



vNN Web Server for ADMET Predictions

Patric Schyman*, Ruifeng Liu, Valmik Desai and Anders Wallqvist*

DoD Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD, United States

OPEN ACCESS

Edited by:

Adriano D. Andricopulo,
São Carlos Institute of Physics,
University of São Paulo, Brazil

Reviewed by:

Fabio Broccatelli,
Genentech, United States
Tero Aittokallio,
Institute for Molecular Medicine
Finland, Finland
Emilio Benfenati,
Istituto Di Ricerche Farmacologiche
Mario Negri, Italy

*Correspondence:

Patric Schyman
pschyman@bhsai.org
Anders Wallqvist
sven.a.wallqvist.civ@mail.mil

Specialty section:

This article was submitted to
Experimental Pharmacology and Drug
Discovery,
a section of the journal
Frontiers in Pharmacology

Received: 22 September 2017

Accepted: 20 November 2017

Published: 04 December 2017

Citation:

Schyman P, Liu R, Desai V and
Wallqvist A (2017) vNN Web Server for
ADMET Predictions.
Front. Pharmacol. 8:889.
doi: 10.3389/fphar.2017.00889

In drug development, early assessments of pharmacokinetic and toxic properties are important stepping stones to avoid costly and unnecessary failures. Considerable progress has recently been made in the development of computer-based (*in silico*) models to estimate such properties. Nonetheless, such models can be further improved in terms of their ability to make predictions more rapidly, easily, and with greater reliability. To address this issue, we have used our vNN method to develop 15 absorption, distribution, metabolism, excretion, and toxicity (ADMET) prediction models. These models quickly assess some of the most important properties of potential drug candidates, including their cytotoxicity, mutagenicity, cardiotoxicity, drug-drug interactions, microsomal stability, and likelihood of causing drug-induced liver injury. Here we summarize the ability of each of these models to predict such properties and discuss their overall performance. All of these ADMET models are publically available on our website (<https://vnnadmet.bhsai.org/>), which also offers the capability of using the vNN method to customize and build new models.

Keywords: ADME, toxicology, QSAR, machine learning, applicability domain, online web platform, open access

INTRODUCTION

Drug discovery is a risky, lengthy, and resource-intensive process with high attrition rates. In recent years, the development of assays and computer-based (*in silico*) models to assess absorption, distribution, metabolism, and excretion (ADME) properties has greatly reduced the attrition rate (Waring et al., 2015). The ability to predict these properties quickly and reliably facilitates the exclusion of compounds with potential ADME issues, and thereby helps investigators prioritize which compounds to synthesize and evaluate. However, toxicity remains a hurdle, with an attrition rate of 40% among new compounds identified in the drug discovery phase (Waring et al., 2015). This necessitates careful selection of compounds during drug development to avoid late-stage attrition. As such, there is an urgent need for *in silico* methods that make fast, easy, and reliable predictions of ADME and toxicity (ADMET) properties, which has resulted in several online tools and web-platforms for ADMET predictions (Walker et al., 2010; Sushko et al., 2011; Cheng et al., 2012; Maunz et al., 2013; Mangano et al., 2016; Daina et al., 2017).

Here we provide an overview of our versatile variable nearest neighbor (vNN) method (Liu et al., 2012) and the 15 models we constructed using this method to predict the ADMET properties of potential target compounds. The vNN method has several advantages over existing *in silico* methods. First, it calculates the similarity distance between molecules in terms of their structure, and uses a distance threshold to define a domain of applicability (i.e., all nearest neighbors that meet a minimum similarity threshold constraint).

This applicability domain, while limiting vNN-based models to making predictions only for molecules that are similar to the reference molecules, ensures that the predictions they generate are reliable. Second, vNN-based models can be built within minutes and require no re-training when new assay information becomes available—an important feature when keeping quantitative structure–activity relationship (QSAR) models up-to-date to maintain their performance levels. Finally, as we show throughout this work, the performance characteristics of our vNN-based models are comparable, and often superior, to those of other more elaborate model constructs.

We have developed a publically available vNN website (<https://vnnadmet.bhsai.org/>). This website provides users with ADMET prediction models that we have developed, as well as a platform for using their own experimental data to update these models or build new ones from scratch. Although we use the vNN method here for predicting ADMET properties, the vNN website can be used to build a variety of classification or regression models.

MATERIALS AND METHODS

The vNN Method

The k-nearest neighbor (k-NN) method is widely used to develop QSAR models (Zheng and Tropsha, 2000). This method rests on the premise that compounds with similar structures have similar activities. The simplest form of the k-NN method takes the average property values of the k nearest neighbors as the predicted value. However, because structurally similar compounds tend to show similar biological activity, it is reasonable to weight the contributions of neighbors so that closer neighbors contribute more to the predicted value. One notable feature of the k-NN method is that it always gives a prediction for a compound, based on a constant number, k, of nearest neighbors no matter how structurally dissimilar they are from the compound. An alternative approach is to use a predetermined similarity criterion. We developed the aforementioned vNN method, which uses all nearest neighbors that meet a structural similarity criterion to define the model's applicability domain (Liu et al., 2012, 2015; Liu and Wallqvist, 2014). When no nearest neighbor meets the criterion, the vNN method makes no prediction.

One of the most widely used measures of the similarity distance between two small molecules is the Tanimoto distance, d , which is defined as:

$$d = 1 - \frac{n(P \cap Q)}{n(P) + n(Q) - n(P \cap Q)}, \quad (1)$$

where $n(P \cap Q)$ is the number of features common to molecules p and q , and $n(P)$ and $n(Q)$ are the total numbers of features for molecules p and q , respectively. The features used to calculate molecular similarity are often based on atom type (connectivity and chemical properties), such as element, charge, donor, acceptor, and aromatic, but they can also be based on

holistic molecular properties, such as molecular weight and partition coefficient (LogP). The predicted biological activity y is then given by a weighted average across structurally similar neighbors:

$$y = \frac{\sum_{i=1}^v y_i e^{-\left(\frac{d_i}{h}\right)^2}}{\sum_{i=1}^v e^{-\left(\frac{d_i}{h}\right)^2}}, \quad d_i \leq d_0 \quad (2)$$

where d_i denotes the Tanimoto distance between a query molecule for which a prediction is made and a molecule i of the training set; y_i is the experimentally measured activity of molecule i ; h is a smoothing factor, which dampens the distance penalty; d_0 is a Tanimoto-distance threshold, beyond which two molecules are no longer considered to be sufficiently similar to be included in the average; and v denotes the total number of molecules in the training set that satisfy the condition $d_i \leq d_0$. The values of h and d_0 are determined from cross-validation studies.

To identify structurally similar compounds, we used Accelrys extended-connectivity fingerprints with a diameter of four chemical bonds (ECFP4) (Rogers and Hahn, 2010). For the vNN website, we chose ECFP4 fingerprints, which have previously been reported to show satisfactory overall performance in retrieving the active compounds of diverse datasets (Hert et al., 2004; Duan et al., 2010; Schyman et al., 2016). We emphasize that h and d_0 are unique, and need to be optimized for each set of fingerprints and training set.

Model Validation

We used the 10-fold cross-validation (CV) procedure to validate the model and determine the values of h and d_0 . We randomly divided the data into 10 sets, 9 of which we used to develop the model and the 10th to validate the model. We repeated this process 10 times, leaving each set of molecules out once. In the next section, we report averages of the 10-fold CV as the performance measures.

Performance Measures

We used the following metrics to assess the quality of the classification models:

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{specificity} = \frac{TN}{FP + TN} \quad (4)$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$\text{kappa} = \frac{\text{accuracy} - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (6)$$

where TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively. The metric kappa assesses the quality of binary classifiers (Dunn and Everitt, 1995). $\text{Pr}(e)$ is an estimate

Abbreviations: Pgp, permeability glycoprotein; MDR, multidrug resistance.

vNN-ADMET Build/Run Models - Biotechnology HPC Software Applications Institute

VARIABLE NEAREST NEIGHBOR ADMET

The vNN-ADMET webserver is a publicly available online platform to predict absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties and to build new models based on our variable nearest neighbor (vNN) methodology. We have developed 15 ADMET prediction models to quickly, and with high confidence assess some of the most important properties, such as cytotoxicity, mutagenicity, cardiotoxicity, drug-drug interactions, microsomal stability, and drug-induced liver injury.

Run ADMET Models

Build Classification Model

Build Regression Model

Run Custom Model

References

1. Liu, R., G. Tawa, and A. Wallqvist. Locally weighted learning methods for predicting dose-dependent toxicity with application to the human maximum recommended daily dose. *Chemical Research in Toxicology*. 2012 October 15; 25(10):2216-2226. [↗](#)
2. Liu, R., and A. Wallqvist. Merging applicability domains for in silico assessment of chemical mutagenicity. *Journal of Chemical Information and Modeling*. 2014 March 24; 54(3):793-800. [↗](#)
3. Liu, R., P. Schyman, and A. Wallqvist. Critically assessing the predictive power of QSAR models for human liver microsomal stability. *Journal of Chemical Information and Modeling*. 2015; 55(8):1566-1575. [↗](#)
4. Schyman, P., R. Liu, and A. Wallqvist. Using the variable-nearest neighbor method to identify P-glycoprotein substrates and inhibitors. *ACS Omega*. 2016 November 16; 1(5):923-929. [↗](#)

© 2017 BHS AI

FIGURE 1 | The vNN-ADMET main page. From this page, users can run ADMET models or build their own models.

of the probability of a correct prediction by chance. It is calculated as:

$$\text{Pr}(e) = \frac{(\text{TP} + \text{FN})(\text{TP} + \text{FP}) + (\text{FP} + \text{TN})(\text{TN} + \text{FN})}{(\text{TP} + \text{FN} + \text{FP} + \text{TN})^2} \quad (7)$$

The sensitivity measures a model's ability to correctly detect true positives, whereas the specificity measures its ability to detect true negatives. Kappa compares the probability of correct predictions to the probability of correct predictions by chance. Its value ranges from +1 (perfect agreement between model prediction and experiment) to -1 (complete disagreement), with 0 indicating no agreement beyond that expected by chance.

The performance measure for regression models is given by the Pearson's correlation coefficient (Adler and Parmryd, 2010):

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (8)$$

where n is the sample size, x_i and y_i are samples, and \bar{x} and \bar{y} are sample means. The correlation coefficient provides a measure of the interrelatedness of numeric properties. Its value ranges from -1 (highly anticorrelated) to +1 (highly correlated), and is 0 when uncorrelated.

We also calculated the coverage, which we define as the proportion of test molecules with at least one nearest neighbor that meets the similarity criterion. For all other molecules that do not meet the criterion, we do not make any predictions. In this case, the coverage is a measure of the size of the applicability domain of a prediction model.

RESULTS

The vNN Platform

The main purpose of the vNN-based platform is to provide users with a tool to make ADMET predictions and a user-friendly environment to build new models. Hence, the platform offers users two main capabilities that are accessible from the main webpage (<https://vnnadmet.bhsai.org/>) (**Figure 1**): (1) to run prebuilt ADMET models and (2) to build and run customized models.

To use prebuilt ADMET models, users need only provide one or more query molecules as the input (**Figure 2**). They can do this either by drawing the molecule, entering the molecular SMILES string (Weininger, 1988) directly on the website, or uploading a text file (csv or txt format) with query molecules in SMILES format. The text file should contain column headers labeled as NAME and SMILES. Once users upload the query molecules, they can submit the job. The application will then automatically

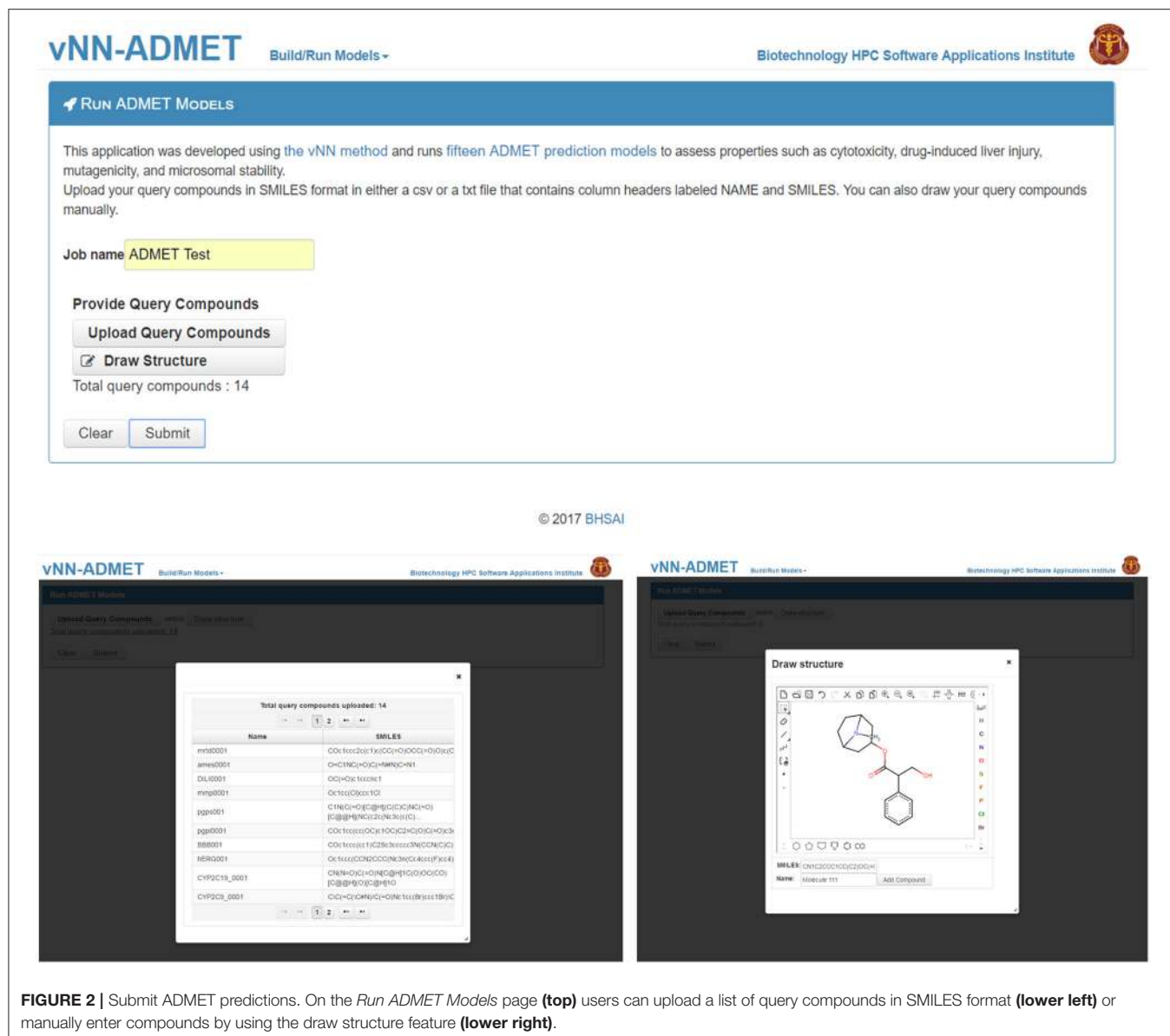


FIGURE 2 | Submit ADMET predictions. On the *Run ADMET Models* page (**top**) users can upload a list of query compounds in SMILES format (**lower left**) or manually enter compounds by using the draw structure feature (**lower right**).

run all ADMET prediction models. The output will be displayed once all predictions are completed and a temporary link to the result page will be sent to the user's e-mail address. The results can be downloaded as a table to the user's computer (**Figure 3**). By default, the user will see the ADMET results for our models, which use a restricted applicability domain. However, there is an option to include the results for the remaining compounds, using our unrestricted applicability domain models. The time required to run 100 query compounds is ~ 5 min on the server. However, this may vary depending on the size of the molecules and whether or not the job has been queued.

Users can build their own models by either selecting *Build Classification Model* or *Build Regression Model* on the main webpage (**Figure 1**). On the *Build Classification Model* page (**Figure 4**), users are asked to upload a list of molecules in SMILES format and the property of interest, with column headers labeled as NAME, SMILES, and PROPERTY. The value of the

property should be set to 1 or 0 for classification models and real numbers for regression models. The vNN platform will then automatically run 10-fold CV by varying the Tanimoto distance (d) from 0.1 to 1.0 in increments of 0.1, and the smoothing factor (h) from 0.1 to 1.0 at each value of d . Once the calculations are completed, a temporary link to the result page will be sent to the user's e-mail address. The results will be displayed on an interactive webpage where users can select the values for d and h (Equation 2), depending on the optimal performance measures and coverage (**Figure 4**). The time required to build a model with a dataset of 1,000 compounds is ~ 10 min.

Users can then select the *Run Custom Model* option to predict the activity of new test molecules (**Figure 5**), using the previously selected values for the *Tanimoto Distance* and *Smoothing Factor*, and add the same molecules as those used to train the model in the *Upload Compounds with Property* data field. They then need to add the new query molecule(s) in SMILES format in the

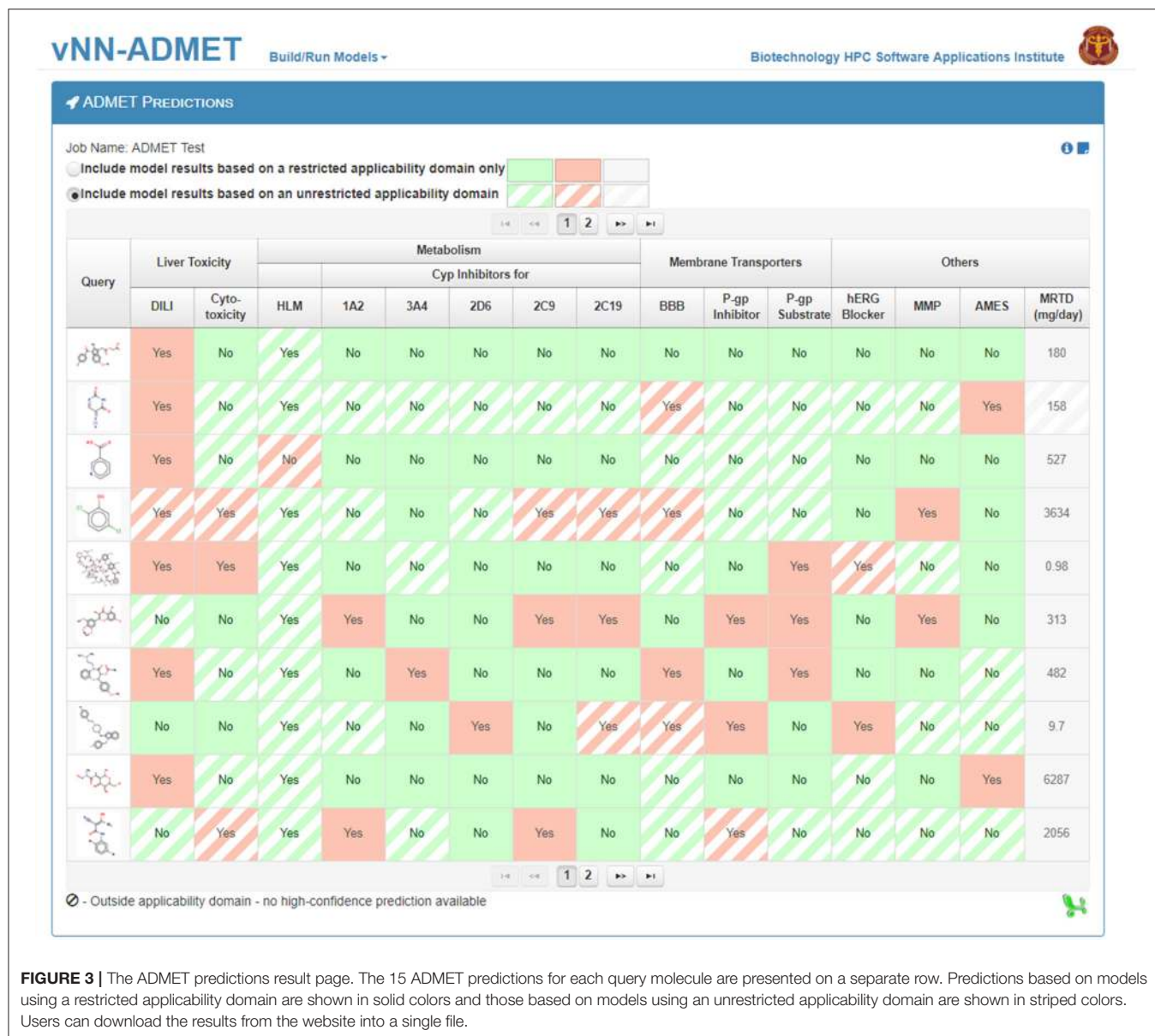


FIGURE 3 | The ADMET predictions result page. The 15 ADMET predictions for each query molecule are presented on a separate row. Predictions based on models using a restricted applicability domain are shown in solid colors and those based on models using an unrestricted applicability domain are shown in striped colors. Users can download the results from the website into a single file.

Upload Query Compounds field. The result will be displayed on a new webpage, and a temporary link to that page will also be sent to the user's e-mail address (Figure 5).

Available ADMET Predictions


The available ADMET prediction models, including their performance measures for the restricted applicability domain model, are summarized in Table 1. The performance measures for the models using an unrestricted applicability domain are presented in Table S1 in the Supplementary Material and on our website (<https://vnnadmet.bhsai.org/>). The 15 models cover a diverse set of ADMET endpoints. We will briefly describe these models and their performance measures, as well as the sources from which we retrieved the data. All datasets are available in SMILES format on the vNN web server or in Structure Data Format (SDF) in the Supplementary Material (Datasheet 1).

Some of the models have already been published (Liu et al., 2012, 2015; Liu and Wallqvist, 2014; Schyman et al., 2016). We also present several new models here for the first time.

Blood-Brain Barrier

The blood-brain barrier (BBB) is a highly selective barrier that separates the circulating blood from the central nervous system (CNS) (Abbott et al., 2006). It allows the passage of water molecules and water-soluble lipid molecules, as well as the selective transport of glucose and amino acids. The benefit of predicting BBB-permeable compounds is two-fold: (1) to identify toxicants that could harm the brain, and (2) to design drug molecules that can pass the BBB and reach their target in the CNS.

We developed a vNN-based BBB model, using 353 compounds whose BBB permeability values ($\log BB$) were obtained from the literature (Muehlbacher et al., 2011; Naef,

vNN-ADMET Build/Run Models - Biotechnology HPC Software Applications Institute 

BUILD CLASSIFICATION MODEL

This application allows you to build a classification model using the vNN method. It will automatically run a 10-fold cross validation and systematically vary the Tanimoto distance and smoothing factor in increments of 0.1.

Upload your training compounds in SMILES format in either a csv or a txt file that contains column headers labeled NAME, SMILES, and PROPERTY. The values in the property column for classification is either 1 or 0. You can also draw your query compounds manually.

Job name

Provide Training Compounds

Draw Structure

Total training compounds : 343



© 2017 BHSAl

FIGURE 4 | Build a classification model. On the *Build Classification Model* page (**top**), users can upload their training data and/or draw structures. On the *Build Classification Model Results* page (**bottom**), users can interactively select/deselect different smoothing factors for comparison. The graph shows accuracy of performance on the 10-fold cross validation test at different Tanimoto distances, where smoothing factors 0.2 and 1.0 are highlighted in green and blue, respectively (strikethrough smoothing factors indicate deselected values). The coverage is shown in gray. The red circle indicates the “best” model performance based on accuracy and coverage, where the black arrows show the corresponding Tanimoto-distance threshold ($d_0 = 0.7$) and smoothing factor ($h = 0.2$). Although the accuracy is reduced to 88 from 90% at $d_0 = 0.6$, the number of compounds predicted increases from 60 to 75%, which may be worth the loss in accuracy.

vNN-ADMET Build/Run Models ▾
Biotechnology HPC Software Applications Institute

⊙ RUN CUSTOM MODEL

This application allows you to use your previously built model to make new predictions. Upload your training compounds in SMILES format in either a csv or a txt file that contains column headers labeled NAME, SMILES, and PROPERTY. You can also draw your training compounds manually. Upload your query compounds in SMILES format in either a csv or a txt file that contains column headers labeled NAME and SMILES. You can also draw your query compounds manually. Enter your Tanimoto distance and smoothing factor determined for this training set in the build model step.

Job name

Select model class
 Regression Classification

Provide Training Compounds

Draw Structure

Total training compounds : 343

Provide Query Compounds

Draw Structure

Total query compounds : 4

Tanimoto Distance

Smoothing Factor

© 2017 BHSAI

vNN-ADMET Build/Run Models ▾
Biotechnology HPC Software Applications Institute

⊙ RUN CUSTOM MODEL RESULTS

Job Name: BBB Model **Model class:**Classification **Tanimoto Distance:**0.7 **Smoothing Factor:**0.2 [Show Training Set](#) !

Name	Compound	Prediction
BBB001		1
BBB002		0
BBB003		⊘
BBB201		⊘

⊘ - Tanimoto Distance outside applicability domain - no high-confidence prediction available

© 2017 BHSAI

FIGURE 5 | Run a customized model. The first step to run a customized model is to upload the training dataset, as well as the selected Tanimoto distance and smoothing factor from **Figure 4**. The second step is to upload query compounds. The results can be downloaded from the *Run Custom Model Results* page (bottom).

TABLE 1 | Performance measures of vNN models in 10-fold cross validation, using a restricted applicability domain.

Model	Data ^a	d_0^b	h^c	Accuracy	Sensitivity	Specificity	Kappa	R^d	Coverage
DILI	1,427	0.60	0.50	0.71	0.70	0.73	0.42		0.66
Cytotox (hep2g)	6,097	0.40	0.20	0.84	0.88	0.76	0.64		0.89
HLM	3,219	0.40	0.20	0.81	0.72	0.87	0.59		0.91
CYP 1A2	7,558	0.50	0.20	0.90	0.70	0.95	0.66		0.75
CYP 2C9	8,072	0.50	0.20	0.91	0.55	0.96	0.54		0.76
CYP 2C19	8,155	0.50	0.20	0.87	0.64	0.93	0.58		0.76
CYP 3A4	10,373	0.50	0.20	0.88	0.76	0.92	0.68		0.78
CYP 2D6	7,805	0.50	0.20	0.89	0.61	0.94	0.57		0.75
BBB	353	0.60	0.20	0.90	0.94	0.86	0.80		0.61
Pgp Substrate	822	0.60	0.20	0.79	0.80	0.79	0.58		0.66
Pgp Inhibitor	2,304	0.50	0.20	0.85	0.91	0.73	0.66		0.76
hERG	685	0.70	0.70	0.84	0.84	0.83	0.68		0.80
MMP	6,261	0.50	0.40	0.89	0.64	0.94	0.61		0.69
AMES	6,512	0.50	0.40	0.82	0.86	0.75	0.62		0.79
MRTD ^e	1,184	0.60	0.20					0.79	0.69

^aNumber of compounds in the dataset; ^bTanimoto-distance threshold value; ^cSmoothing factor; ^dPearson's correlation coefficient; ^eRegression model.

2015). We classified compounds with log *BB* values of < -0.3 and $> +0.3$ as BBB non-permeable and permeable, respectively. To calculate performance measures, we classified BBB permeable and BBB non-permeable compounds as positives and negatives, respectively.

The model predicted whether or not a given compound would pass the BBB, but only for compounds within the applicability domain defined by the training set. The performance measures in **Table 1** were calculated from 10-fold CV. The model showed a high overall accuracy of 90% and a kappa value of 0.80, with a coverage of 61%. The size of the dataset limited the applicability domain of the model. However, if new data become available, they can easily be added to the model to increase the applicability domain.

The model performed on par with the best of the BBB models published thus far. Most of the latter models, which used small datasets, are global models applied to any molecule. However, all models have a finite applicability domain (Cherkasov et al., 2014). Indeed, modeling BBB permeability is complicated because there are different possible routes across the barrier, via passive diffusion or protein transport, and no model singlehandedly accounts for all factors associated with this property. Our vNN model only makes predictions for compounds that are structurally similar enough to the test set molecules to ensure that they have the same type of transport mechanism. Thus, our vNN method accounts for multiple transport routes.

MMP Disruption (Mitochondrial Toxicity)

Given the fundamental role of mitochondria in cellular energetics and oxidative stress, mitochondrial dysfunction has been implicated in cancer, diabetes, neurodegenerative disorders, and cardiovascular diseases (Pieczenik and Neustadt, 2007). Many pharmaceuticals and environmental toxicants cause mitochondrial dysfunction (Meyer et al., 2013). Therefore, the ability to predict the impact of chemicals on mitochondrial

function would be useful. However, predicting mitochondrial toxicants is complicated because mitochondrial dysfunction can result from impairing any of the following: (1) the electron transport chain (ETC), (2) the mitochondrial transport pathway, (3) fatty acid oxidation, (4) the citric acid cycle, (5) mtDNA replication, (6) and mitochondrial protein synthesis.

There are several common experimental techniques to measure mitochondrial function. We used the largest dataset of chemical-induced changes in mitochondrial membrane potential (MMP), based on the assumption that a compound that causes mitochondrial dysfunction is also likely to reduce the MMP. We developed a vNN-based MMP prediction model, using 6,261 compounds collected from a previous study that screened a library of 10,000 compounds (~8,300 unique chemicals) at 15 concentrations, each in triplicate, to measure changes in the MMP in HepG2 cells (Attene-Ramos et al., 2015). The study found that 913 compounds decreased the MMP, whereas 5,395 compounds had no effect. We classified compounds that decreased the MMP as positives and those that did not affect the MMP as negatives.

Our MMP model predicted whether a given compound had the potential to affect the MMP and thereby cause mitochondrial dysfunction. It made predictions for compounds that were well represented in the applicability domain, but not for any other compound. The model showed a high overall accuracy of 89% and a kappa value of 0.61, with a coverage of 69% (**Table 1**).

Cytotoxicity (HepG2)

Cytotoxicity is the degree to which a chemical causes damage to cells. Cytotoxicity assays are widely used to screen compounds for unwanted cell damage, and to identify compounds that could be used, for example, to kill cancer cells. As such, the ability to identify cytotoxic compounds is highly desirable.

We developed a cytotoxicity prediction model, using a training dataset of *in vitro* toxicity against HepG2 cells for

TABLE 2 | Performance measures of vNN DILI models compared with deep learning.

	NCTR ^a 10-fold CV	NCTR ^a Test	Green ^a Test	Xu ^a Test	Combined ^a 10-fold CV
Accuracy (%)	87 (81)	75 (70)	61 (65)	60 (62)	85 (85) ^b
Sensitivity (%)	65 (70)	64 (80)	51 (75)	52 (62)	83 (84) ^b
Specificity (%)	95 (88)	86 (60)	75 (46)	70 (62)	88 (85) ^b
Coverage (%)	40	50	46	41	67
No of Compounds	190	185	320	236	475

^aValues in parentheses are the deep learning results from Xu et al. (2015).

^bValues averaged over 60 runs of 10-fold CV.

6,097 structurally diverse compounds, which we collected from Chemical European Biology Laboratory (ChEMBL) (Bento et al., 2014). In developing our model, we considered compounds with an IC₅₀ of 10 μM or less in the *in vitro* assay as cytotoxic. We classified cytotoxic compounds as positives and non-toxic compounds as negatives.

The cytotoxicity model performed well, with an overall accuracy of 84% and a kappa value of 0.64 (Table 1). Because compounds in the dataset achieved only sparse coverage of the chemical space, the model only predicted compounds that were well represented in the dataset. It did not give predictions for other compounds, and thereby avoided misleading results. When using 10-fold CV, the model reliably predicted 89% of the compounds in our dataset.

Drug-Induced Liver Injury

Over the last 50 years, drug-induced liver injury (DILI) has been the most commonly cited reason for drug withdrawals from the market (Assis and Navarro, 2009). As a result, current drug development efforts are devoted to identifying and eliminating potential DILI compounds. Therefore, a model that predicts at an early stage whether a compound causes liver injury would be highly desirable. However, the mechanisms of DILI are complicated and diverse, making toxicology studies difficult. For example, compounds that cause DILI in humans do not necessarily induce clear liver injury in animal studies.

We collected DILI data from four sources used by Xu et al. (2015): (1) the U.S. FDA's National Center for Toxicological Research (NCTR dataset) (Chen M. et al., 2011), as well as the datasets of (2) Greene (Greene et al., 2010), (3) Xu (Xu et al., 2008), and (4) Liew (Liew et al., 2011). In the first three datasets, which included pharmaceuticals, we classified a compound as causing DILI if it was associated with a high risk of DILI and not if there was no such risk. We excluded low-risk DILI compounds. In the Liew dataset, which contained both pharmaceuticals and non-pharmaceuticals, we classified a compound as causing DILI if it was associated with any adverse liver effect. DILI-associated compounds were classified as positives and non-DILI compounds as negatives.

The performance measures of the vNN model, using 10-fold CV of the entire dataset excluding duplicated compounds, showed an overall accuracy of 71% and a coverage of 66% (Table 1). We also used the same datasets and compared our

models with some previously published deep learning models (Xu et al., 2015; Table 2). Considering the complexity and computational time investment involved in training these deep learning models, our vNN models performed relatively well; they performed on-par with the deep learning models, albeit with a coverage ranging from 40 to 65%.

Cytochrome P450 Inhibition (Drug-Drug Interaction)

Cytochrome P450 enzymes (CYPs) constitute a superfamily of proteins that play an important role in the metabolism and detoxification of xenobiotics (Brown et al., 2008). A drug should not be rapidly metabolized by CYPs if it is to maintain an effective concentration. In addition, it should not inhibit drug-metabolizing CYPs, because such an effect could elevate the concentration of a co-administered drug and potentially lead to drug overdose—an effect known as a drug-drug interaction (Murray, 2006). In drug development, *in vitro* assays are routinely used to assess interactions between drug candidates and CYPs. However, there is a need for *in silico* models that assess potential interactions with CYPs in the early stages of drug development.

We collected data for five main drug-metabolizing CYPs: 1A2, 2D6, 2C9, 2C19, and 3A4. We retrieved CYP inhibitors from ChEMBL (Bento et al., 2014) and classified them as inhibitors if the IC₅₀ was below 10 μM. We removed from the dataset any duplicates or compounds tested multiple times with contradicting results, in which the reported IC₅₀ values were below and above the 10 μM threshold value. For all CYPs, we classified inhibitors and non-inhibitors as positives and negatives, respectively.

The performance measures for the five CYP models are presented in Table 1. All models achieved high accuracy (87–91%) and kappa values (0.54–0.68) while maintaining high coverage (75–78%).

hERG Blockers

The human ether-à-go-go-related gene (hERG) codes for a potassium ion channel involved in the normal cardiac repolarization activity of the heart (Sanguinetti and Tristani-Firouzi, 2006). Drug-induced blockade of hERG function can cause long QT syndrome, which may result in arrhythmia and death (De Ponti et al., 2001). For this reason, hERG liability is one of the toxicology screens that drug candidates must pass during early pre-clinical studies. Therefore, *in silico* models that identify hERG blockers in the early stages of drug design are of considerable interest.

We retrieved 282 known hERG blockers from the literature and classified compounds with an IC₅₀ cutoff value of 10 μM or less as blockers (Wang et al., 2012). We also collected a set of 404 compounds with IC₅₀ values >10 μM from ChEMBL (Bento et al., 2014) and classified them as non-blockers (Czodrowski, 2013). We classified hERG blockers and non-blockers as positives and negatives, respectively.

The hERG model performed with an overall accuracy of 84%, well-balanced sensitivity and specificity values (84 and 83%, respectively), and a kappa value of 0.68 (Table 1). The model reliably predicted 80% of the compounds in our dataset

when using 10-fold CV. However, the coverage of chemical space by the non-hERG blockers in the dataset was sparse, and only compounds well represented in the dataset were predicted with confidence. Because the model did not give predictions for other compounds, it avoided misleading results. Therefore, users should use this model to flag potential hERG blockers rather than to identify non-hERG blockers.

Pgp Substrates and Inhibitors

P-glycoprotein (Pgp) is an essential cell membrane protein that extracts many foreign substances from the cell (Ambudkar et al., 2003). As such, it is a critical determinant of the pharmacokinetic properties of drugs. Cancer cells often overexpress Pgp, which increases the efflux of chemotherapeutic agents from the cell and prevents treatment by reducing the effective intracellular concentrations of such agents—a phenomenon known as multidrug resistance (Borst and Elferink, 2002). For this reason, identifying compounds that can either be transported out of the cell by Pgp (substrates) or impair Pgp function (inhibitors) is of great interest. Therefore, using the vNN method, we developed models to predict both Pgp substrates and Pgp inhibitors.

The Pgp substrate dataset was collected by Hou and co-workers (Li et al., 2014). This dataset included measurements for 422 substrates and 400 non-substrates. To generate a large Pgp inhibitor dataset, we combined two datasets (Broccatelli et al., 2011; Chen L. et al., 2011), and removed duplicates to form a combined dataset consisting of a training set of 1,319 inhibitors and 937 non-inhibitors. We classified the Pgp inhibitors (substrates) and non-inhibitors (non-substrates) as positives and negatives, respectively.

The vNN models for identifying Pgp substrates and inhibitors gave accurate and reliable results, showing overall accuracies of 79 and 85%, respectively, when using 10-fold CV, with corresponding kappa values of 0.58 and 0.66. These models reliably predicted 65 and 76% of the compounds in their datasets to be Pgp substrates and inhibitors, respectively. The performance characteristics of these models were comparable, or at times superior, to those of other model constructs (Schyman et al., 2016).

Chemical Mutagenicity (AMES Test)

Mutagens are chemicals that cause abnormal genetic mutations leading to cancer. A common way to assess a chemical's mutagenicity is the Ames test (Ames et al., 1973). This test has become the standard for assessing the safety of chemicals and drugs, and has been used to test thousands of molecules. We examined whether the vNN method could effectively use existing data to predict mutagenicity.

We retrieved an Ames mutagenicity dataset consisting of 6,512 compounds, of which 3,503 were Ames-positive (Hansen et al., 2009), and developed a vNN Ames mutagenicity prediction model. The model performed well, with an overall accuracy of 82%; sensitivity and specificity values of 86 and 75%, respectively; and a high kappa value of 0.62 (Table 1). The model also reliably predicted 79% of the compounds in the Ames dataset when using 10-fold CV. Further details of the model and its prediction performance can be found elsewhere (Liu and Wallqvist, 2014).

Maximum Recommended Therapeutic Dose

A basic principle of toxicology is that “the dose makes the poison.” For most drugs, the therapeutic dose is limited by toxicity, and the maximum recommended therapeutic dose (MRTD) is an estimated upper daily dose that is safe (Contrera et al., 2004). Investigators carry out toxicological experiments on animals to determine the toxic effects of a drug and the initial dose for human clinical trials. Unfortunately, there is a lack of correlation between animal and human toxicity data. Therefore, we investigated whether the vNN method could predict the MRTD values of new compounds based on known human MRTD data. If so, the values could be used to estimate the starting dose in phase I clinical trials, while significantly reducing the number of animals used in preliminary toxicology studies.

We obtained a dataset of MRTD values publically disclosed by the FDA, mostly of single-day oral doses for an average adult with a body weight of 60 kg, for 1,220 compounds (most of which are small organic drugs). For modeling purposes we converted the MRTD unit from mg/kg-body weight/day to mol/kg-body weight/day via the molecular weight of the compound. However, the predicted values on the website are reported in mg/day based upon an average adult weighing 60 kg. We excluded organometallics, high-molecular weight polymers (>5,000 Da), nonorganic chemicals, mixtures of chemicals, and very small molecules (<100 Da). We used an external test set of 160 compounds, which was collected by the FDA for validation. The total dataset for our model contained 1,184 compounds (Liu et al., 2012).

The MRTD model reliably predicted 69% of the FDA MRTD dataset, with a Pearson's correlation coefficient (*R*) of 0.79 between the predicted and measured $\log(\text{MRTD})$ values, and a mean deviation (mDev) of 0.56 \log units, using 40-fold CV (Liu et al., 2012). For comparison, we used two popular QSAR regression methods—the partial least square (PLS) and support vector machine (SVM) methods—to develop two global models to fit the training dataset. We evaluated the model performance, using 40-fold CV of the training set. The best PLS model achieved an *R*-value of 0.50 and an mDev of 0.79. The results for the SVM model were at best comparable to those of the best PLS model, with an *R*-value of 0.53 and an mDev of 0.63. For further details of the model, we refer the reader to our previous paper (Liu et al., 2012).

Human Liver Microsomal Stability

The human liver is the most important organ for drug metabolism. For a drug to achieve effective therapeutic concentrations in the body, it cannot be metabolized too rapidly by the liver. Otherwise, it would need to be administered at high doses, which are associated with high toxicity. To identify and exclude rapidly metabolized compounds (Di et al., 2003), pharmaceutical companies commonly use the human liver microsomal (HLM) stability assay. This has led to the accumulation of a substantial body of HLM stability data in publicly accessible databases.

However, our knowledge of how enzymes in the HLM assay metabolize drugs remains fragmentary. Therefore, we examined whether the vNN method could effectively predict drugs that are

TABLE 3 | Tox21 assays with PubChem assay identification number.

Assay ID	Assay	PubChem AID
AhR	Aryl hydrocarbon receptor	743122
Aromatase	Aromatase	743139
AR	Androgen receptor	743040
AR-LBD	Androgen receptor LBD	743053
ER	Estrogen receptor alpha	743079
ER-LBD	Estrogen receptor alpha LBD	743077
PPAR-g	Peroxisome proliferator-activated receptor gamma	743140
ARE	Nuclear factor antioxidant responsive element	743219
ATAD5	ATAD5	720516
HSE	Heat shock factor response element	743228
MMP	Mitochondrial membrane potential	720637
p53	p53	720552

rapidly metabolized by the liver. We retrieved HLM data from the ChEMBL database (Bento et al., 2014), manually curated the data, and classified compounds as stable or unstable based on the reported half-life [$T_{1/2} > 30$ min was considered stable, and $T_{1/2} < 30$ min unstable (Liu et al., 2015)]. The final dataset contained 3,219 compounds. Of these, we classified 2,047 as stable and 1,166 as unstable.

The HLM model performed with an overall accuracy of 81%; sensitivity and specificity values of 71 and 87%, respectively; and a high kappa value of 0.60 (Table 1). The HLM model reliably predicted 91% of the compounds in the HLM dataset when using 10-fold CV. We refer the reader to our original paper for further details of the model and its prediction performance (Liu et al., 2015).

Implementation Aspects

The vNN-ADMET web-application is hosted on an Apache Tomcat Web server that is accessible via a secure service over Hypertext Transfer Protocol Secure (https). We developed the application on the basis of a three-tiered architecture, composed of a backend database, controller, and presentation tiers. The first tier consists of a PostgreSQL 9.5.7 database that stores user account information, uploaded files, constructed models, and model predictions. The second (controller) tier provides access to the prediction engine and implements the functionality required to create and manage multiple predictions. We implemented this tier, using Pipeline Pilot protocols hosted on a local Pipeline Pilot server. The third (presentation) tier provides for visualization of the results, with plotting capabilities for multiple predictions. The controller and presentation tiers were developed using Java Platform, Enterprise Edition 7, Spring Framework 4.2.2, JavaServer Faces 2.2, PrimeFaces 6.0, and BootsFaces 1.0.2. The graphical user interface in the presentation tier uses Web standards supported by modern Web browsers, including Microsoft Edge 38, Chrome version 58, and Firefox version 53, without any need for plugins.

To use the system, the user must register for an account at <https://vnnadmet.bhsai.org/>. Once logged in, the user can build custom models, and run pre-built ADMET and custom models.

The data corresponding to a user (login credentials, compounds, models, results, etc.) are not shared with any other user within or outside the system. The uploaded compounds, constructed models, and model predictions are purged from the system every 2 weeks.

DISCUSSION

We have presented a web-based vNN prediction platform, with which a user can build and test models as well as predict the ADMET properties of a compound by using our existing tools.

All vNN models performed well with accuracies of $>71\%$ (see Table 1 for further details). On average, the models predicted 75% of the compounds in their datasets, using 10-fold CV.

Achieving fair comparisons between a new model and a competing model is always difficult because such comparisons require the same training data, validation data, and performance measures. An important advantage of our platform is that it offers an opportunity for developers to compare their methods with our vNN method, using their training and validation data.

For demonstrative purposes, we quantitatively compared our vNN method with the winning method of the Tox21 challenge (Huang et al., 2016). This challenge was issued in 2014 by the U.S. Toxicology in the twenty-first Century (Tox21) program, which aims to improve toxicity prediction methods. The Tox 21 consortium solicited models that could best predict the toxicity of 10,000 compounds it had tested in 12 different assays (Table 3). It used a final evaluation dataset that was concealed to determine the winners.

Table 4 shows the area under the curve for the receiver operating characteristic (AUC-ROC) of the 18 leading research teams with their best-performing model for each of the 12 assays. To compare our models with those in Table 4, we set d to 1.0 so that we could predict all compounds. The vNN method performed reasonably well in predicting most of the Tox21 assays. We note that the grand challenge winner used data from PubChem (Wang et al., 2009) and ChEMBL (Bento et al., 2014), in addition to the Tox21 data, which makes it impossible for us to directly compare our results with their results.

The MMP data we used for our mitochondrial dysfunction model were the same as those used in the Tox21 challenge (Attene-Ramos et al., 2015; Huang et al., 2016). Our MMP model was the seventh best performing model, with an AUC-ROC value of 0.882 (with $h = 0.3$ and $d = 1.0$). This was comparable to the values of more elaborate and computationally time-consuming methods, such as deep learning (Table 4).

Some QSAR methods do not use an applicability domain to determine whether their predictions are reliable. This could lead to the misperception that a model can predict the activity of any molecule. The applicability domain is vital to the vNN method. The user of our platform can adjust it by varying the Tanimoto distance threshold value. Although this could be set to 1 so that the model predicts the activity of any molecule, no model is likely to have an unlimited applicability domain (Liu et al., 2015).

A more reasonable approach to improve a vNN-based model is to increase the applicability domain by adding more reference

TABLE 4 | AUC-ROCs of vNN models and the best 18 models on the final evaluation test of the Tox21 Challenge.

Team	AhR	AR	AR-LBD	ARE	Aromatase	ATAD5	ER	ER-LBD	HSE	MMP	p53	PPARg
GrandWinner	0.928	0.807	0.879	0.840	0.834	0.793	0.810	0.814	0.865	0.942	0.862	0.861
AMAZIZ	0.913	0.770	0.846	0.805	0.819	0.828	0.806	0.806	0.842	0.95	0.843	0.830
dmlab	0.781	0.828	0.819	0.768	0.838	0.800	0.766	0.772	0.855	0.946	0.880	0.831
T	0.913	0.676	0.848	0.801	0.825	0.814	0.784	0.805	0.811	0.937	0.847	0.822
Microsomes	0.901	–	–	0.804	–	0.812	0.785	0.827	–	–	0.826	0.717
FilipsPL	0.893	0.736	0.743	0.758	0.776	–	0.771	–	0.766	0.928	0.815	–
Charite	0.896	0.688	0.789	0.739	0.781	0.751	0.707	0.798	0.852	0.880	0.834	0.7
RCC	0.872	0.763	0.747	0.761	0.792	0.673	0.781	0.762	0.755	0.920	0.795	0.637
Frozenarm	0.865	0.744	0.722	0.700	0.740	0.726	0.745	0.790	0.752	0.859	0.803	0.803
ToxFit	0.862	0.744	0.757	0.697	0.738	0.729	0.729	0.752	0.689	0.862	0.803	0.791
CGL	0.866	0.742	0.566	0.747	0.749	0.737	0.759	0.727	0.775	0.880	0.817	0.738
SuperToX	0.854	–	0.560	0.711	0.742	–	–	–	–	0.862	0.732	–
Kibutz	0.865	0.750	0.694	0.708	0.729	0.737	0.757	0.779	0.587	0.838	0.787	0.666
MML	0.871	0.693	0.660	0.701	0.709	0.749	0.750	0.710	0.647	0.854	0.815	0.645
NCI	0.812	0.628	0.592	0.783	0.698	0.714	0.483	0.703	0.858	0.851	0.747	0.736
VIF	0.827	0.797	0.610	0.636	0.671	0.656	0.732	0.735	0.723	0.796	0.648	0.666
Toxic Avg	0.715	0.721	0.611	0.633	0.671	0.593	0.646	0.640	0.465	0.732	0.614	0.682
Swamidass	0.353	0.571	0.748	0.372	0.274	0.391	0.680	0.738	0.711	0.828	0.661	0.585
vNN	0.883	0.716	0.626	0.727	0.786	0.699	0.738	0.770	0.793	0.882	0.808	0.690
vNN rank	7	12	13	11	6	13	12	9	7	7	10	11

The vNN parameters were set to $h = 0.3$ and $d_0 = 1.0$. Gray cells indicate models showing performance inferior to the vNN models.

compounds. A good test of the power of a model to generate prospective predictions is time-split validation, which divides the data into “old” and “new” data and uses the former to train the model and the latter “new” data for validation (Sheridan, 2013; Liu et al., 2015). We have previously shown in a time-split validation that, whereas the accuracy of a vNN model is roughly maintained, the number of “new” compounds that it can predict is significantly reduced. However, by simply adding a few “new” compounds, the coverage increases significantly (Liu et al., 2015).

The lack of training data poses an important limitation to the vNN approach. When a dataset is too small, there is a high probability that a target molecule will have no qualified near neighbors in the dataset, and hence a high-quality prediction cannot be made. However, the lack of training data is a limitation for all machine learning methods. The difference is that most such methods build a model no matter how small the training dataset, and will always make a prediction for any input molecule without considering the reliability of the predicted result. In our view, it is better not to give a prediction at all if it is unreliable. This also alerts users to use alternative methods, including experimental measurements, to derive a reliable answer. As more experimental data become available over time, the performance of the vNN method will improve without retraining. This is in contrast to most other machine learning methods, which cannot take advantage of new data without retraining a model.

This finding is especially significant for drug discovery labs because the chemical space is restricted by the target candidates they are investigating. For example, when exploring a new drug target, it is crucial to continuously update the model with new data to ensure that the applicability domain is relevant for the

new target. In a vNN-based model, this can be done easily by adding the SMILES strings of the new compounds to the reference dataset. For this reason, we believe that our web-based vNN platform has the potential to greatly accelerate the development of drugs.

AUTHOR CONTRIBUTIONS

PS, RL, and AW developed the method, analyzed the data, and wrote the manuscript. VD designed and implemented the web server.

FUNDING

The authors were supported by the U.S. Army Medical Research and Materiel Command (Fort Detrick, MD), and the Defense Threat Reduction Agency grant CBCall14-CBS-05-2-0007.

ACKNOWLEDGMENTS

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense. This paper has been approved for public release with unlimited distribution.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2017.00889/full#supplementary-material>

REFERENCES

- Abbott, N. J., Rönnbäck, L., and Hansson, E. (2006). Astrocyte-endothelial interactions at the blood-brain barrier. *Nat. Rev. Neurosci.* 7, 41–53. doi: 10.1038/nrn1824
- Adler, J., and Parmryd, I. (2010). Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry A* 77, 733–742. doi: 10.1002/cyto.a.20896
- Ambudkar, S. V., Kimchi-Sarfaty, C., Sauna, Z. E., and Gottesman, M. M. (2003). P-glycoprotein: from genomics to mechanism. *Oncogene* 22, 7468–7485. doi: 10.1038/sj.onc.1206948
- Ames, B. N., Durston, W. E., Yamasaki, E., and Lee, F. D. (1973). Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc. Natl. Acad. Sci. U.S.A.* 70, 2281–2285. doi: 10.1073/pnas.70.8.2281
- Assis, D. N., and Navarro, V. J. (2009). Human drug hepatotoxicity: a contemporary clinical perspective. *Expert Opin. Drug Metab. Toxicol.* 5, 463–473. doi: 10.1517/17425250902927386
- Attene-Ramos, M. S., Huang, R., Michael, S., Witt, K. L., Richard, A., Tice, R. R., et al. (2015). Profiling of the Tox21 chemical collection for mitochondrial function to identify compounds that acutely decrease mitochondrial membrane potential. *Environ. Health Perspect.* 123, 49–56. doi: 10.1289/ehp.1408642
- Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., et al. (2014). The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* 42, D1083–D1090. doi: 10.1093/nar/gkt1031
- Borst, P., and Elferink, R. O. (2002). Mammalian ABC transporters in health and disease. *Annu. Rev. Biochem.* 71, 537–592. doi: 10.1146/annurev.biochem.71.102301.093055
- Broccatelli, F., Carosati, E., Neri, A., Frosini, M., Goracci, L., Oprea, T. I., et al. (2011). A novel approach for predicting P-glycoprotein (ABCB1) inhibition using molecular interaction fields. *J. Med. Chem.* 54, 1740–1751. doi: 10.1021/jm101421d
- Brown, C. M., Reisfeld, B., and Mayeno, A. N. (2008). Cytochromes P450: a structure-based summary of biotransformations using representative substrates. *Drug Metab. Rev.* 40, 1–100. doi: 10.1080/03602530701836662
- Chen, L., Li, Y., Zhao, Q., Peng, H., and Hou, T. (2011). ADME evaluation in drug discovery. 10. Predictions of P-glycoprotein inhibitors using recursive partitioning and naive Bayesian classification techniques. *Mol. Pharm.* 8, 889–900. doi: 10.1021/mp100465q
- Chen, M., Vijay, V., Shi, Q., Liu, Z., Fang, H., and Tong, W. (2011). FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discov. Today* 16, 697–703. doi: 10.1016/j.drudis.2011.05.007
- Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., et al. (2012). admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. *J. Chem. Inf. Model.* 52, 3099–3105. doi: 10.1021/ci300367a
- Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., et al. (2014). QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* 57, 4977–5010. doi: 10.1021/jm4004285
- Contrera, J. F., Matthews, E. J., Kruhlik, N. L., and Benz, R. D. (2004). Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. *Regul. Toxicol. Pharmacol.* 40, 185–206. doi: 10.1016/j.yrtph.2004.08.004
- Czodrowski, P. (2013). hERG me out. *J. Chem. Inf. Model.* 53, 2240–2251. doi: 10.1021/ci400308z
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7:42717. doi: 10.1038/srep42717
- De Ponti, F., Poluzzi, E., and Montanaro, N. (2001). Organising evidence on QT prolongation and occurrence of Torsades de Pointes with non-antiarrhythmic drugs: a call for consensus. *Eur. J. Clin. Pharmacol.* 57, 185–209. doi: 10.1007/s002280100290
- Di, L., Kerns, E. H., Hong, Y., Kleintop, T. A., McConnell, O. J., and Huryn, D. M. (2003). Optimization of a higher throughput microsomal stability screening assay for profiling drug discovery candidates. *J. Biomol. Screen.* 8, 453–462. doi: 10.1177/1087057103255988
- Duan, J., Dixon, S. L., Lowrie, J. F., and Sherman, W. (2010). Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *J. Mol. Graph. Model.* 29, 157–170. doi: 10.1016/j.jmgl.2010.05.008
- Dunn, G., and Everitt, B. (1995). *Clinical Biostatistics: An Introduction to Evidence-based Medicine*. London: E. Arnold.
- Greene, N., Fisk, L., Naven, R. T., Note, R. R., Patel, M. L., and Pelletier, D. J. (2010). Developing structure–activity relationships for the prediction of hepatotoxicity. *Chem. Res. Toxicol.* 23, 1215–1222. doi: 10.1021/tx1000865
- Hansen, K., Mika, S., Schroeter, T., Sutter, A., ter Laak, A., Steger-Hartmann, T., et al. (2009). Benchmark data set for *in silico* prediction of Ames mutagenicity. *J. Chem. Inf. Model.* 49, 2077–2081. doi: 10.1021/ci900161g
- Hert, J., Willett, P., Wilton, D. J., Acklin, P., Azzaoui, K., Jacoby, E., et al. (2004). Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem.* 2, 3256–3266. doi: 10.1039/b409865j
- Huang, R., Xia, M., Nguyen, D.-T., Zhao, T., Sakamuru, S., Zhao, J., et al. (2016). Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. *Front. Environ. Sci.* 3:85. doi: 10.3389/fenvs.2015.00085
- Li, D., Chen, L., Li, Y., Tian, S., Sun, H., and Hou, T. (2014). ADMET evaluation in drug discovery. 13. Development of *in silico* prediction models for P-glycoprotein substrates. *Mol. Pharm.* 11, 716–726. doi: 10.1021/mp400450m
- Liew, C. Y., Lim, Y. C., and Yap, C. W. (2011). Mixed learning algorithms and features ensemble in hepatotoxicity prediction. *J. Comput. Aided Mol. Des.* 25, 855. doi: 10.1007/s10822-011-9468-3
- Liu, R., Schyman, P., and Wallqvist, A. (2015). Critically assessing the predictive power of QSAR models for human liver microsomal stability. *J. Chem. Inf. Model.* 55, 1566–1575. doi: 10.1021/acs.jcim.5b00255
- Liu, R., Tawa, G., and Wallqvist, A. (2012). Locally weighted learning methods for predicting dose-dependent toxicity with application to the human maximum recommended daily dose. *Chem. Res. Toxicol.* 25, 2216–2226. doi: 10.1021/tx300279f
- Liu, R., and Wallqvist, A. (2014). Merging applicability domains for *in silico* assessment of chemical mutagenicity. *J. Chem. Inf. Model.* 54, 793–800. doi: 10.1021/ci500016v
- Manganaro, A., Pizzo, F., Lombardo, A., Pogliaghi, A., and Benfenati, E. (2016). Predicting persistence in the sediment compartment with a new automatic software based on the k-Nearest Neighbor (k-NN) algorithm. *Chemosphere* 144, 1624–1630. doi: 10.1016/j.chemosphere.2015.10.054
- Maunz, A., Gütlein, M., Rautenberg, M., Vorgrimmler, D., Gebele, D., and Helma, C. (2013). lazar: a modular predictive toxicology framework. *Front. Pharmacol.* 4:38. doi: 10.3389/fphar.2013.00038
- Meyer, J. N., Leung, M. C., Rooney, J. P., Sendoel, A., Hengartner, M. O., Kisby, G. E., et al. (2013). Mitochondria as a target of environmental toxicants. *Toxicol. Sci.* 134, 1–17. doi: 10.1093/toxsci/kft102
- Muehlbacher, M., Spitzer, G. M., Liedl, K. R., and Kornhuber, J. (2011). Qualitative prediction of blood–brain barrier permeability on a large and refined dataset. *J. Comput. Aided Mol. Des.* 25, 1095–1106. doi: 10.1007/s10822-011-9478-1
- Murray, M. (2006). Role of CYP pharmacogenetics and drug–drug interactions in the efficacy and safety of atypical and other antipsychotic agents. *J. Pharm. Pharmacol.* 58, 871–885. doi: 10.1211/jpp.58.7.0001
- Naef, R. (2015). A generally applicable computer algorithm based on the group additivity method for the calculation of seven molecular descriptors: heat of combustion, logPO/W, logS, refractivity, polarizability, toxicity and logBB of organic compounds; scope and limits of applicability. *Molecules* 20:18279. doi: 10.3390/molecules201018279
- Piecznik, S. R., and Neustadt, J. (2007). Mitochondrial dysfunction and molecular pathways of disease. *Exp. Mol. Pathol.* 83, 84–92. doi: 10.1016/j.yexmp.2006.09.008
- Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. *J. Chem. Inf. Model.* 50, 742–754. doi: 10.1021/ci100050t
- Sanguinetti, M. C., and Tristani-Firouzi, M. (2006). hERG potassium channels and cardiac arrhythmia. *Nature* 440, 463–469. doi: 10.1038/nature04710
- Schyman, P., Liu, R., and Wallqvist, A. (2016). Using the variable-nearest neighbor method to identify P-glycoprotein substrates and inhibitors. *ACS Omega* 1, 923–929. doi: 10.1021/acsomega.6b00247
- Sheridan, R. P. (2013). Time-split cross-validation as a method for estimating the goodness of prospective prediction. *J. Chem. Inf. Model.* 53, 783–790. doi: 10.1021/ci400084k

- Sushko, I., Novotarskyi, S., Körner, R., Pandey, A. K., Rupp, M., Teetz, W., et al. (2011). Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.* 25, 533–554. doi: 10.1007/s10822-011-9440-2
- Walker, T., Grulke, C. M., Pozefsky, D., and Tropsha, A. (2010). Chembench: a cheminformatics workbench. *Bioinformatics* 26, 3000–3001. doi: 10.1093/bioinformatics/btq556
- Wang, S., Li, Y., Wang, J., Chen, L., Zhang, L., Yu, H., et al. (2012). ADMET evaluation in drug discovery. 12. Development of binary classification models for prediction of hERG potassium channel blockage. *Mol. Pharm.* 9, 996–1010. doi: 10.1021/mp300023x
- Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., and Bryant, S. H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* 37, W623–W633. doi: 10.1093/nar/gkp456
- Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., et al. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* 14, 475–486. doi: 10.1038/nrd4609
- Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* 28, 31–36. doi: 10.1021/ci00057a005
- Xu, J. J., Henstock, P. V., Dunn, M. C., Smith, A. R., Chabot, J. R., and de Graaf, D. (2008). Cellular Imaging predictions of clinical drug-induced liver injury. *Toxicol. Sci.* 105, 97–105. doi: 10.1093/toxsci/kfn109
- Xu, Y., Dai, Z., Chen, F., Gao, S., Pei, J., and Lai, L. (2015). Deep learning for drug-induced liver injury. *J. Chem. Inf. Model.* 55, 2085–2093. doi: 10.1021/acs.jcim.5b00238
- Zheng, W., and Tropsha, A. (2000). Novel variable selection quantitative structure–property relationship approach based on the k-nearest-neighbor principle. *J. Chem. Inf. Comput. Sci.* 40, 185–194. doi: 10.1021/ci980033m

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Schyman, Liu, Desai and Wallqvist. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.