

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/282340936>

# How to rank and discriminate artificial neural networks? Case study: prediction of anticancer activity of 17-picoly and 17-picolinylidene androstane derivatives

ARTICLE in JOURNAL OF THE IRANIAN CHEMICAL SOCIETY · OCTOBER 2015

Impact Factor: 1.09 · DOI: 10.1007/s13738-015-0759-9

---

READS

79

## 7 AUTHORS, INCLUDING:



Strahinja Kovačević

University of Novi Sad

41 PUBLICATIONS 65 CITATIONS

[SEE PROFILE](#)



Lidija Jevrić

University of Novi Sad

58 PUBLICATIONS 153 CITATIONS

[SEE PROFILE](#)



Evgenija A. Djurendić

University of Novi Sad

89 PUBLICATIONS 356 CITATIONS

[SEE PROFILE](#)



Jovana Ajdukovic

University of Novi Sad

22 PUBLICATIONS 88 CITATIONS

[SEE PROFILE](#)

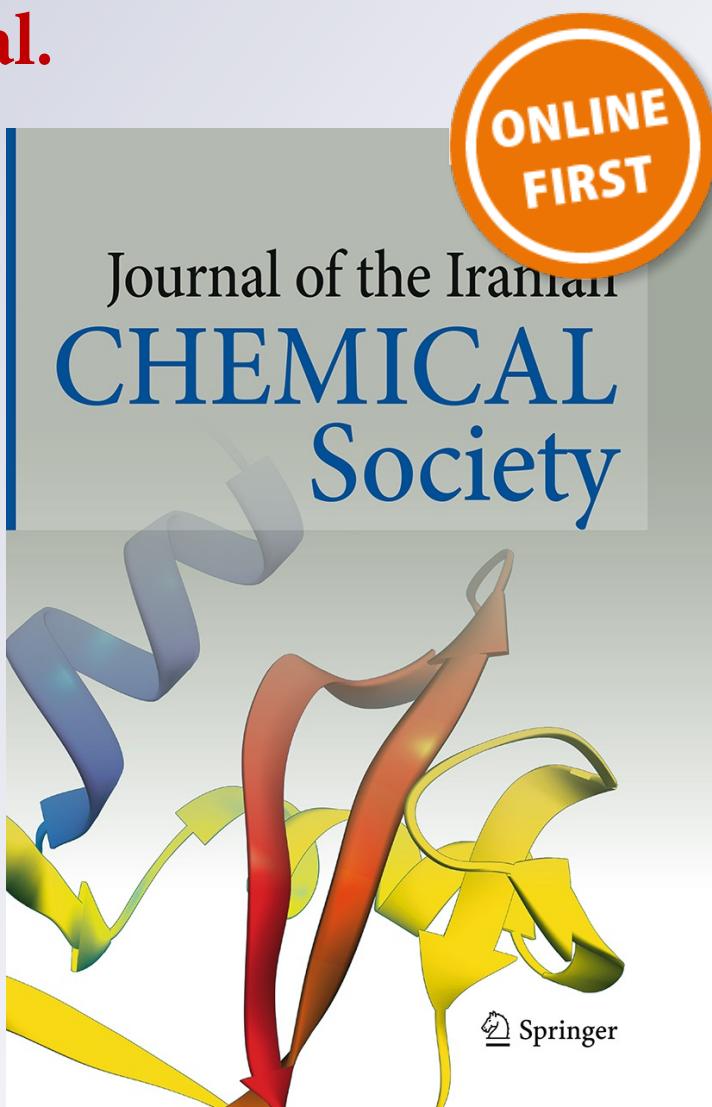
*How to rank and discriminate artificial neural networks? Case study: prediction of anticancer activity of 17-picoly and 17-picolinylidene androstane derivatives*

**Strahinja Z. Kovačević, Sanja O. Podunavac-Kuzmanović, Lidiya R. Jevrić, Evgenija A. Djurendić, Jovana J. Ajduković, et al.**

**Journal of the Iranian Chemical Society**

ISSN 1735-207X

J IRAN CHEM SOC  
DOI 10.1007/s13738-015-0759-9



**Your article is protected by copyright and all rights are held exclusively by Iranian Chemical Society. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# How to rank and discriminate artificial neural networks? Case study: prediction of anticancer activity of 17-picolyl and 17-picolinylidene androstane derivatives

Strahinja Z. Kovačević<sup>1</sup>  · Sanja O. Podunavac-Kuzmanović<sup>1</sup> · Lidiya R. Jevrić<sup>1</sup> ·  
Evgenija A. Djurendić<sup>2</sup> · Jovana J. Ajduković<sup>2</sup> · Slobodan B. Gadžurić<sup>2</sup> ·  
Milan B. Vranes<sup>2</sup>

Received: 22 May 2015 / Accepted: 3 October 2015  
© Iranian Chemical Society 2015

**Abstract** Model discrimination is still not a resolved task. The classical statistical approaches lead to different results (for the same models) and at the same time a lot of models seem to be statistically equivalent. The authors deliberately select such conditions when their algorithm is superior. Hence, it is better to apply different approaches to compare and rank the models fairly. This paper presents the application of methodology called sum of ranking differences (SRD) to rank the artificial neural network models [quantitative structure–activity relationship (QSAR) models] designed for prediction of anticancer activity of 17-picolyl and 17-picolinylidene androstane derivatives toward androgen receptor negative prostate cancer cells (AR-PC-3). The SRD method suggests the consistent models, in terms of compounds order and proximity to the golden standard, which should preferably be used in the prediction of anticancer activity of studied androstane derivatives.

**Keywords** Chemometrics · Mathematical models · Prostate cancer · Quantitative structure–activity relationship · Sum of ranking differences

**Electronic supplementary material** The online version of this article (doi:[10.1007/s13738-015-0759-9](https://doi.org/10.1007/s13738-015-0759-9)) contains supplementary material, which is available to authorized users.

✉ Strahinja Z. Kovačević  
strahko@uns.ac.rs

<sup>1</sup> Department of Applied and Engineering Chemistry, Faculty of Technology Novi Sad, University of Novi Sad, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia

<sup>2</sup> Department of Chemistry, Biochemistry and Environmental Protection, Faculty of Sciences, University of Novi Sad, Trg Dositeja Obradovića 3, 21000 Novi Sad, Serbia

## Introduction

Prostate cancer is considered as the most commonly diagnosed cancer, as well as the second most common cause of cancer-related death in men [1]. Since it is a growing worldwide healthcare problem, there are many researches aimed to improve its treatment. One of the ways to find an effective anticancerogenic compound is its organic synthesis and testing of its potential biological effect in vitro and in vivo. In most cases this way implies trial-and-error approach because it is possible that synthesized compound will not express desired biological effect even if it is very structurally similar to the one which has significant biological response in the applied biological system. It can be assumed that steric, hydrophobic/lipophilic and electronic intermolecular interactions influence the biological activity. Small changes in the molecular structure can significantly change the biological potential of a compound, in negative or positive direction. In order to avoid this problem, chemometrics has taken a significant position in drug discovery procedures. Specifically, quantitative structure–activity relationship (QSAR) analysis, as one of the main chemometric approaches, is widely applicable in the prediction of biological activity of compounds, which can include immunotoxicity, antifungal, antioxidant and anticancer activity, HIV inhibition, etc. [2–7].

Quantitative structure–activity relationship equations can be formed using different simple or multivariate mathematical methods, either linear or non-linear [8–10]. When a QSAR model is obtained, it is very important to define its applicability domain and the conditions under which it is constructed [11]. In every QSAR study the appropriate statistical validation is compulsory and it is a crucial aspect of any QSAR analysis [12]. A simple comparison of the basic

statistical measures and validation parameters of several established QSAR models is the mostly applied procedure for selection of the best model. However, there is another novel statistical approach called sum of ranking differences (SRD), which can be applied for reliable comparison and discrimination of QSAR models. Generally, comparison of models is a difficult and ambiguous task because in some cases models can be statistically indistinguishable or seemingly indistinguishable [13]. Therefore, there is an objective need for a procedure which is able to compare and rank models fairly.

Sum of ranking differences procedure corresponds to the principle of parsimony and provides an easy tool for the evaluation of the models or methods—the smaller the sum, the better the model [13]. The ranking of models is validated by comparison of ranks by random numbers (CRRN procedure), which is a kind of simulation test [13–15]. SRD method has been shown as a very useful tool for comparison of linear QSAR and quantitative structure–retention relationship (QSRR) models [16, 17] and in evaluation of analytical data [18].

In a recently published paper [19], 17 artificial neural networks (ANNs), which can be used for prediction of cytotoxic activity ( $IC_{50}$ ) of 17-picoly and 17-picolinylidene androstane derivatives toward androgen receptor negative prostate cancer cells (AR-, PC-3), were presented. The selection of the best ANNs was carried out on the basis of standard statistical parameters calculated for each network, and it was suggested that the networks named MLP 3-15-1, MLP 3-71-1 and MLP 3-4-1 should preferably be used for the prediction of  $IC_{50}$  values of studied derivatives. With application of non-linear regression techniques, such as ANNs, overtraining and overfitting can be a problem if the modeling is not carefully carried out [20]. In order to ensure a fair selection of the best ANNs, we applied new powerful SRD approach which is completely different from the classical one used in the study by Kovačević et al. [19]. SRD does not use or calculate residual errors and performance parameters. It selects consistent models. The present study gives a significant guidelines to chemists regarding the selection of already established non-linear QSAR (ANN) models in the prediction of anticancer activity of studied biologically active androstane derivatives.

## Material and chemometric-statistical methods

### ANN models

ANN models studied in this paper were established for the prediction of anticancer activity of a series of 17-picoly and 17-picolinylidene androstane derivatives toward PC-3 cell line. The architectures of ANN models, their statistical

characteristics and applied ANN method are fully presented and explained in the literature [19]. For simplicity, in this paper the networks are marked according to the number of neurons (nodes) in hidden layer (4, 7, 7a, 8, 10, 13, 14, 14a, 15, 16, 17, 18, 20, 24, 27, 29, 71).

### Sum of ranking differences (SRD)

Sum of ranking differences procedure is entirely general and it implies that a known reference ranking (benchmark or “golden standard”) is available [14]. In order to rank the ANN models by SRD method, in the first step it is necessary to arrange the data in a matrix form. In the matrix the objects (in this study a set of eighteen 17-picoly and 17-picolinylidene androstane derivatives in the first case or statistical characteristics in the second case) are listed in the rows, while variables (ANNs to be compared) are placed in the columns. The columns of this matrix contain  $IC_{50}$  values ( $\mu M$ ) predicted by 17 ANN models or calculated statistical data of the networks. The experimental data are also included among the columns, since they provide an easy way to select good and bad models. Consistent models rationalize the info in the data better than the experimental values, while consistently “bad” ones rationalize the info in the data (much) worse than the experimental values. Their existence is not justified, we are better off if we use measured values in table form than using “bad” models [13]. For the reference ranking in SRD analysis three types of data can be chosen:

1. averages of  $IC_{50}$  values for all ANN models for each object (row averages), or median if outliers are suspected;
2. minimum or maximum row value (i.e. for statistical parameters); and
3. known reference values.

The average values can be chosen as a reference by consensus. It is likely that the most probable ranking will be provided by the average, but it is not necessarily a bias-free solution [17], but it contains less bias than ranking by any of the individual vectors.

Applying the matrix formed in the first step and the “golden standard”, each individual ANN model can be ranked and compared to the established reference values of  $IC_{50}$  or reference values of statistical parameters. In the next step, the absolute values of the differences between the reference ranking and individual ranking are calculated and summed for each ANN model [14]. The closer is the SRD value of a model to zero, the better is the model (the ideal model has SRD = 0, since in that case it has the same ranking as the “golden standard”) [14]. If two or more models have similar SRD values, they predict  $IC_{50}$  values similarly

**Table 1** The results of Shapiro–Wilks's test (the models in bold have normal distribution)

Networks	Shapiro–Wilks's test ( <i>W</i> ) value; $H_0$ : the residuals are normally distributed; if $W < p$ ( $p = 0.05$ ), $H_0$ is rejected
<b>16</b>	<b>0.0979</b>
20	0.0200
14	0.0008
29	0.0299
8	0.0165
10	0.0230
18	0.0036
15	0.0173
27	0.0070
14a	0.0152
7	0.0426
17	0.0487
24	0.0238
71	0.0293
4	0.0304
13	0.0148
<b>7a</b>	<b>0.0549</b>

(close proximity means close similarity). Therefore, the SRD method can be applied for grouping of models or it can be considered as dissimilarity measure [13, 14].

The SRD modeling of ANNs was carried out on the basis of (1) the  $\text{IC}_{50}$  values and (2) statistical characteristics:  $R_{\text{tr}}$ ,

$R_{\text{v}}$  and  $R_{\text{v}}$  (correlation coefficients for the training, test and validation set, respectively),  $\text{RMSE}_{\text{tr}}$ ,  $\text{RMSE}_{\text{t}}$  and  $\text{RMSE}_{\text{v}}$  (root mean square error for the training, test and validation set, respectively), variation coefficient (VC) and Fisher's value (*F* test).

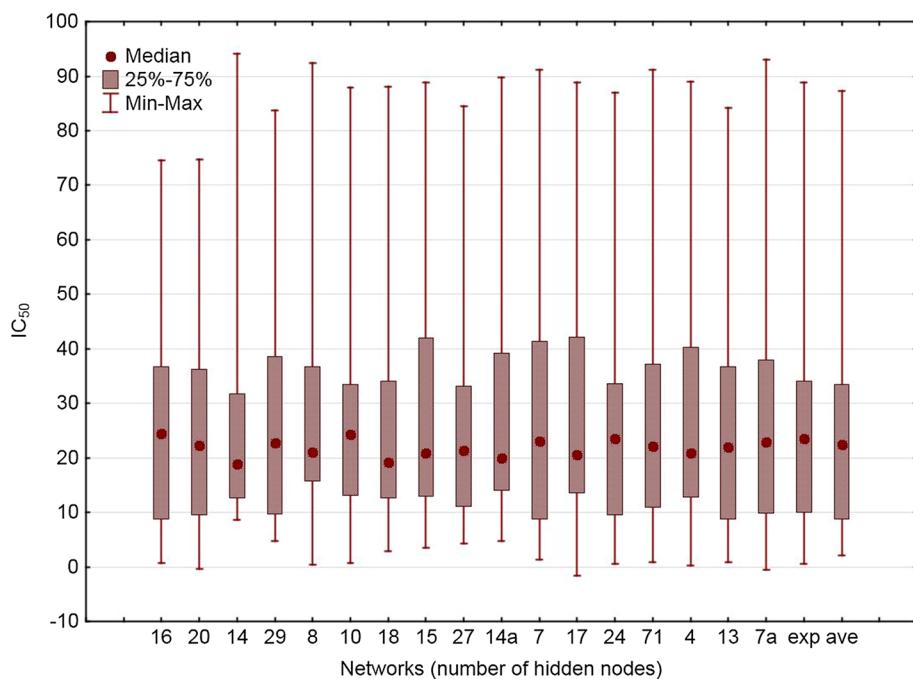
### Validation of SRD procedure: CRRN and cross-validation approaches

Validation of SRD method can be easily conducted by comparison of ranks by random numbers (CRRN) procedure and/or cross-validation (CV).

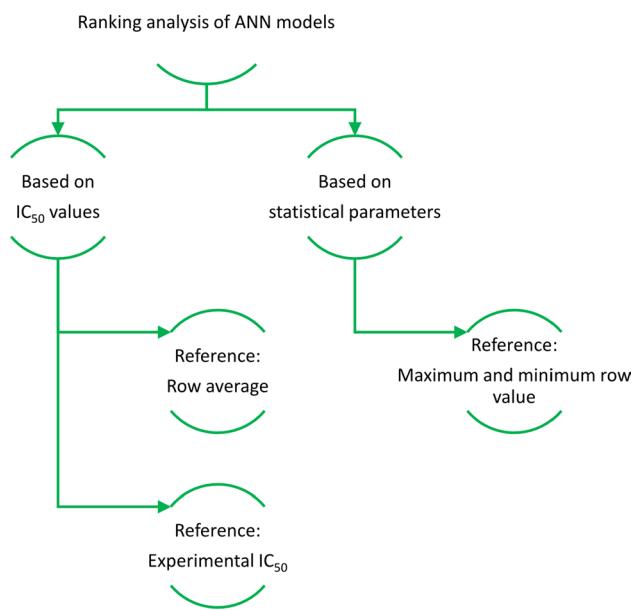
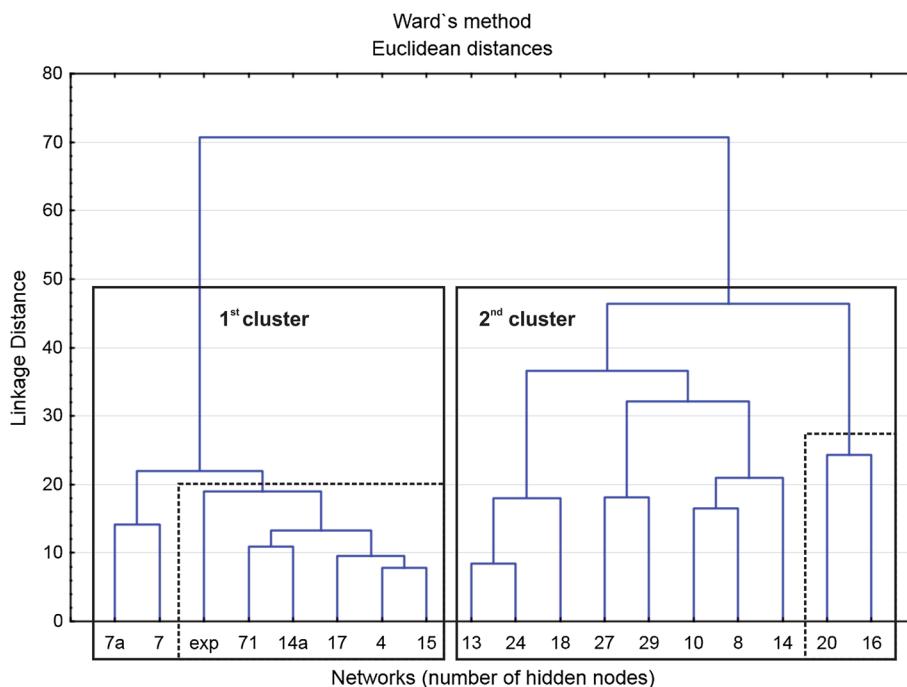
Comparison of ranks by random numbers procedure requires the determination of the theoretical distribution function of SRD values corresponding to the number of  $n$  objects consisted of random numbers [17]. Recursive algorithm is applied for calculation of theoretical distribution function if  $n < 14$ , while the normal distribution is used to approximate the theoretical SRD distribution for random vectors for large number of objects ( $n > 13$ ) [14].

Cross-validation is carried out so that approximately one-seventh of the objects were omitted and the ranking is made on the remained 6/7th number of objects just seven times. If the number of objects is less than 14, then leave-one-out (LOO) CV is carried out. Otherwise, leave-many-out (LMO) CV is completed [15]. CV can estimate uncertainties to the SRD values, which can help to see their clustering [15].

To completely understand the calculation procedures and algorithms of SRD, CRRN and CV methods reading of references [13–15] is recommended.

**Fig. 1** Box–whisker plot of predicted, experimental (exp) and average (ave)  $\text{IC}_{50}$  values

**Fig. 2** Dendrogram as a result of cluster analysis of studied networks and experimental IC<sub>50</sub> values



**Fig. 3** The scheme of the SRD analysis of ANN models applied in this study

### Basic and exploratory statistics

Sign test and Wilcoxon matched pair test can provide statistically valid information about testing the null hypothesis which implies that there is no difference between models (at a certain probability level, i.e. 5 %). Well-known *t* test can be applied only if the data are normally distributed. Sign test is one of the simplest non-parametric

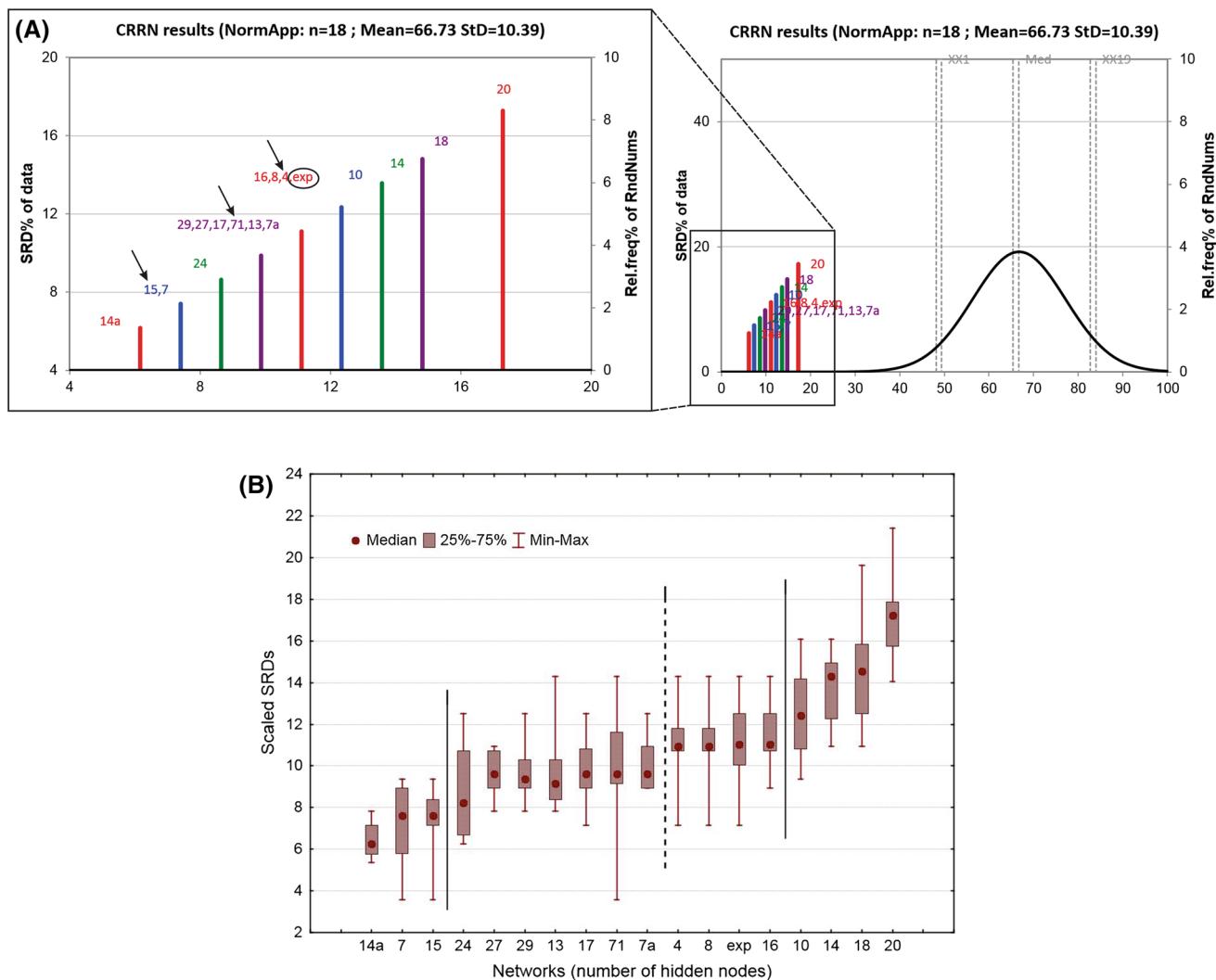
statistical methods which avoids this assumption about normal distribution of the data and is much easier to perform [21]. The Wilcoxon matched pairs test is a non-parametric alternative to the *t* test for dependent samples [22]. It tests the hypothesis that the scores for two variables were drawn from the same distribution. The sign test only uses information about the sign of the difference between two variables. Since the Wilcoxon test takes the relative magnitude of those differences into account, it is more sensitive than the sign test. In this study, these two tests were applied to determine similarities or dissimilarities among the ANNs on the basis of uncertainties of SRDs.

Cluster analysis (CA) is a classical chemometric technique for dividing a group of objects into classes so that similar objects are in the same class (cluster). The groups are not known prior to the mathematical analysis and no assumptions are made about the distribution of the variables [21]. CA searches for objects, which are close together in the variable space. The data in each cluster share some common trait, often proximity according to some defined distance measure. In this study CA was utilized to compare the ANN models and to find their inherent similarity.

## Results and discussion

### Basic and exploratory statistics of ANNs

In the first step, the statistical distribution of the experimental and predicted IC<sub>50</sub> data was determined. According to

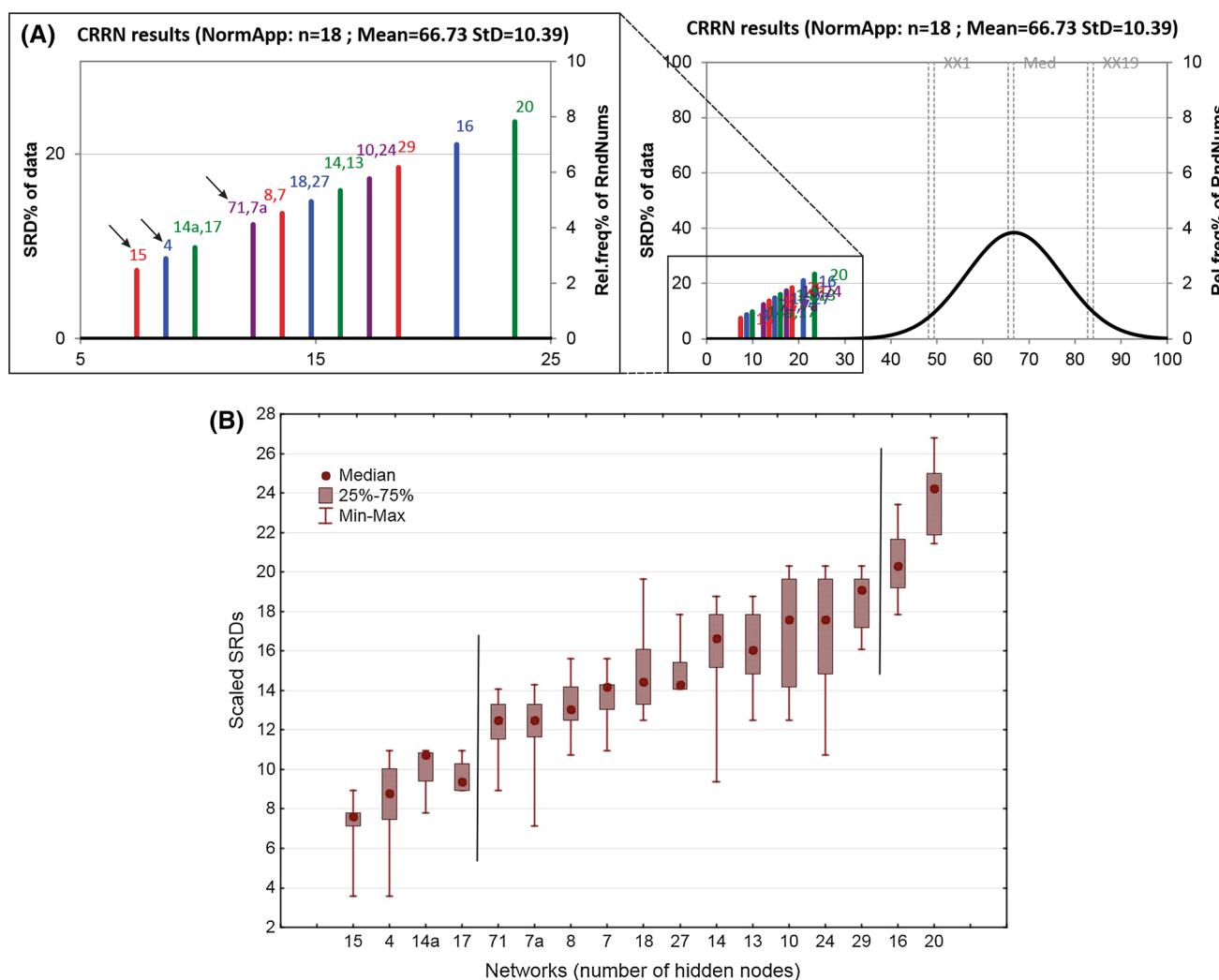


**Fig. 4** **a** The order of 17 ANN models by sum of ranking differences, including the experimental IC<sub>50</sub> values (exp) with row average as a reference. The statistical characteristics of Gaussian fit are the following: first icosaille (5 %), XX1 = 80; first quartile, Q1 = 96;

median, Med = 108; last quartile, Q3 = 120; last icosaille (95 %), XX19 = 136. **b** Box-whisker plot of the seven segments for cross-validation of sum of ranking differences using row average as reference ordering

Shapiro-Wilk's test, there are just two networks with normal distribution. These results are presented in Table 1. As can be seen from the Fig. 1, it is quite difficult to evaluate the models because they seem to have the same distribution and there is no model which is significantly better or worse than the others. Since the majority of the networks does not have normal distribution of the data, *t* test cannot be applied, and instead of it, the sign test and Wilcoxon matched pair test were carried out. The sign test shows significant difference only between the models 16 and 29, while Wilcoxon test distinguishes the models 7 and 13. It is obvious that comparison of the models cannot be achieved by using classical statistical tools, so additional methods are required.

In the next step, the cluster analysis (CA) was conducted on the data matrix which contains the experimental IC<sub>50</sub> values and IC<sub>50</sub> values predicted by the applied ANNs. In CA clustering was based on the Euclidean distances and Ward's linkage algorithm. The obtained dendrogram is presented in Fig. 2. As can be seen from this dendrogram, there are two main clusters. In the first cluster there are the networks which were suggested as the best ones (4, 15 and 71) [19] together with the experimental IC<sub>50</sub> values (exp) and the networks 7 and 7a. The networks with the worst statistical parameters (20 and 16) appeared in the same sub-cluster of the second cluster and can be considered as outliers. The second cluster contains the networks 13, 24, 18, 27, 29, 10, 8 and 14 as well.



**Fig. 5** **a** The order of 17 ANN models by sum of ranking differences with experimental  $IC_{50}$  values as a reference. The statistical characteristics of Gaussian fit are the following: first icosaile (5 %), XX1 = 80; first quartile, Q1 = 96; median, Med = 108; last quartile,

Q3 = 120; last icosaile (95 %), XX19 = 136. **b** Box-whisker plot of the seven segments for cross-validation of sum of ranking differences using experimental  $IC_{50}$  values as reference ordering

## SRD analysis

Sum of ranking differences analysis was conducted in a way schematically presented in Fig. 3. The input matrices for the SRD analyses are given in Supplementary data (Tables S1–S3). The main aim of this analysis was to reveal similarities or dissimilarities among the networks and to select the best ones.

### SRD analysis of ANNs based on experimental and predicted $IC_{50}$ data

The generally accepted rationale behind round-robin tests and the maximum likelihood principle suggest that the average is the best choice as standard ranking when SRD actually measures the differences from the center [13]. The

results are shown in Fig. 4a. The experimental  $IC_{50}$  values (exp) split the ANN models into acceptable and non-acceptable ones. The acceptable models have smaller SRD values than the experimental values. It can be seen that models 15, 71 and 4 belong to this group. The model with the smallest SRD number is 14a and according to the proximity to the reference ranking it can be considered as the best, while the model 20 is the worst.

In order to reveal the uncertainties for SRD values of networks studied, sevenfold cross-validation (CV SRD) was applied following the procedure described in “Validation of SRD procedure: CRRN and cross-validation approaches”. The obtained CV SRD results are presented in the form of box-whisker plot in Fig. 4b. If we compare the obtained box-whisker plot with the graph of SRD analysis (Fig. 4a) and consider the results of sign test and Wilcoxon matched

**Table 2** A simple ranking of the ANN models according to the statistical parameters

Network	R <sup>a</sup>	Network	RMSE <sup>a</sup>	Network	F	Network	VC
<b>4</b>	0.9906	<b>4</b>	3.45	<b>4</b>	837.4	<b>4</b>	0.11
<b>15</b>	0.9896	<b>15</b>	3.59	<b>15</b>	762.1	<b>15</b>	0.12
<b>71</b>	0.9895	<b>71</b>	3.71	<b>71</b>	753.3	<b>71</b>	0.12
<b>17</b>	0.9867	<b>17</b>	4.14	<b>17</b>	588.5	<b>17</b>	0.14
<b>14a</b>	0.9856	<b>14a</b>	4.26	<b>14a</b>	543.8	<b>14a</b>	0.14
<b>7a</b>	0.9845	<b>7a</b>	4.48	<b>7a</b>	503.1	<b>7a</b>	0.15
<b>7</b>	0.9829	<b>7</b>	4.82	<b>7</b>	454.5	<b>7</b>	0.15
<b>10</b>	0.9792	<b>27</b>	5.01	<b>10</b>	372.5	<b>10</b>	0.17
<b>27</b>	0.9790	<b>8</b>	5.02	<b>27</b>	369.4	<b>27</b>	0.17
<b>8</b>	0.9785	<b>10</b>	5.06	<b>8</b>	360.0	<b>8</b>	0.17
<b>24</b>	0.9781	<b>24</b>	5.18	<b>24</b>	353.9	<b>24</b>	0.18
<b>13</b>	0.9768	<b>13</b>	5.37	<b>13</b>	333.5	<b>13</b>	0.19
<b>18</b>	0.9745	<b>29</b>	5.62	<b>18</b>	301.2	<b>29</b>	0.19
<b>29</b>	0.9733	<b>18</b>	5.63	<b>29</b>	288.2	<b>18</b>	0.20
<b>14</b>	0.9678	<b>14</b>	6.42	<b>14</b>	236.9	<b>14</b>	0.21
<b>16</b>	0.9546	<b>16</b>	6.97	<b>16</b>	164.3	<b>16</b>	0.25
<b>20</b>	0.9395	<b>20</b>	8.71	<b>20</b>	120.3	<b>20</b>	0.32

Bold values refer to the number of hidden neurons in the networks

The networks which have different ranking in the columns are italic

<sup>a</sup> These values refer to the whole data set

pair test (Supplementary data, Tables S4–S5), the models can be approximately split into four groups as suggested in Fig. 4b. It was confirmed that the model 14a has the lowest SRD value, while the model 20 has the highest SRD value if the row average is used as the golden standard. The networks 14a, 7 and 15 form a group and their rankings are closest to zero. The networks 15, 71 and 4 are distributed in quite different positions. Since the networks 10, 14, 18 and 20 have the highest SRD values, even higher than the experimental values, their usage should be avoided. Although most models were better than the experimental data, this is not a definitive indication of overfitting, the phenomenon was simply the consequence of random noise [16].

If we compare Figs. 1 and 4b, the superiority of SRD method over the classical statistical tools is obviously manifested. SRD approach has better distinctive ability, better pattern recognition ability, clear division into good and bad models and it indicate clear similarity to experimental values.

In the second approach, where the experimental IC<sub>50</sub> values were used as a benchmark, the order of the models is quite different than in the case of consensus ranking (row average). The SRD and CV SRD results (Fig. 5a, b) indicate three main groups of the networks. The first group contains the networks which are the closest to the experimental IC<sub>50</sub> values (15, 4, 14a and 17). The worst models belong to the third group (models 16 and 20). The combinations of models were confirmed by sign test and Wilcoxon matched pair test (Supplementary data, Tables S6–S7). Very similar grouping is obtained in CA analysis. The model 20 is

placed in a maximum distance from the reference ranking again, hence its usage should be definitely avoided.

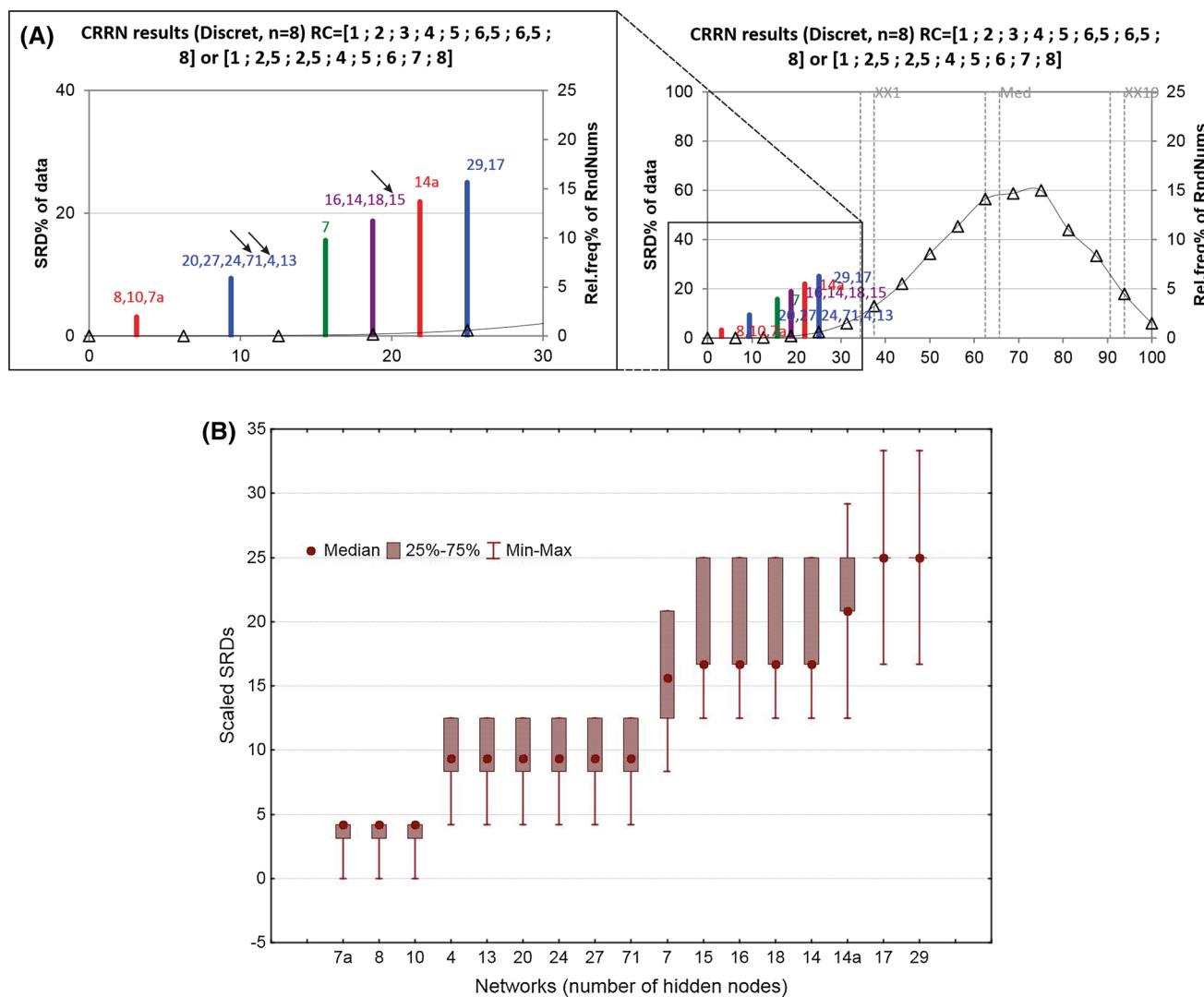
It must be emphasized that the aim of the second approach was not description of experimental errors, but fits.

#### SRD comparison of ANNs based on statistical parameters

Before SRD analysis on the basis of statistical parameters of the ANN models, we listed the models according to the general statistical characteristics R, RMSE, F and VC (Table 2).

Table 2 points out that the model 4 has the best statistical parameters (the highest R and F; the smallest RMSE and VC), while the model 20 has the worst. Since the models 4, 15 and 71 have significantly better statistical parameters than the others, they were suggested as the best ones [19].

In this case, SRD analysis applies completely different approach: it ranks the models according to the maximum row value (for R<sub>tr</sub>, R<sub>t</sub>, R<sub>v</sub> and F parameters) and minimum row value (for RMSE<sub>tr</sub>, RMSE<sub>t</sub>, RMSE<sub>v</sub> and VC parameters) as the reference ranking. The input matrix for this SRD analysis is given in Supplementary data (Table S3). The obtained ranking and CV SRD results are presented in Fig. 6a, b. As can be seen from these results, the best models are 8, 10 and 7a, while the models 29 and 17 have the highest SRD values and therefore can be considered as the worst ones. The networks 4, 13, 20, 24, 27 and 71 are significantly worse than 8, 10 and 7a (Wilcoxon and sign test, Supplementary data, Table S8–S9), but still acceptable and consistent.



**Fig. 6** **a** The order of 17 ANN models by sum of ranking differences based on statistical parameters. The statistical characteristics of the theoretical distribution function are the following: first icosaile (5 %), XX1 = 12; first quartile, Q1 = 17; median, Med = 21; last quartile,

Q3 = 25; last icosaile (95 %), XX19 = 30. **b** Box-whisker plot of the seven segments for cross-validation of sum of ranking differences using statistical parameters of ANN models and their minimum and maximum values of statistical parameters as reference ranking

From the box-whisker plot shown in Fig. 6b the separation of the models into four main groups is more obvious. The networks 7 and 14a are placed between these groups. The first group contains the models 7a, 8 and 10; the second group 4, 13, 20, 24, 27 and 71; the third group 15, 16, 18 and 14; the fourth group 17 and 29.

#### The final question: which ANN models should be used further?

If we consider SRD results only, without any prejudice about the models, the networks which could be recommended as acceptable ones are 14a, 15 and 7 (according to the row average as the reference ranking), 15, 4, 14a and 17 (according to the experimental IC<sub>50</sub> values as the reference

ranking) and 7a, 8 and 10 (according to the maximum and minimum row values of statistic parameters). Obviously, different reference ranking gives different ranking of models. Since the row average is suggested as the optimal choice for reference ranking, the networks 14a, 15 and 7 can be considered as the best ones. In addition, it must be emphasized that taking into account the calculated statistical parameters and SRD procedure, the networks 4 and 71 are also acceptable for application.

#### Conclusions

Sum of ranking differences approach was applied not only for ranking of the ANN models, but also for detection of

similarity or dissimilarity among them. Using SRD method the most similar models can be found easily. SRD analysis based on experimental values as reference ranking revealed similar grouping of the ANN models already obtained by CA. Cross-validation (sevenfolds repeated ranking) of SRD ranking gave uncertainties for each ANN model. The significant differences among the models were tested by sensitive Wilcoxon matched pair test and sign test.

According to row average as reference ranking, the SRD analysis revealed that the usage of the networks 10, 14, 18 and 20 should be avoided due to larger SRD values than the experimental values. The networks 4, 15 and 71, which were recommended by Kovačević et al. [19] as the best, are generally acceptable for application, so these two studies do not cancel each other out. This study takes the row average as the optimal choice for reference ranking, so the networks 14a, 15 and 7 can be considered as the best ones. SRD study certainly gives a new point of view for selection of the best QSAR models based on non-linear artificial neural networks modeling.

**Acknowledgments** This study was performed within the framework of the research projects No. 172012 and No. 172021 supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia and the project No. 114-451-347/2015-02, financially supported by the Provincial Secretariat for Science and Technological Development of Vojvodina.

## References

1. S.J. Ellem, G.P. Risbridger, *Nat. Rev. Cancer* **7**, 621 (2007)
2. E. Tenorio-Borroto, C.G. Peñuelas-Rivas, J.C. Vásquez-Chagoyán, N. Castañedo, F.J. Prado-Prado, X. García-Mera, H. González-Díaz, *Eur. J. Med. Chem.* **72**, 206 (2014)
3. S.Z. Kovačević, S.O. Podunavac Kuzmanović, L.R. Jevrić, *Acta Chim. Slov.* **60**, 756 (2013)
4. N. Minovski, A. Jezierska-Mazzarello, M. Vračko, T. Šolmajer, *Cent. Eur. J. Chem.* **9**, 855 (2011)
5. O. Farkas, J. Jakus, K. Héberger, *Molecules* **9**, 1079 (2004)
6. M. Rybka, A.G. Mercader, E.A. Castro, *Chemom. Intell. Lab. Syst.* **132**, 18 (2014)
7. S. Moonsamy, R.C. Dash, M.E. Soliman, *Molecules* **19**, 5243 (2014)
8. Z. Garkani-Nejad, N. Jalili-Jahani, *Cent. Eur. J. Chem.* **8**, 877–885 (2010)
9. M. Kompany-Zareh, M. Khoshkam, *J. Iran. Chem. Soc.* **9**, 269 (2012)
10. S.B. Gadzuric, S.O. Podunavac Kuzmanović, A.I. Jokic, M.B. Vranes, N. Ajdukovic, S.Z. Kovacevic, *Aust. J. Forensic Sci.* **46**, 166 (2014)
11. T.W. Schultz, M. Hewitt, T.I. Netzeva, M.T.D. Cronin, *QSAR Comb. Sci.* **26**, 238 (2007)
12. A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **20**, 269 (2002)
13. K. Héberger, *TRAC-Trends Anal. Chem.* **29**, 101 (2010)
14. K. Héberger, K. Kollár-Hunek, *J. Chemometr.* **25**, 151 (2011)
15. K. Kollár-Hunek, K. Héberger, *Chemom. Intell. Lab. Syst.* **127**, 139 (2013)
16. M. Vračko, N. Minovski, K. Héberger, *Acta Chim. Slov.* **57**, 586 (2010)
17. K. Héberger, B. Škrbić, *Anal. Chim. Acta* **716**, 92 (2012)
18. K. Héberger, S. Kolarević, M. Kračun-Kolarević, K. Sunjog, Z. Gačić, Z. Kljajić, M. Mitrić, B. Vuković-Gačić, *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **771**, 15 (2014)
19. S.Z. Kovačević, S.O. Podunavac-Kuzmanović, L.R. Jevrić, E.A. Djurendić, J.J. Ajduković, *Eur. J. Pharm. Sci.* **62**, 258 (2014)
20. A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, *J. Med. Chem.* **57**, 4977 (2014)
21. J.N. Miller, J.C. Miller, *Statistics and Chemometrics for Analytical Chemistry*, 6th edn. (Pearson Education Limited, Harlow, 2010), pp. 160–164
22. D.D. Wackerly, W. Mendenhall III, R.L. Scheaffer, *Mathematical Statistics with Applications*, 7th edn. (Brooks/Cole, Cengage Learning, Belmont, 2007), pp. 750–753