Supplementary Information

SMILES2vec: An Interpretable General-Purpose Deep Neural Network for Predicting Chemical Properties

Selection of final SMILES2vec Model

Having selected the CNN-GRU architectural class as the optimal network design, we now examine the hyperparameter details. We observed that the top performing models for the Tox21 and FreeSolv tasks shares limited similarity.

For example, as detailed in Table S1, the majority of the top 10 designs for FreeSolv have 192 filters in the convolutional layer, and 384 units in the second GRU layer, and the designs differ minimally from one another. In contrast, the top 10 designs for Tox21 are more varied, with two noticeable clusters – one has >200 units in the second GRU layer, and the other has <80 units in the second GRU layer.

Tox21				
em_size	**# conv**	**# rnn1**	**# rnn2**	**AUC**
60	176	184	8	0.880
40	160	168	80	0.878
50	124	136	216	0.877
60	104	232	264	0.876
50	88	96	360	0.874
60	52	168	208	0.873
60	132	168	8	0.873
50	144	8	64	0.873
60	176	104	40	0.873
20	120	320	24	0.872
FreeSolv				
em_size	**# conv**	**# rnn1**	**# rnn2**	**RMSE**
40	192	288	384	1.075
50	192	224	384	1.096
50	192	128	384	1.113
50	172	288	384	1.115
50	168	136	384	1.125
40	192	328	384	1.125
50	192	208	384	1.129
40	192	384	112	1.132
50	192	320	384	1.145
50	192	352	384	1.154

Table S1: Top 10 models for the CNN-GRU architectural class for the Tox21 (AUC) and FreeSolv (RMSE) tasks.

In order to ensure convergence to a more common neural network design, we used the top 10 models for each task as additional designs to be evaluated for the other task. We observed that 3 designs, summarized in Table S2 (also highlighted in Table S1), were ranked comparably with the top 10 models identified for each task. Therefore, we selected all 3 designs as potential candidate models for the final SMILES2vec model.

Candidate	em_size	#conv	#rnn1	#rnn2	AUC	RMSE
A	50	192	224	384	0.875	1.096
B	50	192	128	384	0.873	1.113
C	60	104	232	264	0.876	1.156

Table S2: Best performing models for the CNN-GRU architectural class for both Tox21 (AUC) and FreeSolv (RMSE) datasets.

It is also interesting to note that despite having sampled a total of 120 different network designs across 2 different chemical tasks, our initial Bayesian optimization results did not initially converge to a common network design, and we had to use an additional final manual selection step to select candidate models that perform well on both tasks. It possible that additional trials may eventually lead to convergence. In retrospect, modifications to the Bayesian optimizer to include multiple objective functions from both classification and regression tasks may improve the rate of convergence.

Further test of generalizing the candidate SMILES2vec model were performed, where the same optimized network designs were used to predict the full Tox21 dataset and the HIV dataset. As summarized in Table 3, we compared the performance metrics of the three SMILES2vec candidate models, with the best models for each dataset highlighted as indicated. Using the validation performance metrics to evaluate the quality of each model, for the full Tox21 dataset, all 3 models performed similarly (AUC ~0.80). For the HIV dataset, models A and B did slightly better (AUC ~0.78), and model A (RMSE 1.14 kcal/mol) was the top performer for the FreeSolv dataset.

Furthermore, we note that in all models, the difference between the validation and test metrics is small, thus confirming the generalization of the model to compounds it has not seen either during the model training, or in the Bayesian hyperparameter optimization. Considering the consistent good performance of Model A, it was selected as the final SMILES2vec model.

Dataset	Model	Train AUC/RMSE	Val AUC/RMSE	Test AUC/RMSE
Tox21	A	0.856	0.799	0.795
	B	0.861	0.798	0.794
	C	0.855	0.797	0.795
HIV	A	0.868	0.781	0.765
	B	0.847	0.780	0.769
	C	0.844	0.773	0.762
FreeSolv	A	0.624	1.135	1.017
	B	0.9	1.184	1.105
	C	0.858	1.194	1.121

Table S3: Evaluation of the 3 candidate models for SMILES2vec on Tox21, HIV and FreeSolv datasets.