Interpreting Computational Neural Network QSAR Models: A Measure of Descriptor Importance

Rajarshi Guha and Peter C. Jurs*

Chemistry Department, The Pennsylvania State University, University Park, Pennsylvania 16802

Received January 21, 2005

We present a method to measure the relative importance of the descriptors present in a QSAR model developed with a computational neural network (CNN). The approach is based on a sensitivity analysis of the descriptors. We tested the method on three published data sets for which linear and CNN models were previously built. The original work reported interpretations for the linear models, and we compare the results of the new method to the importance of descriptors in the linear models as described by a PLS technique. The results indicate that the proposed method is able to rank descriptors such that important descriptors in the CNN model correspond to the important descriptors in the linear model.

INTRODUCTION

Computational neural networks (CNNs) are an important component of the QSAR practitioner's toolbox for a number of reasons. First, neural network models are generally more accurate than other classes of models. The higher predictive ability of CNNs arises from their flexibility—their ability to model nonlinear functions. Second, a variety of neural networks is available depending on the nature of the problem being studied. One may consider classes of neural networks depending on whether they are trained in a supervised manner (e.g., back-propagation networks) or in an unsupervised manner (e.g., Kohonen map[1]).

Neural network models also have a number of drawbacks. First, neural network models can be overtrained. Thus, it is frequently the case that a neural network model will have memorized the idiosyncrasies of the training set, essentially fitting the noise. As a result, when faced with a test set of new observations, such a model's predictive ability will be very poor. One way to alleviate this problem is the use of cross-validation and use of root-mean-square errors for the training, cross-validation, and prediction sets. A second drawback is the matter of interpretability. Neural networks are generally regarded as black boxes. That is, one provides the input values and obtains an output value, but generally no information is provided regarding how those output values were obtained or how the input values correlate to the output value. In the case of QSAR models, the lack of interpretability forces the use of CNN models as purely predictive tools rather than as an aid in the understanding of structure property trends. As a result, neural network models are very useful as a component of a screening process. But the possibility of using these models in a computationally guided structure design is low. This is in contrast to linear models, which can be interpreted in a simple manner.

A number of reports in the machine learning literature describe attempts to extract meaning from neural network models.[2−7] Many of these techniques attempt to capture the functional representation being modeled by the neural network. In a number of cases these methods require the use of specific neural network algorithms,[8,9] and so generalization can be difficult.

Interpretation can be considered in two forms, broad and detailed. The aim of a broad interpretation is to characterize how important an input neuron (or descriptor) is to the predictive ability of the model. This type of interpretation allows us to rank input descriptors in order of importance. However, in the case of a QSAR neural network model, this approach does not allow us to understand exactly how a specific input descriptor affects the network output (for a QSAR model, predicted activity or property). The goal of a detailed interpretation is to extract the structure−property trends from a CNN model. A detailed interpretation should be able to indicate how an input descriptor for a given input example correlates to the predicted value for that input. The aim of this work is to describe a method to provide a broad interpretation of a neural network model. A method to provide a detailed interpretation of a neural network model is described in a subsequent paper.

METHODOLOGY

In this work we restrict ourselves to the use of three-layer, fully connected, feed-forward computational neural networks. Furthermore, we consider only regression networks (as opposed to classification networks). The input layer represents the input descriptors, and the value of the output layer is the predicted property or activity. The neural network models considered in this work were built using the ADAPT[10,11] methodology which uses a genetic algorithm[12,13] to search a descriptor pool to find the best subset (of a specified size) of descriptors that results in a CNN model having a minimal cost function. The cost function is defined as

$$\text{cost} = \text{TSET}_{\text{RMSE}} + {}^1/_2|\text{TSET}_{\text{RMSE}} - \text{CVSET}_{\text{RMSE}}|$$

where $\text{TSET}_{\text{RMSE}}$ and $\text{CVSET}_{\text{RMSE}}$ represent the root-mean-square errors for the training and cross-validation sets, respectively.

10.1021/ci050022a CCC: $30.25
Published on Web 03/31/2005

The broad interpretation is essentially a sensitivity analysis of the neural network. This form of analysis has been described by So and Karplus.[14] However, the results of a sensitivity analysis have not been viewed as a measure of descriptor importance. It should also be pointed out that the idea behind sensitivity analysis is also the method by which a measure of descriptor importance is generated for random forest models.[15,16]

The algorithm we have developed to measure descriptor importance is as follows. To start, a neural network model is trained and validated. The RMSE for this model is denoted as the base RMSE. Next, the first input descriptor is randomly scrambled, and then the neural network model is used to predict the activity of the observations. Because the values of this descriptor have been scrambled, one would expect the correlation between descriptor values and activity values to be obscured. As a result, the RMSE for these new predictions should be larger than the base RMSE. The difference between this RMSE value and the base RMSE indicates the importance of the descriptor to the model's predictive ability. That is, if a descriptor plays a major role in the model's predictive ability, scrambling that descriptor will lead to a greater loss in predictive ability (as measured by the RMSE value) than for a descriptor that does not play such an important role in the model. This procedure is then repeated for all the descriptors present in the model. Finally, we can rank the descriptors in order of importance.

An alternative procedure was investigated; it consisted of replacing individual descriptors with random vectors. The elements of the random vectors were chosen from a normal distribution with mean and variance equal to that of the original descriptor. The RMSE values of the model with each descriptor replaced in turn by its random counterpart was recorded as described above. We did not notice any significant differences in the final ordering of the descriptors compared to the random scrambling experiments. In all cases the most important descriptor was the same. In two of the data sets the only difference occurred in the ordering of two or three of the least significant descriptors.

DATA SETS AND MODELS

To test this measure of descriptor importance we considered a number of data sets that have been studied in the literature. In two cases, linear regression models had been built and interpreted using the PLS scheme described by Stanton.[17] In all cases neural network models were also reported, but no form of interpretation was provided for these models, as they were developed primarily for their superior predictive ability.

We applied the descriptor importance methodology to these neural network models and compared the resultant rankings of the descriptors to the importance of the descriptors described by the PLS method for the linear models. It is important to note that the linear and CNN models do not necessarily contain the same descriptors (and indeed may have none in common). However, since both types of models should capture similar structure−property relationships present in the data, it is reasonable to expect that the descriptors used in the models should have similar interpretations. Due to the broad nature of the descriptor importance methodology, we do not expect a one-to-one correlation of interpretations of the linear and nonlinear models. However it does allow us to see which descriptors in a CNN model are playing the major role, and by comparison with the interpretations provided for the linear models allows us to confirm that this method is able to capture descriptor importance correctly.

Table 1. Summary of the Linear Regression Model Developed for the DIPPR Data Set

	estimate	std error	*t*	*P*
intercept	−215.092	29.451	−7.30	0.000
PNSA-3	−3.561	0.210	−16.92	0.000
RSHM-1	608.071	21.302	28.55	0.000
V4P-5	19.576	3.308	5.92	0.000
S4PC-12	12.089	1.572	7.69	0.000
MW	0.579	0.061	9.42	0.000
WTPT-2	236.108	16.574	14.25	0.000
DPHS-1	0.198	0.028	7.07	0.000

We considered three data sets. The first data set consisted of a 147-member subset of compounds whose measured activity was a normal boiling point. The whole data set contained 277 compounds, obtained from the Design Institute for Physical Property Data (DIPPR) Project 801 database, and it was studied by Goll and Jurs.[18] Although the original work reported linear as well as CNN regression models we generated new linear[19] and nonlinear models using the ADAPT methodology[20,21] and provide a brief interpretation for the linear case, using the PLS technique,[17] focusing on which descriptors are deemed important.

The second data set consisted of 179 artemisinin analogues. These were studied using CoMFA[22] by Avery et al.[23] The measured activity was the logarithm of the relative activity of the analogues (compared to the activity of artemisinin). We have previously used this data set to build linear and nonlinear models using a 2D-QSAR methodology and provided a PLS based interpretation for the linear model.[24]

The third data set consisted of a set of 79 PDGFR phosphorylation inhibitors described by Pandey et al.[25] The reported activity was $\log(IC_{50})$ and was obtained from a phosphorylation assay. This data set was studied by us,[26] and we reported linear and CNN regression models and also provided an interpretation for the former.

RESULTS

DIPPR Data Set. We first consider the linear and CNN models for the DIPPR data set for modeling boiling points. The statistics of the linear regression model for this data set are summarized in Table 1, and the meanings of the descriptors used in the model are summarized in Table 2. The R^2 value was 0.98, and the F-statistic was 1001 (for 7 and 139 degrees of freedom) which is much greater than the critical value of 2.076 ($\alpha = 0.05$). The model is thus statistically valid. The corresponding PLS analysis is summarized in Table 3. The PLS statistics indicate that the increase in q^2 beyond the fourth component is negligible. Thus, we may consider the most important descriptors in the first three components only. To see which descriptor is contributing the most in a given component, we consider the X weights obtained from the PLS analysis which are displayed in Table 4. In component 1 it is clear that MW and V4P-5 are the most heavily weighted descriptors. Higher values of molecular weight correspond to larger molecules

Table 2. Glossary of the Descriptors Reported in This Paper

descriptor code	description	ref
APAVG	the mean value of all unique atom pairs present in the molecule	36
DPHS-1	the difference between the hydrophobic and hydrophilic surface area	37
ELEC	electronegativity (0.5 × (HOMO + LUMO))	
FLEX-4	molecular mass of rotatable atoms divided by the hydrogen suppressed molecular weight	38
FNSA-3	charge weighted partial negative relative surface area	39
FPSA-3	atom weighted partial positive surface area divided by the total molecular surface area	39
KAPA-6	atom corrected shape index ($^3\kappa$)	40−42
MDE-12	molecular distance edge vector between primary and secondary carbons	43
MDE-13	molecular distance edge vector between primary and tertiary carbons	43
MDEN-23	molecular distance edge vector between secondary and tertiary nitrogens	43
MOLC-6	path of length 4 molecular connectivity index	44, 45
MOLC-8	path-cluster of length 4 molecular connectivity index	44, 45
MOMI-4	ratio of the principal component of the moment of inertia along the *X*-axis to that along the *Y*-axis	46
MW	molecular weight	
NDB-13	number of double bonds	
NSB	number of single bonds	
N5CH-12	number of fifth order chains	47−49
N7CH-20	number of seventh order chains	47−49
PNSA-3	charge weighted partial negative surface area	39
RNHS-3	a hydrophobic surface area descriptor defined as the product of the surface area of the most hydrophilic atom and the sum of the hydrophilic constants divided by the logp value for the molecule	37, 50
RPHS-1	product of the surface area of the most hydrophobic atom and its hydrophobic constant divided by the sum of all hydrophobic constants	37, 50
RSHM-1	fraction of the solvent accessible surface area associated with hydrogens that can be donated in a hydrogen-bonding intermolecular interaction	39
S4PC-12	fourth order simple path cluster	47−49
SURR-5	ratio of the atomic constant weighted hydrophobic (low) surface area and the atomic constant weighted hydrophilic surface area	37, 50
V4P-5	fourth order valence path molecular connectivity index	47−49
WPHS-3	surface weighted hydrophobic surface area	37, 50
WTPT-3	sum of all path lengths starting from heteroatoms	51
WTPT-5	sum of all path lengths starting from nitrogens	51

Table 3. Summary of the PLS Analysis Based on the Linear Regression Model Developed for the DIPPR Data Set

component	X variance	error SS	R^2	PRESS	q^2
1	0.431	94868.5	0.863	99647.6	0.857
2	0.660	26221.6	0.962	29046.7	0.958
3	0.768	16648.8	0.976	19303.3	0.972
4	0.843	14670.8	0.978	17027.6	0.975
5	0.911	14032.5	0.979	16281.3	0.976
6	0.987	13775.9	0.980	15870.6	0.977
7	1.000	13570.9	0.980	15653.0	0.977

Table 4. X-Weights for the PLS Components from the PLS Analysis Summarized in Table 2

	component						
descriptor	1	2	3	4	5	6	7
PNSA-3	−0.303	−0.423	0.202	−0.250	0.254	−0.737	−0.127
RSHM-1	0.190	0.779	0.347	−0.032	0.222	−0.377	0.209
V4P-5	0.485	−0.157	−0.071	−0.664	−0.368	−0.096	0.384
S4PC-12	0.289	−0.079	−0.578	0.531	−0.031	−0.469	0.264
MW	0.499	−0.085	0.368	0.242	−0.397	−0.170	−0.602
WTPT-2	0.483	−0.051	−0.265	−0.221	0.708	0.138	−0.350
DPHS1	0.263	−0.415	0.540	0.322	0.297	0.187	0.487

and thus elevated boiling points. The V4P-5 descriptor characterizes branching in the molecular structure, and higher values indicate a higher degree of branching. Thus, both of the most important descriptors in the first component correlate molecular size to higher values of boiling point. In the second component we see that the most weighted descriptors are RSHM-1 and PNSA-3. RSHM-1 characterizes the fraction of the solvent accessible surface area associated with hydrogens that can be donated in a hydrogen-bonding intermolecular interaction. PNSA-3 is the charge weighted partial negative surface area. Clearly, both these descriptors characterize the ability of molecules to form hydrogen bonds. In summary, the structure−property relationship captured by the linear model indicates that London forces dominate the relationship. Although individual atomic contributions to the trend are small, larger molecules will have more interactions leading to higher boiling points. In addition, attractive forces, originating from hydrogen bond formation, also play a role in the relationship and these are characterized in the second component of the PLS model. We can use the above discussion and information from the PLS analysis to rank the descriptors considered in the PLS analysis in decreasing order of contributions: MW, V4P-5, RSHM-1, and PNSA-3.

The next step was to develop a computational neural network model for this DIPPR data set. The ADAPT methodology was used to search for descriptor subsets ranging in size from 4 to 6. The final CNN model had a 5−3−1 architecture, and the statistics of the model are reported in Table 5. The descriptors in this model were FNSA-3, MOLC-6, WPHS-3, and RPHS-1, which are described in Table 2. The increase in RMSE values for the descriptors in each neural network model are reported in Tables 6−8. In each table the third column represents the increase in RMSE due to the scrambling of the corresponding descriptor over the base RMSE. It is evident that scrambling some descriptors leads to larger increases, whereas others lead to negligible increases in the RMSE. The information contained in these tables is more easily seen in the descriptor

Table 5. Summary Statistics for the Best CNN Model in the DIPPR Data Set[a]

	R^2	RMSE
TSET	0.98	9.92
CVSET	0.99	7.89
PSET	0.98	8.61

[a] The model architecture was 5−3−1.

Table 6. Increase in RMSE Due to Scrambling of Individual Descriptors[a]

	scrambled descriptor	RMSE	difference
1	FNSA-3	30.50	20.58
2	RSHM-1	35.76	25084
3	MOLC-6	51.32	41.39
4	WPHS-3	66.27	56.35
5	RPHS-1	35.75	25.83

[a] The CNN architecture was 5−3−1 and was built using the DIPPR data set. The base RMSE was 9.92.

Table 7. Increase in RMSE Due to Scrambling of Individual Descriptors[a]

	scrambled descriptor	RMSE	difference
1	N5CH-12	0.50	0.20
2	WTPT-3	0.49	0.19
3	WTPT-5	0.39	0.09
4	FLEX-4	0.51	0.21
5	RNHS-3	0.51	0.21
6	SURR-5	0.72	0.42
7	APAVG	0.48	0.18

[a] The CNN architecture was 7−3−1 and was built using the PDGFR data set. The base RMSE was 0.29.

Table 8. Increase in RMSE Due to Scrambling of Individual Descriptors[a]

	scrambled descriptor	RMSE	difference
1	KAPA-6	0.88	0.40
2	N7CH-20	1.97	1.49
3	MOLC-8	0.98	0.50
4	NDB-13	0.99	0.51
5	WTPT-5	0.78	0.30
6	MDE-12	0.85	0.37
7	MDE-13	1.18	0.70
8	MOMI-4	0.77	0.29
9	ELEC	0.93	0.45
10	FPSA-3	0.89	0.41

[a] The CNN architecture was 10−5−1 and was built using the artemisinin data set. The base RMSE was 0.49.

importance plots shown in Figures 1−3. These figures plot the increase in RMSE for each descriptor, in decreasing order.

Considering the DIPPR data set (Table 6 and Figure 1) we see that although the linear and nonlinear models have only one descriptor in common (RSHM-1) the types of descriptors are the same in the both models (topological and charge-related). The CNN model contains charged partial surface area descriptors (RSHM-1 and FNSA-3) as well as hydrophobicity descriptors (WPHS-3 and RPHS-1). One may expect that the structure−activity trends captured by the CNN model are similar to those captured by the linear model. If we look at Figure 1 we see that the most important descriptor is WPHS-3. This descriptor represents the surface weighted hydrophobic surface area. Since this is correlated to the positively charged surface area, this descriptor should characterize hydrogen-bonding ability. The second most important descriptor is MOLC-6, which represents a topological path containing four carbon atoms. This descriptor essentially characterizes molecular size. The next two most important descriptors are RSHM-1, which has been previously described, and RPHS-1, which characterizes the relative hydrophobic surface area. The insignificant separation between these two descriptors along the *X*-axis indicates that these two descriptors are probably playing similar roles in the predictive ability of the CNN model. When compared to the ranking of descriptor contributions in the PLS analysis we see that the CNN descriptor importance places a hydrophobicity descriptor as the most important descriptor, followed by MOLC-6 which, as mentioned, characterizes molecular size. The difference in ordering may be due to the fact that the CNN is able to find a better correlation between the selected descriptors, such that WPHS-3 provides the maximum amount of information. However, in general, the broad interpretation provided by this method does compare well with that of the linear model using PLS.

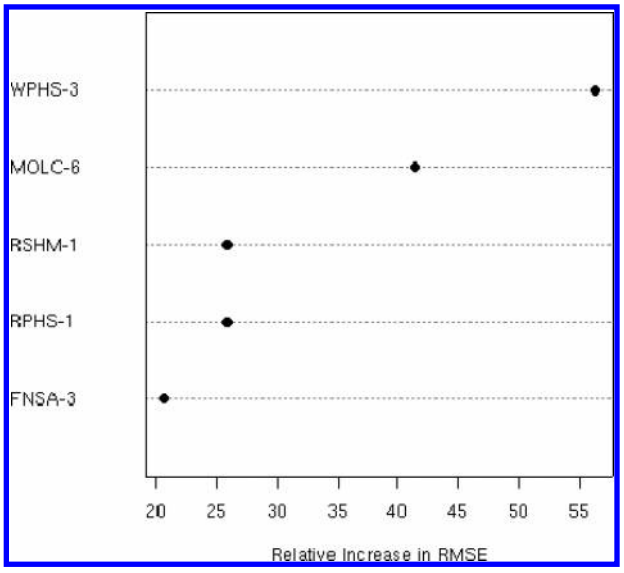

Figure 1. Importance plot for the 5−3−1 CNN model built using the DIPPR data set.

The PDFGR Data Set. We now consider the PDGFR data set. The original work reported a three-descriptor linear regression model. The PLS interpretation of this model indicated that all three descriptors were important. These descriptors were SURR-5, RNHS-3, and MDEN-23. The first two descriptors are hydrophobic surface area descriptors, and the last descriptor is a topological descriptor which represents the geometric mean of the topological path length between secondary and tertiary nitrogens. If we take into account the PLS components in which these descriptors occur, they may be ordered as SURR-5, MDEN-23, and RNHS-3 (decreasing importance). The reported CNN model for this data set contained 7 descriptors: N5CH, WTPT-3, WTPT4, FLEX-4, RNHS-3, SURR-5, and APAVG. A summary of these descriptors may be found in Table 2, and further details are given in the original work.[26] As can be seen, the linear regression and CNN models have two descriptors in common: SURR-5 and RNHS-3.

The descriptor importance plot for the PDFGR data set (Figure 2) shows the most important descriptor to be SURR-

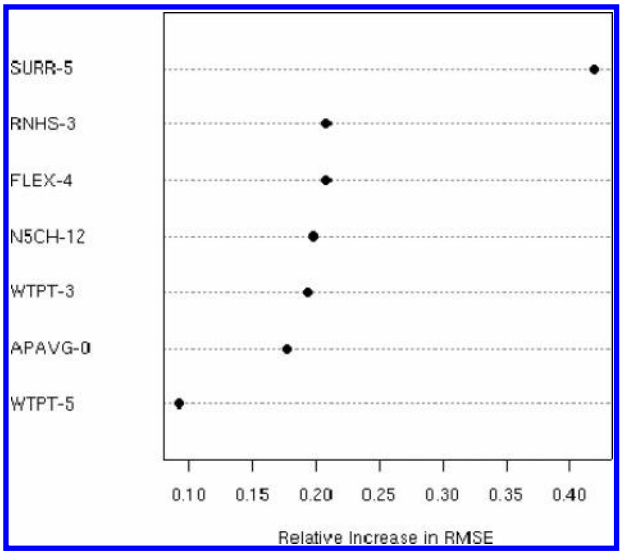

Figure 2. Importance plot for the 7−3−1 CNN model built using the PDGFR data set.

5. The next most important descriptor is RNHS-3. The position of RNHS-3 on the plot indicates that it plays a much less important role than SURR-5 in the model's predictive ability. However, it is notable that the two most important descriptors identified by our method in the CNN model are also the two most important descriptors identified by the PLS analysis of the linear regression model. Given that the CNN model and linear model should capture similar structure property trends in the data set, the similarity between the important descriptors in the two models serve to confirm the validity of this method. It is also interesting to note that FLEX-4, N5CH-12, and WTPT-3 descriptors are relatively closely spaced along the *X*-axis. Though these descriptors are ranked, the relatively small increase between each one might be indicative that these descriptors are performing similar roles within the network. The position of WTPT-5 in the plot indicates that it contributes relatively little to the model's predictive ability.

The Artemisinin Data Set. Finally, we consider the artemisinin data set. The original work[24] presented a 4-descriptor linear regression model. The most important descriptors identified by a PLS analysis of this model were N7CH, WTPT-2, and NSB. N7CH and WTPT-2 are topological descriptors. N7CH is the number of seventh order chains, and WTPT-2 is the weighted path descriptor divided by the total number of atoms. NSB is the count of single bonds. Summaries of the descriptors are presented in Table 2, and further details of these descriptors can be found in the original work.[24] Taking into account the PLS components in which these appear, we may order them as follows: N7CH, NSB, and WTPT-2 (decreasing order of importance). The CNN model reported in the original work had 10 descriptors: KAPPA-6, NDB, MOMI-4, N7CH, MOLC-8, WTPT-5, MDE-12, MDE-13, ELEC, and FPSA-3. Brief summaries of the descriptors can be found in Table 2. In this case, the linear model and the CNN model have only one descriptor in common (N7CH), though both models contain weighted path descriptors (WTPT-2 for the linear model and WTPT-5 for the CNN model). The CNN model also contains a number of topological and geometric descriptors (KAPPA-6, MOLC-8, MDE-12, MDE-13, and MOMI-4) as well as two electronic descriptors (ELEC and FPSA-3). When compared

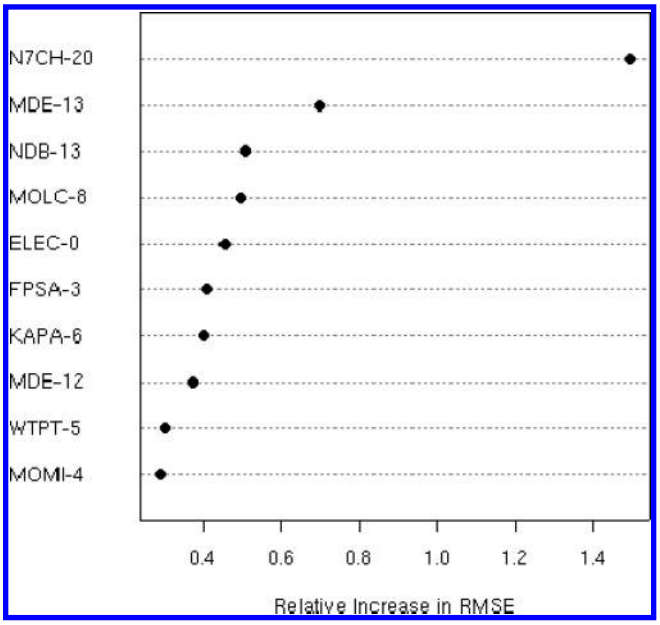

Figure 3. Importance plot for the 10−5−1 CNN model built using the artemisinin data set.

to the linear model, which contained only topological descriptors, it appears that the CNN model has been able to capture a more detailed view of the structure−property relationship by including information from electronic descriptors.

If we now consider the descriptor importance plot in Figure 3 we see that the most important descriptor is N7CH. Its contribution to the model's predictive ability is clearly significant. In contrast, the next most significant descriptor, MDE-13, is much further left than N7CH. MDE-13 is a distance edge descriptor (the number of paths between primary and tertiary carbons) and characterizes molecular size. In this sense, it is similar in nature to the count of single bonds (the second most important descriptor in the linear model). Once again we see that a number of descriptors are relatively closely spaced along the *X*-axis (such as NDB, MOLC-8, FPSA-3, and KAPA-6). It is interesting to note that the electronic descriptors are approximately in the middle of the ranking, whereas the top four descriptors are mainly topological in nature. One may conclude that electronic factors do not play a major role in the structure−property relationship captured by this CNN model but do enhance the predictive ability of the model when compared to the performance of the linear model. It is also interesting to see that the WTPT-5 is ranked quite low in the CNN model and is the least important descriptor in the linear model. As before, if we assume that the linear and nonlinear models capture similar structure−property relationships we see that the CNN descriptor importance method ranks the descriptors such that their order of importance is similar to that in the linear model.

CONCLUSIONS

From the preceding results, we see that the proposed measure of importance is able to characterize the relative importance of descriptors. The resultant ranking of descriptors corresponds well to a ranking of descriptors obtained by the PLS analysis of a linear model using the same data set. The comparison with the linear models is not necessarily

one-to-one since the descriptors in the linear and CNN models will generally be different.

The representation of descriptor importance plots allows easy visual analysis of the descriptors in the model. In addition, apart from merely ranking, the plots provide a qualitative view of how important a given descriptor is relative to others. That is, by looking at the separation between two descriptors on the *X*-axis, one may determine that a certain descriptor plays a much more significant role than another in the model's predictive power. We conjecture that descriptors with little separation along the *X*-axis play similar roles within the CNN architecture. However, confirmation of this requires a more detailed interpretation of the CNN model, which we present in a subsequent paper.

It is clear that the proposed measure of descriptor importance does not allow the user to elucidate exactly how the model represents the information captured by a given descriptor. That is, the methodology does not indicate the sign (or direction) of the effect of each input descriptor. Thus we cannot draw conclusions regarding the nature of the correlation between input descriptors and network output. As mentioned previously, a methodology for the detailed interpretation of a neural network is the subject of a subsequent paper. However, even in the absence of a detailed understanding of the behavior of the input descriptors, the method described here allows the user to determine the most important descriptor (or descriptors) and investigate whether replacement by similar descriptors might lead to an improvement in the model's predictive ability. We investigated this possibility by replacing the SURR-5 descriptor from the 10−5−1 CNN model for the artemisinin data set by other hydrophobic surface area descriptors. In some cases the RMSE of the resultant model did increase, though not significantly. However in a few cases the RMSE of the model decreased, compared to the original RMSE. This is not surprising, as the importance plots show that theoretically related descriptors (such as WTPT-3 and WTPT-5 in Figure 2) may have significantly different contributions. However, the importance measure does allow us to change model descriptors in a guided manner. This procedure is functionally similar to the feature selection, but the main difference is that it is applied to a subset of descriptors that have already been deemed 'good' by a feature selection algorithm (genetic algorithm[13] or simulated annealing[27]). Hence, the result of the 'tweaking' procedure described here is akin to locally fine-tuning a given descriptor subset, rather than selecting whole new subsets.

Finally, we note that the machine learning literature describes a number of approaches to interpretation of CNN models. In general, these methods are closely tied to specific types of neural network models.[8,9] A useful feature of this interpretation methodology is that it is quite general in nature. That is, the methodology is not dependent on the specific characteristics of a neural network and depends only on the input data. This implies that the methodology can be applied to obtain descriptor importance measures for any type of neural network model such as single layer or multilayer perceptrons[28] or radial basis networks.[29,30] Furthermore many interpretation methods are focused on extracting rules or analytical forms of the CNN's internal model.[31−33] In many cases, this necessitates a complex analysis of the model utilizing another pattern recognition[34] or optimization algorithm.[33,35] In contrast, the method described here is simple to carry out and provides easily understandable conclusions.

In summary, we have presented an interpretation methodology that provides a broad view of descriptor importance for a neural network model, thus alleviating the black box nature of the neural network methodology to some extent.

ACKNOWLEDGMENT

We would like to thank Dr. David Stanton for providing the linear regression model for the DIPPR data set.

REFERENCES AND NOTES

(1) Kohonen, T. *Self-organizing maps*; Springer: 1994; Vol. 30.
(2) Ney, H. On the probabilistic interpretation of neural network classifiers and discriminative training criteria. *IEEE Trans. Pat. Anal. Mach. Intell.* **1995**, *17*, 107−119.
(3) Jones, W. T.; Vachha, R. K.; Kulshrestha, A. P. DENDRITE: a system for visual interpretation of neural network data. In *IEEE Proc. Southeastcon*; 1992.
(4) Takahashi, T. An information theoretical interpretation of neuronal activities. In *Ijcnn-91-seattle international joint conference on neural networks*; 1991.
(5) Castro, J. L.; Mantas, C. J.; Benitez, J. M. Interpretation of artificial neural networks by means of fuzzy rules. *IEEE Trans. Neural Networks* **2002**, *13*, 101−116.
(6) Limin, F. Rule generation from neural networks. *IEEE Trans. Sys., Man, Cybern.* **1994**, *24*, 1114−1124.
(7) Taha, I. A.; Ghosh, J. Symbolic interpretation of artificial neural networks. *IEEE Trans. Knowl. Data Eng.* **1999**, *11*, 448−463.
(8) Hervas, C.; Silva, M.; Serrano, J. M.; Orejuela, E. Heuristic extraction of rules in pruned artificial neural network models used for quantifying highly overlapping chromatographic peaks. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1576−1584.
(9) Bologna, G. Rule extraction from linear combinations of DIMLP neural networks. In *Proceedings of the Sixth Brazilian Symposium on Neural Networks*; 2000.
(10) Jurs, P. C.; Chou, J. T.; Yuan, M. Studies of chemical structure biological activity relations using patter recognition. In *Computer assisted drug design*; Olsen, E. C., Christoffersen, R. E., Eds.; American Chemical Society: Washington, DC, 1979.
(11) Stuper, A. J.; Brugger, W. E.; Jurs, P. C. *Computer assisted studies of chemical structure and biological function*; Wiley: New York, 1979.
(12) Wessel, M. D. Computer assisted development of quantitative structure − property relationships and design of feature selection routines. Ph.D. Chemistry, Pennsylvania State University, University Park, 1997.
(13) Goldberg, D. E. *Genetic algorithms in search optimization & machine learning*; Addison-Wesley: Reading, MA, 2000.
(14) So, S.-S.; Karplus, M. Evolutionary optimization in quantitative structure−activity relationship: an application of genetic neural networks. *J. Med. Chem.* **1996**, *39*, 1521−1530.
(15) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification and regression trees*; Wadsworth: 1984.
(16) Breiman, L. Random forests. *Machine Learning* **2001**, *45*, 5−32.
(17) Stanton, D. T. On the physical interpretation of QSAR models. *J. Chem. Inf. Comput Sci.* **2003**, *43*, 1423−1433.
(18) Goll, E. S.; Jurs, P. C. Prediction of the normal boiling points of organic compounds from molecular structures with a computational neural network model. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 974−983.
(19) Stanton, D. personal communication.
(20) Wessel, M. D.; Jurs, P. C. Prediction of reduced ion mobility constants from structural information using multiple linear regression analysis and computational neural networks. *Anal. Chem.* **1994**, *66*, 2480−2487.
(21) Lu, X.; Ball, J. W.; Dixon, S. L.; Jurs, P. C. Quantitative structure−activity relationships for toxicity of phenols using regression analysis and computational neural networks. *Environ. Toxicol. Chem.* **1994**, *13*, 841−851.
(22) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (COMFA). I. effect of shape on binding of steroids to carrier protiens. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(23) Avery, M. A.; Alvim-Gaston, M.; Rodrigues, C. R.; Barreiro, E. J.; Cohen, F. E.; Sabnis, Y. A.; Woolfrey, J. R. Structure activity relationships of the antimalarial agent artemisinin. the development of predictive in vitro potency models using COMFA and HQSAR methodologies. *J. Med. Chem.* **2002**, *45*, 292−303.
(24) Guha, R.; Jurs, P. C. The development of QSAR models to predict and interpret the biological activity of artemisinin analogues. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1440−1449.

(25) Pandey, A.; Volkots, D. L.; Seroogy, J. M.; Rose, J. W.; Yu, J.-C.; Lambing, J. L.; Hutchaleelaha, A.; Hollenbach, S. J.; Abe, K.; Giese, N. A.; Scarborough, R. M. Identification of orally active, potent, and selective 4-piperazinylquinazolines as antagonists of the platelet-derived growth factor receptor tyrosine kinase family. *J. Med. Chem.* **2002**, *45*, 3772−3793.
(26) Guha, R.; Jurs, P. C. The development of linear, ensemble and nonlinear models for the prediction and interpretation of the biological activity of a set of PDGFR inhibitors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2179−2189.
(27) Sutter, J. M.; Dixon, S. L.; Jurs, P. C. Automated descriptor selection for quantitative structure−activity relationships using generalized simulated annealing. *J. Chem. Inf. Comput Sci.* **1995**, *35*, 77−84.
(28) Haykin, S. *Neural Networks*; Pearson Education: Singapore, 2001.
(29) Patankar, S. J.; Jurs, P. C. Prediction of Glycine/NMDA Receptor Antagonist Inhibition from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1053−1068.
(30) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press: 1995.
(31) Ishibuchi, H.; Nii, M.; Tanaka, K. Fuzzy-arithmetic-based approach for extracting positive and negative linguistic rules from trained neural networks. In *Fuzzy systems conference proceedings, ieee international*; 1999.
(32) Gupta, A.; Park, S.; Lam, S. M. Generalized analytic rule extraction for feedforward neural networks. *IEEE Trans. Knowledge Data Eng.* **1999**, *11*, 985−991.
(33) Fu, X.; Wang, L. Rule extraction by genetic algorithms based on a simplified rbf neural network. In *Evolutionary computation, proceedings of the 2001 congress on*; 2001.
(34) Chen, P. C. Y.; Mills, J. K. Modeling of neural networks in feedback systems using describing functions. In *Neural networks, international conference on*; 1997.
(35) Yao, S.; Wei, C.; He, Z. Evolving fuzzy neural networks for extracting rules. In *Fuzzy systems, proceedings of the fifth ieee international conference on*; 1996.
(36) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure−activity studies: definition and application. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64−73.
(37) Stanton, D. T.; Mattioni, B. E.; Knittel, J. J.; Jurs, P. C. Development and use of hydrophobic surface area (HSA) descriptors for computer assisted quantitative structure−activity and structure−property relationships. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1010−1023.
(38) Mosier, P. D.; Counterman, A. E.; Jurs, P. C.; Clemmer, D. E. Prediction of peptide ion collision cross sections from topological molecular structure and amino acid parameters. *Anal. Chem.* **2002**, *74*, 1360−1370.
(39) Stanton, D. T.; Jurs, P. C. Development and use of charged partial surface area structural descriptors in computer assissted quantitative structure property relationship studies. *Anal. Chem.* **1990**, *62*, 2323−2329.
(40) Kier, L. B. A shape index from molecular graphs. *Quant. Struct.-Act. Relat. Pharmacol., Chem. Biol.* **1985**, *4*, 109−116.
(41) Kier, L. B. Shape indexes for orders one and three from molecular graphs. *Quant. Struct.-Act. Relat. Pharmacol., Chem. Biol*. **1986**, *5*, 1−7.
(42) Kier, L. B. Distinguishing atom differences in a molecular graph index. *Quant. Struct.-Act. Relat. Pharmacol., Chem. Biol.* **1986**, *5*, 7−12.
(43) Liu, S.; Cao, C.; Li, Z. Approach to estimation and prediction for normal boiling point (NBP) of alkanes based on a novel molecular distance edge (MDE) vector, lambda. *J. Chem. Inf. Comput Sci.* **1998**, *38*, 387−394.
(44) Balaban, A. T. Higly discriminating distance based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.
(45) Kier, L. B.; Hall, L. H. *Molecular connectivity in chemistry and drug research*; Academic Press: New York, 1976.
(46) Goldstein, H. *Classical mechanics*; Addison-Wesley: Reading, MA, 1950.
(47) Kier, L. B.; Hall, L. H.; Murray, W. J. Molecular connectivity I: relationship to local anesthesia. *J. Pharm. Sci.* **1975**, *64*.
(48) Kier, L. B.; Hall, L. H. *Molecular connectivity in structure activity analysis*; John Wiley & Sons: 1986.
(49) Kier, L. B.; Hall, L. H. Molecular connectivity VII: specific treatment to heteroatoms. *J. Pharm. Sci.* **1976**, *65*, 1806−1809.
(50) Mattioni, B. E. The development of quantitative structure−activity relationship models for physical property and biological activity prediction of organic compounds. Ph.D. Chemistry, Pennsylvania State University, University Park, 2003.
(51) Randic, M. On molecular idenitification numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164−175.

CI050022A