 - \frac{\sum_{i=1}^{N}\left(y^{obs}_i - k' y^{pred}_i\right)^2}{\sum_{i=1}^{N}\left(y^{obs}_i - \bar{y}^{obs}\right)^2} \quad (7)$$

$$m = \frac{\left(r_0^2 - r'^2_0\right)}{r_2} \quad (8)$$

$$n = \frac{\left(r_2 - r_0^2\right)}{r_2} \quad (9)$$

$$r_m^2 = r^2\left(1 - \sqrt{r^2 - r_0^2}\right) \quad (10)$$

$$r_m'^2 = r^2\left(1 - \sqrt{r^2 - r_0'^2}\right) \quad (11)$$

$$\bar{r}_m^2 = \frac{\left(r_m^2 + r'^2_m\right)}{2} \quad (12)$$

$$\Delta r_m^2 = \left[r_m^2 - r'^2_m\right] \quad (13)$$

$$\bar{y}^{obs} = \frac{\sum_{i=1}^{N} y^{obs}_i}{N} \quad (14)$$

$$\bar{y}^{pred} = \frac{\sum_{i=1}^{N} y^{pred}_i}{N} \quad (15)$$

 © 2022 Wiley-VCH GmbH

Table 3. Acceptability criteria (A.C) of a model.

Parameter	m	n	R^2	Q^2	Q^2_{pred}	K	k′	$\bar{r}^2_m$	Δr^2_m	B_f	A_f
A.C value	<0.1		>0.6	>0.5	>0.6	$0.85<$ <1.15		>0.5	<0.2	0.9–1.05	

$$PRESS = \sum_{i=1}^{N} \left(y_i^{obs} - y_i^{pred}\right)^2 \quad (16)$$

$$SSY = \sum_{i=1}^{N} \left(y_i^{pred}\right)^2 \quad (17)$$

$$AARD\% = 100 \bullet \frac{1}{N}\sum_{i=1}^{n}\left|\frac{y_i^{obs} - y_i^{pred}}{y_i^{obs}}\right| \quad (18)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_i^{obs} - y_i^{pred}\right)^2 \quad (19)$$

$$RMSE = \sqrt{\frac{\sum\left(y_i^{obs} - y_i^{pred}\right)^2}{N}} \quad (20)$$

$$MAE = \frac{1}{N} X \sum_{i=1}^{N}\left|y_i^{obs} - y_i^{pred}\right| \quad (21)$$

$$Bf = 10^{\left(\frac{\sum_{i=1}^{N}\log\frac{y_i^{obs}}{y_i^{pred}}}{N}\right)} \quad (22)$$

$$Af = 10^{\left(\frac{\sum_{i=1}^{N}\left|\log\frac{y_i^{obs}}{y_i^{pred}}\right|}{N}\right)} \quad (23)$$

Where, Q^2 is the cross-validated correlation coefficient (also notated CV, q^2), R^2 is the coefficient of determination, r is the correlation coefficient between the observed and predicted value of compounds with interception, k & k′ are the slopes of the corresponding regression lines, r_0^2 is the squared correlation coefficient between the observed and predicted value of compounds without intercept, r'^2_0 it bears the same meaning as r_0^2, but uses the reversed axes, $\bar{r}^2_m \& \Delta r^2_m$ are theaverage and delta of r^2_m, y_i^{obs} is the observed (experimental) value of Y, $y_i^{pred}$is the predicted (calculated) Y-value of training set, test set or validation set, N is the Number of data points in the data set (training, test…), $\bar{y}^{obs}$ is the Average of Y_{obs} , $\bar{Y}_{pred}$ is the Average of $\bar{y}^{pred}$, PRESS is the Predictive Residual Sum of the Squares, SSY is the Sum of the Squares of the response values of the total observations, AARD is the Average Absolute Relative Deviation, MSE is the Mean Squared Error, *RMSE* is the Root-Mean Squared Error, *MAE* is the mean absolute error, *Bf* is the bias factor, and *Af* is the precision factor.

These statistical metrics were used to decide whether the model is sound and unbiased ("good model") or not. These metrics were used as acceptability criteria, as shown in Table 3.

In addition to:

- R^2 -Q^2 indicate whether the model is overfitted or not. It should not be more than 0.3.
- PRESS/SSY ratio smaller than 0.4 indicates that the QSAR model is reasonable. A ratio below 0.1, indicates an excellent model, and proves that the developed models predict better than chance.
- Perfect agreement between the observed values and the predicted values give a *Bf* (or A_f, after ignoring whether the difference between the predicted value and the observed value is positive or negative) of 1. A value greater or less than 1 indicates an intrinsically safe. More detail can be found elsewhere.[78–81]

Another interesting statistic parameter, namely, Akaike's information criterion (AIC)[82,83] was also used:

$$AIC = N\ln\left(\frac{SSE}{N}\right) + 2n_p + \frac{2n_p(n_p+1)}{N-(n_p+1)} \quad (24)$$

$$SSE = \sum_{i=0}^{N}\left(y_i^{obs} - y_i^{pred}\right)^2 \quad (25)$$

where N is the number of experimental samples, Np is the number of the parameters in the model and SSE is the sum of square error. AIC is a statistical metric used to compare different models and detect the best model that represents the experimental results. A small AIC value indicates the better fitted model. Typically, a single AIC value is meaningless; the AIC value of the model studied in this paper was compared to three ANN models.

2.6 Applicability Domain

For this study, William's plot was adopted to verify the domain of applicability of QSAR model. William's plot is a plot of standardized residuals versus leverage values, Hat diagonal values (hi) are commonly used to visualize the applicability domain[84–86] and to detect outliers' response. i.e., compounds with standardized residuals greater than two standard deviation units (Y outliers) and structurally influential chemicals (X outliers) in the model, i.e., hi $>$ h*. In

 © 2022 Wiley-VCH GmbH

this plot, the applicability of domain was established inside a squared area within ±3 standard deviations and a threshold leverage value (h*).

The leverage (h) value of a compound in the original independent variable space and the warning leverage (h*) are defined as follows.[87]

$$hi = x_i(X^TX)^{-1}x_i \quad (26)$$

$$h^* = \frac{3(p+1)}{n} \quad (27)$$

Where, x_i is the descriptor-row vector of the query compound, X k×n is a matrix containing the k descriptor values for each one of the n training compounds, n is the total number of samples in the training set, and p is the number of descriptors involved in the correlation.

2.7 Sensitivity Analysis

ANNs are powerful tools able to predict and analyse easily non-linear relationships. Numerous authors underlined the interest in using ANNs instead of linear statistical models. In fact, starting from input variables, ANNs have the capacity to predict the output variable, but the mechanisms that occur within the network are often ignored. ANNs are often considered as black boxes. Several authors explored these problems and proposed algorithms to illustrate the role of variables in ANN models, the contribution of features. Determination of the contribution of each variable is of prime importance. This explanatory aspect of ANNs was examined in the present study. These methods were used to determine the influence of each input variable and its contribution to the output. They are procedures to estimate the relative contribution of each input variable.

Seven different methods that allow contribution analysis were used. The Garson method to calculate the relative importance of each input variable[88] was considered in this study, as well as the relevance factor *r*, which is in the range of −1 to +1 and is stated by the following equation.[89]

$$r_k = \frac{\sum_{i=1}^{n}(X_{k,i}-\bar{X}_k)(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_{k,i}-\bar{X}_k)^2\sum_{i}^{n}(Y_i-\bar{Y})^2}} \quad (28)$$

Where, $X_{k,I}$ is the i^{th} importing parameter, Y_i is the i^{th} exporting value, $\bar{X}_k$ is the average value of the k^{th} input, $\bar{Y}$ is the average value of exporting parameter, and *n* is the number of sets.

3 Results and Discussion

3.1 Model's Performance

Figure 3 represents the scatter plots of the developed model. i.e., predicted vs. observed values for the training, test, and global set. A close correlation between the values predicted by the ANN model and the observed values of the mole solubility fraction was found, except for some outliers points point, considered as accepted out of range. Table 4 shows the statistical metrics results of the training test and global set, Tables 5 and 6 show the error parameter for the three sets and the AIC values respectively.

Table 4. Validation parameters for the train, test, and global sets for MLP-ANN model.

Statistical values	Train	Test	Global
Q^2	0.9529	0.9528	0.9528
R^2	0.9529	0.9528	0.9528
r	0.9762	0.9761	0.9761
k	1.0000	0.9975	0.9993
k′	0.9976	0.9998	0.9983
R_0^2	1.0000	0.9999	1.0000
R'^2_0	0.9999	1.0000	0.9999
m	−0.0415	−0.0472	−0.0432
n	−0.0414	−0.0473	−0.0432
R_m^2	0.7683	0.7594	0.7687
$R_m'^2$	0.7685	0.7695	0.7688
Bf	1.0014	1.0034	1.0020
Af	1.0379	1.0391	1.0383
$\bar{r}_m^2$	0.7684	0.7694	0.7687
Δr_m^2	0.0002	0.0001	0.0001

Table 5. Different errors for the train, test, and global sets for MLP-ANN model.

Error	Train	Test	Global
AARD%	3.7338	3.8704	3.7748
RMSE	0.5084	0.5339	0.5162
MSE	0.2585	0.2851	0.2665
MAE	0.3651	0.3690	0.3621

In summary, the results above are in agreement with the acceptability criteria. Where the non-linear ANN model gave good results with the highest global correlation coefficients R (0.976), as well as the best robustness (Q2) 0.953, and the coefficients of determination R2 0.953. In addition, to the error parameters that showed highly acceptable results such as the global RMSE (0.516) and AARD% (3.775). The calculation of the coefficient of determination minus the cross-validated correlation coefficient was lower than 0.3, indicating non-over-fitting for

 © 2022 Wiley-VCH GmbH

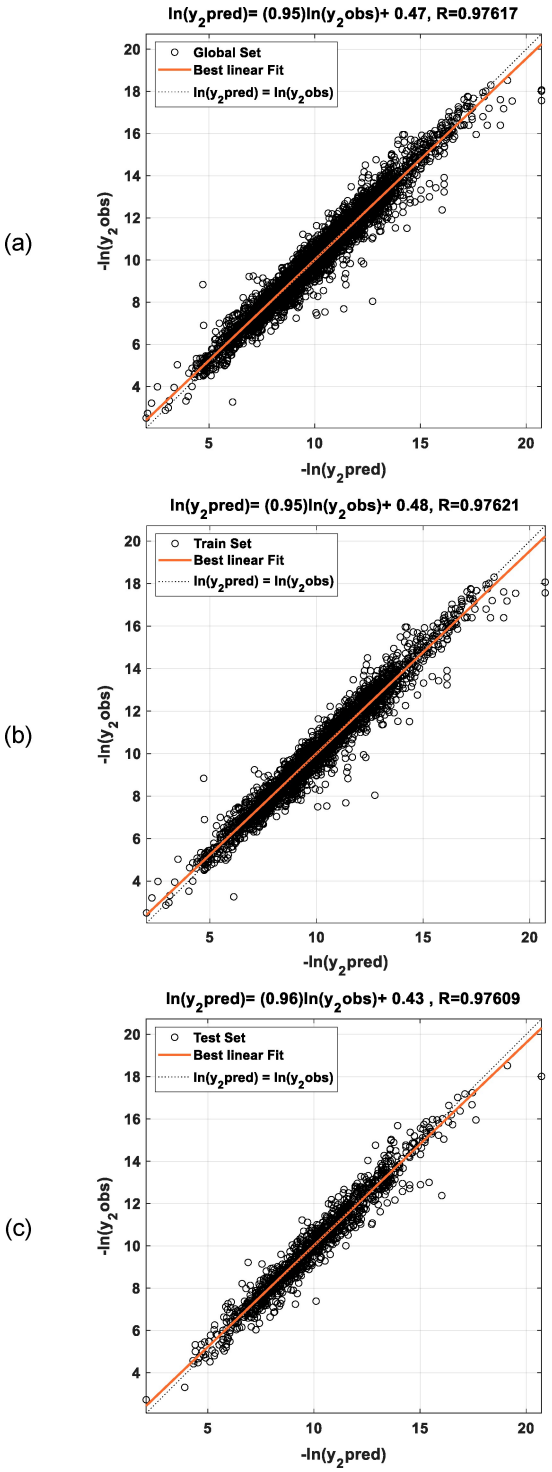


Figure 3. Scatter plots Predicted vs. Observed values of the solubility for the MLP-ANN model: (a) Global Set, (b) Training Set, (c) Test Set, respectively.

this model. The calculated global PRESS value divided on the SSY were lower than 0.1. This proved the efficiency of the developed model, which predicts better than chance. These results indicated that the ANN not only performed well in model development but also had excellent prediction capacity, suggesting a non-linear correlation between the solubility fraction and the relevant descriptors.

$$-\mathit{ln}(solubility) = w_2 \cdot \left(\frac{Inputs}{1 - e^{-w_1 \bullet Inputs}} + b_1 \right) + b_2 \qquad (29)$$

The obtained w_1, w_2, b_1 and b_2 are provided in the supporting information.

In Table 6, the values of AIC for the four ANN models clearly denote that the experimental data is well fitted by corresponding model of this study.

3.2 William's Plot

The applicability domain (AD) of the model was analyzed using William's plot,[87] where the vertical line is the critical leverage value (h*), and the horizontal lines ±3 are the cutoff value for the cutoff value for Y_{space}. In Figure 4 we can

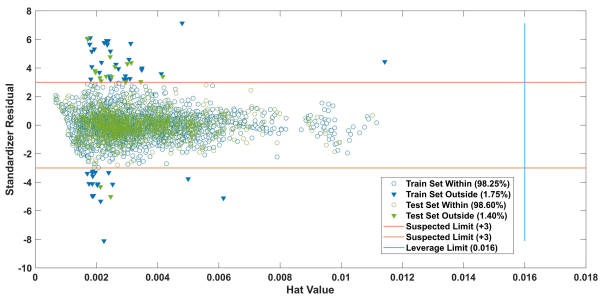


Figure 4. William's plot for MLP-ANN model.

observe that none of the drug-like compounds in the whole dataset have leverage higher than the warning h* value of 0.016. 2633 data points belonging to the train set were in-between the two horizontal lines, which means 98.25% of the train set was covered. Out of 1148 data points of the test set, only 16 data point showed a distribution outside the two horizontal lines (98.6% of the test set was covered). 98.35% of the whole database was covered. It should be noted that only 63 EDP showed a distribution out of the two horizontal lines (1.65% outranges). Thus, these results show that the MLP-ANN model complied with the third principle of the OECD. Accordingly, the model developed in this study provided excellent predictions for 148 drug-like compounds (3828 EDP). It can be used to predict the molar solubility fraction of those compounds, particularly for those that have not been tested as well as new drugs.

 © 2022 Wiley-VCH GmbH

Table 6. AIC values of the four ANN models.

Model	Training function	Activation function Hidden layer	Output layer	AIC
Corresponding Model	trainbr	logsing	purelin	−5034.59
Model 1	trainbr	tansing	purelin	−4895.65
Model 2	trainlm	logsig	purelin	−4926.54
Model 3	trainlm	tansing	purelin	−4527.42

3.3 External Test

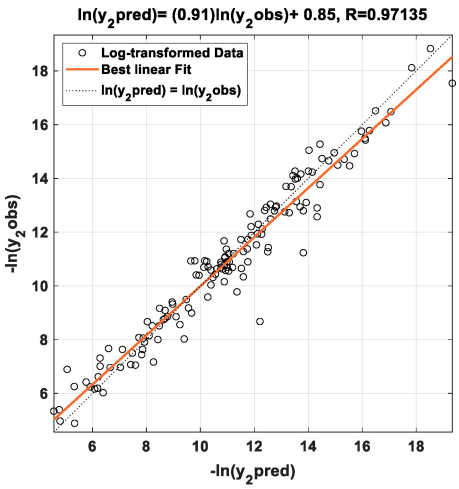
The hidden dataset from the model, which included 143 EDP, for the same selected number of features, was used for the exploitation of the models. i.e., prediction of the solubility fraction of a non-used dataset, known as the validation set or external test. The correlation between both log-transformed and original data is shown in Figure 5, proving that this model can represent the predicted and the observed solubility fraction of validation set with a close positive correlation; while the Table 7 shows an acceptable error for both cases. To this end, the model is valid for prediction of future data.

3.4 Descriptor's Contribution and Interpretation

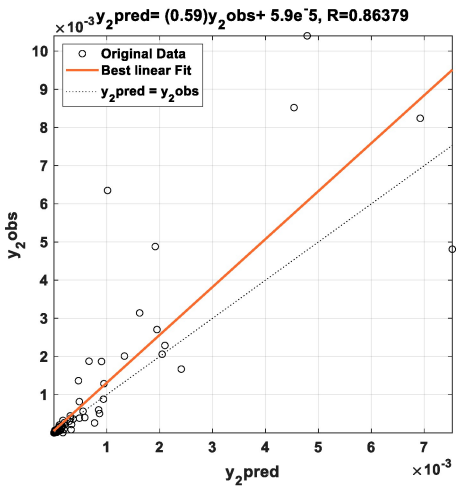
After the selection of the molecule descriptors, 11 pertinent descriptors with two intensive state variables (pressure and temperature) were used to develop the ANN-QSPR model, which were nBondsM, GATS4c, AVP-0, MLFER_S, AATS3v, SHBd, MLogP, GATS4e, GATS3 s, GATS3v, MATS2e, P (MPa) and T (K). To evaluate the scope of each descriptor for the prediction of solid drugs solubility, the relevance and the relative contributions of the 13 descriptors to the ANN model were determined and a description of these descriptors is shown.

The absolute value of the relevance factor has a direct relation to the output. As can be seen in Figure 6, the solubility showed a straight dependency on the descriptors named: AATS3v, nBondsM, SHBd, MLogP, and MLFER_S, and an opposite dependency on the temperature and the pressure, and the rest of descriptors. In addition, MLFER_S, and pressure were the most relevant input variables, with a relevance factor of +0.25 and −0.24 respectively. Contrariety, the temperature and nBondsM were the least relevant, with a relevance factor of −0.03 and +0.01, respectively.

Figure 7 depicts sensitivity analysis for ANN-QSPR model and shows that the GATS4e and GATS3 s share the highest contribution (11.59% and 11.57%, respectively) where the lowest contribution was by T (K) (1.94%). The rest of the inputs had an importance between 4.43% and 11.06%. The relative contribution level of the descriptors was found to be in the following order: GATS4e, GATS3 s, AATS3v, GATS4c, SHBd,MLogP, MLFER_S, AVP-0,nBondsM, GATS3v, MATS2e,P(MPa) and T(K) (Figure 7). In this case, we can see



(a) log-transformed data.



(b) original data.

Figure 5. Scatter plots Predicted vs. Observed values of the solubility for the MLP-ANN model, external test.

that the variables that have the highest relative contribution were successively GATS4e, GATS3 s, AATS3v.

GATS4c, AATS3v, GATS4e, are a 2D auto correlation molecular descriptor, which is known respectively as the geary autocorrelation-lag 4/weighted by charges, autocorrelation of lag 3 weighted by vdw volume, geary autocorre-

 © 2022 Wiley-VCH GmbH

Table 7. Different errors for the validation set for both cases.

Error	Log-transformed Data	original Data
AARD%	5.1982	99.3617
RMSE	0.7507	8.4365e–04
MSE	0.5635	7.1175e–07
MAE	0.5395	0.0013

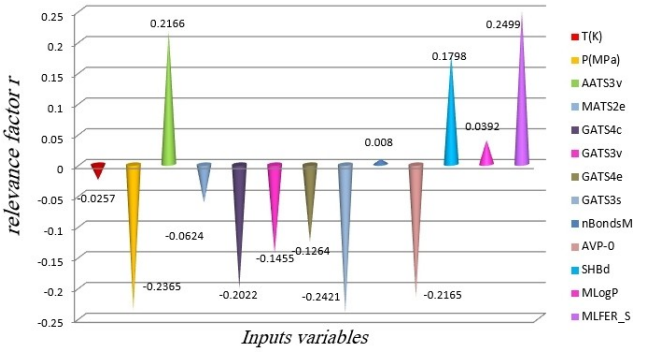


Figure 6. Results of the relevance factor performed on the QSPR-ANN model.

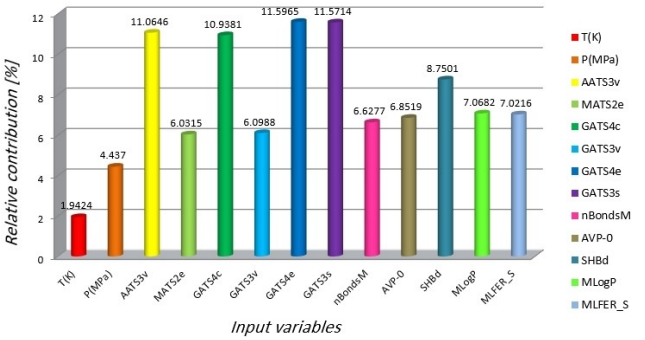


Figure 7. Relative contribution of each input variable using the Garson method.

lation of lag 4 weighted by atomic Sanderson electro-negativity. SHBd describes the electropological state (E-state) in different aspects. SHBd represents the sum of E-states for (strong) hydrogen bond donors. An increase in the value of this descriptor means that the molecule is more and more hydrophobic. AVP-0 is a topological descriptor that gives information regarding the connectivity of various atoms in the molecule. This descriptor is able to take into account the presence of heteroatoms in a molecule, as well as double and triple bonds. GATS3 s is 2D auto correlation molecular descriptors, which describes how the considered property (atomic masses, polarizability, charges and electro-negativities) is distributed along a topological molecular structure. On the other hand, GATS3v and MATS2e are defined as Geary autocorrelation of lag 3 weighted by van der Waals volume and Moran autocorrelation of lag 2 weighted by Sanderson electro-negativity, respectively. The molecular descriptor MLOGP (which is the square root of MLOGP2) is considered as a measure of lipophilicity of a molecule. The MLFER_S descriptor is a descriptor for combined dipolarity/polarizability. It is one of the molecular linear free energy relations. The 2D descriptor "nBondsM" is defined as the dimensionless total number of bonds that have bond order greater than one (for example aromatic bonds have bond of 1.5) and is encoded to an unsaturation. It referred to as the number of non-kekulized structures. In summary, in this model, it is concluded that the lipophilicity, electro-topology and electronic properties (polarizability, electro-negativity, atomic masses, van der Waals volume, unsaturation, and charges) of a molecule are of major importance in the solid drugs solubility, whereas temperature and pressure does not appear to be a major factor. Although both are very important in experiments. The mathematical formula for predicting the molar solubility fraction of drug-like compounds obtained using XLSTAT® (**Addinsoft**, New York, NY,USA) is given by:

$$\begin{aligned} Ln(y_2) = &-(22.4581 - -7.2613 \cdot 10^{-3} \cdot T(k) - \\ &9.0914 \cdot 10^{-2} \cdot p(MPa) + 0.0184 \cdot AATS3v - \\ &2.1674 \cdot MATS2e - 2.9553 \cdot GATS4c - \\ &3.6541 \cdot GATS3v + 1.6024 \cdot GATS4e - \\ &2.3767 \cdot GATS3s - 0.1746 \cdot nBondsM - \\ &11.0664 \cdot AVP0 + 0.6300 \cdot SHBd + 0.5326 \cdot MLogP + \\ &1.0102 \cdot MLFER_S \end{aligned} \quad (30)$$

3.5 The use of QSAR for Regulatory Purposes

This study aims to use QSARs combined with an ANN as an alternative for experimental data to predict and correlate drug solubilisation in supercritical carbon dioxide at specific temperature and pressure values. To facilitate the consideration of a QSAR model for regulatory purposes, the development of QSPR models should comply to the five principles for the QSAR validation drawn by the Organisation for Economic Cooperation and Development (OECD),[90] namely: (1) a defined endpoint; (2) an unambiguous algorithm; (3) a defined domain of applicability; (4) appropriate measures of goodness of fit, robustness, predictability; (5) a mechanistic interpretation, if possible. These criteria were applied to the newly ANN-QSPR model developed in the present work:

1. *OECD criterion 1:* the first step of the QSAR development was the selection of the appropriate endpoint (s). The selected endpoints solubility for drug and drug-like compounds was further defined according to the following endpoints criteria:

 © 2022 Wiley-VCH GmbH

- Target: Molar solubility fraction.
- Solid compounds: drug or drug-like compounds.
- Compounds solubility in SC–CO_2.
- Temperature range: 298–373.15 k.
- Pressure range: 80–500 bars.
- Solubility mole fraction range: 0.000000001–0.131.
- Molecular descriptor: PaDEL Descriptors of the solid compounds.
- Molecular descriptor range: AATS3v (114.80–411.05), MATS2e (−0.25–0.45), GATS4c (0.46–1.66), GATS3v (0.66–1.39), GATS4e (0.38–1.91),GATS3 s (0–1.91), nBondsM (1–27), AVP-0 (0.47–0.73), SHBd (0–2.36), MLogP (1.35–4.98), and MLFER_S (0.22–6.19).

The QSPR-ANN model predicts a defined endpoint.

2. *OECD criterion 2:* the MLP-ANN model developed in this present work is a clear and an easily applicable algorithm carried out under MATLAB® R2019b environment. Both training and activation functions are defined, the number of layers and the number of nodes in each layer is already given, which fulfil the criterion on an unambiguous algorithm. The transparency is ensured.
3. *OECD criterion 3:* QSAR models are only applicable within a specific set of conditions known as applicability domain, which identifies limitations and scope of the model. The applicability domain of the ANN-QSPR model developed was analyzed using a Williams plot, and it is well documented.
4. *OECD criterion 4:* The goodness of the model was reached by verifying the model fitting and the model robustness by calculating various statistical criteria, including the determination coefficient ($R^2 > 0.9$) and the cross-validation coefficient ($Q^2 > 0.9$). Executing Y-scrambling to confirm the existence of a correlation between the selected modelling features and the studied response. The internal performance ends up indicating a well-founded (not by chance) original model. Moreover, the predictivity (as determined by external validation) was reached by characterizing the correlation between predicted and experimental values in the validation set for both log-transformed and original data. In addition, the AARD, RMSE, MSE, MAE that summarizes the overall errors of model, were used to measure and compare prediction accuracy in the training, test and validation sets. The error values were close and there were no significantly large residual values for the validation set, demonstrating that the model had not overfitting. The model exhibited a satisfactory goodness-of-fit, robustness and predictivity.
5. *OECD criterion 5:* the mechanistic interpretation of solid drug solubility in SC–CO_2 shows to be mostly influenced by the pressure, lipophilicity, electro-topology and electronic properties of a molecule, while the temperature has the lowest influence. However, experimental measurements indicate that temperature and pressure have a remarkable influence on the solubility.

For this, the developed ANN-QSPR model for drug and drug-like compounds solubility in SC–CO_2 meets all the OECD criteria for use in a regulatory context.

4 Conclusions

In this work, supercritical carbon dioxide solubility data of 148 drugs collected from the literature were used to develop a combined ANN-QSPR model to predict the solubility of drugs in SC–CO_2. 11 molecular descriptors were retained as relevant after reducing the number of molecular descriptors using various reduction techniques, Then, a total of 3971 experimental data points, representing the solubility (mole fraction) of the drug in SC–CO_2 at different temperatures (K) and pressure (MPa), were used for training and validation of the model. The architecture of the optimized network was obtained by varying different network parameters. The best model was obtained using the training function trainbr "Bayesian regularization backpropagation", logsig "Log-sigmoid transfer function" activation function for the hidden layer, network architecture of {13–10-1}. Statistical studies recommended were tested to assess the predictive performance of the optimized model. The results showed that all the metrics fell within the domain of acceptability, thus confirming the model validity. All validations indicated that the built QSAR model was robust, satisfactory, meets all the OECD principles for QSAR validation and can be used to predict reliably the solubility of drug-like compounds, for interpolation purposes for drugs of the data set within the model space as well as for extrapolation purposes for drugs not included in the data set…

5 Data Availability

The Supplementary data to this article can be found online at (Database.pdf, supporting file 2.xlsx). The software, web servers and computational tools are owned by their respective developers and copyright-holders. All statistical and graphical analyses were obtained with codes we created using the MATLAB® R2019b environment.

Acknowledgements

The authors gratefully acknowledge the Algerian Ministry of Higher Education and Scientific Research (PRFU Project A16N01UN260120220003) and the University Yahia Fares of Medea.

 © 2022 Wiley-VCH GmbH

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

The Supplementary data to this article can be found online at (Database.pdf, supporting file 2.xlsx). The software, web servers and computational tools are owned by their respective developers and copyright-holders. All statistical and graphical analyses were obtained with codes we created using the MATLAB® R2019b environment.

References

[1] G. Sodeifian, S. A. Sajadian, F. Razmimanesh, N. S. Ardestani, *Korean J. Chem. Eng.* **2018**, *35*, 2097–2116.
[2] X.-Q. Bian, Q. Zhang, L. Zhang, J. Chen, *Chem. Eng. Res. Des.* **2017**, *123*, 284–294.
[3] I. Medina, J. L. Bueno, *J. Chem. Eng. Data.* **2000**, *45*, 298–300.
[4] M. Hojjati, Y. Yamini, M. Khajeh, A. Vatanara, *J. Supercrit. Fluids.* **2007**, *41*, 187–194.
[5] A. A. Amooey, *Fluid Phase Equilib.* **2014**, *375*, 332–339.
[6] A. Baghban, A. Jalali, A. H. Mohammadi, S. Habibzadeh, *J. Supercrit. Fluids.* **2018**, *133*, 466–478.
[7] M. Baghbanbashi, N. Hadidi, G. Pazuki, in *Green Sustainable Process for Chemical, Environmental Engineering, Science*), Elsevier Inc. **2020**, pp. 185–254.
[8] D. Jenke, A. Odufu, T. Couch, M. Chacko, S. Strathmann, E. Edgcomb, *PDA J. Pharm. Sci. Technol.* **2012**, *66*, 286–306.
[9] A. B. Hamidreza, B. S. Reza, C. S. Mahdi, D. Y. Mahdi, E. F. Fatemeh, F. H. Hassan, *J. Fundam. Appl. Sci.* **2016**, *8*, 839–859.
[10] R. Parhi, P. Suresh, *J. Adv. Pharm. Sci. Technol.* **2013**, *1*, 13–36.
[11] S. Hitchen, J. Dean, in *Applications of Supercritical Fluids in Industrial Analysis*), Springer. **1993**, pp. 1–11.
[12] R. B. Gupta, J.-J. Shim, *Solubility in supercritical carbon dioxide*, CRC press.
[13] R. E. Fornari, P. Alessi, I. Kikic, *Fluid Phase Equilib.* **1990**, *57*, 1–33.
[14] R. Dohrn, O. Pfohl, *Fluid Phase Equilib.* **2002**, *194*, 15–29.
[15] M. Christov, R. Dohrn, *Fluid Phase Equilib.* **2002**, *202*, 153–218.
[16] R. Dohrn, S. Peper, J. M. Fonseca, *Fluid Phase Equilib.* **2010**, *288*, 1–54.
[17] J. M. Fonseca, R. Dohrn, S. Peper, *Fluid Phase Equilib.* **2011**, *300*, 1–69.
[18] S. Peper, J. M. Fonseca, R. Dohrn, *Fluid Phase Equilib.* **2019**, *484*, 126–224.
[19] Z. Knez, D. Cör, M. Knez-Hrnčič, *J. Chem. Eng. Data.* **2017**, *63*, 860–884.
[20] M. Skerget, Z. Knez, M. Knez-Hrnčič, *J. Chem. Eng. Data.* **2011**, *56*, 694–719.
[21] W. Stiver, *J. Supercrit. Fluids.* **2021**, *177*, 105348.
[22] P. Coimbra, C. Duarte, H. De Sousa, *Fluid Phase Equilib.* **2006**, *239*, 188–199.
[23] H.-W. Wang, C.-M. Hsieh, *J. Supercrit. Fluids.* **2022**, *180*, 105446.
[24] M. Yazdizadeh, A. Eslamimanesh, F. Esmaeilzadeh, *J. Supercrit. Fluids.* **2011**, *55*, 861–875.
[25] K. Adenekan, B. Hutton-Prager, *Fluid Phase Equilib.* **2020**, *507*, 112415.
[26] G. Sodeifian, F. Razmimanesh, S. A. Sajadian, H. S. Panah, *Fluid Phase Equilib.* **2018**, *472*, 147–159.
[27] O. Faraz, M. Poustchi, E. N. Denyani, P. Movahedi, F. R. Kouchi, R. Shahriari, *J. Mol. Liq.* **2022**, *353*, 118809.
[28] Z.-Z. Cai, H.-H. Liang, W.-L. Chen, S.-T. Lin, C.-M. Hsieh, *Fluid Phase Equilib.* **2020**, *522*, 112755.
[29] M. R. Housaindokht, B. Haghighi, M. R. Bozorgmehr, *Korean J. Chem. Eng.* **2007**, *24*, 102–105.
[30] R. G. Bitencourt, A. M. Palma, J. A. Coutinho, F. A. Cabral, A. J. Meirelles, *Fluid Phase Equilib.* **2019**, *482*, 1–10.
[31] S. J. Nejad, H. Abolghasemi, M. Moosavian, M. Maragheh, *Chem. Eng. Res. Des.* **2010**, *88*, 893–898.
[32] A. Dashti, F. Zargari, H. R. Harami, A. H. Mohammadi, Z. Nikfarjam, *Korean J. Chem. Eng.* **2019**, *36*, 1637–1647.
[33] H. Asiabi, Y. Yamini, F. Latifeh, A. Vatanara, *J. Supercrit. Fluids.* **2015**, *104*, 62–69.
[34] B. Mehdizadeh, K. Movagharnejad, *Chem. Eng. Res. Des.* **2011**, *89*, 2420–2427.
[35] J. Deepitha, K. Pitchaiah, C. Brahmmananda Rao, G. Madras, N. Sivaraman, *Sep. Sci. Technol.* **2019**, *54*, 1650–1660.
[36] M. Ardjmand, M. Mirzajanzadeh, F. Zabihi, *Chin. J. Chem. Eng.* **2014**, *22*, 549–558.
[37] H. Zhu, L. Zhu, Z. Sun, A. Khan, *J. Mol. Liq.* **2021**, *338*, 116731.
[38] H. C. Nguyen, F. Alamray, M. Kamal, T. Diana, A. Mohamed, M. Algarni, C.-H. Su, *J. Mol. Liq.* **2022**, *354*, 118888.
[39] A. Baghban, J. Sasanipour, Z. Zhang, *J. CO2 Util.* **2018**, *26*, 262–270.
[40] A. Sadeghi, C.-H. Su, A. Khan, M. L. Rahman, M. S. Sarjadi, S. M. Sarkar, *Arab. J. Chem.* **2022**, *15*, 103502.
[41] T. Wang, C.-H. Su, *J. Mol. Liq.* **2022**, *349*, 118120.
[42] Z. Zhao, P. Liu, Y. Li, S. Zhang, L. Guo, S. Ghazali, A. El-Shafay, *J. Mol. Liq.* **2022**, 118573.
[43] M. Yazdizadeh, A. Eslamimanesh, F. Esmaeilzadeh, *Chem. Eng. Sci.* **2012**, *71*, 283–299.
[44] C. Zhong, H. Yang, *Ind. Eng. Chem. Res.* **2002**, *41*, 4899–4905.
[45] H. Yang, C. Zhong, *J. Supercrit. Fluids.* **2005**, *33*, 99–106.
[46] H. A. A. El, C. Si-Moussa, S. Hanini, M. Laidi, *Chem. Ind. Chem. Eng. Q.* **2013**, *19*, 449–460.
[47] M. H. Anvari, G. Pazuki, *J. Supercrit. Fluids.* **2014**, *90*, 73–83.
[48] Y. Shimoyama, Y. Iwai, *J. Supercrit. Fluids.* **2009**, *50*, 210–217.
[49] L.-H. Wang, S.-T. Lin, *J. Supercrit. Fluids.* **2014**, *85*, 81–88.
[50] L.-H. Wang, C.-M. Hsieh, S.-T. Lin, *Ind. Eng. Chem. Res.* **2015**, *54*, 10115–10125.
[51] Y.-H. Ting, C.-M. Hsieh, *Fluid Phase Equilib.* **2017**, *431*, 48–57.
[52] T. Smejkal, J. Mikyška, *Fluid Phase Equilib.* **2021**, *527*, 112832.
[53] M. Daudon, E. Letavernier, V. Frochot, J.-P. Haymann, D. Bazin, P. Jungers, *Comptes Rendus Chimi.* **2016**, *19*, 1504–1513.
[54] A. Belghait, C. Si-Moussa, M. Laidi, S. Hanini, *Comptes Rendus Chimi.* **2018**, *21*, 494–513.
[55] A. Tabernero, E. M. M. del Valle, M. Á. Galán, *J. Supercrit. Fluids.* **2010**, *52*, 161–174.
[56] M. A. Khansary, F. Amiri, A. Hosseini, A. H. Sani, H. Shahbeig, *Chem. Eng. Res. Des.* **2015**, *93*, 355–365.
[57] X.-Q. Bian, J. Li, J. Chen, M.-J. Li, Z.-M. Du, *Chem. Eng. Res. Des.* **2015**, *104*, 416–428.
[58] Z. J. Baum, X. Yu, P. Y. Ayala, Y. Zhao, S. P. Watkins, Q. Zhou, *J. Chem. Inf. Model.* **2021**, *61*, 3197–3212.
[59] A. KhazaiePoul, M. Soleimani, S. Salahi, *Chin. J. Chem. Eng.* **2016**, *24*, 491–498.

 © 2022 Wiley-VCH GmbH

[60] B. Mehdizadeh, K. Movagharnejad, *Fluid Phase Equilib.* **2011**, *303*, 40–44.
[61] M. Lashkarbolooki, B. Vaferi, M. Rahimpour, *Fluid Phase Equilib.* **2011**, *308*, 35–43.
[62] M. Laidi, C. Si-Moussa, S. Hanini, *Neural Comput. Appl.* **2017**, *28*, 87–99.
[63] B. Vaferi, M. Karimi, M. Azizi, H. Esmaeili, *J. Supercrit. Fluids.* **2013**, *77*, 44–51.
[64] A. Aminian, *J. Supercrit. Fluids.* **2017**, *125*, 79–87.
[65] F. Gharagheizi, A. Eslamimanesh, A. H. Mohammadi, D. Richon, *Ind. Eng. Chem. Res.* **2011**, *50*, 221–226.
[66] Y. Cao, A. Khan, S. Zabihi, A. B. Albadarin, *J. Mol. Liq.* **2021**, *333*, 115942.
[67] J. Leszczynski, *Handbook of computational chemistry*, Springer Science & Business Media.
[68] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, *J. Med. Chem.* **2014**, *57*, 4977–5010.
[69] C. W. Yap, *J. Comput. Chem.* **2011**, *32*, 1466–1474.
[70] M. Yasen, N. Al-Madi, N. Obeid, *in 2018 8th Int. Conf. Comput. Sci. Inf. Technol. CSIT 2018.* **2018**, pp. 71–76.
[71] H. Harandizadeh, D. Jahed Armaghani, M. Khari, *Eng. Comput.* **2021**, *37*, 685–700.
[72] M. Caudill, *Neural networks primer*, Miller Freeman Incorporated.
[73] W. McCelloch, W. Pitts, *Bull. Math. Biophys.* **1943**, *5*, 115–133.
[74] H.-L. Do, *Model. Simul.* **2012**, 2373.
[75] D. J. Armaghani, E. T. Mohamad, M. S. Narayanasamy, N. Narita, S. Yagiz, *Tunn. Undergr. Space Technol.* **2017**, *63*, 29–43.
[76] D. González-Arjona, G. López-Pérez, A. G. González, *Talanta* **2002**, *56*, 79–90.
[77] E. T. Mohamad, R. S. Faradonbeh, D. J. Armaghani, M. Monjezi, M. Z. A. Majid, *Neural Comput. Appl.* **2017**, *28*, 393–406.
[78] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C. P. Varghese, R. K. Agrawal, *Int. J. Drug Des. Discov* **2011**, *3*, 511–519.
[79] K. Roy, R. Das, P. Ambure, R. Aher, **2016**.
[80] K. Roy, I. Mitra, *Comb. Chem. High Throughput Screening* **2011**, *14*, 450–474.
[81] R. Soleimani, A. H. Saeedi Dehaghani, N. A. Shoushtari, P. Yaghoubi, A. Bahadori, *Korean J. Chem. Eng.* **2018**, *35*, 1556–1569.
[82] O. Falyouna, O. Eljamal, I. Maamoun, A. Tahara, Y. Sugihara, *J. Colloid Interface Sci.* **2020**, *571*, 66–79.
[83] F. Korner Nievergelt, T. Roth, S. von Felten, J. Guélat, B. Almasi, P. Korner Nievergelt, *Bayesian data analysis in ecology using linear models with R, BUGS„ STAN.* **2015**, 175–196.
[84] E. De Haas, T. Eikelboom, T. Bouwman, *SAR QSAR Environ. Res.* **2011**, *22*, 545–559.
[85] X. Zhao, Y. Pan, J. Jiang, S. Xu, J. Jiang, L. Ding, *Ind. Eng. Chem. Res.* **2017**, *56*, 4185–4195.
[86] M. Hamadache, S. Hanini, O. Benkortbi, A. Amrane, L. Khaouane, C. S. Moussa, *Chemom. Intell. Lab. Syst.* **2016**, *154*, 7–15.
[87] M. Mansourian, L. Saghaie, A. Fassihi, A. Madadkar-Sobhani, K. Mahnam, *Med. Chem. Res.* **2013**, *22*, 4549–4567.
[88] M. Gevrey, I. Dimopoulos, S. Lek, *Ecol. Modell.* **2003**, *160*, 249–264.
[89] A. Baghban, A. Jalali, M. Shafiee, M. H. Ahmadi, K.-w. Chau, *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 26–39.
[90] K. Roy, *Applications in Pharmaceutical, Chemical, Food, Agricultural, Environmental Sciences; Springer: Cham, Switzerland* **2017**, *555*, 39.
[91] Y.-M. Chen, Y.-P. Chen, *Fluid Phase Equilib.* **2009**, *282*, 82–87.
[92] J.-s. Jin, X. Fan, H. Zhang, Y.-w. Wang, Z.-t. Zhang, *J. Chem. Eng. Data.* **2014**, *59*, 2095–2100.
[93] D. Suleiman, L. A. Estévez, J. C. Pulido, J. E. García, C. Mojica, *J. Chem. Eng. Data.* **2005**, *50*, 1234–1241.
[94] N. Lamba, R. C. Narayan, J. Modak, G. Madras, *J. Supercrit. Fluids.* **2016**, *107*, 384–391.
[95] A. B. Rosa, S. Marceneiro, M. E. Braga, A. M. Dias, H. C. de Sousa, *J. Supercrit. Fluids.* **2015**, *98*, 70–78.
[96] G. Sodeifian, S. A. Sajadian, F. Razmimanesh, *Fluid Phase Equilib.* **2017**, *450*, 149–159.
[97] G. Sodeifian, C. Garlapati, F. Razmimanesh, F. Sodeifian, *J. Chem. Eng. Data.* **2021**, *66*, 1119–1131.
[98] M. Hojjati, A. Vatanara, Y. Yamini, M. Moradi, A. R. Najafabadi, *J. Supercrit. Fluids.* **2009**, *50*, 203–209.
[99] G. Sodeifian, S. A. Sajadian, N. S. Ardestani, *J. Supercrit. Fluids.* **2017**, *128*, 102–111.
[100] Z. Huang, W. D. Lu, S. Kawi, Y. C. Chiew, *J. Chem. Eng. Data.* **2004**, *49*, 1323–1327.
[101] M. Hojjati, Y. Yamini, M. Khajeh, A. Vatanara, *J. Supercrit. Fluids.* **2007**, *41*, 187–194.
[102] Y. Yamini, J. Hassan, S. Haghgo, *J. Chem. Eng. Data.* **2001**, *46*, 451–455.
[103] G. Sodeifian, F. Razmimanesh, N. S. Ardestani, S. A. Sajadian, *J. Mol. Liq.* **2020**, *299*, 112179.
[104] D. L. Sparks, R. Hernandez, L. A. Estévez, N. Meyer, T. French, *J. Chem. Eng. Data.* **2007**, *52*, 1246–1249.
[105] A. Garmroodi, J. Hassan, Y. Yamini, *J. Chem. Eng. Data.* **2004**, *49*, 709–712.
[106] M. Asghari-Khiavi, Y. Yamini, *J. Chem. Eng. Data.* **2003**, *48*, 61–65.
[107] A. Vatanara, A. R. Najafabadi, M. Khajeh, Y. Yamini, *J. Supercrit. Fluids.* **2005**, *33*, 21–25.
[108] M. Johannsen, G. Brunner, *Fluid Phase Equilib.* **1994**, *95*, 215–226.
[109] Y. Yamini, M. Hojjati, P. Kalantarian, M. Moradi, A. Esrafili, A. Vatanara, *Thermochim. Acta* **2012**, *549*, 95–101.
[110] M. Khamda, M. H. Hosseini, M. Rezaee, *J. Supercrit. Fluids.* **2013**, *73*, 130–137.
[111] K. Ongkasin, M. Sauceau, Y. Masmoudi, J. Fages, E. Badens, *J. Supercrit. Fluids.* **2019**, *152*, 104498.
[112] C.-C. Tsai, H.-m. Lin, M.-J. Lee, *J. Supercrit. Fluids.* **2014**, *95*, 17–23.
[113] A. Z. Hezave, A. Mowla, F. Esmaeilzadeh, *J. Supercrit. Fluids.* **2011**, *58*, 198–203.
[114] B. Li, W. Guo, E. D. Ramsey, *J. Chem. Eng. Data.* **2019**, *65*, 153–159.
[115] C.-A. Lee, M. Tang, S.-L. Ho, Y.-P. Chen, *J. Supercrit. Fluids.* **2014**, *85*, 11–16.
[116] M. Lashkarbolooki, A. Z. Hezave, Y. Rahnama, H. Rajaei, F. Esmaeilzadeh, *J. Supercrit. Fluids.* **2013**, *84*, 29–35.
[117] L. Manna, M. Banchero, *J. Chem. Eng. Data.* **2018**, *63*, 1745–1751.
[118] J. T. Paula, I. M. Sousa, M. A. Foglio, F. A. Cabral, *J. Supercrit. Fluids.* **2016**, *112*, 89–94.
[119] C. Zhang, D. Jia, Y. Sun, L. Zhang, *in IOP Conf. Ser. Earth Environ. Sci.* **2018**, pp. 012041.
[120] C.-A. Lee, M. Tang, Y.-P. Chen, *Fluid Phase Equilib.* **2014**, *367*, 182–187.
[121] K. Shi, L. Feng, L. He, H. Li, *J. Chem. Eng. Data.* **2017**, *62*, 4235–4243.
[122] G. Sodeifian, C. Garlapati, F. Razmimanesh, M. Ghanaat-Ghamsari, *Sci. Rep.* **2021**, *11*, 1–16.
[123] C. I. Park, M. S. Shin, H. Kim, *J. Chem. Thermodyn.* **2009**, *41*, 30–34.

 © 2022 Wiley-VCH GmbH

[124] H. Asiabi, Y. Yamini, M. Tayyebi, M. Moradi, A. Vatanara, K. Keshmiri, *J. Supercrit. Fluids.* **2013**, *78*, 28–33.
[125] H. Li, D. Jia, Q. Zhu, B. Shen, *Fluid Phase Equilib.* **2015**, *392*, 95–103.
[126] Y. Yamini, M. Moradi, *J. Chem. Thermodyn.* **2011**, *43*, 1091–1096.
[127] M. H. Hosseini, N. Alizadeh, A. R. Khanchi, *J. Supercrit. Fluids.* **2010**, *52*, 30–35.
[128] J. R. Dean, M. Kane, S. Khundker, C. Dowle, R. L. Tranter, P. Jones, *Analyst.* **1995**, *120*, 2153–2157.
[129] G. Sodeifian, N. S. Ardestani, S. A. Sajadian, H. S. Panah, *Fluid Phase Equilib.* **2019**, *483*, 122–143.
[130] M. Lashkarbolooki, A. Z. Hezave, Y. Rahnama, R. Ozlati, H. Rajaei, F. Esmaeilzadeh, *J. Supercrit. Fluids.* **2013**, *84*, 13–19.
[131] M. Asghari-Khiavi, Y. Yamini, M. A. Farajzadeh, *J. Supercrit. Fluids.* **2004**, *30*, 111–117.
[132] G. Sodeifian, R. S. Alwi, F. Razmimanesh, M. Abadian, *J. Mol. Liq.* **2022**, *346*, 117899.
[133] R. Chim, M. De Matos, M. Braga, A. Dias, H. De Sousa, *J. Chem. Eng. Data.* **2012**, *57*, 3756–3760.
[134] C. Tang, Y.-X. Guan, S.-J. Yao, Z.-Q. Zhu, *J. Chem. Eng. Data.* **2014**, *59*, 3359–3364.
[135] A. Zeinolabedini Hezave, F. Esmaeilzadeh, *J. Chem. Eng. Data.* **2012**, *57*, 1659–1664.
[136] P. Coimbra, D. Fernandes, M. H. Gil, H. C. d. Sousa, *J. Chem. Eng. Data.* **2008**, *53*, 1990–1995.
[137] Y. Yamini, P. Kalantarian, M. Hojjati, A. Esrafily, M. Moradi, A. Vatanara, I. Harrian, *J. Chem. Eng. Data.* **2010**, *55*, 1056–1059.
[138] G. Sodeifian, R. Detakhsheshpour, S. A. Sajadian, *J. Supercrit. Fluids.* **2019**, *154*, 104606.
[139] X. Zha, S. Han, W. Wang, Z. Jiao, *J. Chem. Thermodyn.* **2019**, *131*, 104–110.
[140] A. Z. Hezave, H. Rajaei, M. Lashkarbolooki, F. Esmaeilzadeh, *J. Supercrit. Fluids.* **2013**, *73*, 57–62.
[141] A. R. C. Duarte, P. Coimbra, H. C. de Sousa, C. M. Duarte, *J. Chem. Eng. Data.* **2004**, *49*, 449–452.
[142] S. A. Shojaee, A. Z. Hezave, S. Aftab, M. Lashkarbolooki, F. Esmaeilzadeh, *J. Supercrit. Fluids.* **2013**, *78*, 1–6.
[143] G. Sodeifian, S. M. Hazaveie, F. Sodeifian, *J. Mol. Liq.* **2021**, *330*, 115695.
[144] S.-T. Xiang, B.-Q. Chen, R. K. Kankala, S.-B. Wang, A.-Z. Chen, *J. Supercrit. Fluids.* **2019**, *150*, 147–155.
[145] Y.-M. Chen, P.-C. Lin, M. Tang, Y.-P. Chen, *J. Supercrit. Fluids.* **2010**, *52*, 175–182.
[146] G. Sodeifian, F. Razmimanesh, S. A. Sajadian, *J. Supercrit. Fluids.* **2019**, *146*, 89–99.
[147] E. Jara-Morante, D. Suleiman, L. A. Estévez, *Ind. Eng. Chem. Res.* **2003**, *42*, 1821–1823.
[148] B.-C. Wang, C.-S. Su, *J. CO2 Util.* **2020**, *37*, 285–294.
[149] R. Heryanto, E. C. Abdullah, M. Hasan, *J. Chem. Eng. Data.* **2010**, *55*, 2306–2309.
[150] G. Sodeifian, S. A. Sajadian, F. Razmimanesh, S. M. Hazaveie, **2021**.
[151] M. A. Sabegh, H. Rajaei, F. Esmaeilzadeh, M. Lashkarbolooki, *J. Supercrit. Fluids.* **2012**, *72*, 191–197.
[152] G. Sodeifian, N. S. Ardestani, S. A. Sajadian, H. S. Panah, *Fluid Phase Equilib.* **2018**, *458*, 102–114.
[153] Y. Wang, T. Meng, D. Jia, Y. Sun, N. Li, *J. Solution Chem.* **2017**, *46*, 70–88.
[154] G. Sodeifian, S. A. Sajadian, R. Derakhsheshpour, *Fluid Phase Equilib.* **2020**, *507*, 112422.
[155] Y. Yamini, M. Tayyebi, M. Moradi, A. Vatanara, *Thermochim. Acta* **2013**, *569*, 48–54.
[156] G. Sodeifian, F. Razmimanesh, S. A. Sajadian, H. S. Panah, *Fluid Phase Equilib.* **2018**, *472*, 147–159.
[157] G. Sodeifian, L. Nasri, F. Razmimanesh, M. Abadian, *J. Mol. Liq.* **2021**, *331*, 115745.
[158] K.-W. Cheng, M. Tang, Y.-P. Chen, *Fluid Phase Equilib.* **2002**, *201*, 79–96.
[159] A. Z. Hezave, M. H. Khademi, F. Esmaeilzadeh, *Fluid Phase Equilib.* **2012**, *313*, 140–147.
[160] H. C. de Sousa, M. S. Costa, P. Coimbra, A. A. Matias, C. M. Duarte, *J. Supercrit. Fluids.* **2012**, *63*, 40–45.
[161] Y. Yamini, J. Arab, M. Asghari-Khiavi, *J. Pharm. Biomed.* **2003**, *32*, 181–187.
[162] G. Sodeifian, N. S. Ardestani, F. Razmimanesh, S. A. Sajadian, *Fluid Phase Equilib.* **2020**, *522*, 112745.
[163] C.-S. Su, Y.-P. Chen, *J. Supercrit. Fluids.* **2008**, *43*, 438–446.
[164] P. Kotnik, M. Škerget, Z. e. Knez, *J. Chem. Eng. Data.* **2011**, *56*, 338–343.
[165] I. Medina, J. L. Bueno, *J. Chem. Eng. Data.* **2001**, *46*, 1211–1214.
[166] S. Zhan, L. Cui, Q. Zhao, L. Chen, J. Wang, S. Chen, S. Ding, *J. Solution Chem.* **2015**, *44*, 1–15.
[167] R. Chim, S. Marceneiro, M. E. Braga, A. M. Dias, H. C. de Sousa, *Fluid Phase Equilib.* **2012**, *331*, 6–11.
[168] Z. Tang, J.-s. Jin, Z.-t. Zhang, H.-t. Liu, *Ind. Eng. Chem. Res.* **2012**, *51*, 5515–5526.
[169] G. Sodeifian, S. M. Hazaveie, S. A. Sajadian, F. Razmimanesh, *Fluid Phase Equilib.* **2019**, *493*, 160–173.
[170] Z. Jiao, J. Cheng, M. Guo, S. Han, *J. Chem. Eng. Data.* **2019**, *64*, 4424–4429.
[171] M. Gordillo, M. Blanco, A. Molero, E. M. De La Ossa, *J. Supercrit. Fluids.* **1999**, *15*, 183–190.
[172] M. Ko, V. Shah, P. R. Bienkowski, H. D. Cochran, *J. Supercrit. Fluids.* **1991**, *4*, 32–39.
[173] H. Rajaei, A. Z. Hezave, M. Lashkarbolooki, F. Esmaeilzadeh, *J. Supercrit. Fluids.* **2013**, *75*, 181–186.
[174] G. Sodeifian, R. S. Alwi, F. Razmimanesh, *Fluid Phase Equilib.* **2022**, 113396.
[175] R. D. Weinstein, J. J. Gribbin, K. R. Muske, *J. Chem. Eng. Data.* **2005**, *50*, 226–229.
[176] S. A. Shojaee, H. Rajaei, A. Z. Hezave, M. Lashkarbolooki, F. Esmaeilzadeh, *J. Supercrit. Fluids.* **2013**, *80*, 38–43.
[177] C.-T. Chen, C.-A. Lee, M. Tang, Y.-P. Chen, *J. CO2 Util.* **2017**, *18*, 173–180.
[178] J. García-González, M. J. Molina, F. Rodríguez, F. Mirada, *J. Chem. Eng. Data.* **2001**, *46*, 918–921.
[179] G. Sodeifian, R. S. Alwi, F. Razmimanesh, K. Tamura, *Fluid Phase Equilib.* **2021**, *537*, 113003.
[180] B. Li, W. Guo, E. D. Ramsey, *J. Chem. Eng. Data.* **2017**, *62*, 1530–1537.
[181] G. Sodeifian, S. M. Hazaveie, S. A. Sajadian, N. Saadati Ardestani, *J. Chem. Eng. Data.* **2019**, *64*, 5338–5348.
[182] E. Reverchon, G. Donsi, D. Gorgoglione, *J. Supercrit. Fluids.* **1993**, *6*, 241–248.
[183] G. Sodeifian, S. A. Sajadian, *J. Supercrit. Fluids.* **2019**, *149*, 79–87.
[184] G. Sodeifian, N. Saadati Ardestani, S. A. Sajadian, M. R. Golmohammadi, A. Fazlali, *J. Chem. Eng. Data.* **2020**, *65*, 1747–1760.
[185] G. Sodeifian, F. Razmimanesh, S. A. Sajadian, S. M. Hazaveie, *J. Chem. Thermodyn.* **2020**, *142*, 105998.
[186] A. Z. Hezave, S. Shahnazar, H. Rajaei, M. Lashkarbolooki, F. Esmaeilzadeh, *Fluid Phase Equilib.* **2013**, *355*, 130–134.
[187] G. Sodeifian, C. Garlapati, F. Razmimanesh, F. Sodeifian, *J. Mol. Liq.* **2021**, *335*, 116446.

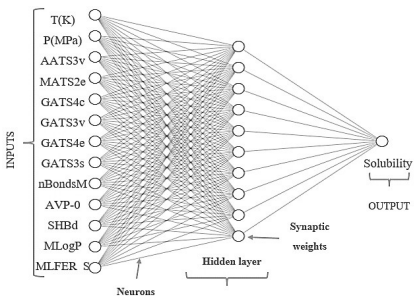
 © 2022 Wiley-VCH GmbH

[188] A. Z. Hezave, S. Aftab, F. Esmaeilzadeh, *J. Supercrit. Fluids.* **2012**, *68*, 39–44.
[189] G. Sodeifian, F. Razmimanesh, S. A. Sajadian, *J. Mol. Liq.* **2020**, *297*, 111740.
[190] S. M. Hazaveie, G. Sodeifian, S. A. Sajadian, *J. Supercrit. Fluids.* **2020**, *163*, 104875.
[191] G. Sodeifian, L. Nasri, F. Razmimanesh, M. Abadian, *J. CO2 Util.* **2022**, *58*, 101931.
[192] Y. Sun, S. Li, C. Quan, *J. Chem. Eng. Data.* **2005**, *50*, 1125–1128.
[193] G. Sodeifian, C. Garlapati, S. M. Hazaveie, F. Sodeifian, *J. Chem. Eng. Data.* **2020**, *65*, 4406–4416.
[194] S. Han, W. Wang, Z. Jiao, X. Wei, *J. Chem. Eng. Data.* **2017**, *62*, 3854–3860.
[195] J.-M. Ciou, B.-C. Wang, C.-S. Su, J.-J. Liu, M.-T. Sheu, *Adv. Powder Technol.* **2018**, *29*, 479–487.

Received: February 2, 2022
Accepted: April 3, 2022
Published online on ■■, ■■■

 © 2022 Wiley-VCH GmbH

RESEARCH ARTICLE



I. Euldji, C. Si-Moussa, M. Hamadache, O. Benkortbi*

1 – 17

QSPR Modelling of the Solubility of Drug and Drug-like Compounds in Supercritical Carbon Dioxide