QSRR modeling of the chromatographic retention behavior for some quinolone and sulfonamide antibacterial agents using firefly algorithm coupled to support vector machine

Marwa A. Fouad (marwa.fouad@pharma.cu.edu.eg)
Cairo University
Enas H. Tolba
National Organization for Drug Control and Research (NODCAR)
Manal A. El-Shal
National Organization for Drug Control and Research (NODCAR)
Ahmed Serag
Al-Azhar University
Ahmed M. El Kerdawy
Cairo University

Research Article

Keywords: Quinolones, Sulfonamides, Quantitative structure-retention relationship, Firefly algorithm, Support vector machine, y-randomization

Posted Date: May 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1619477/v1

Abstract

Quinolone and sulfonamide are two classes of antibacterial agents that have an opulent history of medicinal chemistry features responsible for their improving bacterial spectrum, efficacy, pharmacokinetics, and adverse side effect profiles. The urgent need of their use and escalating rate of their resistance provokes the necessity for developing suitable analytical methods that speed up and facilitate their analysis. In this study, advanced firefly algorithm (FFA) coupled with support vector machine (SVM) were used to select the most significant descriptors and to construct two separate quantitative structure–retention relationship (QSRR) models using a series of 11 selected quinolones and 13 sulfonamide drugs, separately, in order to predict their retention factors in HPLC. Precisely, the effect of different pH range values and acetonitrile composition in the mobile phase on the retention behavior of quinolones and sulfonamides were studied, respectively. The obtained QSRR models showed high performance in both internal and external validation indicating their robustness and predictive ability. Y-randomization validation displayed that the obtained models are not obtained by chance. Besides, the obtained results shed the light on the molecular features that influence the retention behavior of these two classes under the current chromatographic conditions.

1. Introduction

Resistance to antibacterial agents is a major public health threat affecting humans worldwide mainly due to the uncontrolled use of such bioactive compounds, particularly in countries without standard treatment guidelines. Among those antibacterial agents, fluoroquinolones, a fluoro substituent series derived from nalidixic acid, showed escalating rate of resistance after domination over the therapeutic practice for a time particularly against gram-negative pathogens [1–3]. Such class of active compounds needs to be monitored carefully regarding their use and their abundance in the environment. Hence, from the analytical view, the urgent detection and analysis of these drugs become essential considering finding fast, simple, economical and accurate methods for their analysis.

The literature survey revealed that quinolones could be determined thoroughly via high performance liquid chromatography in different matrices viz., biological fluids and tissues [4–10], milk and food of animal origin [11–16], marine products [17], honey [18], waste water [19–21] and in many pharmaceutical formulations [22–27]. Moreover, the relationship between the retention factors and lipophilicity of quinolones using RP-TLC has been assessed [28]. In addition, Wu et al [29] investigated the retention factors-activity relationship of some quinolones using micellar chromatography.

On the other hand, sulfonamides are other synthetic antimicrobial agents, unfortunately with widespread resistance which made them infrequently utilized for medical interventions. However, the application of sulfonamides has been extended from their old capabilities as antimicrobial agents to another medical roles viz., anticancer, antiglaucoma, cyclooxygenase-2 and lipoxygenase inhibitors, anticonvulsant and hypoglycemic activities [30]. Regarding the analytical tools used in their detection, literature survey revealed that the determination of this class was also dominated by reversed phase liquid chromatography [31–34]. In context of their retention mechanisms, Cazenave-Gassiot et al discussed the correlation between sulfonamides retention factors and the proportion of modifier in the mobile phase using supercritical fluid

chromatography [35]. However, like quinolones, the separation behavior of this class on reversed phase liquid chromatography needs to be scrutinized.

Among different models and theories applied to draw an image about the retention manners in reversed phase liquid chromatography, quantitative structure–retention relationship (QSRR) offers some useful insights not only to elucidate how different chemical drugs perform their retention upon analysis, but also to expect their retention chromatographic systems relatively well [36]. Such relationship provides a powerful alternative to the conventional trial-and-error approach with marked improvement in time and cost of experiments.

In these mathematical models, a link between compounds' chemical structures represented by their descriptors and the retention data in different chromatographic systems is built. The number of molecular descriptors that could be obtained for one analyte is enormous where some software could calculate up to 5000 descriptors per analyte [37]. Such massive increase in the dimensionality of the descriptors along with the possible incorporation of some nonempirical features could affect the performance of various QSRR models. Therefore, methods for feature selection are necessary to untangle this problem and decide which descriptors are important regarding the retention of compounds of interest. These methods ranged from classical type as forward and backward elimination to advanced nature inspired ones for example particle swarm optimization (PSO), genetic algorithm (GA) and its descendants (firefly, flower pollination and ant colony algorithms) [38–43].

Furthermore, different chemometric and artificial intelligence methods viz., partial least square (PLS), multiple linear regression (MLR), artificial neural networks (ANN) and support vector regression (SVR) proved to be effective in building reliable QSRR models owing to their ability in extracting maximal chemical information in addition to enhancing the speed and quality of analysis[44]. The application of QSRR models have been reported to different chemical families on reversed-phase liquid chromatography such as non-steroidal anti-inflammatory drugs [45], azole antifungal agents [46] and some pain killers drugs [47]. Support vector machine (SVM), a machine learning algorithm, was firstly published by Vapnik, Chervonenkis and co-workers [48]. The algorithm is based on finding a linear function that explains most of the variation of the response and at the same time links the nonlinear relationship between input and the target data [49]. Compared to conventional regression and neural network methods, SVM displays some advantages, including good generalization ability, global optimization and dimensional independence [50]. Thanks to its capability to model possible nonlinear relations between molecular descriptors and retention time, it has been incorporated in building powerful QSRR models [51, 52].

Previously our group developed two QSRR models aimed to provide some essence of the retention behavior for some β-lactams using multiple linear regression models combined with forward or firefly variable selection algorithms [44]. Our scope in this report is to continue our work regarding QSRR modeling of other antibacterial agents (quinolones and sulfonamides), hopefully to highlight their reversed phase chromatographic retention mechanisms with respect to different ionization states and various percentage of organic modifiers for quinolones and sulfonamides, respectively. Owing to the complexity of the generated data, the use of advanced variable selection technique coupled to a machine learning approach seems imperative. Hence, firefly algorithm coupled to SVR has been employed to develop the QSRR models.

Moreover, the obtained models have been assessed regarding their predictive ability with strict validation criteria, thus could be further employed to predict retention of potential degradation products and even metabolites of these compounds.

2. Experimental

2.1. Solvents, chemicals, sample preparation and instrumentation.

The quinolones (**Fig. S1**) and sulfonamides (**Fig. S2**) under investigation were supplied by different pharmaceutical companies. Pure HPLC-grade acetonitrile, methanol and dimethylsulfoxide were supplied by Scarlau (Barcelona, Spain). Other chemicals used in this work including ortho-phosphoric acid, trifluoroacetic acid, sodium dihydrogen orthophosphate and sodium hydroxide were supplied by Honeywell Riedel-de Haën (Seelze, Germany).

The instruments used in this study include Jenway 3510, Essex-UK, England pH meter equipped with a glass electrode and Agilent 1260 HPLC-UV series.

Stock solutions (2 mg mL^{-1}) of each drug were prepared using a suitable solvent either (methanol, dimethylsulfoxide, water, or acetonitrile). These solutions were stored at 4°C then diluted with the mobile phase to achieve sample concentrations ranging from (0.05–1 mg mL^{-1}) before analysis.

2.2. Chromatographic conditions

The chromatographic elution of the quinolones was achieved using Inertsil® C18 column (250 mm x 4.6 mm, 5 µm) and detection was carried out at 275 nm. In a gradient mode, 5 mobile phases were prepared according to the plan of experiment and chromatographic system applied as programmed in **Table 1**, using acetonitrile and 28 mM sodium dihydrogen orthophosphate buffer prepared at different pHs 2.2, 3.5, 5.2, 6.5 and 8.2 using ortho-phosphoric acid or sodium hydroxide. However, the pH was measured again after mixing the buffer with acetonitrile and was found to be 3.2, 4.4, 5.9, 7.32 and 8.9, respectively. The system flow rate was adjusted at 1 ml min^{-1}. After each injection, the system was reconditioned by going back to the initial ratio and kept constant for 3 min. Data acquisition was performed on Agilent LC Chemstation software. The retention factors of eluted quinolones are listed in **Table 2**.

Table (1) Gradient elution system used in quinolones' separation

Time (min)	Acetonitrile %	Buffer %
0	20	80
3	20	80
5	60	40

Table (2) List of quinolones chromatographic retention factors (k)

Compound name	pH 2.2	pH 3.5	pH 5.2	pH 6.5	pH 8.2
Gatifloxacin	1.580	1.603	1.576	1.566	1.635
Lomefloxacin	1.405	1.560	1.501	0.920	1.035
Moxifloxacin	1.558	1.606	1.592	1.749	1.840
Nadifloxacin	2.121	2.148	2.114	2.036	1.685
Norfloxacin	1.191	1.162	1.192	1.158	0.384
Ofloxacin	1.032	1.176	1.559	1.836	1.885
Ciprofloxacin	1.142	1.363	1.294	1.153	0.646
Gemifloxacin	1.557	1.576	1.578	1.584	1.633
Enrofloxacin	1.622	1.567	1.591	2.468	1.973
Danofloxacin	1.370	1.551	1.572	1.822	1.668
Sparfloxacin	1.560	1.576	1.567	2.486	1.987

Chromatographic separation of sulfonamides was achieved on a hypersil C18 column (150 mm x 4.6 mm, 5 µm) by applying isocratic elution based on a mobile phase consisting of acetonitrile and water acidified with trifluoroacetic acid (1 mL. L^{-1}) in different ratios of 50:50, 45:55 or 30:70, v/v and at a flow rate of 0.8 ml min^{-1}. Ratio of 15:85, v/v was initially included but not considered for further assessment as many compounds were strongly retained onto the column. Analyses were operated at ambient temperature and detection was carried out at 270 nm. Data acquisition was performed on Agilent LC Chemstation software. A list of retention factors of eluted sulfonamides is shown in **Table 3**.

Table (3) List of sulfonamides chromatographic retention factors (k)

Compound name	Acetonitrile%		
	50	45	30
Sulfacetamide Na	0.154	0.203	0.393
Sulfaguanidine	0.170	0.188	0.256
Sulfadiazine	0.174	0.228	0.443
Sulfaclozine	0.549	0.752	2.196
Sulfadimethoxine	0.419	0.567	1.433
Sulfadimidine	0.311	0.389	0.730
Sulfadoxine	0.395	0.524	1.276
Sulfathiazole	0.166	0.221	0.426
Sulfachloropyrazine Na	0.546	0.754	2.177
Sulfanilamide	0.154	0.194	0.295
Sulfamethoxazole	0.421	0.568	1.548
Sulfapyridine	0.306	0.359	0.597
Sulfaquinoxaline	0.519	0.716	2.221

2.3. QSRR modeling

2.3.1 Drawing Structures and molecular descriptors calculation

The major microspecies of the studied quinolones at the pH of interest were estimated using MarvinSketch (6.0.3) [53] which resulted in 21 ions. The canonical smiles of these ions were imported into the Molecular Operating Environment (MOE, 2020.0901) software where they were converted into 3D structures and energy minimized using RMSD gradient of 0.05 kcal.mol^{-1}Å^{-1} with MMFF94x forcefield. The partial charges were automatically calculated. Finally, MOE molecular mechanical descriptors were computed for all the compounds resulting in a descriptor fund of 313 descriptors. The initial descriptor fund was reduced by removing zero values and constant descriptors. This ended to a descriptor fund of 293 descriptors.

As for sulfonamides, PubChem database [54, 55] was used to introduce sulfonamides canonical SMILES into the MOE where they were converted to 3D structures and energy minimized using the same parameters previously mentioned for quinolones. Afterwards, MOE molecular mechanical descriptors were computed for all the compounds and a descriptor fund of 313 descriptors was obtained. The initial descriptor fund was reduced by removing zero values and constant descriptors, resulting in a fund of 112 descriptors in addition to acetonitrile % incorporated as a descriptor.

2.3.2 Training set and test set generation

The 21 quinolones major microspecies were divided into a calibration (training) set of 16 molecules and a test set of 5 molecules. Regarding sulfonamides, a total of 39 experimental retention factors resulted from three different ratios of mobile phase for the 13 compounds were used in building the QSRR model. The total experiments were split into a training set of 30 observations and an external validation test set of 9 observations. The selection of the calibration and the validation samples of quinolones and sulfonamides was based on keeping the distribution value for the retention factor maintained in both sets.

2.3.3 Descriptors selection and modeling

Based on Durbin-Watson (DW) test, linearity of the datasets was tested via augmented partial residual plots (APARP) [56–58]. The test was performed using a custom script written in MATLAB (R2016 a)[59, 60]. The descriptors that continue to exist after the initial filtration were then used to build QSRR models. Firefly algorithm was implemented in MATLAB and applied for descriptor selection as an advanced nature-stimulated algorithm where the $RMSE_{CV}$ of the SVR model was considered as a fitness function inside the algorithm for both datasets. The selected descriptors were finally incorporated in SVR final model building. Parameters of the algorithm was combinatorically optimized. These parameters were varied in intervals of specific increments, keeping in mind that in all optimization iterations, one parameter was always varied, while the others were saved constant.

2.3.4 Model validation

It was requisite to evaluate the applicability of the generated models based on model validation. In the present study, models were validated both internally as well as externally and any chance correlation was tested by the use of a y-scrambling technique: a method frequently used for this purpose.

Internal validation was performed by using leave-one-out cross-validation (CV_{LOO}) in quinolones model while by using leave-10%-out ($CV_{L10\%O}$) in sulfonamides.

External validation was conducted by applying the model on external validation set of 5 microspecies of quinolones and 9 molecules of sulfonamides. The statistical quality of the models was judged by considering the root mean square errors (RMSE) of prediction and the validation correlation coefficients.

For Y-randomization, the output retention factors of the compounds were shuffled randomly, and the resulting dataset was examined by the FFA-SVM model using real (unscrambled) input descriptors to determine the correlation and predictive ability of the resulting models. The whole procedure was repeated 100 times for both datasets.

Hotelling's T2 and William’s plot methods were used to determine the developed models applicability domains (AD) as previously described in our previous work [44].

3. Results And Discussion

3.1 Optimization of the FFA and SVR parameters for the developed QSRR model of quinolones

Firefly algorithm was used as a feature selection method to find the relevant descriptors in order to build reliable QSRR models. However, some parameters were needed to be optimized for proper descriptors' selection. Initially, the selection of the fitness function used to evaluate the performance of the models was a critical step, thus, based on a previous study [44], the $RMSE_{CV}$ was utilized as a fitness function computed by SVR model. Another critical parameter in the FFA is the absorption coefficient parameter "γ" which regulates the light intensity and thus controls the fireflies' attractiveness, thus, this parameter has a powerful impact on the speed of the convergence and the behavior of the whole algorithm. Another valuable parameter is the "α" parameter that prevents sticking to the local optima through providing some sorts of random movements. Finally, the exploration phase of the FFA was controlled via the number of the fireflies used while the exploitation phase was controlled by the number of generations. The adjusted FFA parameters through combinatorial optimization are shown in **Table 4**.

Table (4) Parameters of the firefly algorithm used for variable selection in QSRR modeling

Parameter	Quinolones	Sulfonamides
Number of fireflies	10	20
Generations	100	100
α	0.1	0.15
β_o	1	1
γ	0.01	0.01

Concerning SVR, two parameters determine the quality of the model: Penalty error (C), a parameter that controls the trade-off between complexity of decision rule and frequency of error, and insensitive loss function (□) which is a precision factor expressing the radius of the tube placed around the regression function f(x). Moreover, different types of kernels as basis function expansions were also assessed viz., polynomial, radial basis function (RBF) and sigmoid. Initially, the kernel function was examined through evaluating the performance of developed FFA-SVM models, the RBF was selected as best kernel function to model the nonlinearity of the generated data. The RBF kernel parameter regulates the amplitude of the Gaussian function and influence the generalization ability of SVM. Then, the parameters of C & □ were optimized. To optimize these parameters, their values was systematically varied in the training step through (CV_{LOO}) and ($CV_{L10\%O}$) for quinolones and sulfonamides, respectively, with monitoring the $RMSE_{cv}$ of models. To obtain the optimal □, the SVM with different □ values were trained, during this, we kept value of C as 1 initially, but after finding the optimal value of □, the C value is further optimized. It was found that the best models were obtained using kernel type of (RBF), C = 1 and □ =0.01 for both datasets. The final developed FFA-SVM models were used to predict the retention factors of molecules in test set for quinolones and sulfonamides, respectively.

3.2 QSRR modelling of quinolones in different ionization states

For elucidating the chromatographic behavior of the studied quinolones, it is important to understand the relationship between the pH of the mobile phase and the ionization states of each compound, (**Fig. S3**). Some compounds behave ideally with respect to their ionization state i.e. moxifloxacin exists as a cation (polar) at acidic pHs (2.2 and 3.5) while exists as a neutral compound (hydrophobic) at pH (6.5 and 8.2), so this can rationalize the longer the retention factor of this compound in basic pH rather than the acidic one. Ciprofloxacin, lomefloxacin and norfloxacin exist in different dissociation forms in pHs (5.2 and 8.2) and this describes the fluctuation in their retention factors over these pHs. Nadifloxacin exists as a neutral compound at acidic pHs (2.2, 3.5 and 5.2) what describes the longer the retention factor at these lower pH values while its rapid elution and lower retention factor at basic pH 8.2 as it exists in anionic form. In contrast, ofloxacin and danofloxacin show different behavior, the cationic form of these analytes which appeared at acidic pHs (2.2 and 3.5) show lower retention factors while its anionic forms which are present at basic pHs (6.5 and 8.2) show higher retention factors and so the longer the time it is retained on the column. Additionally, gatifloxacin and gemifloxacin show stability in their retention factors although they can exist in different ionization state along the pH range (2.2–8.2).

From these previous observations, the behavior of quinolone compounds cannot be predicted solely based on their ionization state and a more in-depth analysis is required that can predict their behavior successfully. It is noteworthy to mention that the microspecies of each compound could also be present in various ionic forms and in diverse percentages, thus it will be tricky to predict the retention behavior based on a single microspecies. To tackle this problem, we tried to choose the major microspecies as a representative for each molecule in the given pH taking into consideration not to choose the same microspecies at different pHs or different retention factors for the same ionization state. Considering this approach, we were able to derive a simple, interpretable QSRR model that can predict the retention factors of quinolones in their different ionic forms.

The first step in QSRR model implemented for quinolones is to check the linearity of the data, augmented partial residual plots (APARP) along with Durbin-Watson (DW) test were used to check the residuals correlation [56–58]. The associated probability was found to be 0.045 (> 0.05) revealing the significance of the test and the nonlinearity of the data, therefore, nonlinear models as artificial neural network (ANN) and support vector regression (SVR) were tried for data modeling and the best results were obtained during using SVR.

Five descriptors were chosen by the FFA and combined in building the SVM model. SMR is a 2D descriptor linked to molecular refractivity including implicit hydrogens [61]. This property is an atomic contribution model that assumes the correct protonation state. GCUT_SLOGP_1 is a 2D descriptor that uses atomic contribution to logP in place of partial charge. VSA is a 3D descriptor that is related to surface area, volume and shape of molecules, it represents van der Waals surface area [62]. Vsurf_EWmin 2 is a 3D descriptor and represents the 2nd lowest hydrophilic energy. Vsurf_IW6 is a 3D descriptor that represents the hydrophilic integy moment at (-4.0). Considering the selected descriptors, the model displays that quinolones retention is

based on their size and hydrophobic/hydrophilic nature which are the main elements that influence the retention in reversed phase liquid chromatography.

Regarding the performance of the developed QSRR model, the agreement of the experimental and predicted retention factors shows the good predictive capability of the model as shown in **Table 5**. The nearness between the training set prediction and the cross-validation results point to the robustness of the resulted model and lack of any overfitting. As shown in **Table 6**, the results display the good prediction capability of the obtained model. The correlation between the experimental and predicted retention times for the training set, test set and CV_{LOO} results are presented in (Fig. S4&S5). Spearman ranking correlation coefficient (ρ) was also calculated and found to be 0.976, 0.982, and 0.900 for the training set prediction (ρ_{cal}), CV_{LOO} (ρ_{LOO}) and the external test set (ρ_{pred}), respectively, Table 6. The closeness of ρ to "1" indicates a reasonable accuracy and excellent capability of the generated model to reproduce the experimental retention factor ranking (Fig. 1).

Table (5): Experimental and predicted retention factors (k) of quinolones compounds in training set, cross-validation and test set

Compound name	Buffer pH	Experimental k	Training set prediction	Residuals	Cross-Validation CV_{LOO}	Residuals
Lomefloxacin	6.5	0.920	1.175	0.255	1.247	0.327
Ciprofloxacin	6.5	1.153	1.164	0.011	1.262	0.109
Norfloxacin	3.5	1.162	1.255	0.093	1.342	0.18
Ofloxacin	3.5	1.176	1.261	0.085	1.385	0.209
Ciprofloxacin	3.5	1.363	1.330	-0.033	1.293	-0.07
Lomefloxacin	3.5	1.560	1.550	-0.01	1.401	-0.159
Gatifloxacin	6.5	1.566	1.555	-0.011	1.493	-0.073
Gemifloxacin	3.5	1.576	1.586	0.01	1.593	0.017
Gemifloxacin	6.5	1.584	1.575	-0.009	1.558	-0.026
Gatifloxacin	3.5	1.603	1.613	0.01	1.621	0.018
Moxifloxacin	3.5	1.606	1.616	0.01	1.638	0.032
Danofloxacin	8.2	1.668	1.659	-0.009	1.662	-0.006
Nadifloxacin	8.2	1.685	1.694	0.009	1.763	0.078
Moxifloxacin	6.5	1.749	1.739	-0.01	1.700	-0.049
Enrofloxacin	8.2	1.973	1.819	-0.154	1.773	-0.2
Nadifloxacin	3.5	2.148	1.828	-0.32	1.769	-0.379
Norfloxacin*	6.5	1.158	1.149	-0.009		
Danofloxacin*	3.5	1.551	1.433	-0.118		
Sparfloxacin*	3.5	1.576	1.563	-0.013		
Enrofloxacin*	2.2	1.622	1.493	-0.129		
Ofloxacin*	8.2	1.885	1.603	-0.282		

(*) Test set compound.

Table (6): Quinolones and sulfonamides FFA-SVM model performance evaluation parameters

Parameter	Quinolones FFA-SVM	Sulfonamides FFA-SVM
R^2_{cal}	0.931	0.900
$R^2_{cal\text{-}adj}$	0.926	0.896
q^2_{LOO} / $q^2_{L10\%O}$	0.808	0.812
R^2_{pred}	0.879	0.820
$RMSE_{cal}$	0.114	0.240
$RMSE_{CVLOO}$	0.163	0.328
$RMSE_{pred}$	0.148	0.450
ρ_{cal}	0.976	0.988
ρ_{Loo} / $\rho_{L10\%O}$	0.982	0.941
ρ_{pred}	0.900	0.883

3.3 QSRR modelling of sulfonamides using different organic modifiers

QSRR modeling of sulfonamides was implemented with the aim of studying the associations between the retention factors of the examined compounds eluted upon using different percentages of acetonitrile in the mobile phase composition (**Fig. S6**): 50%, 45% and 30%, and their calculated constitutional, geometrical, physicochemical and electronical descriptors (independent variables),. Linearity of the data were also considered with the same procedures reported in quinolones with associated probability of 3.2^{-17} (> 0.05) indicating the nonlinearity of the generated data. The FFA-SVR model was also utilized in this case resulting in two descriptors plus acetonitrile % in building the QSRR model. The selected features, Vsurf-D2 and vsurf-w2, are 3D descriptors that are related to the molecular hydrophobic and hydrophilic volume, respectively. Moreover, the influence of the third descriptor (acetonitrile % in the mobile phase) indicates that, in the resulted model, the analyte retention is based on its hydrophobic/hydrophilic nature which is a usual element that plays an important role in differential elution of analytes in reversed phase liquid chromatography.

The results also display the good prediction capability of the obtained model as shown in **Tables 6&7**. The model training set, and test set correlation of the experimental and predicted retentions are presented in (**Fig. S7**), whereas the compounds' experimental and predicted retentions in the CV_{LOO} is presented in (**Fig. S8**) indicating the good correlation and the generalized ability of the developed QSRR sulfonamides model.

Spearman ranking correlation coefficient (ρ) was also calculated for the training set prediction (ρ_{cal}), CV $_{L10\%O}$ ($\rho_{L10\%O}$) and the external test set (ρ_{pred}) and was found to be 0.988, 0.941 and 0.883, respectively, (Fig. 2). The closeness of ρ to "1" indicates the capability of the generated model to reproduce the experimental retention factor ranking of the compounds under investigation in a reasonable accuracy.

Table (7): Experimental and predicted retention factors (k) of sulfonamides compounds in training set, cross-validation and test set

Compound name	Acetonitrile %	Experimental k	Training set prediction	Residuals of Training set	Cross-Validation $CV_{L10\%O}$	Residuals of Cross-Validation
Sulfacetamide Na	50%	0.154	0.153	-0.001	0.151	-0.002
Sulfacetamide Na	45%	0.203	0.194	-0.009	0.171	-0.032
Sulfacetamide Na	30%	0.393	0.385	-0.008	0.336	-0.057
Sulfaguanidine	50%	0.170	0.178	0.009	0.212	0.042
Sulfaguanidine	45%	0.188	0.181	-0.006	0.182	-0.006
Sulfaguanidine	30%	0.256	0.267	0.011	0.458	0.202
Sulfadiazine	50%	0.174	0.185	0.011	0.193	0.019
Sulfadiazine	45%	0.228	0.237	0.010	0.236	0.008
Sulfadiazine	30%	0.443	0.454	0.011	0.520	0.078
Sulfaclozine	50%	0.549	0.528	-0.021	0.460	-0.088
Sulfaclozine	45%	0.752	0.701	-0.051	0.657	-0.095
Sulfaclozine	30%	2.196	1.310	-0.886	1.082	-1.114
Sulfadimethoxine	50%	0.419	0.429	0.010	0.427	0.008
Sulfadimethoxine	45%	0.567	0.576	0.009	0.602	0.035
Sulfadimethoxine	30%	1.433	1.278	-0.155	1.006	-0.427
Sulfadimidine	50%	0.311	0.299	-0.012	0.289	-0.022
Sulfadimidine	45%	0.389	0.378	-0.010	0.352	-0.037
Compound name	Acetonitrile %	Experimental k	Training set prediction	Residuals of Training set	Cross-Validation $CV_{L10\%O}$	Residuals of Cross-Validation
Sulfadimidine	30%	0.730	0.643	-0.087	0.555	-0.175
Sulfadoxine	50%	0.395	0.405	0.010	0.419	0.024
Sulfadoxine	45%	0.524	0.540	0.017	0.569	0.046
Sulfadoxine	30%	1.276	1.211	-0.065	0.962	-0.314
Sulfathiazole	50%	0.166	0.224	0.058	0.242	0.076
Sulfathiazole	45%	0.221	0.285	0.064	0.289	0.068
Sulfathiazole	30%	0.426	0.521	0.095	0.574	0.148

Sulfachloropyrazine Na	50%	0.546	0.568	0.022	0.623	0.078
Sulfachloropyrazine Na	45%	0.754	0.742	-0.012	0.698	-0.056
Sulfachloropyrazine Na	30%	2.177	1.233	-0.944	0.937	-1.240
Sulfanilamide	50%	0.154	0.165	0.011	0.231	0.078
Sulfanilamide	45%	0.194	0.183	-0.010	0.167	-0.026
Sulfanilamide	30%	0.295	0.307	0.012	0.409	0.114
Sulfacetamide Na	50%	0.154	0.153	-0.001	0.151	-0.002
Sulfacetamide Na	45%	0.203	0.194	-0.009	0.171	-0.032
Sulfacetamide Na	30%	0.393	0.385	-0.008	0.336	-0.057
Sulfaguanidine	50%	0.170	0.178	0.009	0.212	0.042
Sulfaguanidine	45%	0.188	0.181	-0.006	0.182	-0.006
Sulfaguanidine	30%	0.256	0.267	0.011	0.458	0.202
Compound name	**Acetonitrile %**	**Experimental k**	**Training set prediction**	**Residuals of Training set**	**Cross-Validation $CV_{L10\%O}$**	**Residuals of Cross-Validation**
Sulfadiazine	50%	0.174	0.185	0.011	0.193	0.019
Sulfadiazine	45%	0.228	0.237	0.010	0.236	0.008
Sulfadiazine	30%	0.443	0.454	0.011	0.520	0.078
Sulfaclozine	50%	0.549	0.528	-0.021	0.460	-0.088
Sulfaclozine	45%	0.752	0.701	-0.051	0.657	-0.095
Sulfaclozine	30%	2.196	1.310	-0.886	1.082	-1.114
Sulfadimethoxine	50%	0.419	0.429	0.010	0.427	0.008
Sulfadimethoxine	45%	0.567	0.576	0.009	0.602	0.035
Sulfadimethoxine	30%	1.433	1.278	-0.155	1.006	-0.427
Sulfamethoxazole*	50%	0.421	0.248			
Sulfamethoxazole*	45%	0.568	0.322			
Sulfamethoxazole*	30%	1.548	0.644			
Sulfapyridine*	50%	0.306	0.283			
Sulfapyridine*	45%	0.359	0.306			

Sulfapyridine*	30%	0.597	0.443
Sulfaquinoxaline*	50%	0.519	0.530
Sulfaquinoxaline*	45%	0.716	0.719
Sulfaquinoxaline*	30%	2.221	1.279

-Test set compound (*)

3.4 Y scrambling and applicability domain of both models

Y-randomization or permutation test is another criterion used to validate our findings in this study especially with this low number of observations. In this test, it is suspected that the original QSRR model is significant if there is a solid link between the selected descriptors and the original response variables. Y- randomization was repeated for 100 times. If the statistical attributes of these randomized models are much lower than the original one, it can be decided that the model is sensible and had not been obtained by luck. The below equation was used to evaluate the quality of the obtained models from the 100 randomized matrices and to weigh against the original model quality. $^{c}R_{p}^{2}$ should be above 0.5 to ensure that the original model is not obtained by chance [63].

$$c_{R_p}^{2} = R^{*}\sqrt{R^{2} - R_{y}^{2}}$$

Where ($^{c}R_{p}^{2}$) is the degree of variation in the values of the squared correlation coefficient average of the randomized models R_{y}^{2} and the squared correlation coefficient of the original model R^{2}.
Statistical parameters of the scrambled models gathered around zero in a symmetrical way for both datasets (Fig. 3**)**, representing that the scrambled models are of very low quality. $^{c}Rp^{2}$ values calculated for cross validation was also found to be 0.687and 0.791 (more than 0.5) for quinolones and sulfonamides QSRR models, respectively what denies that the obtained model is the result of a chance correlation.

The applicability domain of a QSPR is the structural, biological space or physicochemical knowledge or information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds. In the William's plot for the FFA-SVM models, the applicability domain is inside a squared area within ± 3 standard deviations and a leverage threshold h* of 1.125 and 0.4 for quinolones and sulfonamides, respectively. Prediction is considered reliable only for those compounds that fall within this AD. Also, it can be seen that all compounds (training and test sets) are inside this area without any outlier (Fig. 4**)**. Moreover, the residual plots for both classes show the differences between the predicted and the experimental retention factor (residuals) for the different compounds. The random dispersion of the residuals around the horizontal axis confirmed the predictability of the model, (**Fig. S9&S10**).

4. Conclusion

Two QSRR models were built for prediction of the retention factors of quinolones and sulfonamides in HPLC system. The influence of the pH of the mobile phase on the retention factors and the ionization state of each

quinolone and the effect of acetonitrile composition in the mobile phase on the retention factors of sulfonamides were studied resulting in selection of 21 major microspecies of quinolones and 39 sulfonamides compounds. In both classes, significant descriptors that are related to the retention behavior in the chromatographic system were selected using the advanced FFA and then incorporated in building the QSRR models using SVM algorithm. In both, FFA-SVM models displayed that the analyte retention is dependent on its hydrophobic/hydrophilic nature and/or its size. The two models showed high performance on both the training level and the validation level. In quinolones, the regression coefficients of the training set prediction (R^2_{cal}), CV $_{LOO}$ (q^2_{LOO}) and the external test set (R^2_{pred}) were 0.931 ($R^2_{adjusted}$ = 0.926), 0.808 and 0.879, respectively, and with root mean square errors (RMSE) of 0.114, 0.163 and 0.148, respectively. In sulfonamides, the regression coefficients of the training set prediction (R^2_{cal}), CV $_{L10\%O}$ ($q^2_{L10\%O}$) and the external test set (R^2_{pred}) were 0.900 ($R^2_{adjusted}$ = 0.896), 0.812 and 0.820, respectively, and with root mean square errors (RMSE) of 0.240, 0.450 and 0.328, respectively. In the Y-randomization validation test, the two models showed $^cR_p^2$ values of 0.687 and 0.791 for quinolones and sulfonamides, respectively indicating that both models are significant and not obtained by chance.

5. Declarations

5.1. Ethics approval and consent to participate

Not applicable

5.2. Consent for publication

Not applicable

5.3. Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

5.4. Competing interests

The authors have no competing of interest

5.5. Funding

No funding received

5.6. Authors' contributions

Marwa A. Fouad: Conceptualization, Writing - Review & Editing, Supervision **Enas H. Tolba**: Methodology, Investigation, Writing - Original Draft **Manal A. El-Shal**: Supervision **Ahmed Serag:** Methodology, Software, Writing - Original Draft and **Ahmed M. El Kerdawy:** Conceptualization, Methodology, Software, Writing - Review & Editing

5.7. Acknowledgements

Not applicable

References

1. Doble A. Quinolones. xPharm: The Comprehensive Pharmacology Reference. 2007;:1–3.
2. Aldred KJ, Kerns RJ, Osheroff N. Mechanism of quinolone action and resistance. Biochemistry. 2014;53:1565–74.
3. Pham TDM, Ziora ZM, Blaskovich MAT. Quinolone antibiotics. MedChemComm. 2019;10:1719–39.
4. Gauhar S, Ali SA, Shoaib H, Shyum Naqvi SB, Muhammad IN. Development and Validation of a HPLC Method for Determination of Pefloxacin in Tablet and Human Plasma. Mashhad University of Medical Sciences. 2009;12:33–42.
5. Holtzapple CK, Buckley SA, Stanker LH. Determination of fluoroquinolones in serum using an on-line clean-up column coupled to high-performance immunoaffinity–reversed-phase liquid chromatography. Journal of Chromatography B: Biomedical Sciences and Applications. 2001;754:1–9.
6. Sultana N, Arayne MS, Shafi N, Naz A, Naz S, Shamshad H. A RP-HPLC Method For The Simultaneous Determination Of Diltiazem and Quinolones in Bulk, Formulations and Human Serum. Journal of the Chilean Chemical Society. 2009;54:358–62.
7. Haritova AM, Petrova DK, Stanilova SA. A Simple HPLC Method For Detection Of Fluoroquinolones in Serum Of Avian Species. Journal of Liquid Chromatography & Related Technologies. 2012;35:1130–9.
8. Nemutlu E, Kır S, Özyüncü Ö, Beksaç MS. Simultaneous Separation and Determination of Seven Quinolones Using HPLC: Analysis of Levofloxacin and Moxifloxacin in Plasma and Amniotic Fluid. Chromatographia. 2007;66:15–24.
9. Cavazos-Rocha N, Carmona-Alvarado I, Vera-Cabrera L, Waksman-de-Torres N, Salazar-Cavazos M de la L. HPLC Method for the Simultaneous Analysis of Fluoroquinolones and Oxazolidinones in Plasma. Journal of Chromatographic Science. 2014;52:1281–7.
10. Yu H, Mu H, Hu Y-M. Determination of fluoroquinolones, sulfonamides, and tetracyclines multiresidues simultaneously in porcine tissue by MSPD and HPLC–DAD. Journal of Pharmaceutical Analysis. 2012;2:76–81.
11. Gili M, Marchis D, Stella P, Olivo F, Ostorero F, Franzoni M, et al. Multiresidue confirmatory method for determination of quinolones in milk by HPLC: method development and validation according to the criteria of Commission Decision 2002/657/EC. Italian Journal of Food Safety. 2013;2:9.
12. Christodoulou EA, Samanidou VF. Multiresidue HPLC analysis of ten quinolones in milk after solid phase extraction: Validation according to the European Union Decision 2002/657/EC. Journal of Separation Science. 2007;30:2421–9.
13. Stoilova N, Surleva A, Stoev G. Determination of quinolonones in food of animal origin by liquid chromatography coupled with fluorescence and mass spectrometric detection. Acta Chromatographica. 2014;26:599–614.

14. Sijun Zhao, Haiyang Jiang, Xuelian Li, Tiejun Mi, Cun Li and, Shen* J. Simultaneous Determination of Trace Levels of 10 Quinolones in Swine, Chicken, and Shrimp Muscle Tissues Using HPLC with Programmable Fluorescence Detection. :J Agric Food Chem . 2007;55:3829–34.
15. Chonan T, Fujimoto T, Inoue M, Tazawa T, Ogawa H. [Multiresidue determination of quinolones in animal and fishery products by HPLC]. Shokuhin eiseigaku zasshi Journal of the Food Hygienic Society of Japan. 2008;49:244–8.
16. Stoilova N, Petkova M. Developing and Validation Of Method For Detection of Quinolone Residues in Poultry Meat. Trakia Journal of Sciences. 2010;8:64–9.
17. Jang J, Lee K, Kwon K, Bae S, Kim HS. Simultaneous determination of thirteen quinolones in livestock and fishery products using ultra performance LC with electrospray ionization tandem mass spectrometry. Food Science and Biotechnology. 2013;22:1–9.
18. Du W, Yao J, Li Y, Hashi Y. Rapid Determination of Residual Quinolones in Honey Samples by Fast HPLC with an On-Line Sample Pretreatment System. American Journal of Analytical Chemistry. 2011;02:200–5.
19. Turiel E, Bordin G, Rodríguez AR. Determination of quinolones and fluoroquinolones in hospital sewage water by off-line and on-line solid-phase extraction procedures coupled to HPLC-UV. Journal of Separation Science. 2005;28:257–67.
20. Ibrahim HK, Abdel-Moety MM, Abdel-Gawad SA, Al-Ghobashy MA, Kawy MA. Validated electrochemical and chromatographic quantifications of some antibiotic residues in pharmaceutical industrial waste water. Environmental Science and Pollution Research. 2017;24:7023–34.
21. Eva M. Golet, Alfredo C. Alder, Andreas Hartmann, Thomas A. Ternes and, Giger W. Trace Determination of Fluoroquinolone Antibacterial Agents in Urban Wastewater by Solid-Phase Extraction and Liquid Chromatography with Fluorescence Detection. Anal Chem. 2001;73:3632–8.
22. Scherer R, Pereira J, Firme J, Lemos M, Lemos M. Determination of Ciprofloxacin in Pharmaceutical Formulations Using HPLC Method with UV Detection. Indian journal of pharmaceutical sciences. 2014;76:541–4.
23. Chamseddin C, Jira TH. Comparison of the chromatographic behavior of levofloxacin, ciprofloxacin and moxifloxacin on various HPLC phases. Die Pharmazie. 2011;66:244–8.
24. Sversut RA, Serrou do Amaral M, César de Moraes Baroni A, Rodrigues PO, Rosa AM, Galana Gerlin MC, et al. Stability-indicating HPLC-DAD method for the simultaneous determination of fluoroquinolones and corticosteroids in ophthalmic formulations. Anal Methods. 2014;6:2125–33.
25. Puranik M, Bhawsar D, Rathi P, Yeole P. Simultaneous determination of ofloxacin and ornidazole in solid dosage form by RP-HPLC and HPTLC techniques. Indian Journal of Pharmaceutical Sciences. 2010;72:513.
26. e Souza MJ, Bittencourt CF, Morsch LM. LC determination of enrofloxacin. Journal of pharmaceutical and biomedical analysis. 2002;28:1195–9.
27. John P, Azeem W, Ashfaq M, Khan IU, Razzaq SN. Stability indicating RP-HPLC method for simultaneous determination of piroxicam and ofloxacin in binary combination. Pakistan journal of pharmaceutical sciences. 2015;28:1713–21.

28. Hubicka U, Żuromska-Witek B, Komsta Ł, Krzek J. Lipophilicity study of fifteen fluoroquinolones by reversed-phase thin-layer chromatography. Analytical Methods. 2015;7:3841–8.
29. Wu L-P, Chen Y, Wang S-R, Chen C, Ye L-M. Quantitative retention–activity relationship models for quinolones using biopartitioning micellar chromatography. Biomedical Chromatography. 2008;22:106–14.
30. Parasca OM, Gheaţă F, Pânzariu A, Geangalău I, Profire L. Importance of sulfonamide moiety in current and future therapy. Revista medico-chirurgicala a Societatii de Medici si Naturalisti din Iasi. 117:558–64.
31. Ye S, Yao Z, Na G, Wang J, Ma D. Rapid simultaneous determination of 14 sulfonamides in wastewater by liquid chromatography tandem mass spectrometry. Journal of Separation Science. 2007;30:2360–9.
32. Malintan NT, Mohd MA. Determination of sulfonamides in selected Malaysian swine wastewater by high-performance liquid chromatography. Journal of Chromatography A. 2006;1127:154–60.
33. Saini B, Bansal G. Degradation Study on Sulfasalazine and a Validated HPLC-UV Method for its Stability Testing. Scientia pharmaceutica. 2014;82:295–306.
34. Ozkorucuklu SP, Sahin Y, Alsancak G. Determination of sulfamethoxazole in pharmaceutical formulations by flow injection system/HPLC with potentiometric detection using polypyrrole electrode. Journal of the Brazilian Chemical Society. 2011;22:2171–7.
35. Cazenave-Gassiot A, Boughtflower R, Caldwell J, Coxhead R, Hitzel L, Lane S, et al. Prediction of retention for sulfonamides in supercritical fluid chromatography. Journal of Chromatography A. 2008;1189:254–65.
36. Kaliszan* R. QSRR: Quantitative Structure-(Chromatographic) Retention Relationships. 2007. https://doi.org/10.1021/CR068412Z.
37. Mauri A, Consonni V, Pavan M, Match RT-, 2006 undefined. Dragon software: An easy approach to molecular descriptor calculations. academia.edu.
38. Fister I, Yang X-S, Brest J, Fister D. A Brief Review of Nature-Inspired Algorithms for Optimization. ELEKTROTEHNISKI VESTNIK. 2013;80:1–7.
39. Yang X-S. Flower Pollination Algorithm for Global Optimization. Springer, Berlin, Heidelberg; 2012. p. 240–9.
40. Yang X-S. Firefly Algorithms for Multimodal Optimization. Springer, Berlin, Heidelberg; 2009.
41. Qi Shen, Jian-Hui Jiang, Jing-chao Tao, Guo-li Shen and, Ru-Qin Yu. Modified Ant Colony Optimization Algorithm for Variable Selection in QSAR Modeling: QSAR Studies of Cyclooxygenase Inhibitors. J Chem Inf Model. 2005;45:1024–9.
42. Gonzalez M, Teran C, Saiz-Urra L, Teijeira M. Variable Selection Methods in QSAR: An Overview. Current Topics in Medicinal Chemistry. 2008;8:1606–27.
43. Shahlaei M. Descriptor Selection Methods in Quantitative Structure–Activity Relationship Studies: A Review Study. Chemical Reviews. 2013;113:8093–103.
44. Fouad MA, Tolba EH, El-shal MA, El AM. QSRR modeling for the chromatographic retention behavior of some -lactam antibiotics using forward and firefly variable selection algorithms coupled with multiple linear regression. 2018;1549:51–62.

45. Carlucci G, D'Archivio AA, Maggi MA, Mazzeo P, Ruggieri F. Investigation of retention behaviour of non-steroidal anti-inflammatory drugs in high-performance liquid chromatography by using quantitative structure–retention relationships. Analytica Chimica Acta. 2007;601:68–76.
46. Golubović J, Protić A, Zečević M, Otašević B, Mikić M, Živanović L. Quantitative structure–retention relationships of azole antifungal agents in reversed-phase high performance liquid chromatography. Talanta. 2012;100:329–37.
47. Ghasemi J, Saaidpour S. QSRR Prediction of the Chromatographic Retention Behavior of Painkiller Drugs. Journal of Chromatographic Science. 2009;47:156–63.
48. Vapnik N V. The Nature of Statistical Learning Theory. 2nd edition. New York: Springer science & business media; 2000.
49. Louis B, Agrawal VK, Khadikar P V. Prediction of intrinsic solubility of generic drugs using MLR, ANN and SVM analyses. European Journal of Medicinal Chemistry. 2010;45:4018–25.
50. Du H, Wang J, Yao X, Hu Z. Quantitative Structure-Retention Relationship Models for the Prediction of the Reversed-Phase HPLC Gradient Retention Based on the Heuristic Method and Support Vector Machine. Journal of Chromatographic Science. 2009;47:396–404.
51. Du H, Wang J, Yao X, Hu Z. Quantitative structure-retention relationship models for the prediction of the reversed-phase HPLC gradient retention based on the heuristic method and support vector machine. Journal of chromatographic science. 47:396–404.
52. Li X, Luan F, Si H, Hu Z, Liu M. Prediction of retention times for a large set of pesticides or toxicants based on support vector machine and the heuristic method. Toxicology Letters. 2007;175:136–44.
53. ChemAxon - Software Solutions and Services for Chemistry & Biology.
54. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and Compound databases. Nucleic Acids Research. 2016;44:D1202–13.
55. The PubChem Project. https://pubchem.ncbi.nlm.nih.gov/. Accessed 20 Nov 2017.
56. Olivieri AC. Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial. Analytica Chimica Acta. 2015;868:10–22.
57. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. Wiley; 2012.
58. Centner, V.; De Noord, O.; Massart D. Detection of nonlinearity in multivariate calibration. Analytica Chimica Acta. 1998;376:153–68.
59. Martinez WL, Martinez AR, Solka JL. Exploratory data analysis with MATLAB. 2nd edition. CRC Press; 2011.
60. Menke W, Menke JE (Joshua E. Environmental data analysis with MatLab. 2nd edition. 2016.
61. Sharma MC, Kohli D V. Insight into the structural requirement of substituted quinazolinone biphenyl acylsulfonamides derivatives as Angiotensin II AT1 receptor antagonist: 2D and 3D QSAR approach. Journal of Saudi Chemical Society. 2014;18:35–45.
62. Guha R, Willighagen E. A Survey of Quantitative Descriptions of Molecular Structure. Current Topics in Medicinal Chemistry. 2020;12:1946–56.

63. Gramatica P. Principles of QSAR models validation: internal and external. QSAR & Combinatorial Science. 2007;26:694–701.

Figures

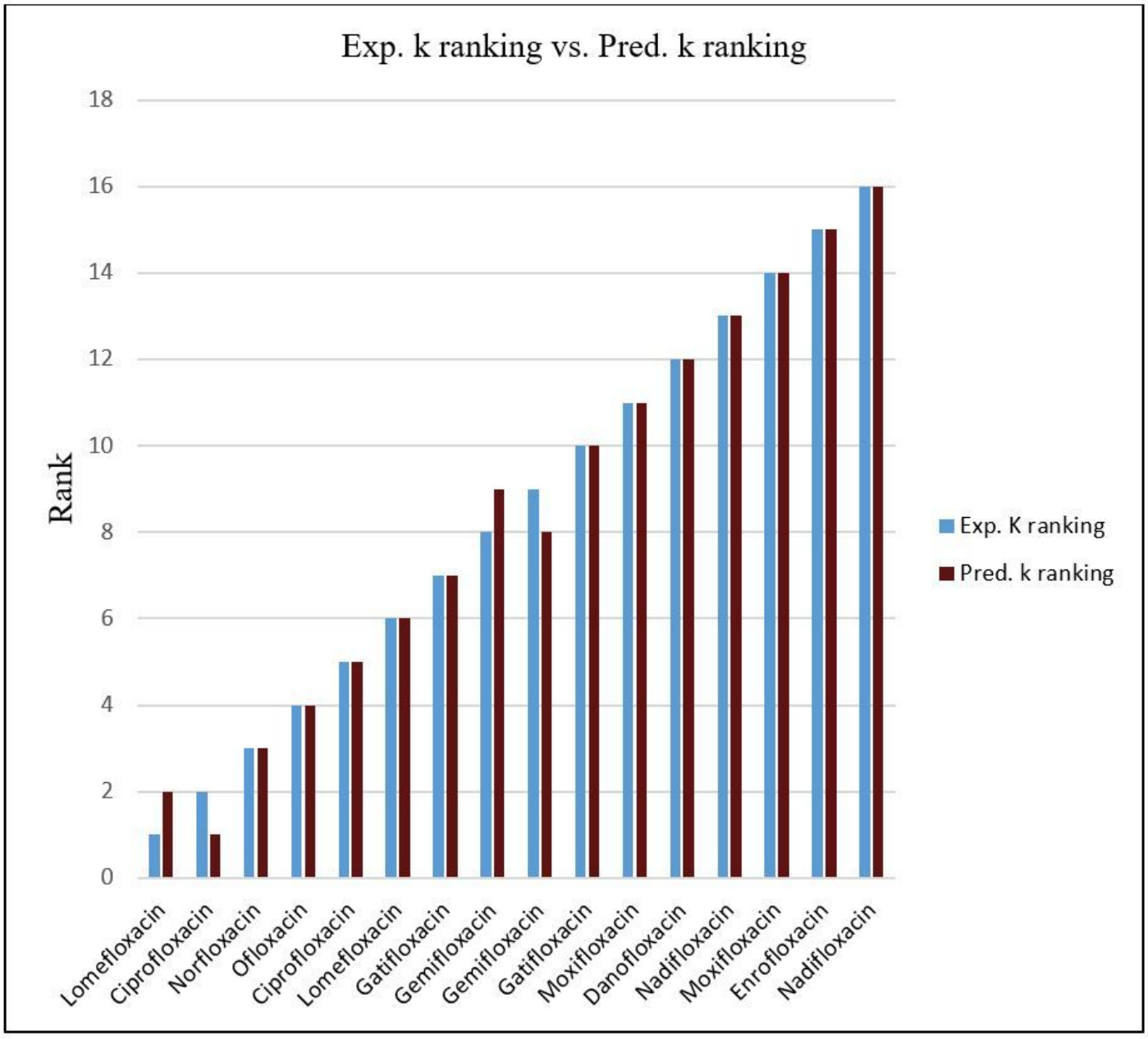


Figure 1

FFA-SVM model experimental k ranking vs predicted k ranking in Quinolones training set prediction.

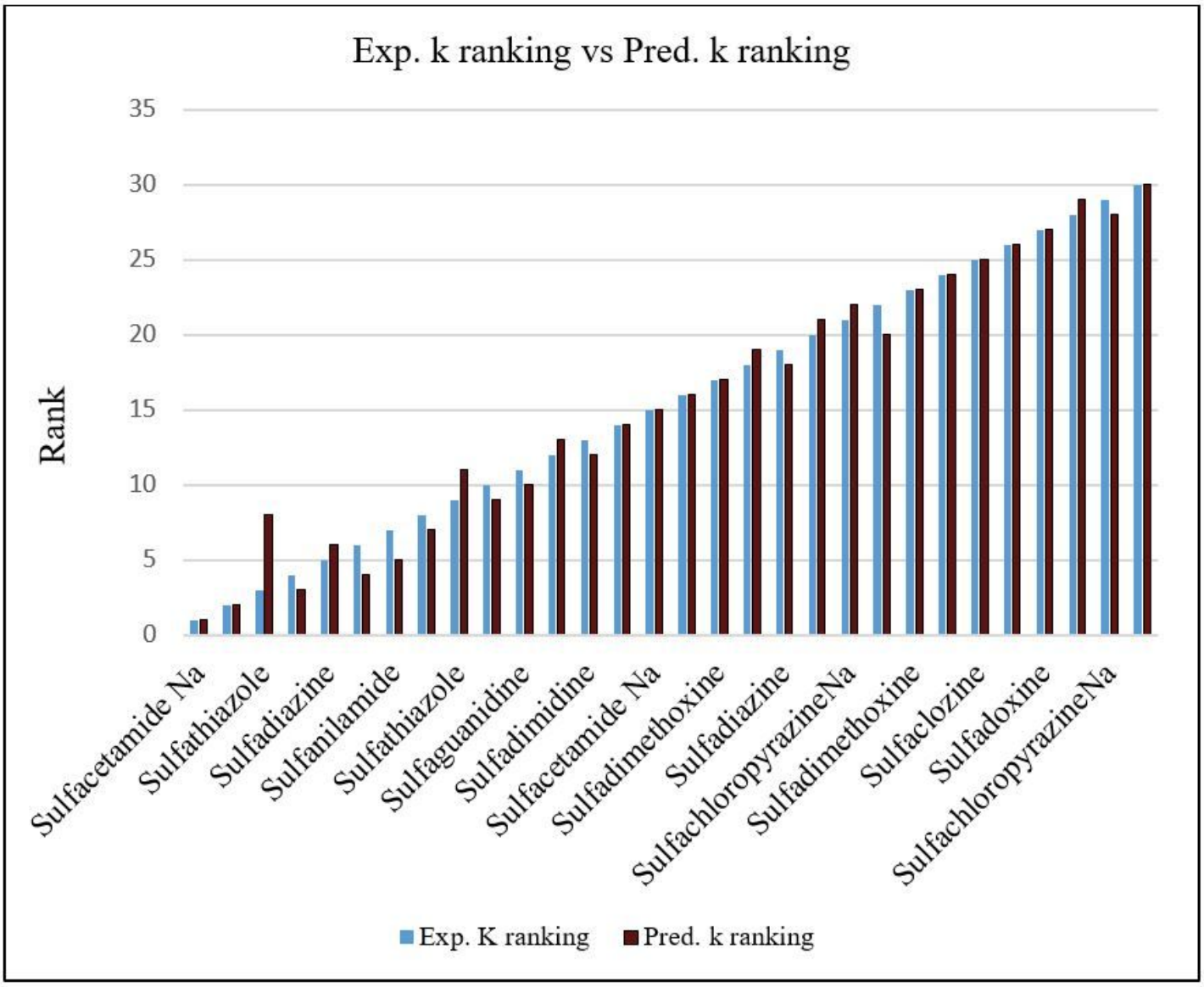
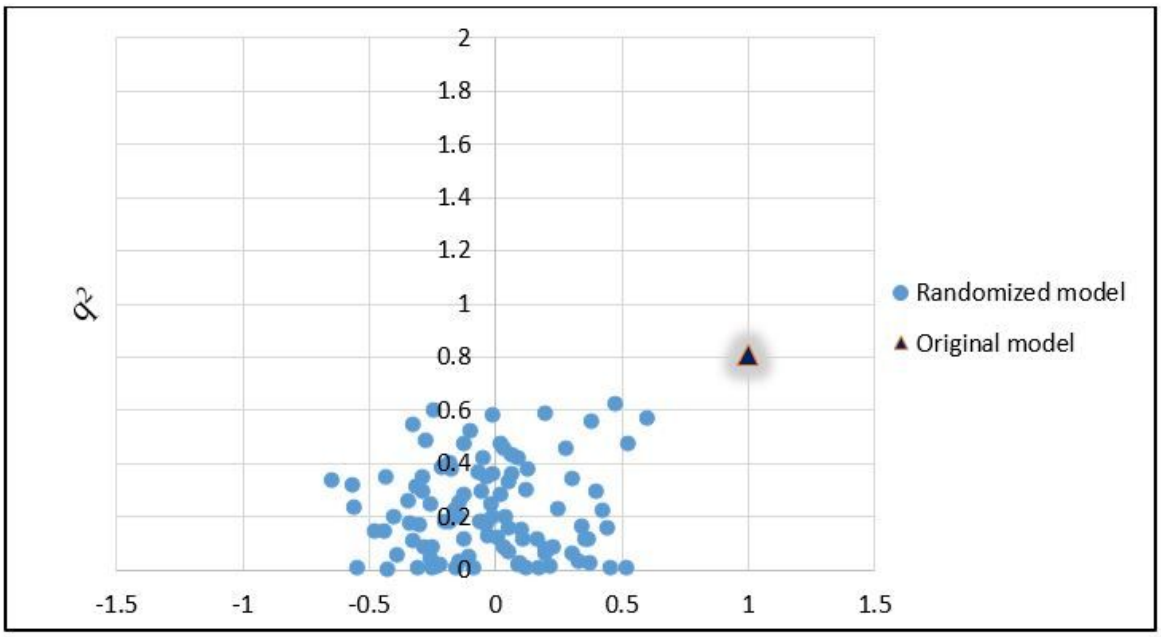
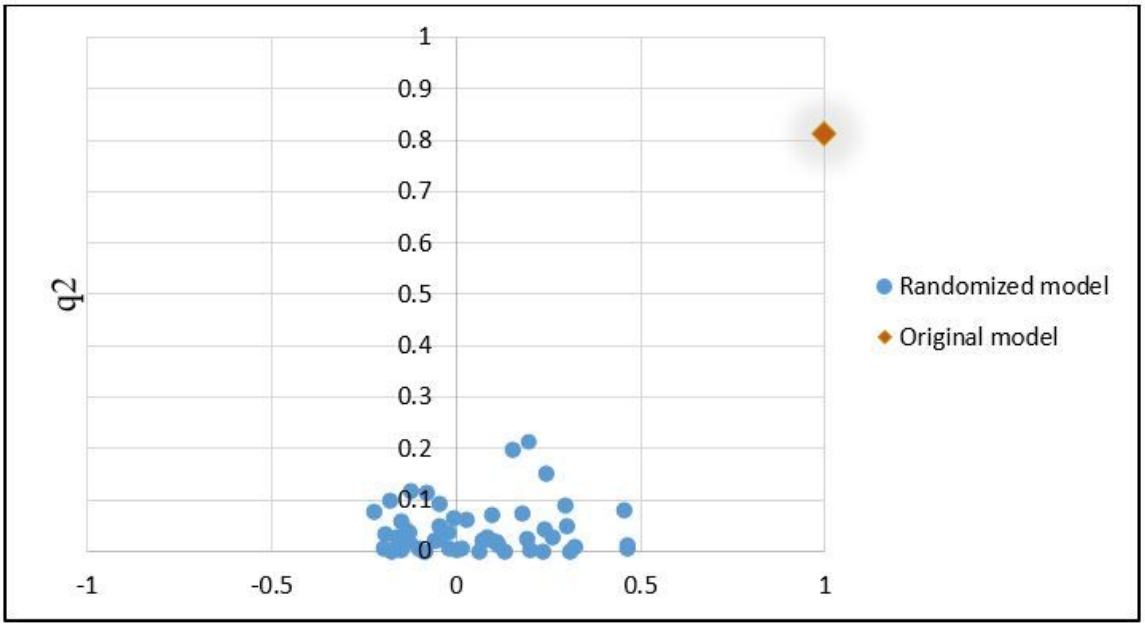


Figure 2

FFA-SVM model experimental k ranking vs predicted k ranking in Sulfonamides training set prediction



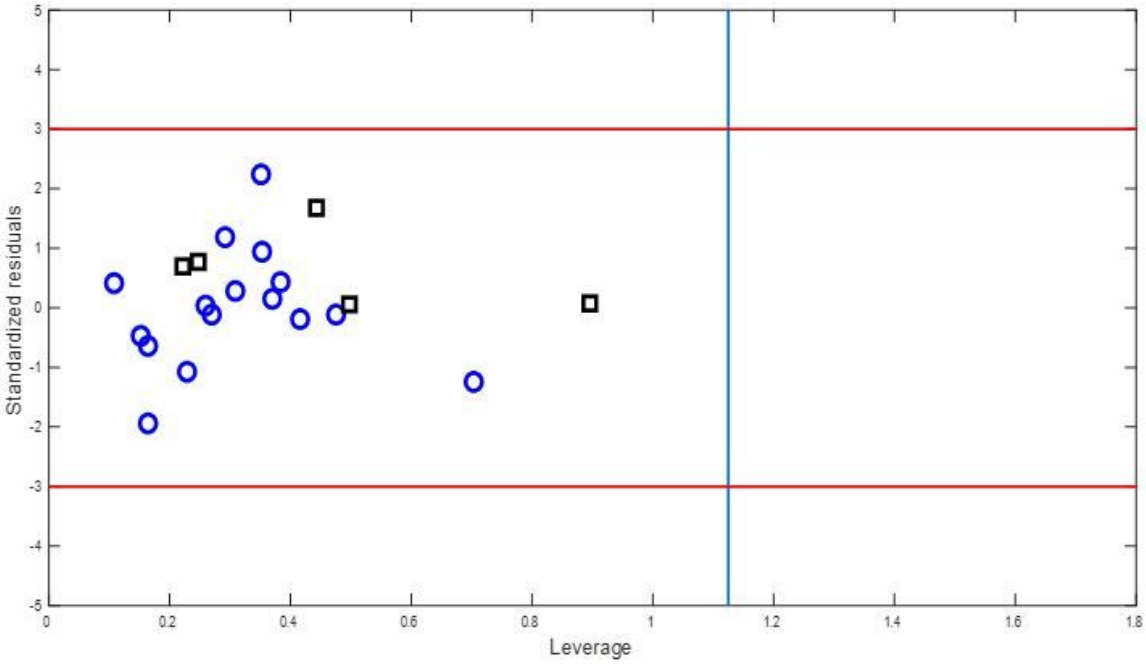
(A)



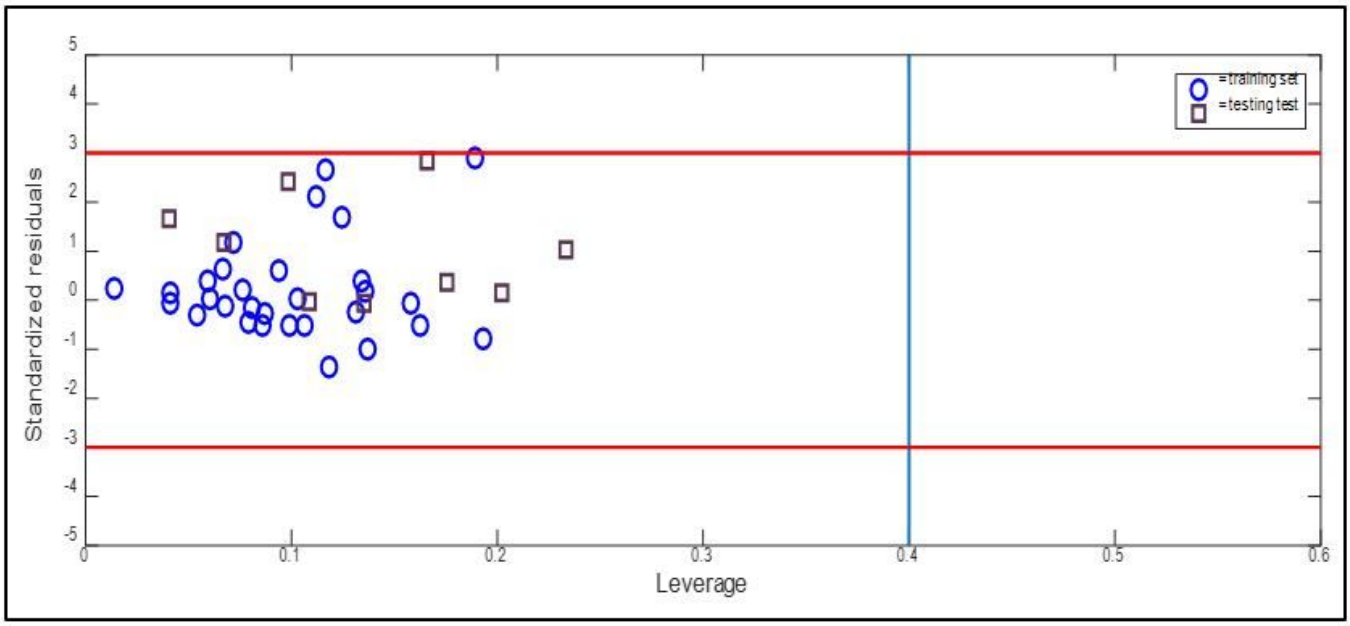
(B)

Figure 3

Y-randomization validation results for the FFA-SVM for (A) quinolones (B) sulfonamides modeling.



(A)



(B)

Figure 4

Williams plot for FFA-SVM models of (A) quinolones with ±3 standard deviations, and a leverage threshold h* of 1.125 as warning limits and (B) sulfonamides with ±3 standard deviations, and a leverage threshold h* of 0.4 as warning limits. Circles represent training set cross-validation prediction and diamonds represent test set prediction.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx
- quinolones.xlsx
- sulfonamides.xlsx