Supplementary/Supporting Information

Room temperature ionic liquids viscosity prediction from deep-learning models

Zafer Acar[1,3], Phu Nguyen[2,3], Xiaoqi Cui[3], Kah Chun Lau[3]

[1]Department of Physics, Michigan Technological University, Houghton, MI 49931, USA.

[2]Department of Computer Science, San José State University, San José, CA 95192-0249, USA.

[3]Department of Physics and Astronomy, California State University Northridge, Northridge, CA 91330, USA.

Correspondence to: Prof./Dr. Kah Chun Lau, Department of Physics and Astronomy, California State University Northridge, Northridge, CA 91330, USA. E-mail: kahchun.lau@csun.edu
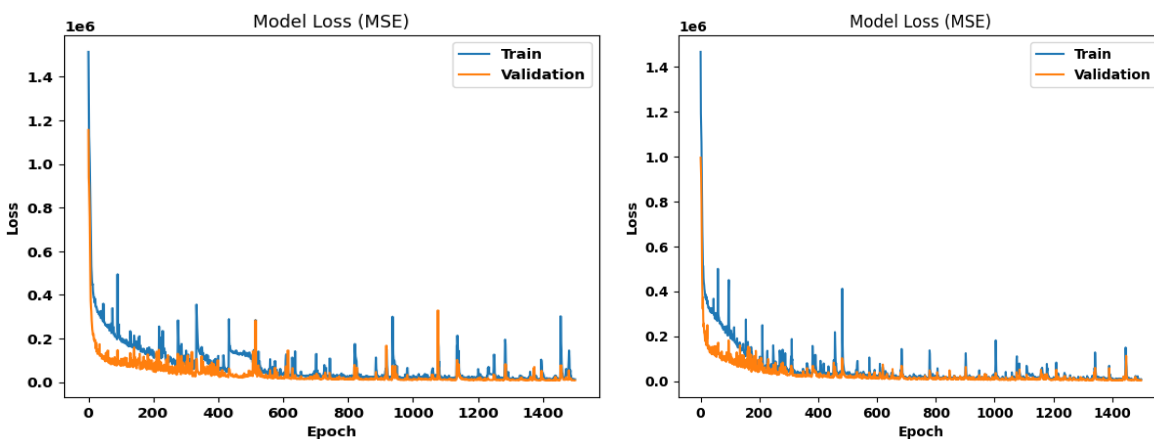
Figure S1: The convergence test of the error rate vs. the number of epochs in DL models: (left) DNN model; (right) CNN model.

Table S1: The top 20 most important molecular descriptors obtained from DNN and CNN models based on Pearson correlation for regression analysis of IL viscosity.

Deep Neural Network (DNN)		Convolutional Neural Network (CNN)	
Descriptors	**Descriptors Type**	**Descriptors**	**Descriptors Type**
viscosity	IL viscosity	viscosity	IL viscosity
MPC10	walk and path counts	MPC10	walk and path counts

B02[C-S]	2D atom pairs	C-011	atom-centered fragments
C-011	atom-centered fragments	B02[C-S]	2D atom pairs
D/Dtr06	ring descriptors	D/Dtr06	ring descriptors
CATS2D_07_DL	CATS2D	B07[C-O]	2D atom pairs
B03[C-O]	2D atom pairs	P_VSA_e_4	P_VSA descriptors
CATS2D_07_AL	CATS2D	JGI9	2D autocorrelations
B05[O-O]	2D atom pairs	CATS2D_07_DL	CATS2D
piPC09	walk and path counts	piPC09	walk and path counts
B04[N-S]	2D atom pairs	B06[C-O]	2D atom pairs
F04[N-S]	2D atom pairs	IDDE	information indices
CATS2D_03_DA	CATS2D	F05[C-N]	2D atom pairs
CATS2D_08_NL	CATS2D	B05[O-O]	2D atom pairs
CATS2D_08_DL	CATS2D	piPC10	walk and path counts
F06[C-O]	2D atom pairs	CATS2D_03_AL	CATS2D
CATS2D_03_AL	CATS2D	CATS2D_06_DL	CATS2D
NdssC	atom-type E-state indices	F04[N-S]	2D atom pairs
F05[C-N]	2D atom pairs	B04[N-S]	2D atom pairs
SAdon	molecular properties	NdssC	atom-type E-state indices

| dense_20_input | input: | [(None, 179)] |
| InputLayer | output: | [(None, 179)] |

| dense_20 | input: | (None, 179) |
| Dense | output: | (None, 4814) |

| dense_21 | input: | (None, 4814) |
| Dense | output: | (None, 2048) |

| dense_22 | input: | (None, 2048) |
| Dense | output: | (None, 512) |

| dense_23 | input: | (None, 512) |
| Dense | output: | (None, 256) |

| dense_24 | input: | (None, 256) |
| Dense | output: | (None, 128) |

| dense_25 | input: | (None, 128) |
| Dense | output: | (None, 64) |

| dense_26 | input: | (None, 64) |
| Dense | output: | (None, 32) |

| dense_27 | input: | (None, 32) |
| Dense | output: | (None, 16) |

| dense_28 | input: | (None, 16) |
| Dense | output: | (None, 8) |

| dense_29 | input: | (None, 8) |
| Dense | output: | (None, 1) |

A)

| dense_26_input | input: | [(None, 20)] |
| InputLayer | output: | [(None, 20)] |

| dense_26 | input: | (None, 20) |
| Dense | output: | (None, 4814) |

| dense_27 | input: | (None, 4814) |
| Dense | output: | (None, 2048) |

| dense_28 | input: | (None, 2048) |
| Dense | output: | (None, 512) |

| dense_29 | input: | (None, 512) |
| Dense | output: | (None, 256) |

| dense_30 | input: | (None, 256) |
| Dense | output: | (None, 128) |

| dense_31 | input: | (None, 128) |
| Dense | output: | (None, 64) |

| dense_32 | input: | (None, 64) |
| Dense | output: | (None, 32) |

| dense_33 | input: | (None, 32) |
| Dense | output: | (None, 16) |

| dense_34 | input: | (None, 16) |
| Dense | output: | (None, 8) |

| dense_35 | input: | (None, 8) |
| Dense | output: | (None, 1) |

B)

```python
model = Sequential()
model.add(Dense(4814, activation='relu'))
model.add(Dense(2048, activation='relu'))
model.add(Dense(512, activation='relu'))
model.add(Dense(256, activation='relu'))
model.add(Dense(128, activation='relu'))
model.add(Dense(64, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(16, activation='relu'))
model.add(Dense(8, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile(loss='binary_crossentropy', optimizer="adam", metrics=["accuracy"])
```
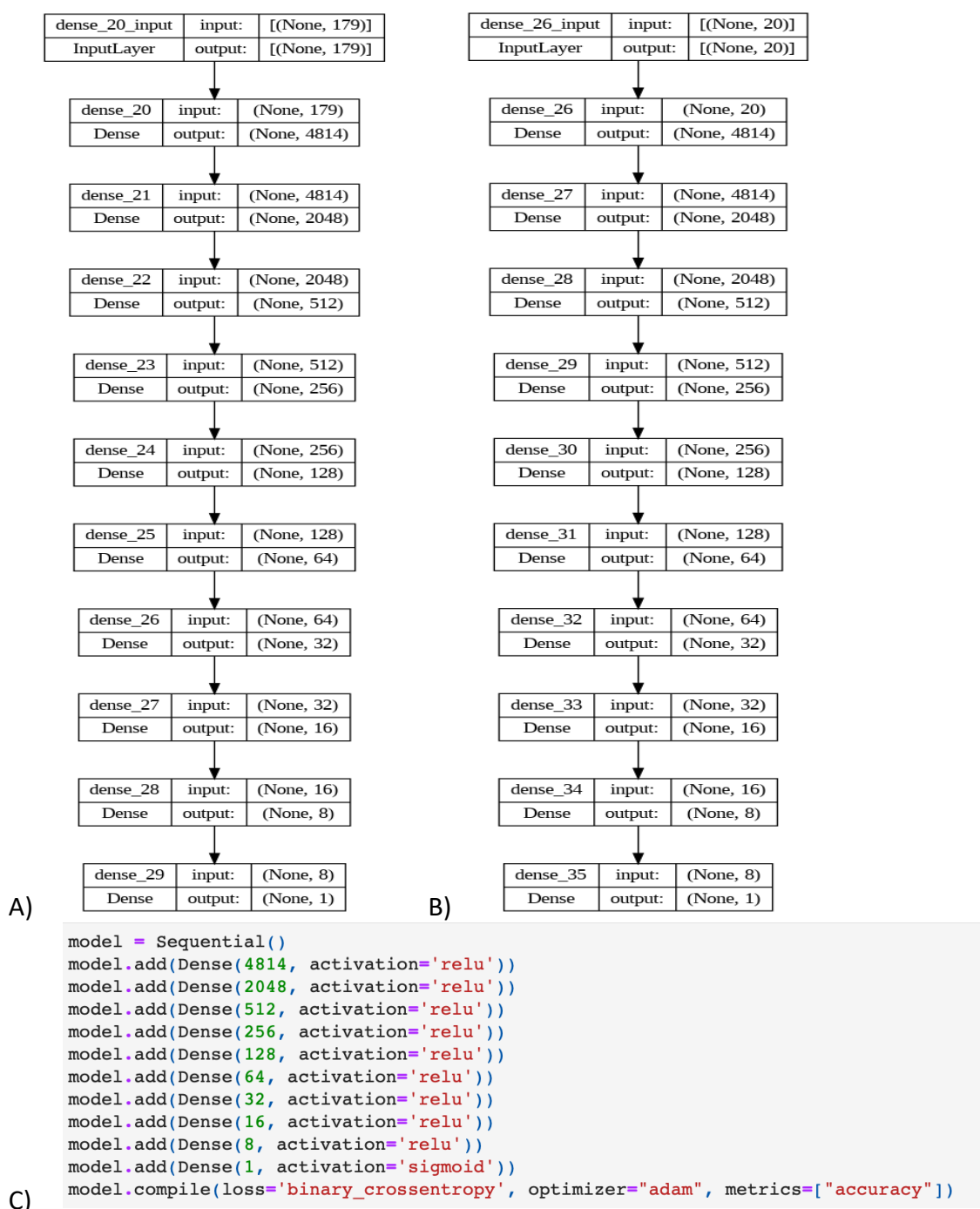
C)

Figure S2. The schematic plot of the low/high-η binary classification model based on a deep neural network (DNN) that uses (A) 179 and (B) 20 selected molecular features/descriptors. This DNN model for low/high-η binary classification consists of one input layer (179 neurons), 1st hidden layer (4814 neurons), 2nd hidden layer (2048 neurons), 3rd hidden layer (512 neurons), 4th hidden layer (256 neurons), 5th hidden layer (128 neurons), 6th hidden layer (64 neurons), 7th hidden layer (32 neurons), 8th hidden layer (16 neurons), 9th hidden layer (8 neurons), and an output layer (1 neuron). (C) A portion of the Python code for this DNN model based on 20 selected molecular descriptors shown in (B).
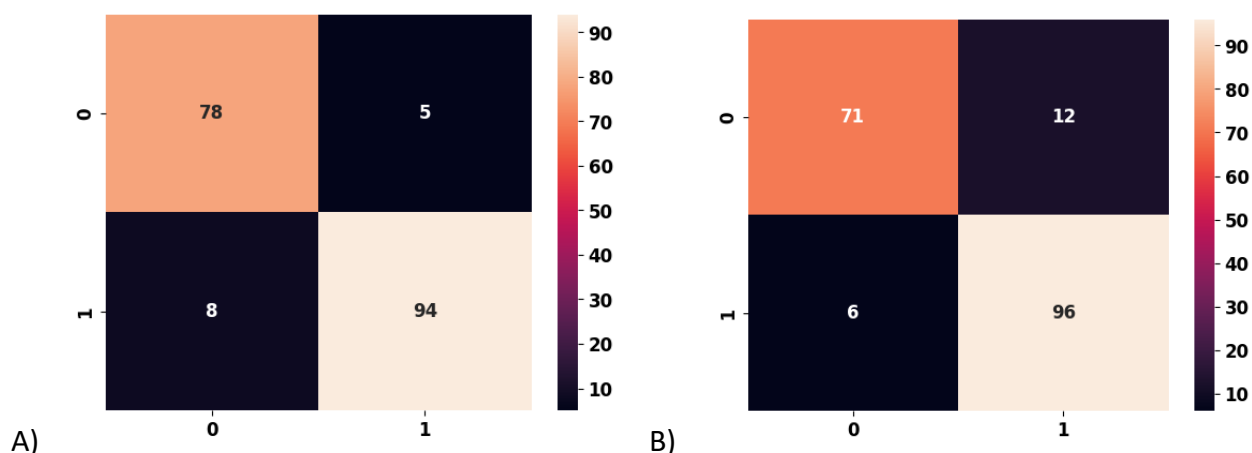
A)   B)

Figure S3: For the test dataset (i.e., 20% of the dataset or 185 types of ILs), the computed confusion matrix heatmap is generated based on the DNN model (Figure S1) by using the selected (A) 179 and (B) 20 important molecular descriptors. For both Figures A and B, the negative (low-η) and positive (high-η) cases are represented by "0" and "1" separately. The number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) can be found in the segment of (1,1), (0,0), (0,1), and (1,0) separately.

Table S2: The top 20 most important molecular descriptors and their description obtained from the DNN model for low/high-η binary classification based on Pearson correlation.

Deep Neural Network (DNN)		
Descriptors Name	Description	Descriptors Type
viscosity	viscosity	IL viscosity
ATS8m	Boto-Moreau autocorrelation of lag 8 (log function) weighted by mass	2D autocorrelations
MPC10	Molecular path count of order 10	walk and path counts
piPC10	Molecular multiple paths of order 10	walk and path counts
piPC09	Molecular multiple paths of order 9	walk and path counts
CATS2D_06_AL	CATS2D Acceptor-Lipophilic at lag 6	CATS2D
MPC09	Molecular path count of order 9	walk and path counts
MPC08	Molecular path count of order 8	walk and path counts
P_VSA_ppp_L	P_VSA like on potential pharmacophore points, L − lipophilic	P_VSA-like descriptors
ATSC7i	Boto-Moreau autocorrelation of lag 7 (log function) weighted by ionization potential	2D Autocorrelations

P_VSA_e_4	P_VSA like on Sanderson electronegativity, bin 4	P_VSA-like descriptors
NdssC	Number of atoms of type ddC (-=C<)	atom-type E-state indices
B07[F-F]	Presence/absence of F-F at topological distance 7	2D atom pairs
F-083	F-attached to C3(sp3)	atom-centered fragments
B04[N-F]	Presence/absence of N-F at topological distance 4	2D atom pairs
F05[S-F]	Frequency of S-F at topological distance 5	2D atom pairs
P_VSA_LogP_3	P_VSA-like on LogP, bin 3	P_VSA-like descriptors
CATS2D_03_DA	CATS2D Donor-Acceptor at lag 03	CATS2D
F06[O-F]	Frequency of O-F at topological distance 6	2D atom pairs
F07[F-F]	Frequency of F-F at topological distance 7	2D atom pairs