

**Supporting information to the article 'A
Generalized Machine Learning Model for
Predicting Ionic Conductivity for Ionic Liquids'**

Pratik Dhakal and Jindal K.Shah*

E-mail: jindal.shah@okstate.edu

Equations

Mean Absolute Error

$$\text{MAE} = \frac{\sum_{i=1}^N |y_i^{\text{pred}} - y_i^{\text{exp}}|}{N} \quad (1)$$

Root Mean Squared Error

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i^{\text{pred}} - y_i^{\text{exp}})^2}{N}} \quad (2)$$

Residual Deviation

$$RD = (y^{\text{exp}} - y^{\text{pred}}) \quad (3)$$

Where:

N is the number of data points; y^{exp} is experimental data, and y^{pred} is predicted data using model.

Cation Type

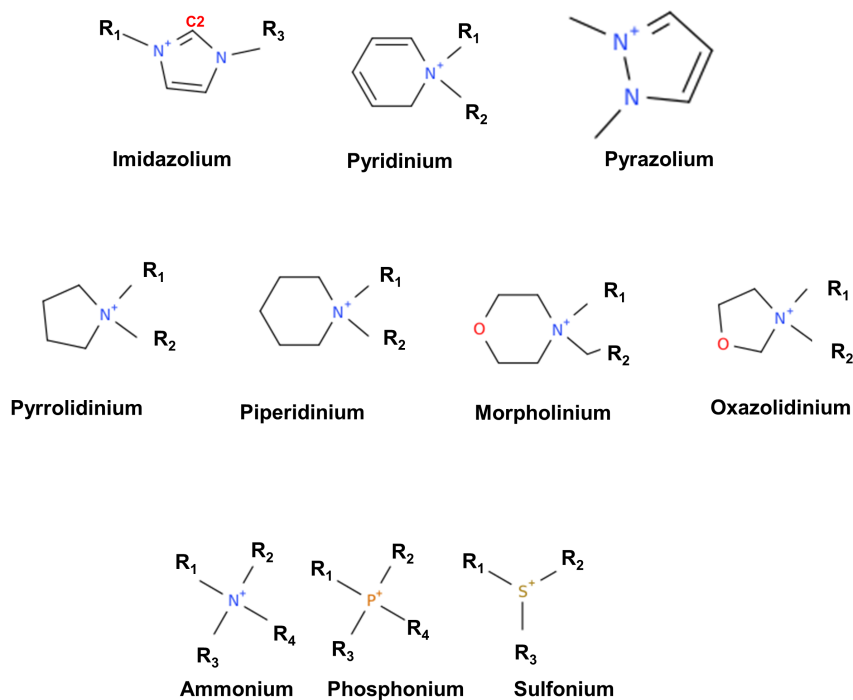


Figure S1: Schematics of cation type used in this study for model development. R₁, R₂, R₃, and R₄ refer to substitutes in the cation such as alkyl, ether, cyano and hydroxyl functional group. Images were drawn using OPSIN package.¹

Table S1: A detailed breakdown of the number of experimental ionic conductivity data for various cation type in the model development data set.

Cation type	Data points
Imidazolium	1440
Pyridinium	493
Ammonium	280
Pyrrolidinium	273
Phosphonium	161
Sulfonium	74
Piperidinium	67
Morpholinium	61
Pyrazolium	11
Oxazolidinium	9
Total	2869

Hyperparameter Feature Space

Random Forest

```
param_grid = {  
    'bootstrap': [True],  
    'max_depth': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,  
                 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35,  
                 36, 37, 38, 39, 40],  
    'n_estimators': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000,  
                   1100, 1200, 1300, 1400, 1500]  
}
```

Figure S2: Hyper-parameter space for Random Forest model. 'max_depth' indicates the maximum depth of a decision tree, while 'n_estimators' denotes the number of decision trees generated for computing the average of the outputs to yield a prediction.

Final hyper-parameter : max_depth = 35, n_estimators = 500

XGBoost

```
params = { 'max_depth': [3, 4, 5, 6, 7, 8, 9, 10, 15, 17, 20, 23, 25, 27, 30, 33, 35, 37, 40],  
          'learning_rate': [0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.0065, 0.007,  
                           0.0072, 0.0075, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09,  
                           0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9],  
          'subsample': np.arange(0.1, 1.0, 0.1),  
          'colsample_bytree': np.arange(0.1, 1.0, 0.1),  
          'colsample_bylevel': np.arange(0.1, 1.0, 0.1),  
          'n_estimators': [50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 650, 700,  
                          750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500]}
```

Figure S3: Hyper-parameter space for XGBoost model. 'max_depth' indicates the maximum depth of an individual tree; 'learning_rate' refers to the step size for the gradient descent method; 'subsample' denotes the fraction of data chosen at random to train an individual tree; 'colsample_bytree' controls the fraction of features chosen at random to train an individual tree; 'colsample_bylevel' identifies the fraction of features selected at random to train each node in a tree. Finally, 'n_estimators' is the total number of trees for computing the average of outputs to yield a prediction.

Final hyper-parameters: n_estimators = 900, max_depth = 17, learning_rate = 0.07, colsample_bytree = 0.8, subsample = 0.4, colsample_bylevel = 0.1

Model Validation

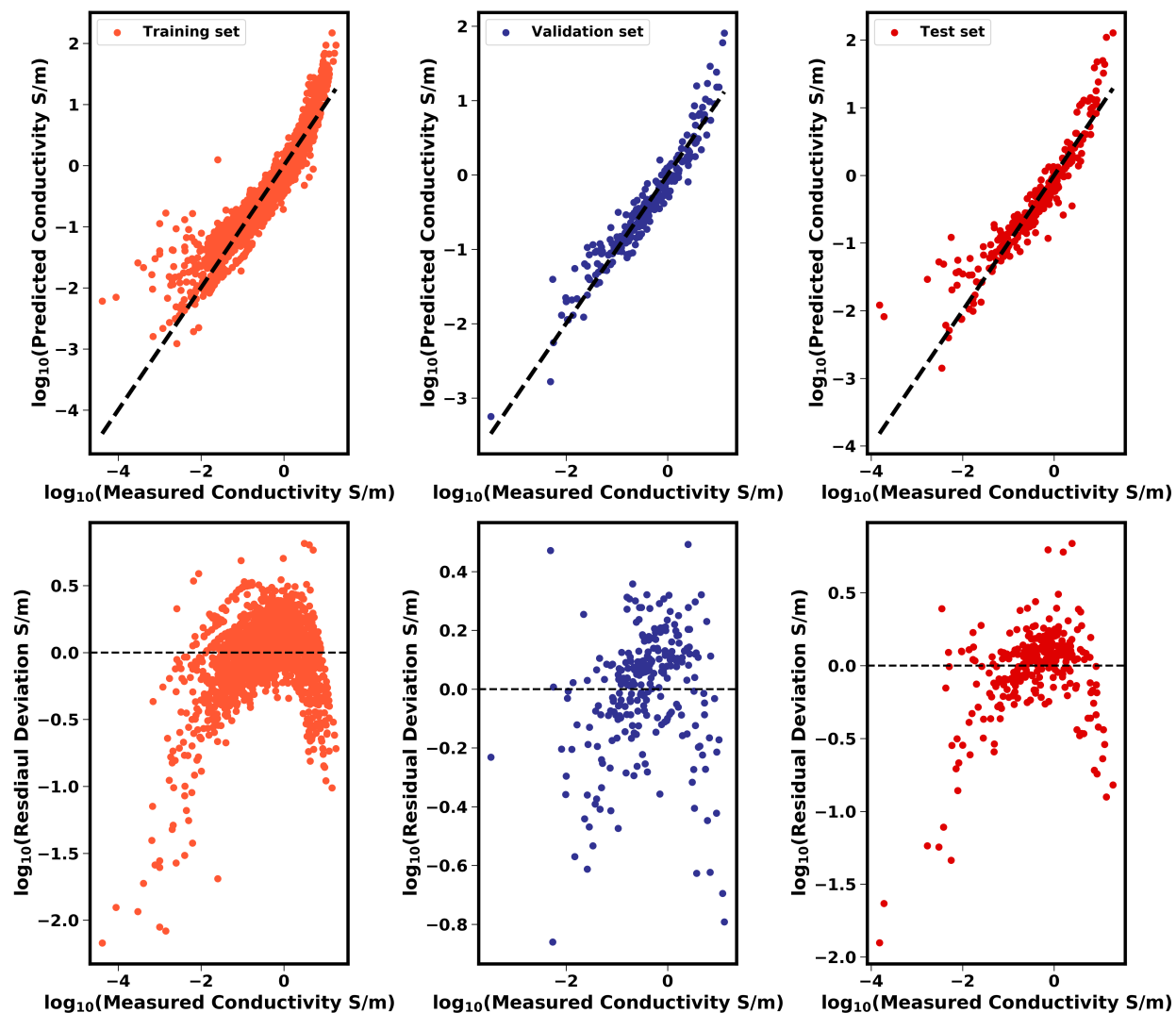


Figure S4: Comparison of model predictions with the experimental data on \log_{10} scale for the training set (left pane), validation set (middle pane) and test set (right pane) using Multiple Linear Regression model.

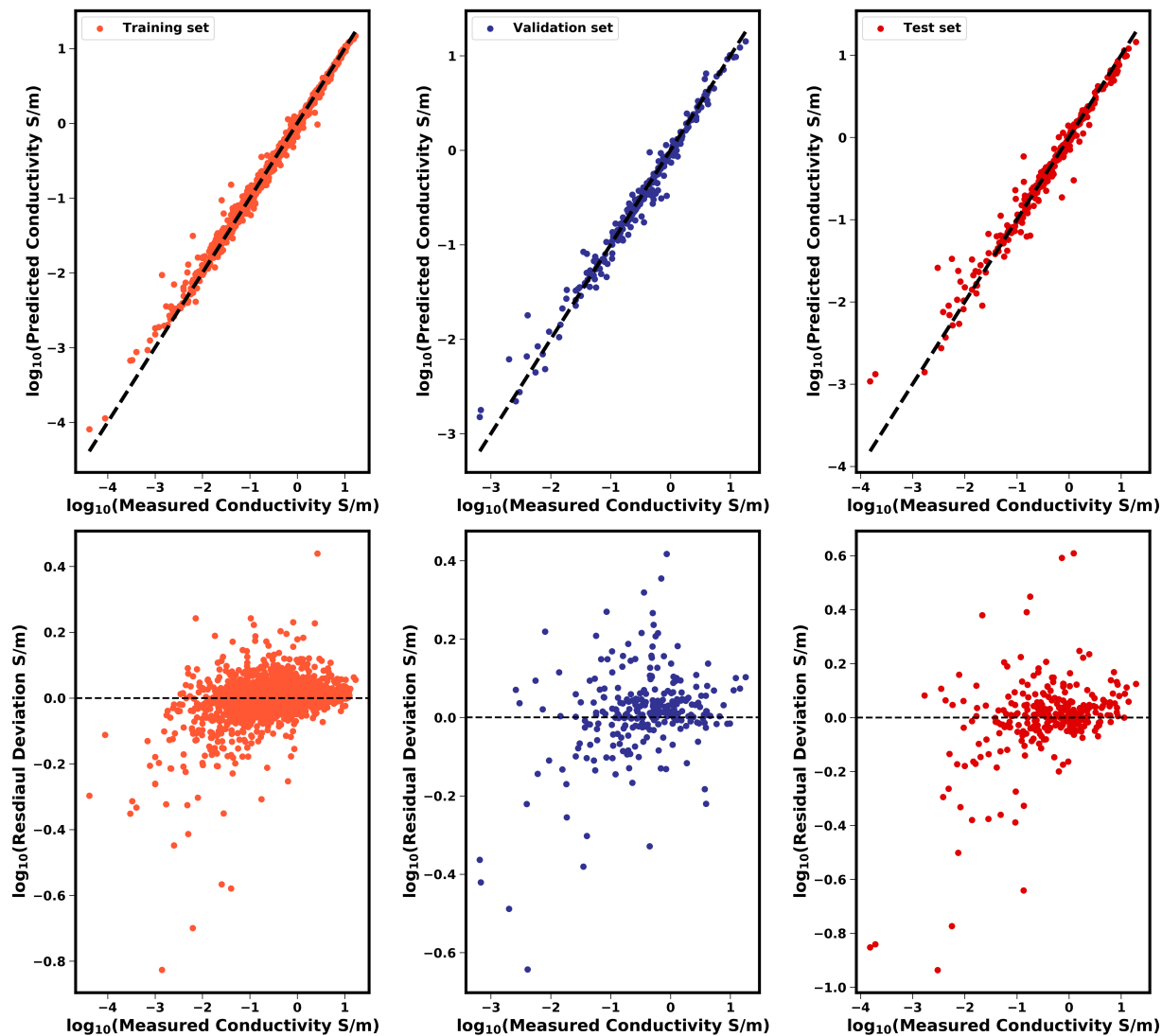


Figure S5: Comparison of model predictions with the experimental data on \log_{10} scale for the training set (left pane), validation set (middle pane) and test set (right pane) using Random Forest model.

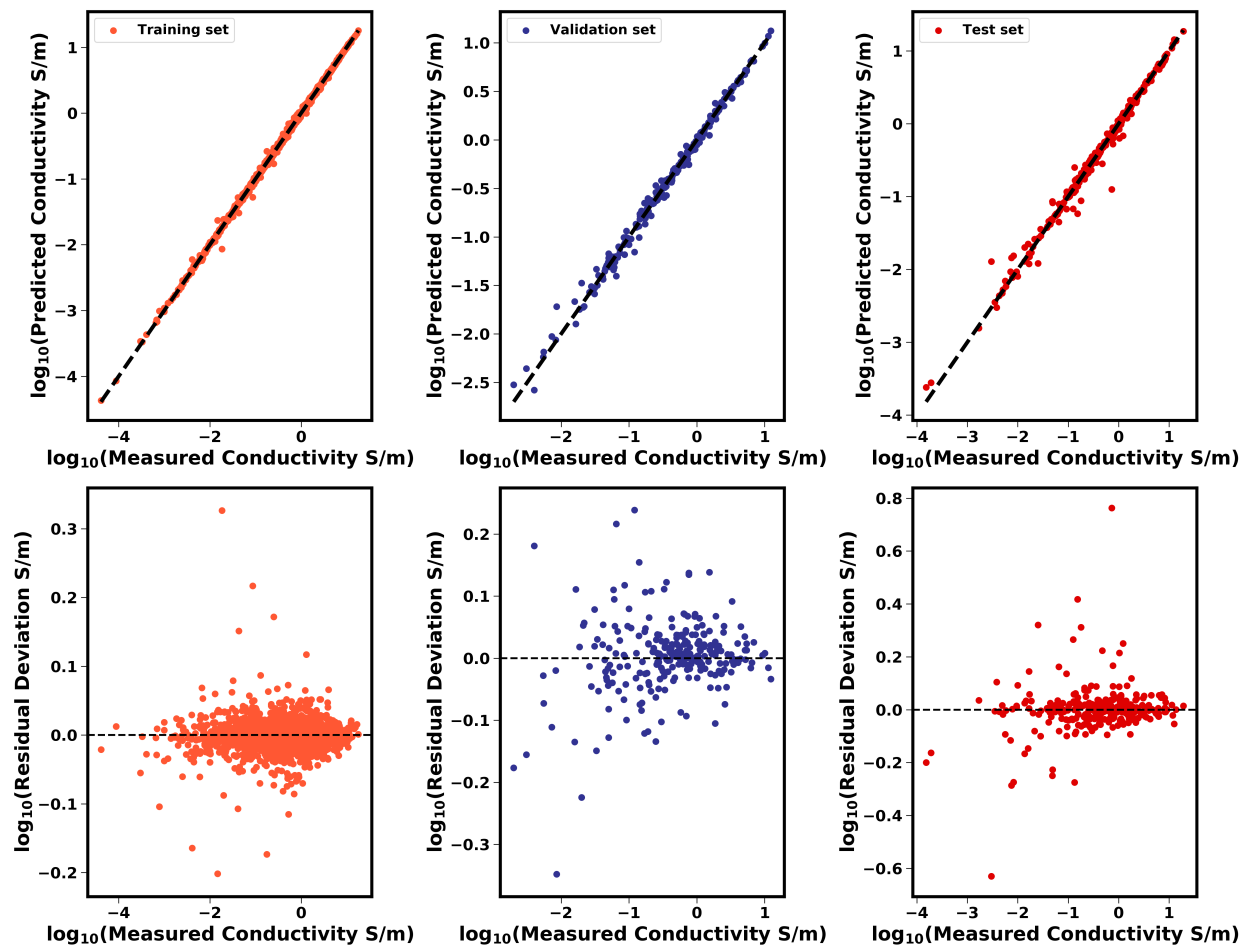


Figure S6: Comparison of model predictions with the experimental data on \log_{10} scale for the training set (left pane), validation set (middle pane) and test set (right pane) using XGBoost model.

Performance by cation type

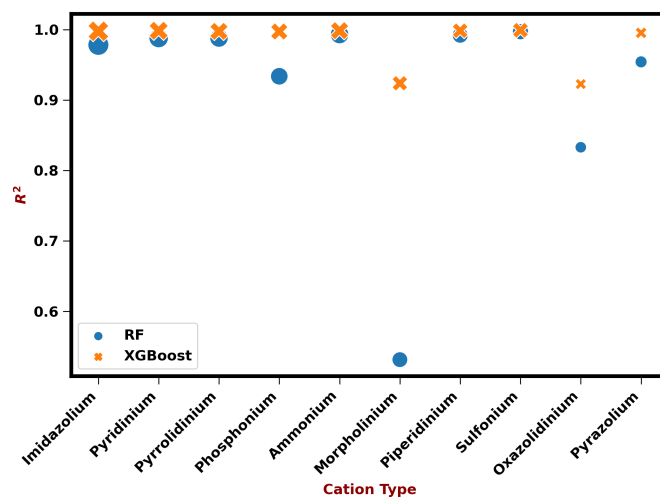


Figure S7: Correlation coefficient (R^2) for RF (circles) and XGBoost (crosses) models for the entire data set as a function of cation family. Size of the marker indicates the relative proportion of data points present for the given cation type.

External Test set

This section is focused on evaluating the model's performance on the external test case that consists of ionic liquids missing from the model development. Figure S8 shows the comparison in prediction between the model and experimental data with R^2 of 0.80, RMSE of 0.20 S/m, and MAE of 0.14 S/m for the 30 data set on the normal ionic conductivity scale.

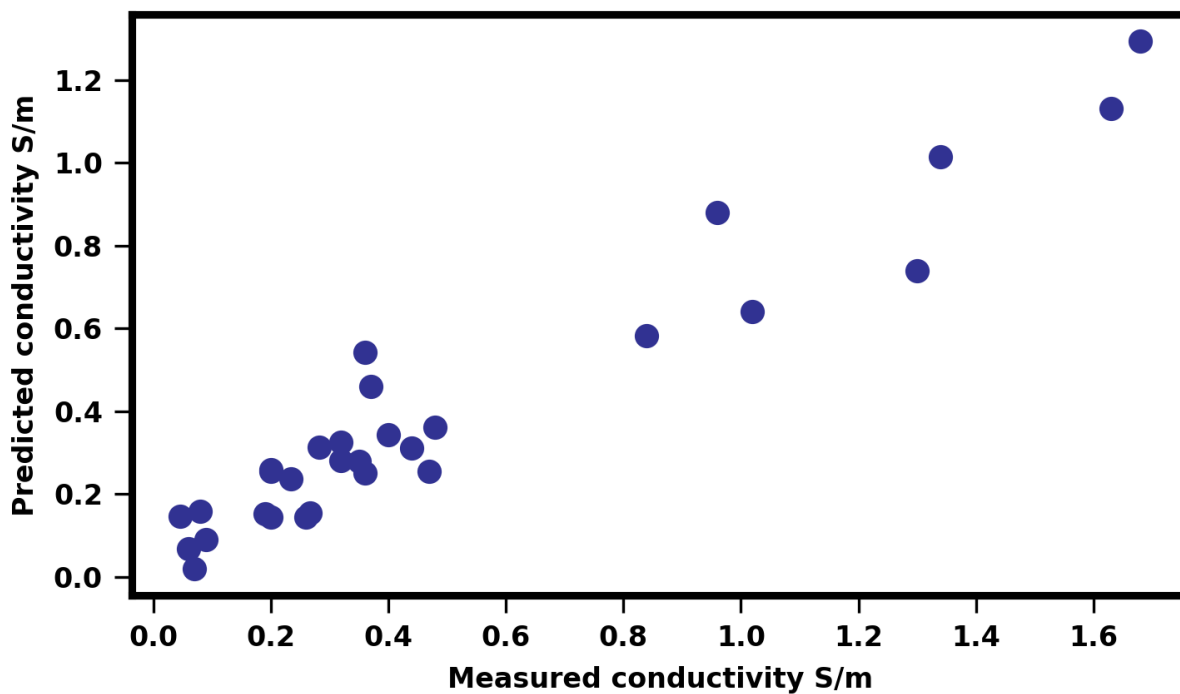


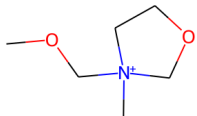
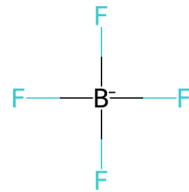
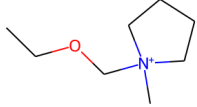
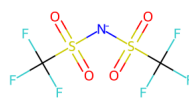


Figure S8: Comparison of experimental external data set and XGBoost model prediction. The ionic conductivity values are reported in S/m.

Table S2 depicts the unique ionic liquid combination evaluation. The first in the list are the two oxazolidinium cation which is rarely studied with limited data available in the literature. Out of eleven data points from this family collected from literature, nine of them were included in the training set, while two of the data points were added to this test case. The first cation 3-methyl-3-methoxyethyloxazolidinium cation is not part of the training set, while the cation 3-methyl-3-methoxymethyloxazolidinium is present in the training data with other anions but not with tetrafluoroborate $[\text{BF}_4]^-$ anion. Thus, this test case was added to validate the model's ability to predict ionic conductivity for cation families with a very small fraction of representation in the training set as seen in Table S2.

Table S2: Comparison of ionic conductivity between experimental data (Exp) and XGBoost predictions for unique ionic liquids for which either the cation or the anion is a part of the training data set but not the combinations shown here. Value inside the square bracket denotes the temperature at which the measurement was taken.

No.	Cation Structure	Anion Structure	Exp	XGBoost	Ref
1			0.09[298.15]	0.09	2
2			0.08[298.15]	0.16	2
3			0.37[298.15]	0.46	2

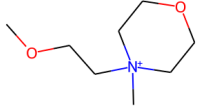

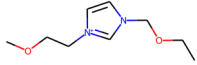

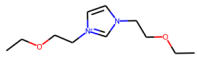

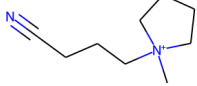

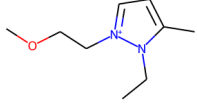

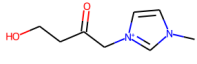
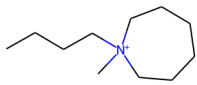

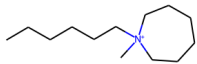

4			0.06[298.15] 0.07	2
5			0.28[298.15] 0.31	3
6			0.27[298.15] 0.15	3
7			0.05[298.15] 0.15	4
8			0.24[298.15] 0.24	5
9		Cl ⁻	0.07[293.15] 0.02	6

Table S3 demonstrates the model’s ability to generalize prediction beyond the cation training set. The list of the cations shown in the table are cations that were not part of the model development. However, they are structurally similar to some of the cations present in the model database. Cations (1) and (2) belong to the azepanium family, similar to the piperidinium-based cation. The piperidinium cation is a six-ring cyclic structure, while the azepanium cation is a seven-ring cyclic structure. The remaining cations belong to the pyrrolinium cation family, which are found to be more stable, with better transport properties than the common pyrrolidinium-based cations.⁷ These cations have a minimal resemblance to any cations present in the set besides pyrrolidinium cations which are still far from similar to them. The pyrrolidinium cations present in the training set are cyclic cations with no double bonds and oxygen functional groups attached. This could be why the model’s quantitative prediction has a relatively significant deviation compared to other test cases seen earlier. However, the qualitative trends have close agreement compared to experimental data.

Table S3: Comparison of ionic conductivity between experimental data (Exp) and XGBoost predictions for cations that bear close resemblance to the cation types investigated in this work. Value inside the square bracket denotes the temperature at which the measurement was taken.

No.	Cation Structure	Anion Structure	Exp	XGBoost	Ref
1			0.36[333.15]	0.54	8
2			0.20[333.15]	0.25	8

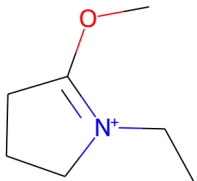
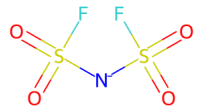
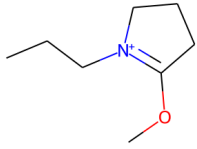
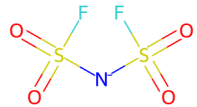
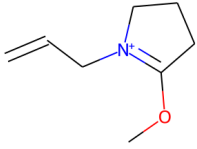
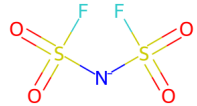
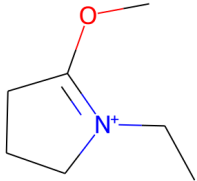
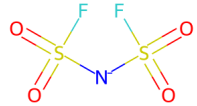
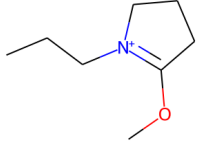
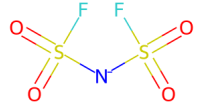
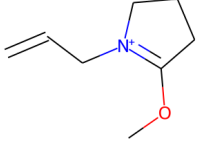
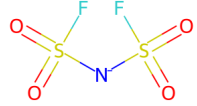
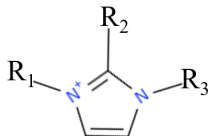
3			1.30[298.15]	0.73	7
4			0.84[298.15]	0.58	7
5			1.02[298.15]	0.64	7
6			1.68[323.15]	1.29	7
7			1.34[323.15]	1.01	7
8			1.63[323.15]	1.13	7

Table S4 demonstrates the predictive capability of the model when the functional groups are

present at various positions in the imidazolium cation; the anion is represented by $[\text{NTf}_2]^-$. One of the unique advantages of ionic liquids is the design flexibility that allows a large number of possible cations with different functional group attachments at various positions. Thus it is important to know whether a given model can predict ionic conductivity correctly as the functional group location is varied. This data set originally contained fifteen data points. Out of which, one ether-functionalized cation and one allyl-functionalized cation were added to the training set to ensure the model has seen such cations with functional groups located at R_3 position. The rest of the cations are separated as test cases to gauge the model’s ability in predicting ionic conductivity.

Table S4: Comparison of ionic conductivity at 298 K between experiment⁹ and XGBoost prediction for cations paired with bis(trifluoromethylsulfonyl)imide $[\text{NTf}_2]^-$ anion.



No.	R_1	R_2	R_3	Experiment(S/m)	XGBoost
1	Methyl	H	Ethyl	0.96	0.88
2	Methyl	Methyl	Ethyl	0.47	0.25
3	Methyl	Methyl	n-Propyl	0.36	0.25
4	Methyl	Methyl	n-Butyl	0.26	0.14
5	Methyl	Methyl	n-Pentyl	0.20	0.14
6	Methyl	Methyl	Methoxyethyl	0.35	0.28
7	Methyl	Methyl	Ethoxyethyl	0.32	0.33
8	Allyl	Methyl	Ethyl	0.48	0.36
9	Allyl	Methyl	n-Propyl	0.40	0.34
10	Allyl	Methyl	n-Butyl	0.20	0.26
11	Allyl	Methyl	n-Pentyl	0.19	0.15
12	Allyl	Methyl	Allyl	0.44	0.31
13	Allyl	Methyl	Ethoxyethyl	0.32	0.28

Feature Importance

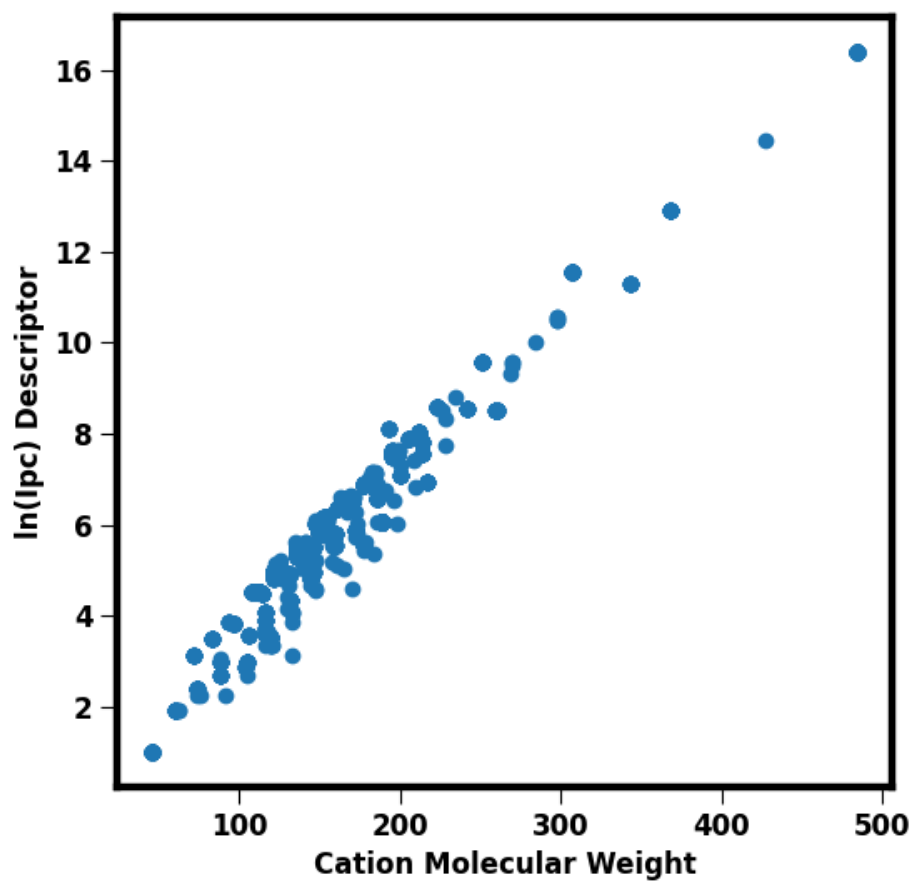


Figure S9: Cation feature $\ln(I_{pc})$ vs molecular weights of the cations for which experimental data were included in the model development.

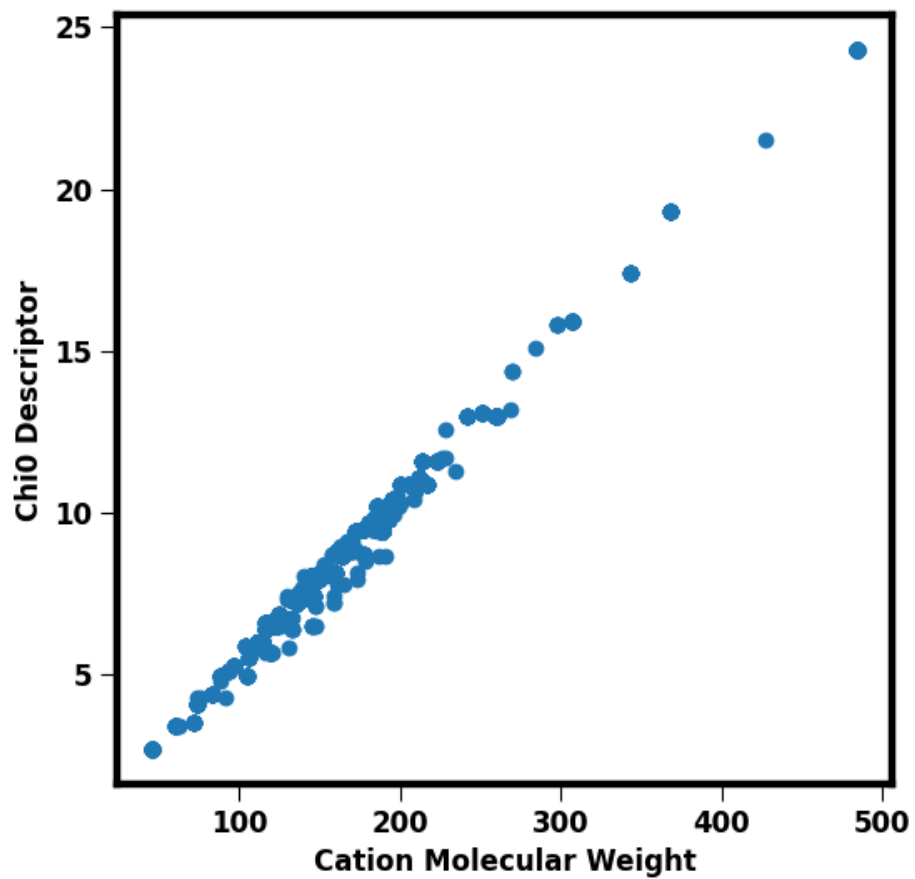


Figure S10: Cation feature Chi0 vs molecular weight of the cations for which experimental data were used in the model development.

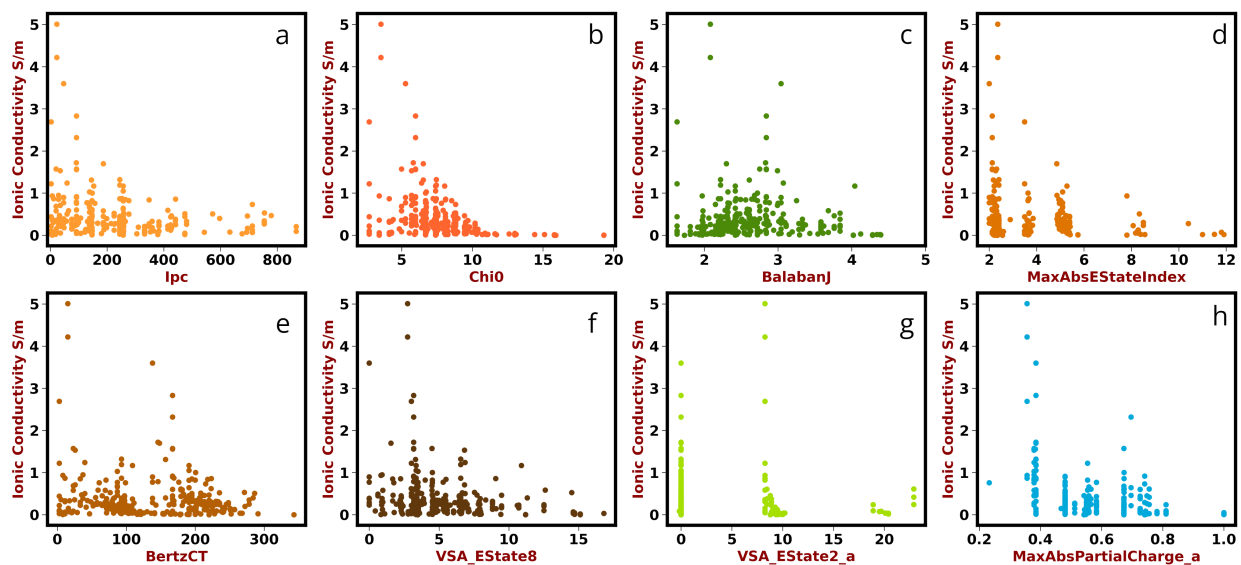


Figure S11: Ionic conductivity vs. cation and anion features deemed to impact the ionic conductivity the most in the XGBoost model. The ionic conductivity data are plotted for the experimental data at 298.15 K. Features ending with 'a' indicates features for anions.

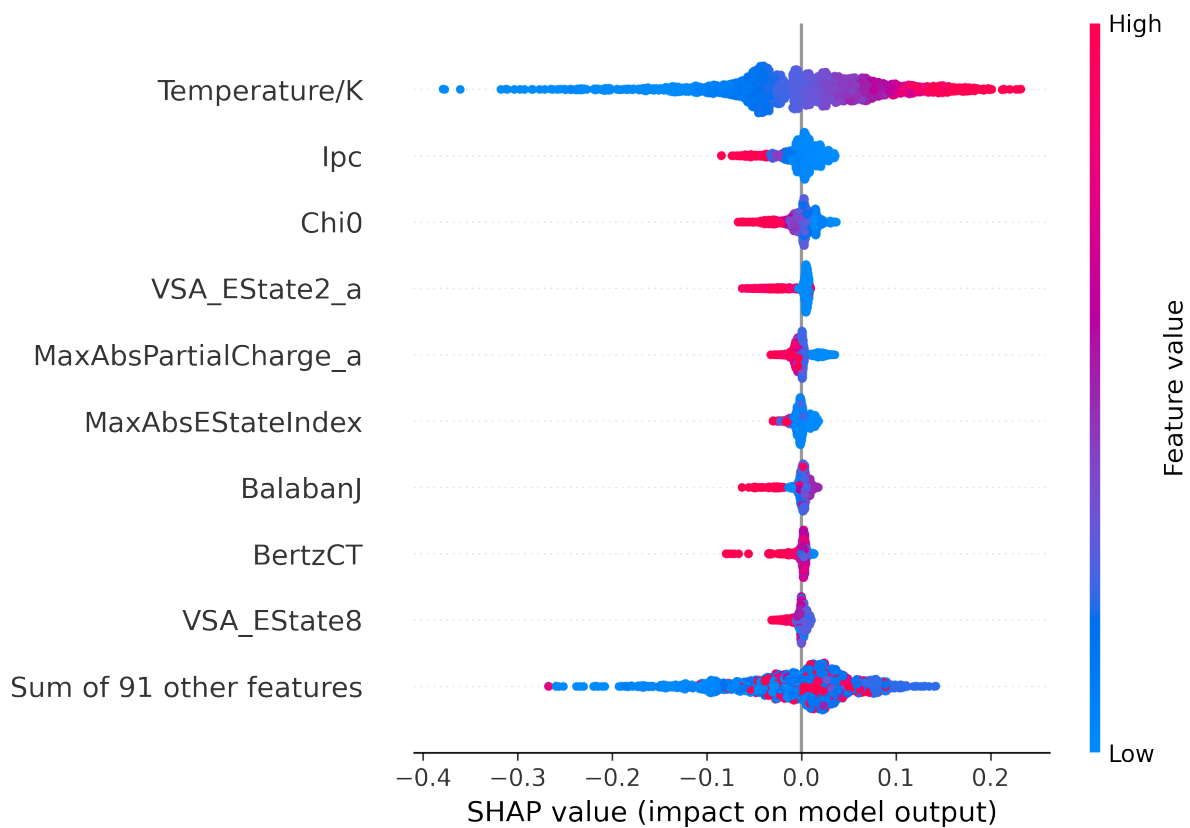


Figure S12: SHAP feature importance for the training set data. Features ending with 'a' indicates features for anions.

External Set Classification

Based on the SHAP analysis, it is clear that some of the features have a very high influence on the ionic conductivity compared to the rest. In this section, we attempt to build a decision tree-based classification model to leverage insights generated from the SHAP analysis. The primary objective here is to determine the accuracy of such a model using a few selected features as inputs for the model development. We selected experimental data at 298 K to construct the classification model, as the number of ionic conductivity data (337 in total) at this temperature is the highest. Furthermore, we only considered ionic liquids that contained the bis(trifluoromethanesulfonyl)imide $[\text{NTf}_2]^-$. In selecting this anion, we took into consideration two points: (a) it is one of the most commonly studied anion and it featured in 137 ionic liquids at the selected temperature; (b) a model based solely on the cation descriptors would fail to capture the ionic conductivity changes if the data set contained multiple anions. Although such a model may appear restrictive, it can actually provide a reference point. For example, if a novel cation paired with $[\text{NTf}_2]^-$ anion is classified to have a 'high' conductivity value, then it is very likely that the cation when paired with anions with faster dynamics than the bulky $[\text{NTf}_2]^-$ anion.

The boundary for separating the ionic conductivity between 'high' and 'low' is set to the median value of experimental ionic conductivity, which is 0.265 S/m, where any ionic liquids below median value fall under the 'low' ionic conductivity category. Although alternative values of the ionic conductivity can be assigned to demarcate the two classes, the choice of the median eliminates any bias in the classification boundary as there is an equal number of cations for both categories. The decision tree (DT) model is built using Scikit-learn¹⁰ with 90% for the data set aside for training the model and 10% left for test purposes. The features for the model are the six cation features shown in the SHAP plot (Figure S12). The classification model achieves 98% accuracy for the training set and 92% for the test set. The accuracy is very high considering that the model is built only with six descriptors

and an artificial classification boundary. Furthermore, we evaluated the performance of the classification model for a few ionic liquids in an external data set with $[\text{NTf}_2]^-$ as the anion (Table S5 and S6). The model correctly classified 63% of the ionic liquids with seven ionic liquids mislabeled out of 19 data points. Four of the mislabeled ionic liquids have ionic conductivity in the 0.20-0.27 S/m range close to the boundary separating the two categories. Thus, this demonstrates the possibility of rapidly screening ionic liquids with a model that is built using only six cation descriptors.

Table S5: Classification of external test case cations paired with $[\text{NTf}_2]^-$ at 298 K. Low refers to ionic conductivity less than 0.265 S/m.

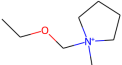
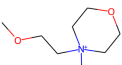
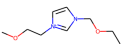
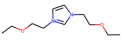
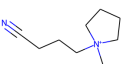
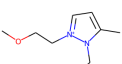
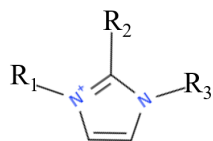
No.	Cation Structure	Exp(S/m)	Type _{exp}	Type _{pred}	Ref
1		0.37	High	Low	2
2		0.06	Low	Low	2
3		0.28	High	High	3
4		0.27	High	Low	3
5		0.05	Low	Low	4
6		0.24	Low	High	5

Table S6: Classification of external test case cations paired with $[\text{NTf}_2]^-$ at 298 K. Low refers to ionic conductivity less than 0.265 S/m. Experiment data taken from ref⁹.



No.	R ₁	R ₂	R ₃	Experiment(S/m)	Type _{exp}	Type _{pred}
1	Methyl	H	Ethyl	0.96	High	High
2	Methyl	Methyl	Ethyl	0.47	High	High
3	Methyl	Methyl	n-Propyl	0.36	High	High
4	Methyl	Methyl	n-Butyl	0.26	High	High
5	Methyl	Methyl	n-Pentyl	0.20	Low	High
6	Methyl	Methyl	Methoxyethyl	0.35	High	Low
7	Methyl	Methyl	Ethoxyethyl	0.32	Low	Low
8	Allyl	Methyl	Ethyl	0.48	High	High
9	Allyl	Methyl	n-Propyl	0.40	High	High
10	Allyl	Methyl	n-Butyl	0.20	Low	High
11	Allyl	Methyl	n-Pentyl	0.19	Low	Low
12	Allyl	Methyl	Allyl	0.44	High	High
13	Allyl	Methyl	Ethoxyethyl	0.32	High	Low

Unique Ionic Liquids

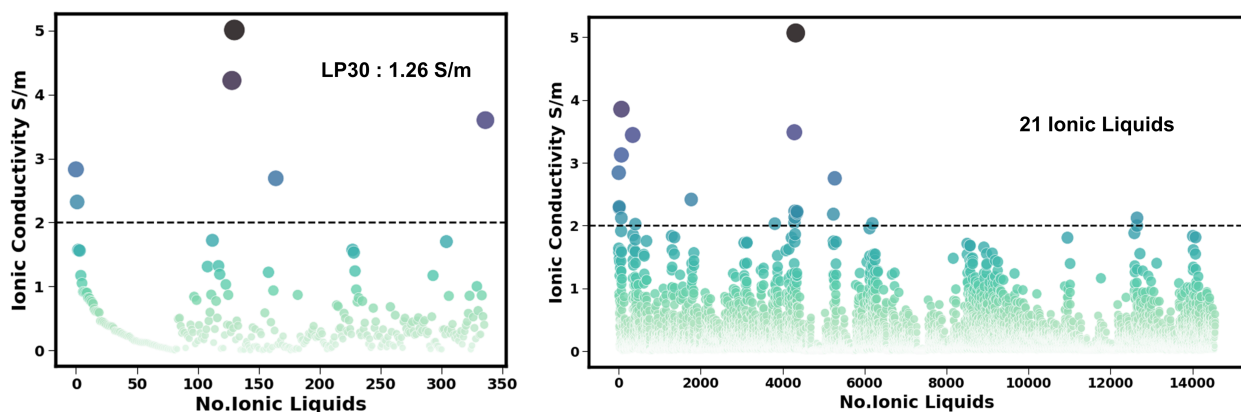


Figure S13: (a) Experimental data at 298.15 K. (b) Unique ionic liquid predictions at 298.15 K using XGBoost model. LP30 here refers to the commonly used Li-ion electrolyte with a high ionic conductivity of 1.26 S/m at 298.15 K^{11,12}. Addition of Li⁺ salts to ionic liquids is known to reduce ionic conductivity by 30-50%.⁷

Table S7: Ionic Liquids with ionic conductivity greater than 2.0 S/m at 298.15 K calculated using unique ionic liquid method based on XGBoost model.

Cation	Anion	Ionic Conductivity (S/m)
pyrrolidinium	nitrate	5.07
1-ethyl-3-methylimidazolium	cyanoborohydride	3.85
pyrrolidinium	dicyanamide	3.49
1,3-dimethylimidazolium	dicyanamide	3.44
1-ethyl-3-methylimidazolium	dicyanoborohydride	3.13
1-ethyl-3-methylimidazolium	dicyanamide	2.84
ethylammonium	nitrate	2.75
1-methylimidazolium	dicyanamide	2.42
1-ethyl-3-methylimidazolium	thiocyanate	2.30
1-ethyl-3-methylimidazolium	tricyanomethane	2.28
pyrrolidinium	tricyanomethane	2.23
pyrrolidinium	dicyanoborohydride	2.22
pyrrolidinium	tricyanoborohydride	2.20
ethylammonium	dicyanamide	2.18
pyrrolidinium	thiocyanate	2.13
1-ethyl-2-methylpyrazolium	cyanoborohydride	2.12
1-ethyl-3-methylimidazolium	tricyanoborohydride	2.12
pyrrolidinium	bis(fluorosulfonyl)imide	2.05
diethylmethylammonium	dicyanamide	2.03
1-methylpyridinium	dicyanamide	2.03
1,3-dimethylimidazolium	dicyanoborohydride	2.02

References

- (1) Lowe, D. M.; Corbett, P. T.; Murray-Rust, P.; Glen, R. C. Chemical name to structure: OPSIN, an open source solution. 2011.
- (2) Zhou, Z.-B.; Matsumoto, H.; Tatsumi, K. Cyclic quaternary ammonium ionic liquids with perfluoroalkyltrifluoroborates: synthesis, characterization, and properties. *Chemistry—A European Journal* **2006**, *12*, 2196–2212.
- (3) Zhang, J.; Fang, S.; Qu, L.; Jin, Y.; Yang, L.; Hirano, S.-i. Synthesis, characterization, and properties of ether-functionalized 1, 3-dialkylimidazolium ionic liquids. *Industrial & Engineering Chemistry Research* **2014**, *53*, 16633–16643.

- (4) Baek, B.; Lee, S.; Jung, C. Pyrrolidinium cation-based ionic liquids with different functional groups: butyl, butyronitrile, pentenyl, and methyl butyrate. *Int. J. Electrochem. Sci* **2011**, *6*, 6220–6234.
- (5) Chai, M.; Jin, Y.; Fang, S.; Yang, L.; Hirano, S.-i.; Tachibana, K. Low-viscosity ether-functionalized pyrazolium ionic liquids as new electrolytes for lithium battery. *Journal of Power Sources* **2012**, *216*, 323–329.
- (6) Tot, A.; Podlipnik, Č.; Bešter-Rogač, M.; Gadžurić, S.; Vraneš, M. Influence of oxygen functionalization on physico-chemical properties of imidazolium based ionic liquids—Experimental and computational study. *Arabian Journal of Chemistry* **2020**, *13*, 1598–1611.
- (7) Kim, H.-T.; Kwon, O. M.; Mun, J.; Oh, S. M.; Yim, T.; Kim, Y. G. Novel pyrrolinium-based ionic liquids for lithium ion batteries: effect of the cation on physicochemical and electrochemical properties. *Electrochimica Acta* **2017**, *240*, 267–276.
- (8) Pohlmann, S.; Olyschläger, T.; Goodrich, P.; Vicente, J. A.; Jacquemin, J.; Balducci, A. Azepanium-based ionic liquids as green electrolytes for high voltage supercapacitors. *Journal of Power Sources* **2015**, *273*, 931–936.
- (9) Kakibe, T.; Yoshimoto, N.; Egashira, M.; Morita, M. Optimization of cation structure of imidazolium-based ionic liquids as ionic solvents for rechargeable magnesium batteries. *Electrochemistry communications* **2010**, *12*, 1630–1633.
- (10) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
- (11) Tsurumaki, A.; Agostini, M.; Poiana, R.; Lombardo, L.; Lufrano, E.; Simari, C.; Matic, A.; Nicotera, I.; Panero, S.; Navarra, M. A. Enhanced safety and galvanostatic performance of high voltage lithium batteries by using ionic liquids. *Electrochimica Acta* **2019**, *316*, 1–7.

- (12) Tarascon, J.; Guyomard, D. New electrolyte compositions stable over the 0 to 5 V voltage range and compatible with the $\text{Li}^{1+} \text{xMn}_2\text{O}_4$ /carbon Li-ion cells. *Solid State Ionics* **1994**, *69*, 293–305.