Full Paper

www.molinf.com

molecular informatics

SPECIAL ISSUE

DOI: 10.1002/minf.201600118

Performance of Deep and Shallow Neural Networks, the Universal Approximation Theorem, Activity Cliffs, and QSAR

David A. Winkler*[a, b, c, d] and Tu C. Le[a]

Abstract: Neural networks have generated valuable Quantitative Structure-Activity/Property Relationships (QSAR/QSPR) models for a wide variety of small molecules and materials properties. They have grown in sophistication and many of their initial problems have been overcome by modern mathematical techniques. QSAR studies have almost always used so-called "shallow" neural networks in which there is a single hidden layer between the input and output layers. Recently, a new and potentially paradigm-shifting type of neural network based on Deep Learning has appeared. Deep learning methods have generated impressive improvements in image and voice recognition, and are now being applied to QSAR and QSAR modelling. This paper describes the differences in approach between deep and shallow neural networks, compares their abilities to predict the properties of test sets for 15 large drug data sets (the KAGGLE set), discusses the results in terms of the Universal Approximation theorem for neural networks, and describes how DNN may ameliorate or remove troublesome "activity cliffs" in QSAR data sets.

Keywords: deep learning · deep neural network · shallow neural network · Bayesian regularized neural network · universal approximation theorem · activity cliff

1 Introduction

Neural networks have become an important component in the toolkit of quantitative structure-activity relationships (QSAR) practitioners because of their ability to generate robust models of the chemical structure–biological activity relationships. They are one element of a suite of machine learning and statistical regression methods that aim to make robust and quantitative predictions of properties for new molecules, when trained on properties for a set of existing data. They are well matched to the increasing number and size of data sets generated by automated chemical synthesis and high throughput screening technologies. Although originating in the drug design arena, their use has been very successfully adapted to model and predict properties of materials generally.[1] The linear free energy relationships origin of QSAR models has also substantially broadened to include chemically diverse, noncongeneric molecules and biological activities and properties that are clearly not directly free energy related.[2] A recent paper has elucidated the main uses to which QSAR models, particularly those generated from large diverse data set using machine learning methods, are now applied.[2] Another recent review has highlighted advances in the application of neural networks to QSAR and quantitative structure-property relationships (QSPR) modeling, and indicated the most significant advances in the field that will impact on future QSAR studies.[3]

QSAR modelling attempts to fit a multidimensional response surface to data that is often sparsely sampled. The response surface describes the relationship between the molecular features of members of the training set and the biological or other types of responses or properties that the members generate. Response surfaces have been of interest since the beginnings of QSAR and from the early work of Kauffman and Eigen, who viewed them as fitness landscapes that could be explored by evolutionary methods,[4] a concept that is now beginning to have practical value given the ability to generate large data sets using automated experiments.[5–6]

One of these new developments that has had very important impact in areas other than QSAR/QSPR is deep learning. This machine learning method employs a neural network with more than two hidden layers and a very large number of hidden layer nodes (many thousands). Although

[a] *D. A. Winkler, T. C. Le*
CSIRO Manufacturing, Clayton 3168, Australia
phone +61 3 9545 2477; fax: +61 3 9545 2446
**e-mail: dave.winkler@csiro.au*

[b] *D. A. Winkler*
Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Australia

[c] *D. A. Winkler*
Latrobe Institute for Molecular Science, Latrobe University, Bundoora 3082, Australia

[d] *D. A. Winkler*
School of Chemical and Physical Science, Flinders University, Bedford Park 5042, Australia

multilayer neural networks were explored in early QSAR studies, they were found to offer no advantages over networks with a single hidden layer and were largely abandoned.[7] The large number of adjustable weights relative to a single hidden layer neural network meant that they were more difficult to train and could easily be overfitted, and the chain rule that is used to adjust the weights in back propagation algorithms was unreliable because derivatives tended to become smaller as you moved from the output layer towards the input layer.[3] A recent paper by LeCun and Hinton[8] revived interest in deep neural networks (DNN, those with several hidden layers containing a large number of nodes) compared to shallow neural networks (those with a single hidden layer with a small number of nodes). The main difference between deep neural networks and earlier multilayer neural networks was that way the majority of weights in the hidden layer nodes were initialized and updated, the nature of the transfer function they contained, and the method of regularizing the network so that overfitting was minimized.[8] Many reviews are now appearing on the application of deep learning to drug discovery, computational biology, and QSAR generally.[9–12]

Paradoxically, the Universal Approximation theorem states that QSAR/QSPR models generated by shallow neural networks should be as capable of predicting the properties of new compounds as well as those derived from deep neural networks, although experimental error in the data may play a role in prediction accuracy as well. The purpose of this paper is to explore the implications of the Universal Approximation theorem and to evaluate, using large, 'real world' drug data sets, whether deep neural networks do indeed make better predictions than shallow neural networks. We also discuss the potential for deep neural networks to ameliorate the effects of activity cliffs that are sometimes problematic in drug design.

1.1 New Developments/Applications

Shallow neural network (previously just called neural networks) are versatile machine learning algorithms for modelling relationships between structure and properties or biological activities in complex data.[13–15] In QSAR applications they generally consist of a three-layered architecture containing an input layer (used to input molecular descriptors), a hidden layer containing a relatively small number of nodes, and a single output layer often consisting of a single node. The hidden layer nodes are fully connected to the nodes in the input and output layers by weights and, for regression, only the hidden layer nodes contain nonlinear transfer functions (Figure 1). The network generates an output by receiving descriptors for a given molecule at the input layer, distributing them to the hidden layer nodes via the weights, the weights at each hidden layer node are summed and presented to the transfer function to generate a non-linearly transformed version of the sum. The hidden layer node outputs are then distributed to the output layer nodes where they are summed and used to generate the output that is compared to the measured value of the property being modelled. In many cases, the backpropagation algorithm is used to move the residual error between the known and predicted output back to the input, using the chain rule to adjust the weights. After the training data are presented many times and the weights adjusted, the network learns the relationship between the molecular structure, as represented by the descriptors, and the property being modelled.

Deep learning is a branch of machine learning that models complex data using high-level representations of input descriptors using a neural network with multiple processing (hidden) layers composed of multiple linear and non-linear transfer functions, all (or most) of which are subject to learning (Figure 1). Each hidden layer transforms its input to increase both the selectivity and the invariance of the input descriptors (representation).[8] As they contain a very large number of weights compared to shallow

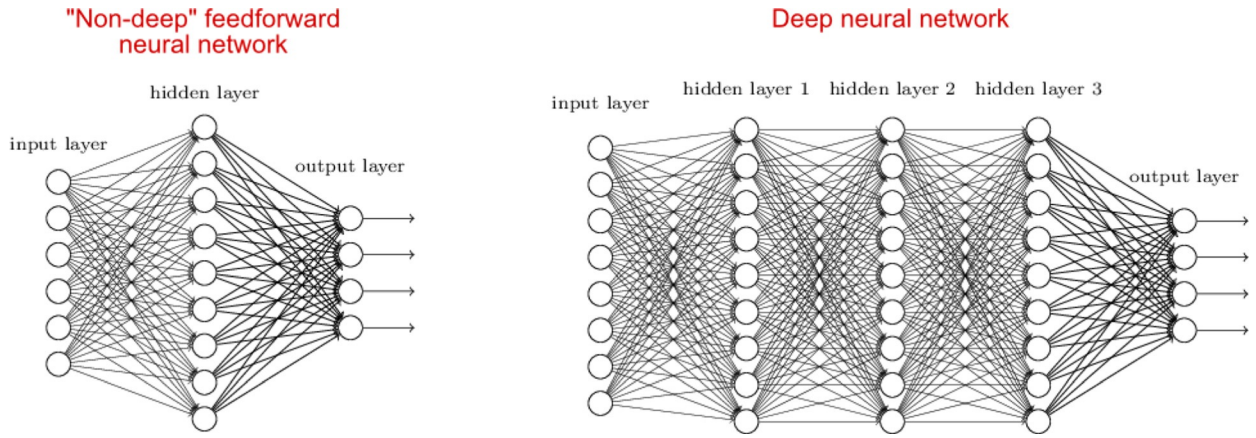

Figure 1. Fundamental differences in architecture of shallow (left) and deep (right) neural networks. (Adapted from Nielsen[16] under Creative Commons Attribution-NonCommercial 3.0 Unported License)

neural networks, they are very prone to overfitting if the weights are all adjusted during training. This is overcome by using a linear rectifier transfer function[17] $f(x) = \max(0,x)$ which imposes sparsity on the network, the use of L1 regularizer, $\lambda\Sigma|w_i|$, on the hidden layer nodes (penalizing the sum of the modulus of the weights, w_i), and by the use of a novel regularization method, random node drop out.[18]

Like shallow neural networks, DNNs model complex non-linear relationships. The additional hidden layers, and the additional nodes they contain, combine and transform features from input and lower hidden layers, potentially allowing modelling of complex data better than a similarly performing shallow network. This may translate into models with higher and more robust predictivity. The hidden layers essentially distort the input descriptors in a non-linear way so that, for instance, categories become linearly separable by the last layer.[8] A similar process is exploited by Support Vector Machines[19] and Relevance Vector Machines[20] to transform a linearly inseparable space into a higher dimensional, linearly separable space (Figure 2).

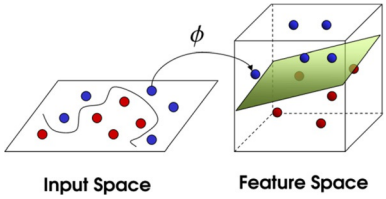

Figure 2. Using higher dimensional representations of descriptors to convert a linearly inseparable problem into a linearly separable one.

1.2 Universal Approximation Theorem

A very important theorem, not widely known in the drug and materials design community, is the Universal Approximation theorem.[21] With respect to artificial neural networks, the universal approximation theorem states that a feed-forward neural network with a *single hidden layer* containing a *finite* number of neurons, can approximate *any continuous function* on compact (closed and bounded) subsets of n-dimensional Euclidian space, under mild assumptions on the activation function. It is a corollary of the famous Kolmogorov theorem that states that any multivariate continuous function can be represented by the superimposition of a small number of univariate continuous functions. However, this theoretically important possibility theorem, does not and cannot explain the successes achieved by shallow neural networks.[22]

The theorem was first proven for neural networks by George Cybenko[23] for sigmoid activation functions, and the proof was extended to any specific activation function by Hornik.[22] He also showed that it was the multilayer feedforward architecture itself that allows neural networks to be universal approximators. Although the proof was only for the single output node (linear activation function) it can easily be generalized to more than one output. Indeed, Kůrková extended this to two layer neural networks and used Kolmogorov's theorem to estimate the number of hidden layer nodes required based on the properties of the response surface function being approximated.[24]

1.2.1 Are Descriptor Spaces Euclidian or Euclidian-like?

Euclidean space is an abstraction not dependent on specific reference frames, units, etc. Euclidian distance is a mathematical space where axes are not required to be orthogonal. A Euclidean plane is a two-dimensional real vector space equipped with an inner product, and can be similarly generalized to higher dimension objects and spaces. One of the basic tenets of Euclidean geometry is that two objects of the plane should be considered equivalent (congruent) if one can be transformed into the other by some sequence of translations, rotations and reflections. We do a similar transformation of axes in QSAR when we do principal components analysis or partial least squares modelling.

1.2.2 Are QSAR SAR or SPR Response Surface Continuous and Differentiable Functions?

Case for yes: The success of the QSAR method, with many thousands of published studies, suggests that response surfaces (the multidimensional surface describing the relationship between descriptors and biological activity) is complex but continuous and differentiable. Published studies show that many QSAR/QSPR response surfaces are linear and those that are nonlinear are not highly so.[1,25–27] A plethora of nonlinear mapping algorithms, including neural networks,[13–15,28] random forests,[29] support and relevance vector machines,[20] are capable of generating models with good to excellent predictivity of training and test sets.

Case for no: The apparent smoothness of such response surface may be significantly dependent on the choice of descriptors used to model the relationship between structures and activity. For example, if a substituent makes an important salt bridge with a protein target, and the descriptors employed do not capture the presence of this substituent or its physicochemical properties, an apparent discontinuity in the QSAR response surface may appear. Clearly, poor choice of descriptors, where critical molecular properties are not captured, will lead to poor models. Response surface discontinuities correspond to the so-called "activity cliffs", an important issue for drug design and optimization. Activity cliffs occur when small changes to active drug molecules can cause large changes in biological activity. Maggiora suggested that the existence of activity cliffs has several important implications for QSAR modeling.[30] First, linear models QSAR models are unlikely to satisfactorily account for significantly nonlinear or complex activity landscapes with significant numbers of cliffs. More importantly, the presence of activity cliffs in activity surfaces is

also a function of the molecular representation used, and different representations can lead to dramatic changes in the nature of activity cliffs or their disappearance entirely.[30–31]

It is not entirely clear whether activity cliffs are due largely or entirely to inherent discontinuity or lack of differentiability in the activity surface, or are artefacts of the types of descriptors used (e.g. integer or class membership versus continuous) or the mapping method. It has been proposed that activity cliffs may become less frequent, or may disappear if higher dimensional composite descriptors were used.[3,32]

We also expect that the ability of deep learning to create multiple levels of data representations with different complexity could provide fundamentally new ways of analyzing structure-activity relationships and solving the problems of great importance for drug discovery, such as the problem of activity cliffs.[32] Indeed, very rugged and bumpy activity landscapes with numerous activity cliffs with respect to input descriptors using low-level representations might appear to be very smooth and simple when transformed to high-level representations in deep-learning systems; this is the essence of representation learning. The ability of metric learning, a kind of linear representation learning, to eliminate activity cliffs in activity landscapes has recently been demonstrated.[33] One can expect that non-linear representation learning provided by neural networks should give an even greater effect.

The bottom line is that most surfaces may be differentiable and a few are not for a given set of descriptors. Non-differentiable surface may be made smoother and differentiable by high dimensional transformations inherent in deep learning methods. However, a fundamental question is – does the universal approximation theorem mean the deep neural networks can never be significantly better than shallow neural networks if the conditions are met?

1.3 Comparing DNN and State-of-the-art BNN Using Test Cases

This essential question is difficult to answer at this early stage of the application of deep learning methods to QSAR and QSPR. As few DNN QSAR studies have been reported so far, and comparisons of deep and shallow machine learning methods should ideally be done with a relatively large number of 'real world' data sets containing large numbers of compounds with high molecular diversity. To this end, the Merck Molecular Activity KAGGLE competition data set was well suited to such a comparison.[34] The competition allowed a wide range of QSAR modelling approaches, including DNN, to be tested with a consistent, fixed set of molecular descriptors. Differences in predictive performance were therefore due to the abilities of the different algorithms and not descriptors. A recent paper that also compared the performance of structure-activity modeling algorithms and sets of descriptors concluded that descriptor choice had much larger effect on prediction accuracy than the modeling algorithm used.[35] Ma et al.[36] showed that DNN generated test set predictions that were superior to those of the current state-of-the-art method, random forests (RF). In this paper we compared the performance of DNN against a state-of-the-art shallow neural network, the Bayesian regularized neural network (BNN), proven to give good predictivity and robust performance for QSAR and QSPR.

2 Methods

2.1 Data Sets

The KAGGLE data sets (https://www.kaggle.com/c/MerckActivity)[34] surprisingly did not provide the same set of descriptors for the training and test sets for each data set. Consequently, we selected the set of descriptors common to training and test sets for each data set, so that the test set values could be predicted from the models. Although the study of Ma et al.[36] compared QSAR modelling methods using only the test set r^2 value, this is not the most robust metric to use. A recent paper by Alexander et al.[37] showed that measures of dispersion like the root-mean-square (RMS) error or the standard error of prediction (SEP) for the test set are preferred, as they do not depend on the size of the data set or number of parameters in the model like r^2. Consequently, we used SEP to compare the Merck KAGGLE data set prediction performance of BNN and DNN with recommended parameter values (dispersion measures for the tests sets kindly provided by R. Sheridan, private communication).

2.2 Modelling Methods

The Bayesian regularized neural network consisted of a three layer feed forward neural network as has been described in previous publications.[15,28,38–42] Sparse feature selection was used to remove low relevance descriptors for the model to improve prediction performance. The input layer had the same number of nodes as the sparse set of descriptors chosen by feature selection. The number of nodes in the single hidden layer was between 2 and 5 as it has been shown that Bayesian regularized neural networks asymptote to a fixed number of effective adjustable parameters when more than the minimal number of nodes are used. The output layer contained a single node that provided the predicted dependent variable value for the model. The hidden layer nodes used sigmoidal transfer functions, the input and output nodes used linear transfer functions, and the training and test set data were scaled to zero mean and normalized using the standard deviation of each descriptor. The training of models used the gradient descent algorithm. The Bayesian regularization employed a Gaussian prior for larger data sets and a Laplacian prior

for smaller data sets. The network was trained on the KAGGLE training sets and the models were used to predict the properties of the KAGGLE test sets. As described, a consensus set of descriptors was used for the training and test sets by removing descriptors not common to both sets for each data set.

3 Results and Discussion

The SEP values for the DNN reported by Ma et al. using recommended parameter values, and those for the BNN models are summarized in Table 1. Clearly, given experimental measurement uncertainties, small differences between SEP values are unlikely to be statistically significant. We have focused on those data sets where the differences in SEP values for test set predictions were relatively large and likely to be statistically important. Cases where DNN gave better predictions than RF are coloured blue, and those where BNN gave better predictions than DNN are colored green. For the cases where DNN performs better than RF, BNN also performs better than RF. Clearly, BNN and DNN had similar prediction accuracy on average across the 15 data sets. BNN was significantly better at predicting the test set for the HIVPROT data set (SEP of 1.04 compared to 1.66 for DNN), and the THROMBIN data set (SEP of 1.53 compared to 2.17 for DNN). On the other hand, DNN made significantly better predictions for the METAB data set (SEP 21.8 compared to 23.9 for BNN). The similar predictive power of models generated by a deep learning method (DNN) and shallow learning method (BNN) is broadly consistent with the universal approximation theorem. This in turn suggests that SAR/SPR surfaces are continuous and differentiable, at least for the data sets analyzed in this paper. Capuzzi et al. very recently reported the results of modeling the data from the Tox21 Challenge that also supports our conclusions.[43] They compared the performance of DNN with random forests using consensus models that combined predictions of several types of models. Although they used the area under the Receiver Operating Curve as the measure of performance, not dispersion measures, they found that shallow prediction methods performed similarly to DNNs across the data sets. They also stated that much larger data sets would be required to fully take advantage of deep learning methods, a significant limitation in many cases.

Although the ability to predict the properties of new data is similar for models generated by deep and shallow learning methods, there may still be cases where one performs better than the other, or one is preferred for reasons of coding speed, computational efficiency, efficacy in running on graphical processing unit (GPU) or other cluster machines etc. Shallow neural networks in particular may be faster to train, and can be run on a machine with smaller resources. The automatic optimization of regularization afforded by the Bayesian framework allows the architecture of the network to be easily optimized, lower relevance weights in the network to be pruned, and the most important descriptors to be selected automatically and objectively. However, models built using shallow learning methods may be more sensitive to the quality of the descriptors used. Shao et al. conducted a study comparing the type of modeling algorithm versus the types of descriptors for the same data sets[35] and found the latter was a much larger contributor to model quality and predictivity than the type of modelling algorithm.

The mathematics of DNN is simpler to implement using matrix algebra so may be faster to code and easier to implement on multi-core machines. When DNN have their parameters optimized, they may be as effective as shallow neural networks at predicting the properties of new data. They may also resolve many of the activity cliff issues that

Table 1. Comparison of standard errors of prediction of the KAGGLE test sets using DNN and BNN. Blue: DNN better than random forest. Green: BNN better than DNN.

Data set	Description	Size of data set		Test set SEP	
		Training	Test	DNN	BNN
3A4	CYP P450 3A4 inhibition pIC_{50} [†]	37241	12338	0.48	0.50
CB1	binding to cannabinoid receptor 1 pIC_{50}	8716	2907	1.25	1.14
DPP4	inhibition of dipeptidyl peptidase 4 pIC_{50}	6148	2045	1.30	1.27
HIVINV	inhibition of HIV integrase pIC_{50}	1815	598	0.44	0.46
HIVPROT	inhibition of HIV protease pIC_{50}	3212	1072	1.66	1.04
LOGD	logD measured by HPLC method	37388	12406	0.51	0.53
METAB	% remaining after 30m microsomal incubation	1569	523	21.78	23.89
NK1	inhibition of neurokinin1 receptor pIC_{50}	9965	3335	0.76	0.72
OX1	inhibition of orexin 1 receptor pK_i [#]	5351	1769	0.73	0.79
OX2	inhibition of orexin 2 receptor pK_i M	11151	3707	0.95	1.08
PGP	transport by p-glycoprotein log(BA/AB)	6399	2093	0.36	0.40
PPB	log(bound/unbound) to human plasma protein	8651	2899	0.56	0.58
RATF	log(rat bioavailability) at 2 mg/kg	6105	1707	0.54	0.49
TDI	time dependent 3A4 inhibition[&]	4165	1382	0.40	0.39
THROMBIN	human thrombin inhibition pIC_{50}	5059	1698	2.04	1.53

[†] $pIC_{50} = -\log(IC_{50})$ M; [#] $pK_i = -\log(K_i)$ M; [&] $\log(IC_{50}$ without NADPH/IC_{50} with NADPH)

arise sporadically in QSAR problems,[44] albeit at the expense of requiring more computational resources. Further studies are required that employ DNN to model data sets containing activity cliffs in order to determine how useful DNN are in solving this difficult problem.

4 Conclusion

Clearly, for the large and diverse drug data sets modelled here using DNN and shallow regularized neural networks (BNN), the predictive power of the models is similar. This is completely consistent with the Universal Approximation theorem that states that shallow neural networks with a single hidden layer should be capable of modelling any continuous function given sufficient data. The ability of models to generalize to new data is driven more strongly by the choice of descriptors than by the modelling method. Although DNN are more computationally demanding, they may generate better models in some circumstances, for example, with sparse data sets, and their abilities to transform input descriptors into higher dimensional spaces may make them superior in classification problems and where QSAR anomalies, such as activity cliffs, occur. It is likely that deep and shallow neural networks will work together synergistically to solve a wide range of QSAR and QSPR problems in the future. A watching brief on deep learning methods is required and it will be interesting to see where they find their strongest sphere of application in drug and materials design and optimization.

Conflict of Interest

None declared.

Acknowledgments

The authors would like to thank Prof. Robert Sheridan for providing the results of the DNN KAGGLE predictions as RMS values, subsequently converted to SEP values, and for advice on the computational time required to train the DNN models.

References

[1] T. Le, V. C. Epa, F. R. Burden, D. A. Winkler, *Chem. Rev.* **2012**, *112*, 2889–2919.
[2] T. Fujita, D. A. Winkler, *J. Chem. Inf. Mod.* **2016**, *56*, 269–274.
[3] I. Baskin, D. A. Winkler, I. Tetko, *Expert Opin. Drug Dis.* **2016**, *11*, 785–795.
[4] S. A. Kauffman, W. G. Macready, *J Theor Biol* **1995**, *173*, 427–440.
[5] T. C. Le, D. A. Winkler, *ChemMedChem* **2015**, *10*, 1296–1300.
[6] T. C. Le, D. A. Winkler, *Chem. Rev.* **2016**, *116*, 6107–6132.
[7] T. Nakama, *Lect. Notes Comput. Sci.* **2011**, *6675*, 270–279.
[8] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436–444.
[9] C. Angermueller, T. Parnamaa, L. Parts, O. Stegle, *Mol. Syst. Biol.* **2016**, *12*.
[10] Y. F. Chen, Y. Li, R. Narayan, A. Subramanian, X. H. Xie, *Bioinform.* **2016**, *32*, 1832–1839.
[11] S. Ekins, *Pharm. Re.s* **2016**, *33*, 2594–2603.
[12] E. Gawehn, J. A. Hiss, G. Schneider, *Mol Inform.* **2016**, *35*, 3–14.
[13] D. Winkler, *Drug Discov. Today* **2001**, *6*, 1198–1199.
[14] D. A. Winkler, *Drugs Future* **2004**, *29*, 1043–1057.
[15] D. A. Winkler, *Mol. Biotechnol.* **2004**, *27*, 139–167.
[16] M. A. Nielsen, *Neural Networks and Deep Learning*, Determination Press, **2015**.
[17] X. Glorot, A. Bordes, Y. Bengio, in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, **2011**, pp. 315–323.
[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
[19] V. Vapnik, *Proc. IEEE Int. Symp. Circuits Syst.* **2000**, *II*, 749–752.
[20] F. R. Burden, D. A. Winkler, *J. Chem. Inf. Model.* **2015**, *55*, 1529–1534.
[21] B. C. Csáji, MSc thesis, Eötvös Loránd University (Hungary), **2001**.
[22] K. Hornik, *Neural Netw.* **1991**, *4*, 251–257.
[23] G. Cybenko, *Math. Control Signals Syst.* **1989**, *2* 303–314.
[24] V. Kurkova, *Neural Netw.* **1992**, *5*, 501–506.
[25] D. Dobchev, M. Karelson, *Expert Opin. Drug Dis.* **2016**, *11*, 627–639.
[26] L. Michielan, S. Moro, *J. Chem. Inf. Model.* **2010**, *50*, 961–978.
[27] T. Scior, J. L. Medina-Franco, Q. T. Do, K. Martinez-Mayorga, J. A. Y. Rojas, P. Bernard, *Curr. Med. Chem.* **2009**, *16*, 4297–4313.
[28] F. Burden, D. Winkler, *Methods Mol. Biol.* **2008**, *458*, 25–44.
[29] B. Chen, R. P. Sheridan, V. Hornak, J. H. Voigt, *J. Chem. Inf. Model.* **2012**, *52*, 792–803.
[30] G. M. Maggiora, *J. Chem. Inf. Model.* **2006**, *46*, 1535–1535.
[31] D. Stumpfe, J. Bajorath, *J. Med. Chem.* **2012**, *55*, 2932–2942.
[32] M. Cruz-Monteagudo, J. L. Medina-Francos, Y. Perez-Castillo, O. Nicolotti, M. N. D. S. Cordeiro, F. Borges, *Drug Discov. Today* **2014**, *19*, 1069–1080.
[33] N. V. Kireeva, S. I. Ovchinnikova, S. L. Kuznetsov, A. M. Kazennov, A. Y. Tsivadze, *J. Comput. Aid. Mol. Des.* **2014**, *28*, 61–73.
[34] J. Bentzien, R. Bharadwaj, D. C. Thompson, *Drug Discov Today* **2015**, *20*, 874–883.
[35] C. Y. Shao, S. Z. Chen, B. H. Su, Y. F. J. Tseng, E. X. Esposito, A. J. Hopfinger, *J. Chem. Inf. Model.* **2013**, *53*, 142–158.
[36] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, V. Svetnik, *J. Chem. Inf. Model.* **2015**, *55*, 263–274.
[37] D. Alexander, A. Tropsha, D. A. Winkler, *J. Chem. Inf. Model.* **2015**, *55*, 1316–1322.
[38] F. R. Burden, D. A. Winkler, *J. Med. Chem.* **1999**, *42*, 3183–3187.
[39] F. R. Burden, D. A. Winkler, *J. Chem. Inf. Comp. Sci.* **1999**, *39*, 236–242.
[40] F. R. Burden, D. A. Winkler, *QSAR Comb. Sci.* **2009**, *28*, 1092–1097.
[41] F. R. Burden, D. A. Winkler, *QSAR Comb. Sci.* **2009**, *28*, 645–653.
[42] D. Winkler, F. Burden, J. Halley, *Drugs Future* **2007**, *32*, 26–26.
[43] S. J. Capuzzi, R. Politi, O. Isayev, S. Farag, A. Tropsha, *Front. Environ. Sci.* **2016**, *4*.
[44] D. Stumpfe, J. Bajorath, *J. Chem. Inf. Model.* **2012**, *52*, 2348–2353.

Received: September 1, 2016
Accepted: October 4, 2016
Published online: October 26, 2016

