

LEVERAGING ATOMISTIC SIMULATIONS AND MACHINE LEARNING
FOR THE DESIGN OF IONIC LIQUIDS AS ELECTROLYTES FOR
BATTERY APPLICATION

By

PRATIK DHAKAL

Bachelor of Science in Chemical Engineering
Shandong University of Science and Technology
Qingdao, Shandong Province, China
2016

Master of Science in Chemical Engineering
Miami University
Oxford, Ohio, United States
2018

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May 2022

LEVERAGING ATOMISTIC SIMULATIONS AND MACHINE LEARNING
FOR THE DESIGN OF IONIC LIQUIDS AS ELECTROLYTES FOR
BATTERY APPLICATION

Dissertation Approved:

Jindal K. Shah

Dissertation Adviser

Omer Ozgur Capraz

Yu Feng

Sandip P. Harimkar

ACKNOWLEDGEMENTS

I want to thank my family for their support and love throughout this entire journey, without which this would not have been possible. I also would like to express my gratitude towards Rina Harianto for all her support throughout the years.

I would also like to thank my advisor, Dr. Jindal K. Shah, for all his support, mentoring, and feedback since I joined the research group. I always felt truly blessed and extremely grateful for the opportunity to work in such a productive and healthy environment within his research group. As an aspiring researcher, I could not have asked for a better role model professionally throughout this journey. Beyond research, I am also grateful for his understanding and support on a personal level.

I am also highly thankful to my committee members, Dr. Omer Ozgur Capraz, Dr. Yu Feng, and Dr. Sandip P. Harimkar, for agreeing to be a part of my committee and providing valuable feedback for this work. I would also like to thank all the faculty members and administrative personal at the chemical engineering department for all their support and assistance throughout this journey.

I would also like to acknowledge OSU for making high-performance computing easily accessible, which has been an extremely valuable resource for this work. In addition, I would also like to thank Jesse Shaffer from the OSU high performance computing center for providing continuous technical support during the time of need. I am also thankful for all of my previous research colleagues, Dr. Utkarsh Kapoor, for his guidance, and Dr. Atiya Banerjee for his support and assistance throughout this journey. I also want to express my sincere thanks to my previous

research colleagues, Reagen Black, Wyatt Gassaway and Phillip Pergande, and my current research colleagues, Brinkli Abbitt, Sudip K. Das, and Amey Thorat, for their support and feedback on my presentations.

Lastly, I would also like to give a special thanks to my master's advisor, Dr. Andrew S. Paluch, for his mentoring and support through graduate college and my undergraduate research advisor, Dr. Gao Jun, for his support towards my research interests.

Acknowledgements reflect the views of the author and are not endorsed by committee members or Oklahoma State University

Name: PRATIK DHAKAL

Date of Degree: MAY, 2022

Title of Study: LEVERAGING ATOMISTIC SIMULATIONS AND MACHINE LEARNING FOR THE DESIGN OF IONIC LIQUIDS AS ELECTROLYTES FOR BATTERY APPLICATION

Major Field : CHEMICAL ENGINEERING

Abstract: Ionic liquids are classes of salts that are often found in a liquid state composed entirely of ions. They have gained widespread interest in the research community because of several unique and desirable features, such as negligible vapor pressure, environmental friendliness, and high thermal stability. They are currently studied for various industrial applications as a replacement for conventional solvents. Among them, it has caught the interest of the energy community as a potential electrolyte for battery applications. The current electrolytes found in lithium-ion batteries are based on carbonate solvents known for their excellent performance and low material cost. However, they are plagued with numerous safety concerns as the solvent is highly volatile and prone to flammability during thermal runaway or short circuit. Growing demand for lithium-ion batteries for technology such as electric vehicles has mandated the need for safer and more sustainable batteries. This has made ionic liquids a potential electrolyte candidate as they have impeccable thermal and chemical stability with negligible vapor pressure, eliminating any concerns related to safety. However, the performance of ILs is still far behind in matching the performance of current carbonate electrolytes. Finding the appropriate ionic liquid candidate with high stability and performance can be challenging because of the vast ionic liquid chemical space, as synthesizing and testing each ionic liquid would be expensive and unfeasible. Running atomistic simulations to complement the experimental techniques would be tedious as the ionic liquid space is estimated to be in billions, which can be computationally expensive. Instead, machine learning methods can be an excellent tool to search for and design ionic liquids suited for battery applications as they can be easily trained on existing data to generate additional new data at a very low cost in a short time. Thus, the work in this dissertation is focused on developing machine learning models to correlate properties of ionic liquids geared towards battery application. The developed models are then used to generate additional data to search for high-performance ionic liquids that are on par with conventional organic electrolytes, thus expanding the list of potential electrolyte candidates. The latter part of the work utilizes an advanced deep learning method to discover an entirely new family of ionic liquid cations in search of candidates that can operate at high voltage conditions.

TABLE OF CONTENTS

1	INTRODUCTION	1
1.1	Overview	1
1.1.1	Ionic Liquids	2
1.1.2	Application	8
1.1.3	Current Limitation	10
1.1.4	Research Focus	12
1.1.5	Research Scope	12
1.1.6	Dissertation Outline	13
2	LITERATURE REVIEW	16
2.1	Overview	16
2.1.1	QSPR Method	16
2.1.2	Group Contribution Method	18
2.1.3	Machine Learning Approach	20
2.1.4	Molecular Dynamics	22
2.1.5	Electrochemical Stability	23
3	METHODOLOGY	26
3.1	Machine Learning	26
3.1.1	Multiple Linear Regression	26
3.1.2	Cost Function	27
3.1.3	Decision Tree	29
3.1.4	Random Forest	34
3.1.5	XGBoost	35
3.1.6	Support Vector Machine	35
3.1.7	Neural Network	38

3.1.8	Variational AutoEncoder	42
3.1.9	Principal Component Analysis	45
3.1.10	Data Processing	47
3.1.11	Feature Reduction	47
3.1.12	Hyperparameter Tuning	48
3.2	Molecular Dynamics	49
3.2.1	Molecular Dynamics Simulation	49
3.2.2	Self Diffusion Constant	52
3.2.3	Ionic Conductivity	52
3.2.4	Radial Distribution Function	53
3.2.5	Coordination Number	54
3.2.6	Hydrogen Bonding	54
3.3	Electrochemical Stability	56
4	Developing Machine Learning Models for Ionic Conductivity of Imidazolium-Based Ionic Liquids	61
4.1	Abstract	61
4.2	Introduction	62
4.3	Methodology	65
4.3.1	Data Collection and Processing	65
4.3.2	Feature Generation and Elimination	67
4.3.3	Data Normalization	68
4.3.4	Model Development	68
4.4	Results and Discussion	70
4.4.1	Model Validity	70
4.4.2	Unique Ionic Liquid Combination	75
4.4.3	Binary Ionic Liquid Mixtures	78

4.4.4	Comparison of Experimental and FFANN-predicted Ionic Conductivity of Binary ILs	79
4.5	Conclusion	84
5	Revealing Hydrogen Bond Dynamics between Ion Pairs in Binary and Reciprocal Ionic Liquid Mixtures	86
5.1	Abstract	86
5.2	Introduction	87
5.3	Methodology	91
5.3.1	Force Fields	91
5.4	Simulation Details	92
5.5	Results and discussion	93
5.5.1	Liquid Densities	93
5.5.2	Radial Distribution Function (RDF)	94
5.6	Hydrogen Bonding Lifetimes	100
5.6.1	Self Diffusion Constant	103
5.7	Conclusion	106
6	A Generalized Machine Learning Model for Predicting Ionic Conductivity for Ionic Liquids'	108
6.1	Abstract	108
6.2	Introduction	109
6.3	Methodology	112
6.3.1	Data collection and processing	112
6.3.2	Feature generation and processing	114
6.3.3	Model Development	114
6.3.4	Multiple Linear Regression	115
6.3.5	Random Forest	116

6.3.6	XGBoost	116
6.3.7	Cross Validation and Model Evaluation	116
6.4	Results and Discussion	118
6.4.1	Model Performance Metrics	118
6.4.2	Model Interpretation	121
6.4.3	Unique ILs	123
6.5	Conclusion	125
7	Accelerated Discovery of Novel Ionic Liquid Cations using a Continuous Latent Space	
	Representation of Chemical Space	127
7.1	Abstract	127
7.2	Introduction	128
7.3	Methodology	132
7.3.1	Variational AutoEncoder (VAE)	132
7.3.2	Data Generation	133
7.3.3	Model Development	133
7.3.4	Post Processing	134
7.3.5	Electrochemical Window	136
7.4	Results and Discussion	138
7.4.1	Model Validation	138
7.4.2	Data Generation	140
7.4.3	Stability Test	140
7.4.4	Non-Aromatic/Non-Cyclic Cations	141
7.4.5	Aromatic/Cyclic Cations	144
7.4.6	Electrochemical Window	146
7.4.7	Conclusion	148

8	Mapping the Frontier Orbital Energies of Imidazolium-Based Cations Using Machine Learning	150
8.1	Abstract	150
8.2	Introduction	150
8.3	Methodology	155
8.3.1	Data generation	155
8.3.2	DFT calculations	157
8.3.3	Machine learning	158
8.4	Results and discussion	159
8.4.1	Model validation	159
8.4.2	External test validation	159
8.4.3	Feature importance	161
8.4.4	Entire isomer chemical space mapping	163
8.5	Conclusion	165
9	CONCLUSION	166
9.0.1	Conclusions	166
9.0.2	Future Work	169
Appendix A	Developing Machine Learning Models for Ionic Conductivity of Imidazolium-Based Ionic Liquids	172
Appendix B	Revealing Hydrogen Bond Dynamics between Ion Pairs in Binary and Reciprocal Ionic Liquid Mixtures	199
Appendix C	A Generalized Machine Learning Model for Predicting Ionic Conductivity for Ionic Liquids?	212
Appendix D	Accelerated Discovery of Novel Ionic Liquid Cations using a Continuous Latent Space Representation of Chemical Space	236
Appendix E	Mapping the Frontier Orbital Energies of Imidazolium-Based Cations	251

Bibliography 257

LIST OF FIGURES

1.1	Commonly studied cations, anions and functional groups structure drawn using OPSIN [10]. R_n indicates some of the commonly attached location for functional groups on the cation. The tunability of the cations is derived by changing the functional group attached to the cations.	3
1.2	Experimental ionic conductivity data at 298.15 K. LP30 here refers to the conventional electrolyte used in Li-ion batteries [71, 72].	9
1.3	Experimental ionic conductivity data based on cation type distribution. Data collected from NIST IL thermo database and other literature sources referenced above.	11
3.1	Schematic diagram of a decision tree model.	30
3.2	Schematic diagram of a random forest model.	34
3.3	Hyperplane for separating the data points into respective classes for support vector machine.	36
3.4	Schematic diagram of a neural network model.	39
3.5	Schematic diagram of a variational autoencoder model.	42
3.6	Comparison of ECW between TDC [42], HOMO/LUMO approach (this work), one electron approach (this work) and experimental data for various imidazolium ILs.	58
3.7	Correlation of ECW between HOMO/LUMO and one electron (Li/Li+) electrode approach.	60

4.1	The number of ILs for a given cation-anion combination with more than five data points in the NIST ILThermo database is shown.	67
4.2	Description of model development followed in this study.	70
4.3	(a) Comparison of FFANN model predictions with the experimental data on a \log_{10} scale for the training set. A perfect prediction would fall on the $y = x$ dotted line; (b) comparison for the test set (c) Residual deviation on the log-10 scale calculated as $(\sigma_{\text{experiment}} - \sigma_{\text{prediction}})$ where σ refers to the ionic conductivity for the training set; (d) Residual deviation for the test set.	72
4.4	(a) Comparison of SVR model predictions with the experimental data on a log-10 scale for the training set. A perfect prediction would fall on the $y = x$ dotted line; (b) comparison for the test set (c) Residual deviation on the log-10 scale calculated as $(\sigma_{\text{experiment}} - \sigma_{\text{prediction}})$ where σ refers to the ionic conductivity for the training set; (d) Residual deviation for the test set.	73
4.5	(a) Ionic conductivity comparison between experiment (open circle in red) and model prediction(green) using FFANN for all those data at 298 K. (b) Unique ionic liquid predictions using FFANN for 1102 ionic liquid obtained by combining 29 unique cations and 38 anions at 298 K. The ionic conductivity for these ILs at 350 and 400 K appear in the supporting information (Figure A.15.)	77
4.6	Comparison between experiment and FFANN model for $[\text{C}_4\text{mim}]^*[\text{Azide}]_{x_1} + [\text{C}_4\text{mim}]^*[\text{BF}_4]_{1-x_1}$ at (a) 303.15 K, (b) 323.15 K and (c) 368.15 K. The dashed lines connecting the pure end points are only guide to the eye. $[\text{C}_4\text{mim}]^*$ stands for 1-butyl-2,3-dimethylimidazolium cation [209]. The dashed line with \star is obtained by a logarithmic combining rule for ionic conductivity $\ln \sigma_{mix} = x_1 * \ln \sigma_1 + (1 - x_1) * \ln \sigma_2$, while the dashed line without symbol indicates estimates with a linear combining rule.	79

4.7	Percentage enhancement and suppression in ionic conductivity for binary ionic liquid mixtures at 298 K. Binary_A_Max stands for a binary mixture sharing a common cation showing maximum enhancement; Binary_C_Max stands for a binary cation mixture displaying maximum enhancement; Binary_A_Min denotes a binary anion mixture exhibiting minimum suppression, and Binary_C_Min represents a binary cation mixture producing minimum suppression.	83
5.1	Chemical structures of the cations and anions used in this study.	90
5.2	Center of mass (COM) radial distribution plot of all the cation-anion interactions in reciprocal mixture.	95
5.3	COM RDF plot for [C ₂ mim]--[DCA] interactions in pure and various mixtures.	97
5.4	RDF plots for site-site interaction in pure and various mixtures. (a) H _a - -N _c site-site interactions between [C ₂ mim]--[DCA]. (b) H _w - -N _c site-site interaction between [C ₂ mim]--[DCA].	98
5.5	RDF plots for site-site interaction in pure and various mixtures. (a) H _a - -O site-site interactions between [C ₂ mim]-[NTf ₂]. (b) H _w - -O site-site interaction between [C ₂ mim]-[NTf ₂].	99
5.6	Correlation between hydrogen bonding lifetime and RDF peak height of the first solvation shell of hydrogen bonding sites. (a) [C ₂ mim]--[DCA]. (b) [C ₂ mim]--[NTf ₂]. Binary_A refers to [C ₂ mim][DCA][NTf ₂] system, Binary_C refers to [C ₂ mim][C ₆ mim][DCA] or [C ₂ mim][C ₆ mim][NTf ₂] system and Reciprocal refers to [C ₂ mim][C ₆ mim][DCA][NTf ₂] system.	102
5.7	Self Diffusion constant of all the systems calculated using equation equation 5.12, 5.14, 5.13 and 5.15.	106
6.1	Experimental ionic conductivity data distribution % by cation type for the model development data set after data cleaning. The percentage for each of the cation family is calculated using the number of data points collected in Table C.1. . .	113

6.2	Workflow for developing machine learning models for correlating ionic conductivity of ILs.	115
6.3	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. . . .	120
6.4	SHAP feature importance for the training set data. Features ending with '_a' indicates features for anions.	122
6.5	Categorical data of unique ILs based on cation family type at 298.15 K.	124
7.1	Variational Autoencoder schematic representation. The encoder takes in the input information and compresses it to low dimensional space on the latent space (z). The decoder takes the compressed information from the latent space and reconstructs it back to the form of the original input. The latent space is represented in the form of a Gaussian distribution with a mean (μ) and variance (σ). $z_1, z_2..z_n$ represents the attributes of the latent space with (n) dimensions. . . .	133
7.2	(a) % Representation of different cation type in the VAE model development. (b) One hot encoding vector of a sample organic molecule with a vocabulary size of 6 and length of 100. 'E' denotes the end of SMILE string. (c) Neural network representation of the encoder and decoder for the VAE model.	136
7.3	Projection of latent space into two dimensional axis using the first two principal components. Each of the subplot indicates the region for each of the cation type on the latent space.	139

7.4	(a) Bar chart of the cation type generated using the VAE model for non-aromatic/non-cyclic based cations. (b) Two-dimensional PCA plot of the known cation chemical space with their respective label generated using features from RDKit. (c) A two-dimensional PCA plot of the new VAE generated non-cyclic/non-aromatic cations on the cation space in addition to the classification training data. The features for these cations are generated using RDKit. Here "Unknown" refers to cations that do not belong to known cation families. (d) Maximum Tanimoto Similarity index for the "Unknown" cation type compared to the rest of the known cation type.	143
7.5	(a) Bar chart of the cation type generated using the VAE model for aromatic/cyclic-based cations. (b) Two-dimensional PCA plot of the known cation chemical space generated using features from RDKit with their respective label. (c) A two-dimensional PCA plot of the new VAE generated cyclic/aromatic cations on the cation space in addition to the classification training data. Here "Unknown" refers to cations that do not belong to known cation families. (d) Maximum Tanimoto Similarity index for the "Unknown" cation type compared to the rest of the known cation type	145
7.6	Electrochemical window values of all the "Unknown" cyclic/aromatic cations paired with (a) [DCA] ⁻ , (b) [NTf ₂] ⁻ , (c) [FSI] ⁻ and (d) [BF ₄] ⁻ anion.	147
8.1	Number of isomers for [C _n C _m mim] ⁺ cation in log10 scale. Cations denoted by (*) indicates data used for model development. Cations denoted by (-) and (+) indicate data used for external test set validation.	156
8.2	HOMO and LUMO energy distribution of cations present in the model development data set calculated using DFT.	157

8.3	Performance metrics for the external set using HOMO/LUMO energy model. The cations listed here are not present in the model development data. The size of the marker varies based on the amount of branched isomer data present for the individual cation set.	160
8.4	SHAP feature importance for (a) HOMO energy model and (b) LUMO energy model.	162
8.5	Energy gap of the entire $[C_nC_m\text{mim}]^+$ cations using (a) DFT for the 1399 data set for model development and (b) XGBoost model for 317,382 cations.	164
A.1	Temperature distribution profile of all the experimental data points used in the model development.	173
A.2	Ionic conductivity distribution profile of all the experimental data points used in the model development.	173
A.3	Comparison between experimental data in normal scale and transformed data using Min Max Scaler to scale the data between 0-1 before model development.	174
A.4	Comparison between experimental data in \log_{10} scale and transformed data using Min Max Scaler to scale the data between 0-1 before model development.	175
A.5	Heat map correlation of cation features with experimental ionic conductivity data.	176
A.6	Heat map correlation of anion features with experimental ionic conductivity data.	177
A.7	FFANN model topology. Total number of data points in the training set : 1190. Ratio of weight and bias to the total number of data points in the training set : 10.35	181
A.8	(a) Comparison of SVR model predictions with the experimental data on normal scale for the training set. A perfect prediction would fall on the $y = x$ dotted line; (b) comparison for the test set (c) Residual deviation on the normal scale calculated as $(\sigma_{\text{experiment}} - \sigma_{\text{prediction}})$ where σ refers to the ionic conductivity for the training set; (d) Residual deviation for the test set.	183

A.9	(a) Comparison of SVR model predictions with the experimental data on normal scale for the training set. A perfect prediction would fall on the $y = x$ dotted line; (b) comparison for the test set (c) Residual deviation on the normal scale calculated as $(\sigma_{\text{experiment}} - \sigma_{\text{prediction}})$ where σ refers to the ionic conductivity for the training set; (d) Residual deviation for the test set.	184
A.10	Comparison between experiment and model for ten ionic liquids at the bottom of the ionic conductivity values at 298 K using FFANN model. [C ₄ mim]*: 1-butyl-3-methyl-1H-imidazolium; [C ₃ mim]*: 1-propyl-3-methyl-1H-imidazolium; [C]*:3-(2-(butylamino)-2-oxoethyl)-1-ethyl-1H-imidazolium; [Amim] stands for 1-allyl-3-methylimidazolium.	185
A.11	Comparison between experiment and model for top 10 ionic liquids with the highest ionic conductivity at 298.0 K using FFANN model. [AMIm] stands for 1-allyl-3-methylimidazolium.	186
A.12	Comparison between experiment and model for ionic conductivity of alkyl and alkenyltrifluoroborate anions paired with [C ₂ mim] ⁺ cation at 298.15 K using FFANN model. Experiment data were taken from ref [359].	188
A.13	Comparison between experimental data and model outputs for ionic conductivity of hydrogen sulfate and alkylsulfates [C _n SO ₄] ⁻ anions (n = 1, 2, 4, 6, and 8) paired with [C ₂ mim] ⁺ cation at 298.15 K and 323.15 K using FFANN model. Experiment data were taken from ref [360].	190
A.14	Predicted ionic conductivity as a function of alkyl chain for the [C _n mim] cation series for all the available anions at 298.0 K using FFANN model.	192
A.15	Predicted ionic conductivity for all the unique ionic liquid combination(1102) as a function of temperature using FFANN model. Note:The numbering of ILs in x-axis is the same for all three sub-plots	193
A.16	Binary Mixture of [C ₂ mim][NTf ₂] _{x1} + [C ₂ mim][CF ₃ SO ₃] _{1-x1} at 298.0 K.	193

A.17	Heat map of anions based on frequency of occurrence in binary anion common cation causing maximum enhancement at 298.0 K.	194
A.18	Heat map of anions based on frequency of occurrence in binary anion common cation causing negative suppression at 298.0 K.	195
A.19	Heat map of cations based on frequency of occurrence in binary cation common anion causing maximum enhancement at 298.0 K.	196
A.20	Heat map of cations based on frequency of occurrence in binary cation common anion causing negative suppression at 298.0 K.	197
B.1	Liquid phase density of all the system studied in this paper at 323.0 K using ± 0.8 charge scaling.	200
B.2	Center of mass (COM) radial distribution plot of the cation ring and anion interactions in reciprocal mixture.	201
B.3	Comparison of center of mass (COM) radial distribution plot of all the cation-anion interactions in reciprocal mixtures between two charge scaling. ± 1.0 charge scaling denoted by black color and ± 0.8 charge scaling denoted by red color. . .	202
B.4	COM RDF plot for $[\text{C}_2\text{mim}]^-$ - $[\text{NTf}_2]$ interactions in various system containing $[\text{C}_2\text{mim}][\text{NTf}_2]$	203
B.5	COM RDF plot for $[\text{C}_6\text{mim}]^-$ - $[\text{NTf}_2]$ interactions in various system containing $[\text{C}_6\text{mim}][\text{NTf}_2]$	203
B.6	COM RDF plot for $[\text{C}_6\text{mim}]^-$ - $[\text{DCA}]$ interactions in various system containing $[\text{C}_6\text{mim}][\text{DCA}]$	204
B.7	Spatial distribution function (SDF's) of anions around the $[\text{C}_2\text{mim}]^+$ cation for $[\text{C}_2\text{mim}][\text{C}_6\text{mim}][\text{DCA}][\text{NTf}_2]$ reciprocal mixture. Red represents $[\text{NTf}_2]^-$ anion, cyan represents $[\text{DCA}]^-$ anion. Isosurface density was set to 2.2 times the bulk density.	205
B.8	Hydrogen bonding interaction sites for the cation and anion.	205

B.9	RDF of site-site interaction for hydrogen (H_a) - acceptor (N_c) and com interaction between the cation and anion in $[C_2mim][DCA]$	206
B.10	RDF of site-site interaction for hydrogen (H_a) - acceptor (O) and com interaction between the cation and anion in $[C_2mim][NTf_2]$	207
B.11	RDF of site-site interaction for ydrogen (H_a) - acceptor (N_c) and com interaction between the cation and anion in $[C_6mim][DCA]$	208
B.12	RDF of site-site interaction for ydrogen (H_a) - acceptor (O) and com interaction between the cation and anion in $[C_6mim][NTf_2]$	209
B.13	Mean Square Displacement of $[C_2mim]^+$ cation and $[DCA]^-$ anion for pure $[C_2mim][DCA]$ system.	210
B.14	Mean Square Displacement of $[C_2mim]^+$ cation and $[NTf_2]^-$ anion for pure $[C_2mim][NTf_2]$ system.	210
B.15	Mean Square Displacement of $[C_6mim]^+$ cation and $[DCA]^-$ anion for pure $[C_6mim][DCA]$ system.	211
B.16	Mean Square Displacement of $[C_6mim]^+$ cation and $[NTf_2]^-$ anion for pure $[C_6mim][NTf_2]$ system.	211
C.1	Schematics of cation type used in this study for model development. R_1 , R_2 , R_3 , and R_4 refer to substituents in the cation. Images were drawn using OPSIN package [10].	213
C.2	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.	214

C.3	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.	215
C.4	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.	216
C.5	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.	217
C.6	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.	218
C.7	Comparison of experimental external data set and XGBoost model prediction. The ionic conductivity values are reported in S/m.	219
C.8	Cation feature $\ln(I_{pc})$ vs molecular weights of the cations for which experimental data were included in the model development.	226
C.9	Cation feature Chi0 vs molecular weight of the cations for which experimental data were used in the model development.	227
C.10	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.	228

C.11	SHAP feature importance for the training set data. Features ending with '_a' indicates features for anions.	229
C.12	(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 an high ionic conductivity of 1.26 S/m at 298.15 K [71,291]. Addition of Li+ salts to ionic liquids is known to reduce ionic conductivity by 30-50% [278].	234
D.1	(a) % Representation of different cation type in the VAE model development. (b) One hot encoding vector of a sample organic molecule with a vocabulary size of 6 and length of 100. 'E' denotes the end of SMILE string. (c) Neural Network representation of the encoder and decoder for the VAE model.	237
D.2	Heat map of the the reconstruction accuracy % for the validation set during hyper-parameter search. X-axis indicates the latent space dimension. Y-axis indicates the total number of data points out of which 10% is kept aside as test data, the remaining 90% is further divided into training purposes and validation (90:10). Red square denotes the highest validation set reconstruction accuracy %.	239
D.3	Heat map of the reconstruction accuracy % for the training set during the hyper-parameter search. The X-axis indicates the latent space dimension. Y-axis indicates the total number of data points, of which 10% is kept aside as test data. The remaining 90% is further divided into training purposes and validation (90:10). Red squares denote the highest training set reconstruction accuracy %.	240
D.4	Confusion matrix prediction of the test set using the classification model. . . .	242
D.5	Principal component analysis (PCA) of the latent space for the training data labeled by cation type.	243
D.6	Kernel Density Estimation (KDE) of each of the latent space attribute.	244
D.7	VAE generated "Unknown" cation type as labeled by the classification model. The numbers below the cation denote the maximum of Tanimoto similarity index.	245

D.8	Top 10 least similar aromatic/cyclic cations from the "Unknown" cation family based on Tanimoto similarity index. The numbers below the cations denote the maximum of Tanimoto similarity index.	246
D.9	Benchmarking electrochemical window calculations using the methodology adopted in this work against (a) The work of Kazemiabnavi et al. [42] as indicated by the red circle denoting thermodynamic cycle (TDC) approach calculated against Li ⁺ /Li reference electrode. Experiment data are also added for some of the ionic liquids. [11]. (b) Asha et al. [104] work using TDC approach also calculated against Li ⁺ /Li reference electrode for pyrrolidinium ionic liquids and experimental data. [22, 301]	247
D.10	Cations and anion involved in MD simulation.	248
D.11	Self diffusion constant and center of mass radial distribution function for four ionic liquids calculated using MD simulation at 323.15 K.	249
E.1	Energy gap (eV) of [C _n C _m im] ⁺ cations calculated using DFT.	252
E.2	XGBoost hyperparameter grid space for the HOMO model.	253
E.3	XGBoost hyperparameter grid space for the LUMO model.	254
E.4	Predicted vs DFT calculations of HOMO energy for training, validation and test set. Residual deviation % is the difference in DFT and model divided the DFT value.	255
E.5	Predicted vs DFT calculations of LUMO energy for training, validation and test set. Residual deviation % is the difference in DFT and model divided the DFT value.	256

LIST OF TABLES

4.1	Comparison of the predictions results for FFANN and SVR for the training set, test set and the entire dataset. MSE is the mean squared error, MAE is the mean absolute error, RMSD is root mean square deviation and R^2 is the squared correlation between experiment and predicted data. \log_{10} scale refers to ionic conductivity scaled to \log_{10}	71
4.2	Root mean-squared deviation of the prediction and experimental data for binary ionic liquid mixtures. N.D stands for number of datapoints present in the dataset. [C ₄ mim] ⁺ stands for 1-butyl-2,3-dimethylimidazolium cation.	80
5.1	Detailed description of the system composition and equilibrium densities at 323 K	94
5.2	Average lifetime τ_{lt} of hydrogen bonds. C _{a2} and H _{a2} atom are attached to the [C ₂ mim] ⁺ cation, C _{a6} and H _{a6} atom are attached to the [C ₆ mim] ⁺ cation, N _c is attached to the [DCA] ⁻ anion and O is attached to the [NTf ₂] ⁻ anion.	101
5.3	Self Diffusion constant ($D \times 10^{-7}$ cm ² /sec) of ions for pure, binary and reciprocal mixtures at 323 K.	104

6.1	Average and standard deviation of the performance metrics for the training and validation set using MLR, RF, XGBoost model. RMSE is the root mean squared error, MAE is the mean absolute error, and R^2 is the correlation coefficient between experiment and predicted data. Shuffle-Split indicates random data shuffle into 100 different training/validation splits. The model with the best performance on the validation set during shuffle-split becomes the final choice of model for test set evaluation. Note: The RMSE and MAE values are for ionic conductivity in the \log_{10} scale.	119
7.1	Reconstruction accuracy % of the SMILE data using the VAE model for different data set called upon 100 different instances. Reconstruction accuracy % is calculated by taking the total SMILE generated by the VAE model that matches the input SMILE to the total number of input SMILE.	138
8.1	Performance metrics of the XGBoost Model for HOMO (eV) and LUMO (eV) energies. (R^2) is the correlation coefficient, (RMSE) is root mean squared error and (MAE) is mean absolute error.	159
A.1	Comparison of the predictions results for FFANN and SVR for the training set and test set by varying the partition of the test/train split. This was done by changing the random state parameter in sklearn. MSE is the mean squared error, MAE is the mean absolute error, RMSD is root mean square deviation and R^2 is the squared correlation between experiment and predicted data. \log_{10} scale refers to ionic conductivity scaled to \log_{10} , while normal scale refers to ionic conductivity data without any scaling in S/m.	182

A.2	Comparison of the predictions results for FFANN and SVR for the training set, test set and the entire dataset. MSE is the mean squared error, MAE is the mean absolute error, RMSD is root mean square deviation and R^2 is the squared correlation between experiment and predicted data. MAPE is mean absolute percentage error in %.	185
A.3	Comparison with experimental data for the GC method and FFANN model for the $[C_{n=(2,4,6,8)}\text{mim}][\text{NTf}_2]$ subset series at various temperatures. The experimental data and GC data for this predictions were taken from the supporting information of Chen et al. work [113].	186
A.4	Comparison between experiment and model for ionic conductivity of alkyl and alkenyltrifluoroborate anions paired with $[C_2\text{mim}]^+$ cation at 298.15 K using FFANN model and SVR model. Experiment data were taken from ref [359].	188
A.5	Comparison between experimental data and model outputs for ionic conductivity of hydrogen sulfate and alkylsulfates $[C_n\text{SO}_4]^-$ anions ($n = 1, 2, 4, 6, \text{ and } 8$) paired with $[C_2\text{mim}]^+$ cation at 298.15 K and 323.15 K using FFANN model and SVR model. Experiment data were taken from ref [360].	189
A.6	Comparison of ionic conductivity for $[R_1R_2R_3\text{im}][\text{NTf}_2]$ ionic liquids at 298.15 K between experiment and models. RD stands for Relative Deviation. Experiment data taken from ref [277].	198
B.1	Comparison of liquid densities for pure ionic liquids against experimental data using two different charge scaling at 323 K.	199
B.2	Cut-off distances (nm) used in the hydrogen bonding average lifetime calculations. The hydrogen-acceptor cut-off was set to the first solvation denoted by $H_a\text{-}N_c/\text{O}$. The distance for the donor-acceptor was limited to the first solvation shell of the center of mass (cation)-center of mass (anion) denoted by Cation-Anion.	209

C.1	A detailed breakdown of the number of experimental ionic conductivity data for various cation type in the model development data set.	214
C.2	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.	220
C.3	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.	222
C.4	Comparison of ionic conductivity at 298 K between experiment [277] and XGBoost prediction for cations paired with bis(trifluoromethylsulfonyl)imide [NTf ₂] ⁻ anion.	225
C.5	Classification of external test case cations paired with [NTf ₂] ⁻ at 298 K. Low refers to ionic conductivity less than 0.265 S/m.	232
C.6	Classification of external test case cations paired with [NTf ₂] ⁻ at 298 K. Low refers to ionic conductivity less than 0.265 S/m. Experiment data taken from ref [277].	233
C.7	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.	235
E.1	1-alkyl-3-methylimidazolium ([C _m C _n mim] ⁺) branched isomer cation chemical space for alkyl chain length from 1 to 10.	251

CHAPTER 1

INTRODUCTION

1.1 Overview

The discovery and use of fossil fuels as a source of energy may have forever altered the course of human civilization as we know it. Aside from providing energy for all industrial processes and transportation, the use of fossil fuels has also had an indirect impact on every aspect of modern life, including some notable ones such as material synthesis for housing, clothing, and infrastructure development, fertilizer manufacturing for food production, and drug development. This increase in the quality of life has enabled humans to be more productive, healthier, and live longer lives, resulting in tremendous economic and societal growth in a relatively short period. The rapid advancement, however, has come at a cost to the environment, as the burning of fossil fuels emits a massive amount of CO₂ into the atmosphere, causing a rise in global temperature. This could eventually cause irreversible damage to our planet and living beings, with rapidly melting glaciers flooding the coastline, intense heatwaves causing drought and famine, and climate extremes destroying livelihoods worldwide.

As a result, a global effort is underway to mitigate this disaster by limiting greenhouse gas emissions, which might require a complete shift in energy sources from nonrenewable to renewable energy sources such as solar, wind, and geothermal. However, changing the entire energy grid is not an easy task as renewable energies

are intermittent energy sources that are only available at certain times during the day. Harvesting these energy sources and storing the energy for later requires energy storage mediums such as batteries. Aside from the energy grid sector, there is a rapidly growing demand for electric vehicles powered by battery technology. As demand increases, the need for safer, more sustainable, high-energy-density batteries to power devices ranging from portable electronics to electric vehicles grows. However, meeting this demand would require significant improvements in current battery components and electrolyte chemistry [1,2].

The scope of this work is limited to electrolytes because, despite their excellent performance, the currently used organic solvent electrolytes are plagued by safety issues due to high volatility and flammability during thermal runaway or short-circuit [3–5]. Several alternative candidates are being considered, including solid-state electrolytes, polymer-based electrolytes, and ionic liquids, to improve safety while matching the performance of current organic solvent electrolytes [6].

1.1.1 Ionic Liquids

Ionic liquids (ILs) are classes of salts composed entirely of ions with a melting point below 100 °C. They have gained tremendous interest in the research community as a replacement for conventional solvents because of features such as negligible vapor pressure, which solves the issue of volatile organic contaminants (VOC) emission during the solvent usage [7]. Many of the ILs also have very high thermal and chemical stability, with some of them reaching a thermal decomposition temperature of 450 °C [8] which is an improvement over conventional solvents accompanied by low thermal and chemical stability. The most significant advantage of ILs over conventional solvents is the design flexibility, which allows the design of ILs with desired properties based on the requirements of a given application. There are a large number of cations and anions available that one can choose from based on the application requirement, as seen in Figure 1.1 with estimates of up to 10^{14} different ILs [9].

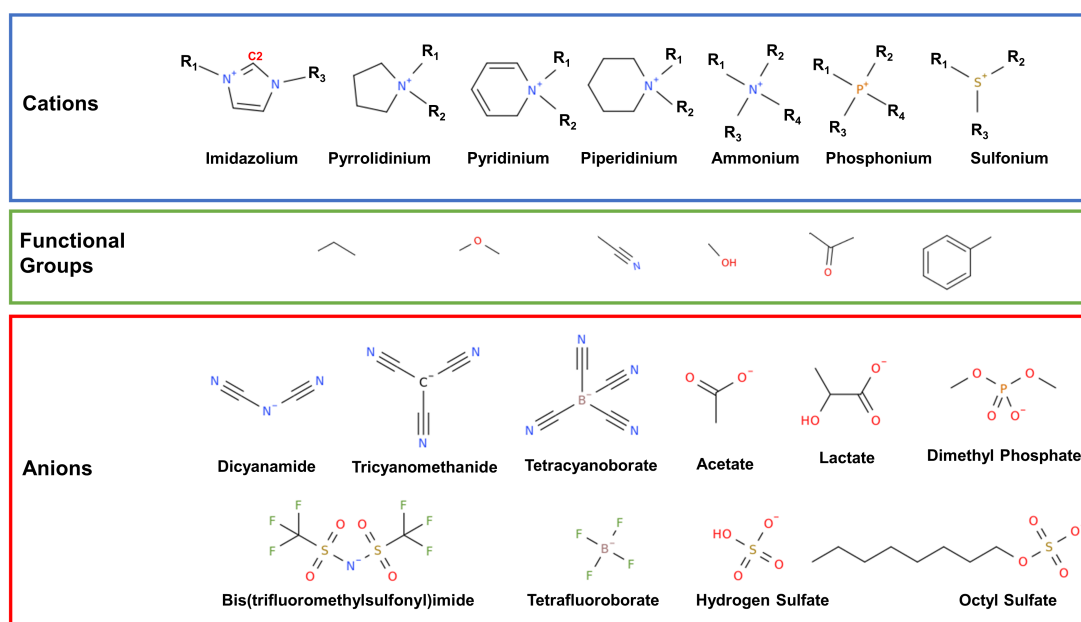


Figure 1.1: Commonly studied cations, anions and functional groups structure drawn using OPSIN [10]. R_n indicates some of the commonly attached location for functional groups on the cation. The tunability of the cations is derived by changing the functional group attached to the cations.

The selection of the cation and anion is guided by considering their unique advantages and challenges. For battery applications, properties such as low viscosity, high ionic conductivity, and high electrochemical stability are a few of the most important attributes considered before selecting an ionic liquid of interest. Among all of the ILs, the most popular class is based on the imidazolium group, as ILs derived from this moiety are often associated with very low viscosity, and high ionic conductivity [11]. However, a major limitation of imidazolium cations is the small electrochemical window (3.0 – 4.5 V) because of the presence of an acidic proton at the C2 position (Figure 1.1) on the imidazolium ring that gets easily reduced. Replacing the hydrogen at the C2 position with an alkyl group resolves the low electrochemical stability issue [12], but any changes made to the C2 position have shown to increase viscosity and lower ionic conductivity dramatically [13].

The next closest relative to the imidazolium cations are the pyridinium-based cations that are more sluggish, accompanied by high viscosity and low ionic conductivity compared to the imidazolium cations, which is why there is a limited amount of studies done on exploring their application as electrolytes for battery applications [14–16]. For instance the ionic conductivity of 1-butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide is about 0.401 S/m at 298.15 K [17] compared to 0.272 S/m for 1-butyl-3-methylpyridinium bis(trifluoromethylsulfonyl)imide at the same temperature [16].

Beyond the aromatic cations, cyclic cations such as pyrrolidinium and piperidinium cations have generated tremendous interest as they offer high biodegradability and low toxicity [18, 19]. The pyrrolidinium cation also offers low viscosity and high ionic conductivity, and, unlike imidazolium cations, they have a very high electrochemical window, with a majority of them reaching above 4.5 - 5.0 V [11]. Besides faster dynamics, the pyrrolidinium cations also have very high stability toward lithium metal, making them an ideal candidate for battery application as potential electrolytes [20, 21].

Modifying the five-ring pyrrolidinium structure to a six-ring structure gives rise to the piperidinium cations. Similar to pyridinium cations, piperidinium cations have slower dynamics than pyrrolidinium cations because of the bulky nature of the cation. For instance at 298.15 K the ionic conductivity of 1-butyl-1-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide is reported to be around 0.277 S/m [17] as compared to 0.139 S/m for 1-butyl-1-methylpiperidinium bis(trifluoromethylsulfonyl)imide [22]. As such, there are relatively few studies that have explored the possibility of piperidinium cations as electrolytes for battery application [22–24].

Besides cyclic and aromatic cations, other central atom-based cations such as phosphonium, ammonium, and sulfonium are also extensively studied for various applications [25–27]. The ammonium-based cations have a very high electrochemical

window compared to imidazolium but are also associated with very high viscosity, and low ionic conductivity [28]. Similarly, phosphonium cations also exhibit high electrochemical stability but have almost twice the ionic conductivity compared to their ammonium counterpart [29]. For instance, the ionic conductivity of tributyl(hexyl)ammonium bis(trifluoromethylsulfonyl)imide is reported to be around 0.28 S/m at 298.15 compared to 0.43 S/m for tributyl(hexyl)phosphonium bis(trifluoromethylsulfonyl)imide at the same temperature [30]. Molecular dynamic studies have alluded to this difference in transport properties between the two cations primarily because of the disparity in electrostatic charge concentration around the central atom [29, 31].

The sulfonium cations have also generated tremendous interest in battery applications as they have very low viscosity and high ionic conductivity compared to other central-based cations because of the presence of a small-sized central atom [27, 32, 33]. Several studies have reported the ionic conductivity of sulfonium to be greater than phosphonium and twice as high as ammonium cations [32, 33]. Furthermore, as compared to the phosphonium cation, the sulfonium cation is found to perform better for energy storage applications [34].

Beyond the standard alkyl chain substitute on the cation, it is also possible to add other functional groups (Figure 1.1) at various positions on the cation. One of the commonly studied functional groups is the ether functional group, known to enhance transport properties by lowering viscosity and increasing ionic conductivity [35–39]. Neale et al. [37] conducted a comprehensive experimental study on the thermophysical and transport properties of three cations (1-methyl-1-butylpyrrolidinium, 1-methyl-1-butylpiperidinium, and 1-methyl-1-butylazepanium) paired with bis(trifluoromethylsulfonyl)imide anion. The authors reported an increase in ionic conductivity and a decrease in viscosity roughly by a factor of 1.50 with the replacement of the butyl chain ($\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3$) with an ether group ($\text{CH}_2\text{CH}_2\text{OCH}_3$). However the trend is reversed as the ether group ($\text{CH}_2\text{CH}_2\text{OCH}_3$) is extended to ($\text{CH}_2\text{CH}_2\text{OCH}_2\text{CH}_2\text{OCH}_3$). This is because of the presence of

the long functional chain length that reduces the movement of the ions, slowing down the dynamics. The dynamics get severely sluggish as both the methyl and butyl groups are replaced with the ether group. Yoshi et al. [38] reported similar trends with 1-methyl-1-propylpyrrolidinium-based cations with the addition of ether group as a substitute for the alkyl chain paired with bis(trifluoromethylsulfonyl)imide anion. The authors found the substitution of propyl group ($\text{CH}_2\text{CH}_2\text{CH}_3$) with (CH_2OCH_3) enhanced transport properties by roughly a factor of 1.50.

Besides cations, the choice of anion also plays a crucial role in shaping the properties of an ionic liquid. Cyano-based anions such as dicyanamide and tetracyanoborate are some of the most popular anions that are often associated with faster dynamics [40,41]. However, these cyano-based anions are easily susceptible to oxidation, which is why they have very low electrochemical stability [11]. When combined with a cation, fluoro-based anions such as bis(trifluoromethylsulfonyl)imide, tetrafluoroborate, and hexafluorophosphate provide high electrochemical stability, with the majority reaching electrochemical windows greater than 4.0 V [42,43]. However, several studies have found the negative impact of fluoro and cyano-based ILs on the environment as they are hard to degrade, tend to persist, and accumulate in the water sources for a more extended period [44,45]. On the other hand, organic anions such as acetate, lactate, saccharinate, and amino-based anions are more desirable from an environmental perspective as they are readily biodegradable with very low toxicity [46]. Compared to cyano and fluoro-based anions, these anions have high viscosity, low ionic conductivity, and low electrochemical stability [47,48].

In addition to pure ILs, several studies have demonstrated that combining cations and anions to form binary IL mixtures can improve existing drawbacks such as high viscosity and low ionic conductivity [49–51]. Kunze et al. [49] in an experimental study found that the mixture of 1-butyl-3-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide and 1-propyl-3-methylpyrrolidinium

bis(fluorosulfonyl)imide increased the ionic conductivity of the mixture by 10^3 at $-40\text{ }^\circ\text{C}$ when compared to the two pure ILs. The author noted that this enhancement was possible because mixing the two ILs prevented them from forming crystals at such a low temperature, which reduces the conductivity to an almost negligible value. Similarly, Evra et al. [50] reported an increase in ionic conductivity as a function of composition based on experimental measurement for the mixture of 1-methyl-3-ethylimidazolium trifluoromethanesulfonate and 1-methyl-3-ethylimidazolium bis(trifluoromethylsulfonyl) imide.

Such a deviation in properties for IL mixtures from ideality is dictated by the size difference and the hydrogen bonding ability of the ions. Kapoor et al. [52] showed that the difference in volume and hydrogen bonding accepting ability between the anions played a driving force in determining whether a given mixture shows non-ideal behavior. The authors conducted molecular dynamics simulations for 16 different binary anion mixtures and found that pure ILs with a molar volume difference greater than $60\text{ cm}^3/\text{mol}$ and a hydrogen bonding basicity difference greater than 0.4 when mixed are presumed to show non-ideal behavior. Before running an experiment or simulation, this rule of thumb could serve as a screening tool to identify ionic liquid mixtures that show non-ideal behavior. Several other binary anions common cation mixture studies have found similar results where the difference in hydrogen bonding ability of the anions and the difference in anion size induces competition between the anions that causes the system to show non-ideal behavior. The strongly coordinating anions prefer to form hydrogen bonds with the hydrogen located at the C2 position on the imidazolium cation, displacing the less basic anion, inducing a non-ideal structural transition [53, 54].

Besides mixing ILs with themselves, they are also compatible with conventional solvents in the form of mixtures primarily aimed at reducing overall solvent cost while having some of the desirable features of ILs [55].

1.1.2 Application

Because of their several unique desirable features and flexibility, as described above, ILs have gained wide-spread attention in the research community as a potential replacement for conventional solvents. Various pure ILs and mixtures are being studied for gas capture [56], catalysis [57], extraction agent for LLE separation [58–60], cellulose dissolution [61] and energy storage application [62].

The energy community’s interest in ILs has multiplied over the years as the sector moves toward the development of next-generation Li-ion batteries. Current battery technologies based on Li-ion batteries have made energy storage possible at our convenience. It has helped pave the way for renewable energy to be considered a potential energy source for reducing the consumption of nonrenewable energy sources, which are the major contributors to greenhouse gas emissions. Meeting the global supply of Li-ion batteries as the primary energy storage for everyday devices, however, would be difficult as demand for safer, more sustainable, and high-energy-density batteries for a variety of applications increases [63].

Developing such Li-ion batteries would require significant improvement in current battery technology based on materials that are more sustainable along with improving the stability and performance of major components such as electrodes, separators, and electrolytes [64–66]. In this work, however, the focus is solely on electrolytes as there are growing safety concerns with the current state-of-the-art electrolytes based on organic carbonate solvents [67]. These organic carbonate solvents are added along with Li salts to form an electrolyte that transports Li-ions between the electrodes during the charging and discharging of the batteries. The carbonate solvents have been a tremendous success in the Li-ion battery movement as they are incredibly cheap and offer low viscosity and high ionic conductivity. However, there are growing safety issues because of their highly volatile nature and low stability that can easily catch fire during a short circuit or thermal runaway [3–5, 68]. Safety concerns for current Li-ion battery electrolytes have grown

even more as the demand soars for large-scale applications such as electric vehicles [69].

Thus, a tremendous effort is underway to find the next generation of electrolytes to solve the safety issue by offering high thermal, chemical stability and similar performance to the organic carbonate solvents [70]. As discussed earlier, ILs have excellent thermal and chemical ratings regarding stability and safety, solving some of the concerns associated with conventional electrolytes. However, they are still far from matching the performance of organic carbonate solvents as they have very high viscosity and low ionic conductivity at room temperature compared to conventional electrolytes [6].

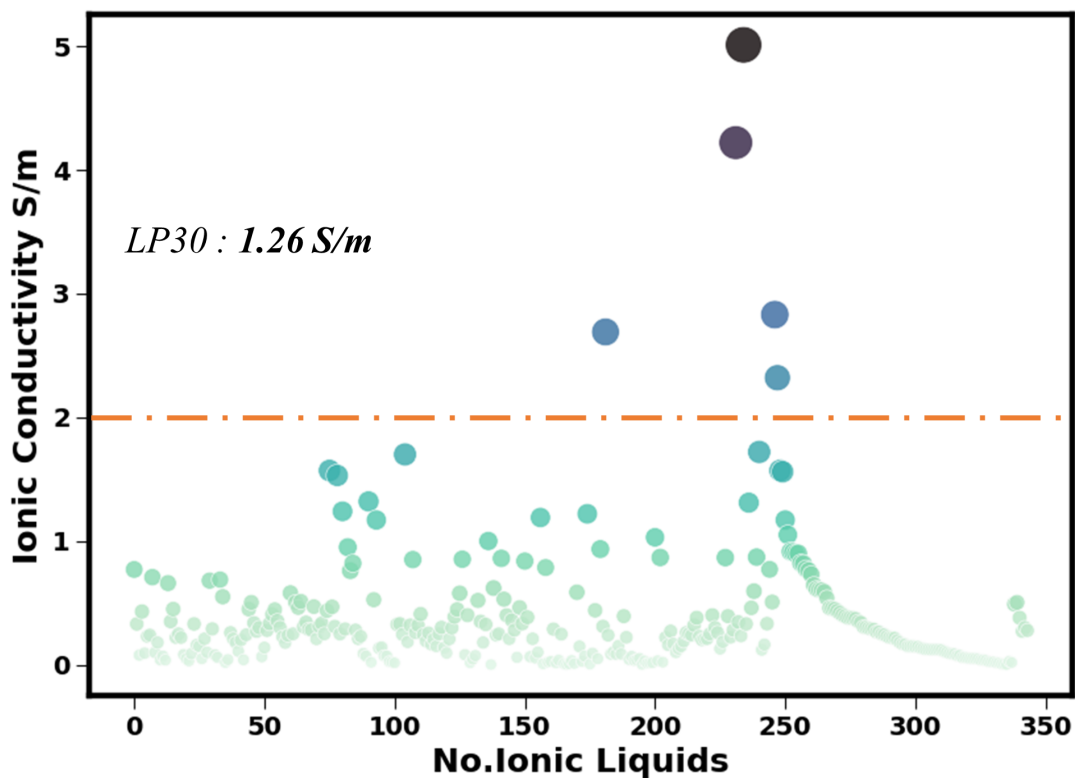


Figure 1.2: Experimental ionic conductivity data at 298.15 K. LP30 here refers to the conventional electrolyte used in Li-ion batteries [71, 72].

Figure 1.2 depicts the experimental ionic conductivity of all the ILs at 298.15

K obtained from NIST IL Thermo database [73, 74] and other literature sources [36, 75–94]. Based on the figure, it is evident that the amount of ionic liquid data for ionic conductivity is minuscule compared to the number of ILs that are possible. Secondly, it is also evident that the majority of the ILs have very low ionic conductivity, as alluded to earlier.

The current commercial electrolytes based on organic carbonate LP30 (Ethylene carbonate + Dimethyl carbonate + LiPF_6) has an ionic conductivity of 1.26 S/m at room temperature and electrochemical stability up to 4.5 V [71, 72]. For ILs to have similar performance, the ionic conductivity needs to be at least around 2.0 S/m as the addition of Li-salts significantly increases the viscosity and decreases the ionic conductivity by 30 - 40% based on the molar concentration of the salt [95, 96].

That leaves us with only five ILs (Figure 1.2), of which one of them is 1-ethyl-3-methylimidazolium thiocyanate ionic liquid with an ionic conductivity of 2.32 S/m but an electrochemical window of less than 3.0 V [97]. The second ionic liquid is the primary ammonium ionic liquid (ethyl ammonium nitrate) with an ionic conductivity of 2.69 S/m but with low electrochemical stability of 1.5 V [98]. The third one is 1-ethyl-3-methylimidazolium dicyanamide anion ionic liquid with an ionic conductivity of 2.83 S/m that is known for its extremely fast dynamics but with a very small electrochemical window of 3.2 V [99, 100]. The top two ILs that cross the 4.0 S/m barrier are the same pyrrolidinium nitrate at two different pressures with a low electrochemical window of 1.7 V [101, 102]. Despite the high ionic conductivity, these five ILs are not suitable for battery application as they have very low electrochemical stability making them highly susceptible to oxidation and reduction during their usage.

1.1.3 Current Limitation

For ILs to be considered as potential electrolyte candidates, some of the limitations discussed above have to be resolved, such as the generation of more ionic conductivity data to thoroughly explore the ionic conductivity property space to

identify a potential list of ionic liquid candidates that also have high electrochemical stability at the same time. To put that into perspective, in our preliminary research phase, we explored about 4000 experimental data points, out of which we found around only 350 data points at room temperature, as seen in Figure 1.2 with barely five ILs reaching an ionic conductivity that are close in range compared to the current commercial electrolytes.

Moreover, as discussed above, most top-performing ILs possess high ionic conductivity but low electrochemical stability, making them undesirable from a commercial application point of view. It is also evident that there is a limited amount of experimental ionic conductivity data available to work with to explore new ILs with high ionic conductivity and high electrochemical stability.

Figure 1.3 depicts the distribution of experimental ionic conductivity data based on cation type. As previously stated, the majority of experimental ionic conductivity data for imidazolium cations are available, with very few data points for pyrrolidinium-based cations, despite numerous studies reporting pyrrolidinium cations to be more stable and environmentally friendly than imidazolium [103, 104].

Generating more data for other cation types using experimental techniques would be very expensive and time-consuming as there is a need for IL synthesis, which

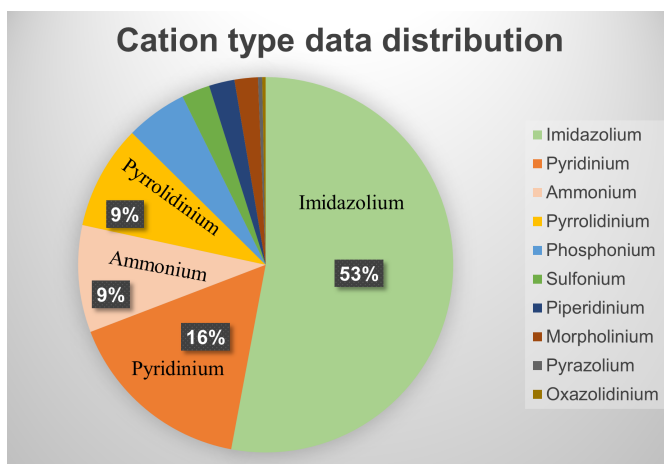


Figure 1.3: Experimental ionic conductivity data based on cation type distribution. Data collected from NIST IL thermo database and other literature sources referenced above.

could involve numerous reaction steps. Similarly, using atomistic simulation to generate a large amount of new data would be highly time-consuming and computationally expensive. Furthermore, as discussed earlier, some of the drawbacks of pure ILs are resolved by mixing ILs to form binary mixtures with non-ideal behavior. However, very few studies have looked at measuring the ionic conductivity of binary mixtures to understand the trends and identify cations and anions that could induce such behavior [51]. Conducting experiments on generating new binary ionic liquid mixture data would be extremely time-consuming and costly, as one now has to consider the composition dependency of the mixture.

1.1.4 Research Focus

This data scarcity challenge and the difficulty narrowing the chemical space provide a unique opportunity to use machine learning techniques to build a robust and accurate model capable of generating large ionic conductivity data by learning patterns within the available experimental data.

In addition, as mentioned earlier, the potential ILs are in the range of 10^{14} [9], yet there are only a handful of cation types known over all these years. Besides correlating properties, this work focuses on using deep learning ML methods to help accelerate the discovery of new ionic liquid structures beyond the existing ones to expand the potential pool of candidates with the desired property.

1.1.5 Research Scope

1. Develop a machine learning model based on existing data to predict ionic conductivity and expand the ionic conductivity data space.
2. Explore the possibility of generating binary mixture data using a model derived from pure ionic liquid data.
3. Understand and provide insights into molecular-level interactions and dynamics for a binary and reciprocal mixture using molecular dynamics.

4. Expand the model for ionic conductivity beyond imidazolium to include pyrrolidinium, sulfonium, and other widely studied ionic liquid cations.
5. Accelerated discovery of new cation types beyond the existing ones to expand the potential list of candidates as electrolytes for battery applications.
6. Explore the possibility of building a machine learning model to predict density functional theory (DFT) properties such as HOMO/LUMO.

1.1.6 Dissertation Outline

Following the introductory chapter focused on the introduction of some of the most widely studied ILs, their properties, advantages, and limitations compared to conventional electrolytes for energy application, the rest of the dissertation is outlined as follows:

Chapter 2 provides a brief overview of some of the work done using a computational prediction-based approach to correlate properties and expand the ionic liquid space. This includes some of the most widely used prediction approaches, from simple linear models to molecular dynamics simulation, deep learning methods, and density functional theory (DFT) approaches. The scope of the literature review is focused on ionic conductivity and electrochemical stability of ILs, as those are the two properties of interest for this work.

Chapter 3 walks through the background and methodology of the machine learning approach, including the theory behind some of the widely used machine learning algorithms. Next, a lengthy discussion on the machine learning pipeline includes data gathering protocol, data formatting, feature generation, feature reduction, and hyper-parameter tuning for model development. The chapter also focuses on the theory behind molecular dynamic simulation and the derivation of some of the properties that are often calculated from molecular dynamic simulations. Lastly, the chapter ends with an overview of the concept behind calculating electrochemical stability from DFT calculations.

Chapter 4 discusses the development of a machine learning model for correlating the ionic conductivity of imidazolium ILs. The paper expands the ionic conductivity of imidazolium at room temperature to 1102 unique ILs by combining all the cations and anions present in the database. Lastly, the possibility of predicting the ionic conductivity of binary mixtures using a model derived from pure ionic liquid data is evaluated by comparing the prediction against all available experimental mixture data. Analysis of the ionic conductivity mixture data reveals a large number of systems to show enhancement or suppression in ionic conductivity as a function of composition.

Chapter 5 employs molecular dynamic simulation to study the molecular level interaction that drives such non-ideal behavior in mixtures, as seen in the previous chapter. The paper focuses on binary cation mixtures (two cations and one anion), binary anion mixtures (two anions and one cation), and reciprocal mixture (two cations and two anions). Based on the analysis, the results indicate strong preferential interactions between particular cation and anion, leading to significant enhancement in the hydrogen bond dynamics for the mixtures.

Chapter 6 expands the machine learning model to cations beyond the imidazolium ILs to cover nine different cation families. Instead of utilizing the neural network model, which is considered a 'black box model', this paper evaluates three different machine learning approaches that offer more insights on the feature importance in the model decision making. The paper also examines the possibility of using some of these feature insights to build a simple classification model that aims to classify cations into different ionic conductivity categories.

Chapter 7 explores the use of advanced deep learning methods to accelerate the discovery of new cations to expand the potential list of candidates as electrolytes for battery application. The paper discusses the methodology behind the generative algorithm to discover new cations and the post-processing steps involved in

processing the data for duplication, structure validation, and stability tests. Some of the cations discovered in this work occupy a new region of distinct space in the cation chemical space, signifying the presence of a new cation family. Lastly, the electrochemical stability of these cations is computed using density functional theory (DFT) calculations.

Chapter 8 evaluates the possibility of developing a machine learning model to correlate two important DFT properties. The hope is to build a robust model capable of mimicking DFT prediction, eliminating the need to further use computationally expensive DFT calculations for the rest of the chemical space.

Chapter 9 summarizes the findings of the work presented in the previous chapters and potential future directions.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

Several studies in the ionic liquid community have used a statistical prediction-based approach to fill gaps in thermophysical data of ILs by correlating properties using available experimental measurements. However, the majority of the work has focused on density, viscosity, melting point, and gas solubility using various predictive methods [105–108] with a limited amount of study done on ionic conductivity and electrochemical window predictions. This section discusses each of the commonly used predictive methods for correlating ionic conductivity and electrochemical window of ILs.

2.1.1 QSPR Method

The Quantitative-Structure-Property-Relationship (QSPR) method is one of the most basic and effective form of predictive models. It is a structure-property mapping technique that connects chemical descriptors or features to the output in a linear form. This is represented as follows:

$$\hat{y} = constant + x_1 feature_1 + x_2 feature_2 + x_n feature_n \quad (2.1)$$

where \hat{y} is the predicted data, x_1 , x_2 and x_n are the weight of the features for *feature*₁, *feature*₂ and *feature*_n. One of the major advantages of the QSPR method over other predictive forms is the qualitative insight it can provide based on the descriptor weights indicating the importance of each descriptor concerning output. Several studies have attempted to use this prediction technique to correlate ionic conductivity of ILs [109–111].

Tochigi et al. [109] developed a linear quantitative structure-property relationship (QSPR) to predict ionic conductivity for eight different cation families, and sixteen different anions. The authors reported an overall R^2 of 0.91 and a maximum error of 4.191 for 139 data points. One of the significant advantages of the QSPR method is the ability to infer insights on feature importance concerning the output. For instance, the positive sign of coefficients indicates the positive influence on ionic conductivity, such as temperature, which is known to accelerate dynamics as the temperature is increased. For the anions, all of them have negative coefficients of different magnitude, with the least for $[\text{BF}_4]^-$ and $[\text{CF}_3\text{BF}_3]^-$ as cations paired with these two anions have higher ionic conductivity than the rest.

Johansson et al. [110] found the molar volume of cation-anion calculated using the *ab initio* method to have an excellent correlation with molar ionic conductivity. However, this correlation is only limited to six pyrrolidinium ILs, and there is no mention of whether this correlation would extend to other ILs. Eiden et al. [111] demonstrated the use of Gibbs solvation energy, molecular radius, ion volume, and symmetry number as input parameters to correlate the dielectric constant, viscosity, and ionic conductivity of ILs. The authors' rationale for using Gibbs solvation energy was to consider the ion's interaction with its nearby charged species that could affect the macroscopic transport properties. The ionic conductivity model showed a correlation coefficient R^2 of 0.74 and RMSE of 0.25.

QSPR methods are often preferred for small data sets as they can generate qualitative insights and identify the importance of each of the descriptors. However, these

methods fail to capture complex non-linear trend structure-property relationships resulting in low correlation accuracy.

2.1.2 Group Contribution Method

Another prevalent prediction technique is the group contribution (GC) method which relies on the functional group as input descriptors instead of chemical descriptors, as we saw earlier. Similar to QSPR with descriptors, the functional groups can also provide valuable insights into the structure-property relationships.

$$y = constant + \sum_i n_i W_i \quad (2.2)$$

where W_i is the weight of the group (i), n_i is the number of instances a given group is repeated throughout the molecule. Several studies have attempted to correlate ionic conductivity of ILs using a GC-based approach [112–114]. Coutinho et al. [112] used a three-parameter GC method equation to estimate the ionic conductivity for pure ILs.

$$\ln\lambda = \ln A_\lambda + \frac{B_\lambda}{(T - T_{0\lambda})} \quad (2.3)$$

where λ is ionic conductivity, A_λ , B_λ and $T_{0\lambda}$ are parameters for the equation. The two parameters (A_λ and B_λ) can be derived using GC approach as follows:

$$A_\lambda = \sum_{i=1}^k n_i a_{i\lambda} \quad (2.4)$$

$$B_\lambda = \sum_{i=1}^k n_i b_{i\lambda} \quad (2.5)$$

where n_i is the number of instances the group (i) is repeated, $a_{i\lambda}$ is the weight of the group. The parameter $T_{0\lambda}$ is a temperature dependency parameter that takes into account the relation between temperature and ionic conductivity derived during model fitting. The database consists of five cation family groups, seven anion groups, and one functional group. Overall the author reported an R^2 of 0.997.

Similarly, Wooley et al. [113] applied a four-parameter GC-based approach to estimate the ionic conductivity of ILs.

$$\ln \frac{\epsilon}{R_{0\epsilon}} = A_\epsilon + B_\epsilon \frac{100}{T} + C_\epsilon \left(\frac{100}{T}\right)^2 \quad (2.6)$$

where ϵ is ionic conductivity, $R_{0\epsilon}$ an adjustable parameter, A_ϵ , B_ϵ and C_ϵ are parameters based on group contribution.

$$A_\epsilon = \sum_{i=1}^k n_i a_{i\epsilon} \quad (2.7)$$

$$B_\epsilon = \sum_{i=1}^k n_i b_{i\epsilon} \quad (2.8)$$

$$C_\epsilon = \sum_{i=1}^k n_i c_{i\epsilon} \quad (2.9)$$

The GC database contained eight cation groups, 34 anion groups, and four functional side groups derived from 1578 data points. The authors reported an average absolute relative deviation (AARD %) of 6.83 for the test set and 3.30 for the training set.

Gharagheizi et al. [114] employed a least-squares support vector machine GC method to estimate ionic conductivity consisting of a dataset with 54 different

unique ILs with an AARD of 3.3%. As one can imagine, the GC method is only applicable as long as there are functional groups present in the database beyond which it is not applicable.

2.1.3 Machine Learning Approach

However, with more experimental data and computational power, the focus has shifted towards using machine learning methods to build complex structure-property relationships to improve the model’s predictive capability.

Hezave et al. [115] deployed a neural network to predict electrical conductivity of ternary mixture constituting of 1-butyl-3-methylimidazolium hexafluorophosphate + water + ethanol and 1-butyl-3-methylimidazolium hexafluorophosphate + water + acetone. The model utilized temperature, molecular weight, and mixture composition as input descriptors trained on 78 data points. The authors reported a R^2 of 0.99 for both the test and training sets. Similarly, Nordness et al. [116] in a recent study, explored the use of support vector regression (SVR) to correlate the thermophysical properties of ILs that included density, viscosity, and ionic conductivity based on descriptors calculated using conductor-like screening model for real solvents (COSMO-RS). Using the SVR model, the authors trained 1305 data points to correlate ionic conductivity with a correlation coefficient R^2 of 0.955 and AARD of 10.2%. Koi et al. [117] also utilized (COSMO-RS) descriptors in a recent paper to develop an MLR and SVR model to correlate ionic conductivity of imidazolium-based ILs. The authors used a data set of 239 data points that consisted of 25 different ILs. The SVM method obtained higher accuracy with a R^2 of 0.990 than the MLR method with R^2 of 0.855, indicating a non-linear structure-property relationship that the MLR fails to consider.

Despite the high accuracy of all the regression models in correlating ionic conductivity, none of the papers explore the possibility of generating ionic conductivity data using the cations and anions available in the model data set. This is impor-

tant, as discussed in the introduction chapter (Figure 1.2), as there is a definite need for more ILs with ionic conductivity greater than 2.0 S/m to match the performance of conventional electrolytes. Besides generating pure ionic liquid data, there is also a lack of study based on the literature review that has explored the possibility of generating binary ionic liquid mixture ionic conductivity data using machine learning techniques to examine any non-ideal behavior such as enhancement in ionic conductivity as a function of mixture composition.

Recently, deep learning methods have gained tremendous popularity across all domains with the availability of powerful computational resources to accelerate material discovery. Beckner et al. [118] in their recent study, used QSPR based deep neural network model to train models to predict heat capacity and density of ILs. The cations present in the data set were broken down into functional groups and used for genetic mutation of new cations. These mutated ILs were then passed through the heat capacity and density model to filter ILs with the highest value. Quantum mechanical (QM) calculations and molecular dynamics simulation analysis were performed on these mutated cations to validate the property prediction and ensure the structures were stable and in the liquid state. Once identified as a stable liquid, the heat capacity and density model were re-formulated by including these high-performing ILs to exceed the existing data. The authors were able to identify several high-performance novel ILs with high capacity and density data.

Beckner et al. in another paper utilized a generative-based model known as variational autoencoder (VAE) along with a QSPR-based model to generate and predict properties for new ILs. The authors employed a transfer learning approach that initially learns the chemical space from an extensive list of organic molecule SMILES followed by a smaller set of ionic liquid SMILES to generate new ionic liquid structures with desired property range [119]. The central concept behind VAE is to store information about the input data in a low-dimensional latent space that serves as a medium to generate new data with the desired property. Besides this work, there are no studies based on the literature review that have explored

the use of generative machine learning algorithms such as VAE to generate new cations and anions to expand the chemical space for ILs with desired properties.

2.1.4 Molecular Dynamics

Several studies have also utilized molecular dynamics (MD) simulation to predict the thermophysical properties of ILs. Properties such as density and heat capacity are relatively less intensive calculations and easier to predict with high accuracy. However, calculating transport properties such as viscosity, self-diffusion, and ionic conductivity has been challenging as it requires a long simulation time and is under-predicted by a large magnitude.

Lee et al. [120] calculated the ionic conductivity of 1-n-butyl-3-methylimidazolium paired with five fluoro-based anions. The authors calculated ionic conductivity using the Nernst-Einstein equation with values twice as low in magnitude as the experimental data. Reddy et al. [121] in a 2020 study conducted an MD simulation on hydroxylammonium-based cations paired with lactate and formate anions. The authors computed several thermophysical properties along with ionic conductivity. The Einstein equation for calculating ionic conductivity was found to deviate significantly compared to experimental data.

The current non-polarizable forcefields used for IL MD simulation require charge scaling to induce polarizability to improve transport properties. The charge scaling is often scaled by a factor of ± 0.8 for most studies computing transport properties [122]. However, even that is not enough in some cases as it requires further scaling to match experimental data. Amir et al. [123] calculated the thermophysical properties using a function of electrostatic charge scaling. The authors found the charge scaling of ± 0.65 to have an excellent agreement compared to property prediction, including ionic conductivity.

Zeindlhofer et al. [124] calculated the ionic conductivity of 1-ethyl-3-methylimidazolium

dicyanamide and tetrafluoroborate binary mixture as a function of composition using MD simulations. The authors employed a united atom model with a charge scaling of ± 0.78 to induce non-polarization. Using MD simulation, calculating transport properties for ILs are often computationally expensive as a long simulation time is required to ensure the system is well equilibrated. The authors simulated the system for about 380 ns. The paper reported an excellent agreement in ionic conductivity value calculated using the Einstein equation compared to experimental data.

Generating an extensive amount of ionic conductivity data using MD simulation would not be feasible because of the high computational cost as the simulation needs to run for an extended period. Instead, MD simulations can be used to gather insights on molecular-level interactions for a few selected systems to explain the macroscopic level property.

2.1.5 Electrochemical Stability

Besides ionic conductivity, the electrochemical stability of an electrolyte is also an essential property as it characterizes the electrochemical window (ECW) at which the ionic liquid is neither susceptible to oxidation nor reduction. Several experimental studies have looked at measuring ECW for several ILs using different techniques [11]. Xue et al. [125] compiled an extensive list of experimental ECW data for a large number of ILs with various cations and anions. The authors reported the phosphonium cations followed by pyrrolidinium to be the most stable, while imidazolium is at the bottom of the list as it gets easily reduced. However, relying on experimental techniques to measure ECW for every ionic liquid can be tedious, time-consuming, and expensive as the potential can vary with the type of cation, anion present, and length of the functional group attached to the cation/anion. Thus several studies have explored the possibility of using computational techniques to calculate the ECW of ILs.

The ECW calculations using computational techniques are performed using: (a) Thermodynamic cycle (TDC) approach, (b) One-electron approach, and (c) Highest occupied molecular orbital (HOMO)/ Lowest occupied molecular orbital (LUMO) approach. Asha et al. [104] reported a maximum deviation of 26% in ECW for pyrrolidinium ILs using the HOMO/LUMO approach compared to the TDC method, which had a maximum deviation of 9.0%. Similarly, Kazemiabnavi et al. [42] found a similar trend for the ECW of imidazolium ILs using the HOMO/LUMO method with deviation up to 50% compared to 15% using the TDC approach.

Liang et al. [126] calculated the HOMO/LUMO of 42 cations and 42 anions that were then combined to form 1764 unique ILs. The authors reported the cations paired fluoro and borate-based anions to have high ECW as these anions do not get oxidized easily. Similarly, Ilawe et al. [127] used the HOMO/LUMO approach to calculate the electrochemical and radiation stability of the 42 cations and 42 anions using various levels of DFT theory. The authors reported the HOMO/LUMO calculation accuracy to vary based on the functional and basis set.

Panidian et al. [128] did an extensive study on calculating the ECW of ILs using various methodologies. The authors found the Δ SCF Self Consistent Field approach of calculating ECW in a vacuum to have a deviation of 48 - 54.1% compared to the experiment data depending on the basis set. Changing the vacuum environment to an SMD solvent (dichloroethane) and using the one-electron approach reduced the deviation to 2.4%. The one-electron addition calculates the ECW using the reduction potential of the cations and the oxidation potential of the anions. The reduction potential is the difference in Gibbs free energy between V_C and V_{C+e^-} . Similarly, the oxidation potential is the Gibbs free energy difference of V_A and V_{A-e^-} . However, no studies based on literature review have calculated electrochemical window using the HOMO/LUMO approach in the presence of a solvent rather than a vacuum to see if that reduces the deviation in electrochemical window calculations compared to experimental measurement.

To summarize, the work in this dissertation focuses on expanding the ionic conductivity database using a machine learning approach to investigate the search for high ionic conductivity ILs. Aside from the widely studied imidazolium cation, the study also examines the possibility of correlating ionic conductivity for other cation types with limited data to expand the list of potential candidates. Beyond the known cation space, the latter part of the work aims to accelerate the discovery of new cations with high electrochemical stability. The electrochemical stability is calculated using the HOMO/LUMO approach, including solvent effects, to improve ECW predictions. Lastly, this work also demonstrates the possibility of building a machine learning model to correlate DFT properties such as HOMO/LUMO energies with high accuracy to reduce the computational cost and time in running DFT calculations for the enormous ionic liquid space.

CHAPTER 3

METHODOLOGY

3.1 Machine Learning

3.1.1 Multiple Linear Regression

Multiple linear regression (MLR) is a simple regression-based model that attempts to model the relationship between input data and output predictor by fitting a best fit line through the data. The model assumes the relationship between the dependent and independent variables to be in a linear form, thus failing to take into account any non-linear relationship between input and output. The equation for MLR can be written as follows:

$$\hat{y} = b + a_1x_1 + a_2x_2 \dots + a_ix_i \quad (3.1)$$

where \hat{y} is the predicted data, b is the intercept, a_1 , a_2 and a_i are the coefficients of the input data (x_1 , x_2 and x_i). Thus the objective function of the MLR model can be written as:

$$J(b, a_i) = \frac{1}{2n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3.2)$$

where (J) is the cost function that needs to be minimized, (n) in the total number of data points, (y_i) is the original data, and (\hat{y}_i) is the predicted data.

3.1.2 Cost Function

The cost function is often minimized using the gradient descent algorithm, widely used across all machine learning algorithms to optimize the model parameters [129,130]. This can be derived using an example as shown below. Let us consider a two-parameter equation:

$$\hat{y} = mx + b \quad (3.3)$$

The cost function would then be:

$$J(m, b) = \frac{1}{2n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3.4)$$

Substituting equation 3.3 in equation 3.4 for \hat{y}_i :

$$J(m, b) = \frac{1}{2n} \sum_{i=1}^n ((mx + b) - y_i)^2 \quad (3.5)$$

Gradient descent algorithm initially starts with a random value of (m and b). The parameters are then updated using the following equation:

$$b = b - \alpha \frac{\delta}{\delta b} J(m, b) \quad (3.6)$$

$$m = m - \alpha \frac{\delta}{\delta m} J(m, b) \quad (3.7)$$

where α is known as the learning rate (LR) parameter. The α term determines how much the parameters are scaled every time (m and b) are updated. A small α parameter means the model converges extremely slowly as the descent is minimal from one point to another. In contrast, a high α can lead to a big jump in the descent, missing the global minimum. Thus, a balanced α parameter is key to the success of converging the gradient descent algorithm to a global minimum. It is worth noting that the weights of (m and b) in equations 3.6 and 3.7 are simultaneously updated.

Substituting equation 3.5 in equation 3.6,

$$b = b - \alpha \frac{\delta}{\delta b} \frac{1}{2n} \sum_{i=1}^n ((mx + b) - y_i)^2 \quad (3.8)$$

Solving the partial differential equation:

$$b = b - \alpha \frac{1}{n} \sum_{i=1}^n ((mx + b) - y_i) \quad (3.9)$$

Similarly, substituting equation 3.5 in equation 3.7,

$$m = m - \alpha \frac{\delta}{\delta m} \frac{1}{2n} \sum_{i=1}^n ((mx + b) - y_i)^2 \quad (3.10)$$

Solving the partial differential equation:

$$m = m - \alpha \frac{1}{n} \sum_{i=1}^n ((mx + b) - y_i) \cdot x \quad (3.11)$$

The objective of the gradient descent algorithm is to find the global minimum of a given function. In this case, the updated (m and b) parameters are substituted back into equation 3.5 to calculate the cost function. The algorithm stops once the cost function reaches a given threshold or until it exceeds the number of iterations specified.

3.1.3 Decision Tree

Decision tree (DT) is a supervised machine learning technique that uses a tree-based approach to make a prediction. This technique is widely used for both regression and classification problems. The model's decision-making is based on if-else conditions by splitting data based on input attributes [131].

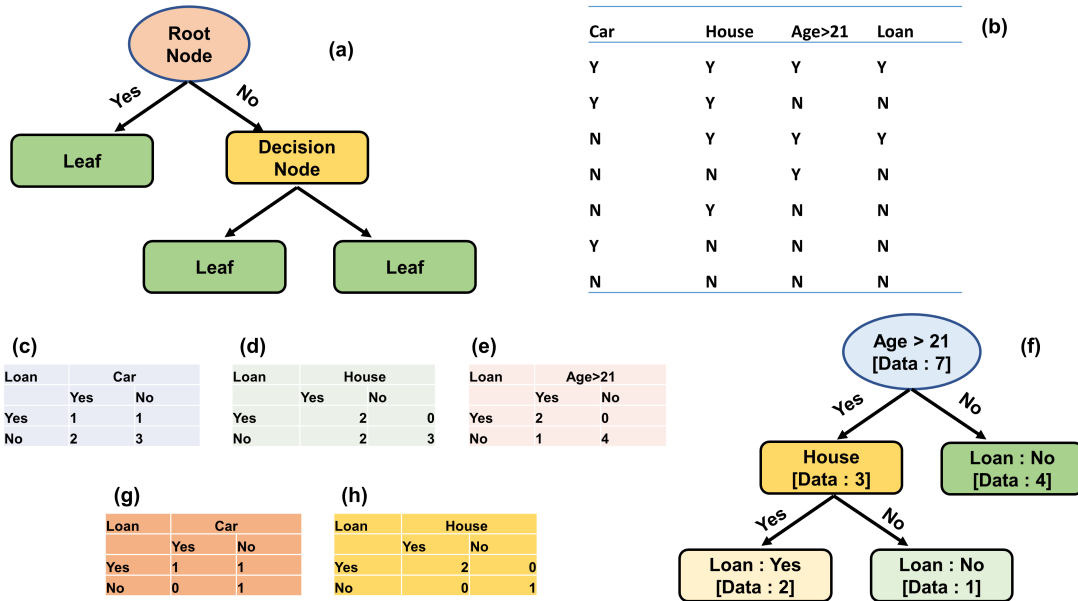


Figure 3.1: Schematic diagram of a decision tree model.

The top of each DT is called the node that begins with an attribute that asks a yes/no question, as seen in Figure 3.1 (a). Each decision (yes/no) is followed by tree branches that further split various attributes into (yes/no) conditions. The branches get divided until they can no longer be divided, which becomes the final node or leaf of the tree.

The gini index $H(T)$, is used to determine which attribute becomes the tree's node. The gini index term measures each attribute's impurity concerning the output value. A gini index of 0 means the data is easily separable into different classes.

$$H(T) = 1 - \sum_{i=1}^n p_i^2 \quad (3.12)$$

Consider the example shown in Figure 3.1 (b) for a loan decision based on three different attributes (car, house, age > 21). First, the gini index is calculated for each feature to determine which one becomes the node for the tree. This is done by separating the data for each of the attributes concerning the output (loan) as

seen in Figure 3.1 table (c), (d), and (e).

For car attributes, there is one data point for someone with a car and loan approved. However, two of the data points had loans rejected despite having a car. So the gini index for car attributes concerning loans is 0.44. Similarly, there is one data point with a loan approved despite having no car, while three of the data points have loans rejected for having no car. Thus, the gini index for having no car becomes 0.37.

$$H(T) = 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2 = 0.44 \quad (3.13)$$

$$H(T) = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2 = 0.37 \quad (3.14)$$

Thus, the weighted gini index for car attribute becomes:

$$H(T) = \frac{3}{7} * 0.44 + \frac{4}{7} * 0.37 = 0.40 \quad (3.15)$$

Similarly, for house attributes, there are two data points with house and loan approved, while two data points with house have loan rejected. There are zero data points for loans being accepted with no house, while three of them have had loans denied for having no house.

$$H(T) = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.50 \quad (3.16)$$

$$H(T) = 1 - \left(\frac{0}{3}\right)^2 - \left(\frac{3}{3}\right)^2 = 0 \quad (3.17)$$

Thus, the weighted gini index for house attribute becomes:

$$H(T) = \frac{4}{7} * 0.50 + \frac{3}{7} * 0 = 0.28 \quad (3.18)$$

Lastly, for age>21 attribute:

$$H(T) = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 = 0.44 \quad (3.19)$$

$$H(T) = 1 - \left(\frac{0}{4}\right)^2 - \left(\frac{4}{4}\right)^2 = 0 \quad (3.20)$$

Thus, the weighted gini index for age>21 attribute becomes:

$$H(T) = \frac{3}{7} * 0.44 + \frac{4}{7} * 0 = 0.19 \quad (3.21)$$

Based on the weighted gini index, the attribute (age > 21) has the lowest gini index, which means this attribute has the highest leverage for splitting the output data. As seen in Figure 3.1 (f) the attribute (age > 21) becomes the node of the tree. Splitting the attribute to yes/no, from the table, there are four data points whose loan is denied (No) if their age is less than 21. Thus, one can see why this attribute effectively splits the data by being at the top of the node. There are three remaining data points on the right side that need to be separated. To decide the next attribute for the decision node, the weighted gini index is calculated again for the remaining data points.

For car attribute again:

$$H(T) = 1 - \left(\frac{1}{1}\right)^2 - \left(\frac{0}{1}\right)^2 = 0 \quad (3.22)$$

$$H(T) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.50 \quad (3.23)$$

Thus, the weighted gini index for car attribute becomes:

$$H(T) = \frac{1}{3} * 0 + \frac{2}{3} * 0.50 = 0.33 \quad (3.24)$$

For house attribute:

$$H(T) = 1 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2 = 0 \quad (3.25)$$

$$H(T) = 1 - \left(\frac{1}{1}\right)^2 - \left(\frac{0}{1}\right)^2 = 0 \quad (3.26)$$

Thus, the weighted gini index for house attribute becomes:

$$H(T) = \frac{2}{2} * 0 + \frac{1}{1} * 0 = 0 \quad (3.27)$$

Based on the weighted gini index, it is clear that the house attribute is best suited to splitting the remaining tree as it has a weighted index of 0, meaning it can separate all the remaining data. This is evident from Figure 3.1 (f) as two of the data points with houses get the loan, while the one data point with no house does not get the loan. As one can see, the DT method is pretty straightforward, which

makes it easier to follow through and make predictions.

However, a common drawback of the DT model is that it is very prone to overfitting as it often fails to generalize data beyond the training set. Furthermore, even a slight change in the data can completely change the entire split, resulting in a different tree with different outputs. Thus, the DT model is generally avoided for a large set of data with multiple attributes.

3.1.4 Random Forest

Random forest (RF) is an ensemble-based supervised-based learning method that improves upon some of the significant weaknesses of DT-based methods. Instead of a single tree, RF consists of many individual trees that run in parallel. Initially, each tree is constructed by randomly selecting a subset of bootstrapped data with replacement. Next, the algorithm selects a random set of features from the bootstrapped data, creating a unique tree. The central idea is to have an ensemble of trees with different attributes to make predictions for a given dataset instead of a single tree predicting the output, thus removing any biases.

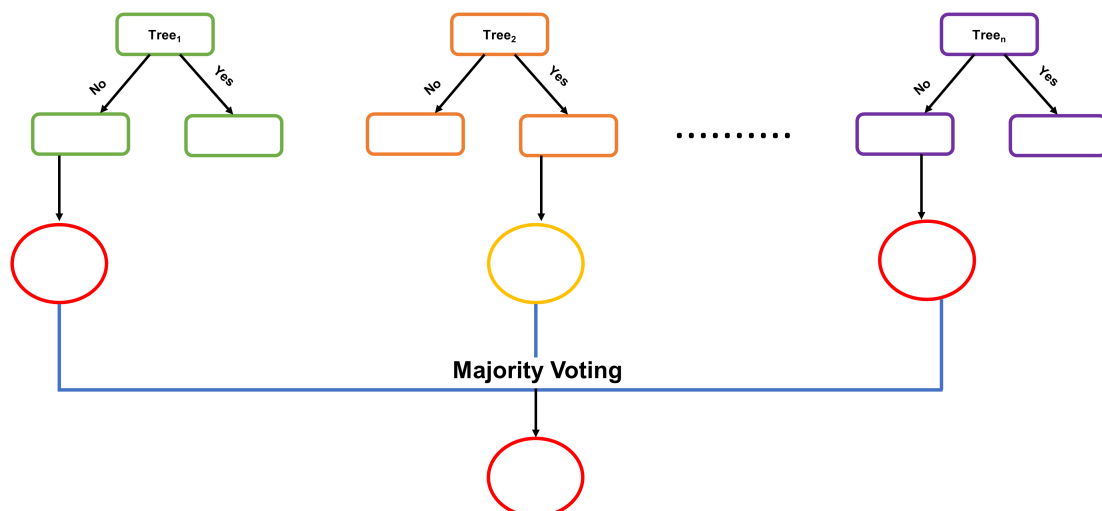


Figure 3.2: Schematic diagram of a random forest model.

The final output for a given dataset is obtained by majority-based voting for a classification model, where the highest Yes/No becomes the final output, while the prediction is the average of all the ensemble predictions for regression. The RF model has been shown to drastically improve the robustness and accuracy of the model as it is significantly less sensitive to the change in data and overfitting compared to DT.

3.1.5 XGBoost

Extreme gradient boosting (XGBoost) is another ensemble-based supervised learning technique that uses gradient boosting to predict properties [132]. Similar to the RF model, XGBoost contains an ensemble of unique decision trees that are utilized to make a prediction. Unlike RF, however, these trees are built sequentially, where the error made by the previous tree is added to the loss function of the next tree and so on until the prediction error is within the desired threshold. And this concept of boosting the prediction is at the heart of the XGBoost model.

For any given model, the objective function is to minimize the difference in prediction and actual value, as shown below:

$$OBJ = \sum_i^n loss(y_i, \hat{y}_i^{t-1} + \hat{y}_i) \quad (3.28)$$

where \hat{y}_i is the predicted value. In XGBoost [132] however the loss function is modified to add the error made by previous tree (\hat{y}_i^{t-1}) as discussed above.

3.1.6 Support Vector Machine

Another supervised learning algorithm that is widely used for classification and regression problems is the support vector machine (SVM). SVM uses the concept of a hyper-plane, which is a line to separate the data into respective classes. For

instance, in Figure 3.3 (a), the hyper-plane (H1, H2, H3) aims to separate the two shapes into respective classes. As for which hyper-plane leads to the highest accuracy in separating the two classes, it is determined by maximizing the distance between points and the hyper-plane, also known as the maximum margin [133–138].

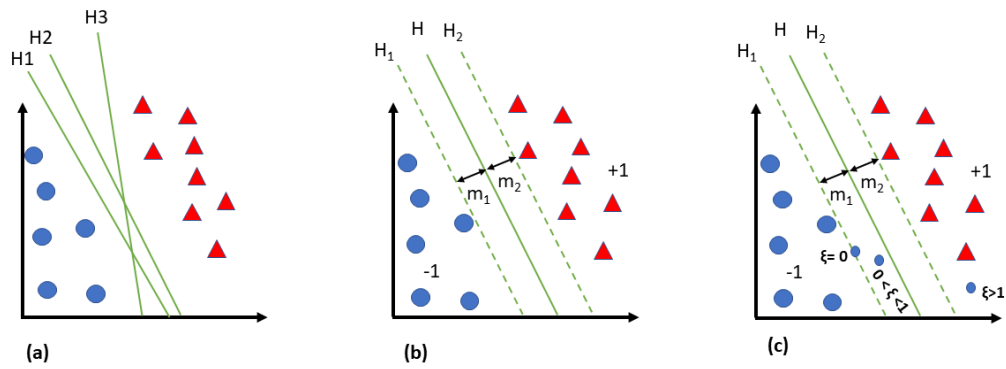


Figure 3.3: Hyperplane for separating the data points into respective classes for support vector machine.

The equation to define the hyperplane can be written as:

$$y_1 = ax_1 + b \quad (3.29)$$

or,

$$ax_1 - y_1 + b = 0 \quad (3.30)$$

Let, $w = (a, -1)$ and $x = (x_1, y_1)$. The equation then can be written as:

$$w^T x + b = 0 \quad (3.31)$$

Now, let us assume the class above the hyperplane is denoted by +1, and the category below the hyperplane is indicated by -1, separated by two hyper-plane margins (H_1) and (H_2) as seen from Figure 3.3 (b). The equation for the two hyperplane margins can be written as:

$$w^T x + b \geq 1, \text{ for class} = +1 \quad (3.32)$$

$$w^T x + b < -1, \text{ for class} = -1 \quad (3.33)$$

Thus, the objective function of the SVM would be to maximize the margin (M), which can be written as: $\max (M)$ given $y_i(w^T x + b) \geq 1$. The margin (M) here refers to the distance between the two margin classes (m1 and m2). Maximizing the margin (M) allows enough room between the two margin lines to separate the data into classes with very few misclassifications easily. However, most real-world data does not exhibit such linear separable patterns as there can be outliers making it challenging to separate them into classes. In such a case, the above equation cannot be satisfied. Thus, to consider real-world scenario cases with outliers leading to misclassification, an additional variable, ϵ is added to the equation:

$$y_i(w^T x + b) \geq 1 - \epsilon \quad (3.34)$$

where if a data point is on the correct side of the margin, ϵ takes a value of 0, leaving the equation unchanged. However, if the data point is between the hyperplane and the margin, then this is considered a margin violation. The value of ϵ is taken to be between 0 and 1, serving as a penalty parameter if there is a margin violation. Lastly, if the data is on the wrong side of the hyperplane, then ϵ takes a value greater than 1, indicating misclassification. This can be visualized in Figure 3.3 (c) for more clarity. This ϵ parameter is a hyper-parameter for the

SVM that must be determined using a k-fold cross-validation parameter search. The last important hyper-parameter that needs to be searched and tuned is the 'C' parameter, the penalty parameter for margin violation. Thus, the objective function of the SVM can be written as:

$$\max(M) + C \sum_i^N \epsilon_i \quad (3.35)$$

Smaller value of 'C' allows for higher tolerance for margin violation data points, while a higher value of 'C' adds severe punishment for margin violation. Thus, an optimum value of 'C' is required as a very high value of 'C' can lead to overfitting as the model may not be able to generalize beyond the training set. In contrast, a lower value of 'C' could lead to underfitting.

3.1.7 Neural Network

Neural network (NN) techniques are currently regarded as state-of-the-art machine learning algorithms, modeling complex functions with very high robustness and accuracy. Because of their capability to identify and learn high-level abstract patterns is now being used in day-to-day activities such as self-driving cars, facial recognition, image classification, natural language processing, and more.

In general, NN consists of three essential layers: the input layer that receives the information, the hidden layer that processes the input information and passes it to the output layer, and finally, the output layer that gets the data from the hidden layer and converts it to a meaningful output, as seen in Figure 3.4. The size of the input and output layers is user-specified, while the hidden layer needs to be tuned to get the best performance. Relatively simple hidden layer architecture can lead to underfitting where its architecture limits the model's ability to learn complex functions. In contrast, highly complex model architecture can lead to severe overfitting, where the model fails to generalize predictions outside of the training set.

Parameters such as the number of hidden layers, the number of neurons in the hidden layer, learning rate, and the activation function are some of the essential parameters that need to be tuned using hyperparameter tuning techniques such as random search or grid search combinations to find the optimum hyper-parameters.

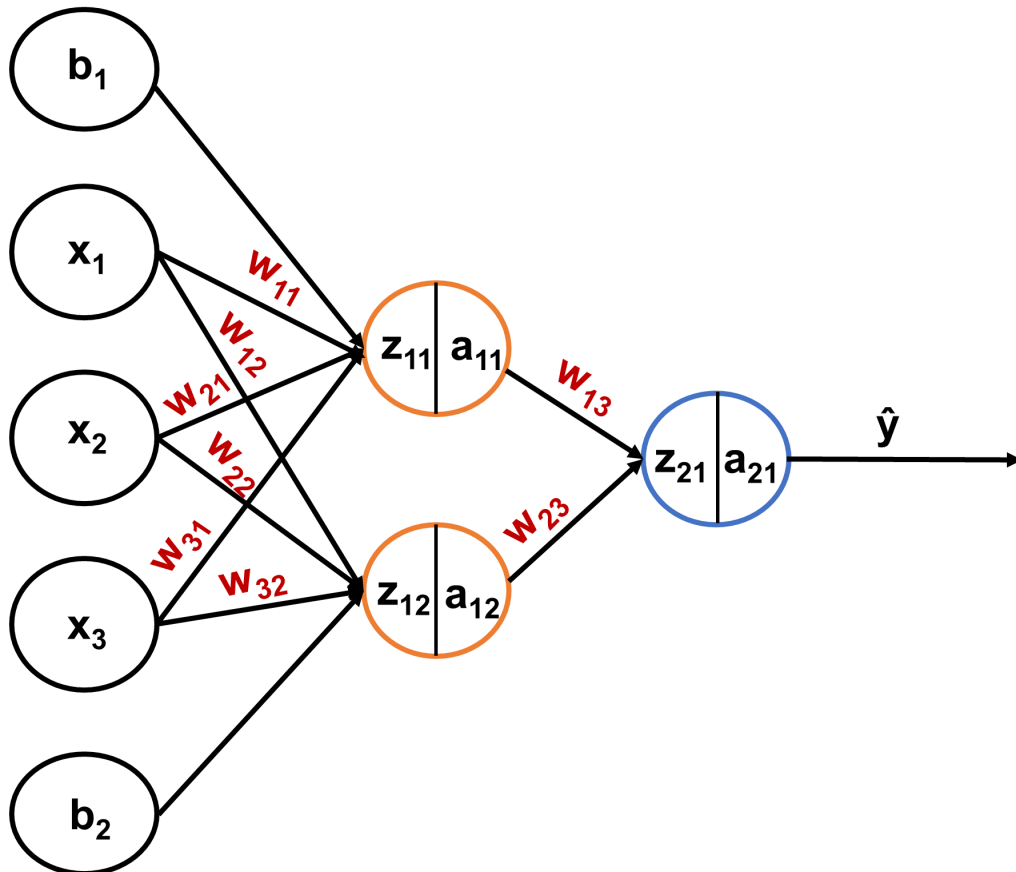


Figure 3.4: Schematic diagram of a neural network model.

Consider a neural network with three input features (x_1, x_2, x_3) with a single hidden layer that consists of two neuron perceptron (z_{11} and z_{12}) receiving the input information as seen from Figure 3.4. Then the equation of the hidden layer can be written as:

$$z_{11} = w_{11}x_1 + w_{21}x_2 + w_{31}x_3 + b_1 \quad (3.36)$$

$$z_{12} = w_{12}x_1 + w_{22}x_2 + w_{32}x_3 + b_2 \quad (3.37)$$

where (w_{ij}) is the weights of the input features for neuron (z_{11}) , (z_{12}) and (b_i) is the bias function for each of the neurons.

The passage from the hidden layer to the output layer is determined by a key NN component known as the activation function. This function acts as a switch to turn "on" or "off" a given neuron based on the input data. There are several kinds of activation functions one can choose from based on the problem at hand. For instance, the most straightforward activation function is the linear function that converts the input data to a linear form. In such a case, the NN becomes a simple linear regression model. However, approximating non-linear relations requires non-linear activation functions such as sigmoid, relu, and tanh, which are some of the most popular ones.

The sigmoid activation function, widely used for classification problems, approximates any given input information to an output in the range between $[0, 1]$, as shown in equation 3.38. Another commonly used activation function for regression problems is the relu activation function that transforms the data with the function $\max [0, z]$, converting any negative value to 0, effectively switching "off" a given neuron. The tanh activation function is widely used for handling negative output with a mapping range of $[-1, 1]$.

$$a_{11} = \frac{1}{1 + e^{-z}} \quad (3.38)$$

Passing the information from the hidden layer (z_{11}) , (z_{12}) through the activation

function:

$$a_{11} = g(z_{11}) \quad (3.39)$$

$$a_{12} = g(z_{12}) \quad (3.40)$$

The output from the activation node is multiplied by its weight, and a bias term is added to the equation:

$$z_{21} = a_{11}w_{13} + a_{12}w_{23} + b_3 \quad (3.41)$$

Often, the output layer also has an activation function that aims at transferring the output value in a given range.

$$a_{21} = g(z_{21}) \quad (3.42)$$

where (a_{21}) refers to final predicted data (\hat{y}). Thus the cost function then becomes:

$$J = \frac{1}{2n} \sum_{i=1}^m J(\hat{y}_i, y_i) \quad (3.43)$$

where (n) is the total number of data points, (y_i) is the original data. Here the objective function would be to minimize the cost function by updating all the weights parameters accordingly using a backpropagation algorithm [139].

3.1.8 Variational AutoEncoder

VAE is a generative-based machine learning model representing the information mapping between input and output data using a probability distribution. It consists of two deep neural networks: an encoder network that learns to encode the input information in a compressed form on a low-dimensional latent space, learning only the specific attributes of the input data, and a decoder network that learns to decompress that latent space information back to reconstruct the original input data, as shown in Figure 3.5 (c). The VAE can generate new data by sampling the low dimensional latent space passing through the decoder.

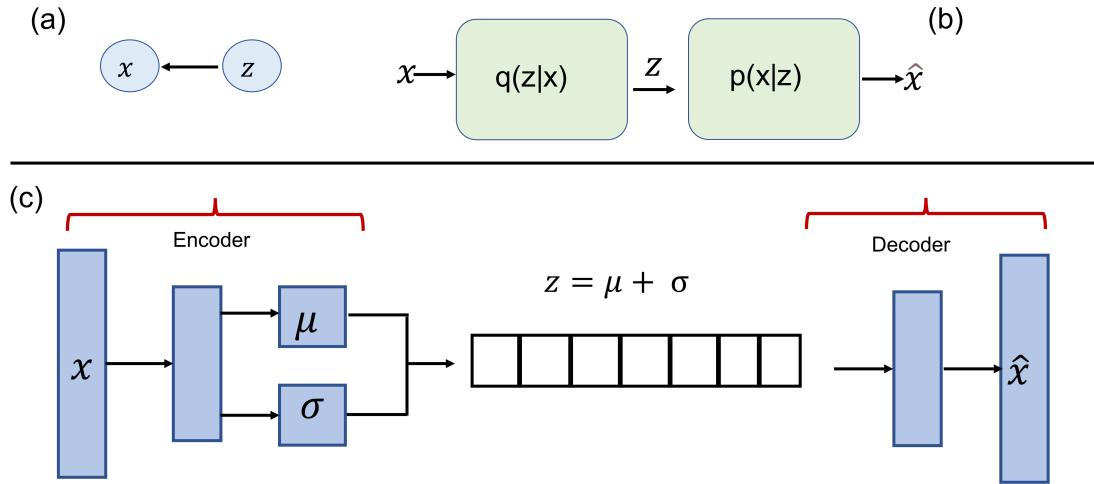


Figure 3.5: Schematic diagram of a variational autoencoder model.

The mathematical representation of VAE starts with a variable (x) that represents the input information. Assume there is a hidden variable (z) in the lower dimensional space that contains the details about (x) that are required to reconstruct it back to its original form (\hat{x}). To infer the characteristics of latent space (z), given the input observation using Bayes' theorem [140]:

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x, z)}{p(x)} \quad (3.44)$$

The challenge, however, arises when calculating the probability distribution of x , $p(x)$ using the latent space information:

$$p(x) = \int p(x|z)p(z)dz \quad (3.45)$$

where, $p(z)$ and $p(x|z)$ follows a standard Gaussian distribution. Maximizing $p(x)$ to reconstruct the information back to the output state with very high accuracy can be challenging and tedious. As it needs to search over the entire latent space (z). The workaround to this problem is by assuming $p(z|x)$ to another traceable distribution $q(z|x)$, where (q) is a known Gaussian distribution with a mean of (g) and standard deviation of (h).

$$q_x(z) = N(g(x), h(x)) \quad (3.46)$$

The parameters (g) and (h) becomes the output parameters for the encoder model seen in Figure 3.5 (c) as it tries to find parameters for (q) that closely match the distribution (p).

The loss between the two Gaussian distributions (p) and (q) is calculated using Kullback-Leibler (KL) divergence [141] that quantifies the amount of information loss that occurs using the approximated distribution (q).

$$KL(q(z|x)||p(z|x)) = - \sum q(z|x) \log \frac{p(z|x)}{q(z|x)} \quad (3.47)$$

Replacing $p(z|x)$ using equation 3.44:

$$KL(q(z|x)||p(z|x)) = - \sum q(z|x) \log \frac{p(x|z)p(z)}{q(z|x)} + \log(p(x)) \quad (3.48)$$

moving $\log(p(x))$ to the left hand side:

$$\log(p(x)) = KL(q(z|x)||p(z|x)) + \sum q(z|x) \log \frac{p(x|z)p(z)}{q(z|x)} \quad (3.49)$$

where the second term is known as Evidence Lower Bound or (ELBO).

$$ELBO = \sum q(z|x) \log \frac{p(x|z)p(z)}{q(z|x)} \quad (3.50)$$

Expanding the ELBO equation:

$$ELBO = \sum q(z|x) \log(p(x|z)) + \sum q(z|x) \log \frac{p(z)}{q(z|x)} \quad (3.51)$$

The second term in equation 3.51 represents the KL divergence.

$$KL(q(z|x)||p(z)) = - \sum q(z|x) \log \frac{q(z|x)}{p(z)} \quad (3.52)$$

Equation 3.51 then becomes,

$$ELBO = \sum q(z|x) \log(p(x|z)) - KL(q(z|x)||p(z)) \quad (3.53)$$

Or,

$$ELBO = E_{q(z|x)}[\log(p(x|z))] - KL(q(z|x)||p(z)) \quad (3.54)$$

where the first term is the expectation of how well (x) is reconstructed from the latent space (z) using the distribution (p). The second term is a deterministic function that minimizes the difference between the Gaussian distribution (q) and the prior distribution (p).

The shape of the attribute (z), which represents some information about the input variable (x), is determined based on the mean and the standard deviation from the encoder. The decoder then samples the latent space distribution (z) to form the distribution $p(\hat{x}|z)$ that is close to the distribution $q(z|x)$ to construct the output \hat{x} accurately.

To recap everything, we started with an input (x) which is inferred by a hidden variable (z) that contains the necessary information to reconstruct the original input (\hat{x}). The mapping function from (x) to (z) is given by the known Gaussian distribution (q), which matches another Gaussian distribution (p) that maps (z) back to (\hat{x}).

3.1.9 Principal Component Analysis

Principal component analysis (PCA) is a dimensionality reduction technique that is extremely useful for compressing large feature spaces into smaller dimensions. This technique often comes in handy when there is a need to reduce the features to speed up the model building process or explore and visualize the large feature

space.

The first step in PCA involves feature normalization to ensure all the features are on the same scale to avoid any bias towards a given feature. This is usually done by scaling the input data with the mean and standard deviation, converting the scaled data to have a mean of zero. In the next step, a covariance matrix is calculated between the feature itself and other features to quantify the variance using:

$$cov(X, Y) = \frac{1}{n-1} \sum_i^n (X_i - \bar{x})(Y_i - \bar{y}) \quad (3.55)$$

where X and Y are the feature columns, X_i and Y_i is the feature value at position i , \bar{x} and \bar{y} is the mean value of the standardized column. The dimension of the covariance matrix (A) is based on the number of features (f) present, which equals $f \times f$. Next, the eigenvector (v) and the eigenvalues (λ) of the covariance matrix is calculated using the identity matrix relation ($Av - \lambda Iv = 0$).

After the calculation of eigenvalues (λ) and eigenvector (v), the values for (v) are sorted based on the (λ) by descending order. The eigenvector (v) becomes the new feature column regarded as the principal component. Finally, the original data is transformed into the new principal component feature space:

$$X' = X \cdot v \quad (3.56)$$

where (X) is the original feature space data, (v) is the principal component vector. The transformed data (X') dimension equals the number of principal component vector dimensions.

For instance, if the user specifies a three-dimensional feature space, after calculating eigenvalues (λ), it is sorted in descending order, and the respective first three (v) are selected as the principal component vectors, transforming the data to a three-dimensional feature space.

3.1.10 Data Processing

The NIST IL Thermo database, which maintains a collection of experimental data for ILs [73, 74], provided the majority of the ionic conductivity data used in the study to build machine learning models. The data is downloaded using a web scraping script implemented in the python package [142]. The script for processing the raw data and formatting is written using bash scripting with the python programming language. Chemical features for the SMILE structure were generated using an open-source cheminformatics package called RDKit, which generates 196 chemical descriptors [143].

The input chemical features along with the output data were normalized and scaled before any model development to avoid any kind of bias for a given feature during model development. This was done using Min-Max Scaling [144] that scales the features between $[0, 1]$.

$$\hat{X} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.57)$$

3.1.11 Feature Reduction

Generating a large number of features for a given molecule can be quickly done using available cheminformatics packages such as RDKit [143]. However, not all the features might be necessary for model development. Having many features does come at a cost, as it requires more computational time to train a large multidimensional model and can also cause the model to overfit as the model puts more effort into learning some of the noise within the data. Arbitrarily removing

features or descriptors without knowing how it affects the model's ability can lead to poor model performance. One of the easiest ways to reduce features is to employ a feature correlation method. If feature A is highly correlated with feature B, then one can drop either of the features as the other feature is sufficient to represent the information.

Another popular technique to eliminate unnecessary features is through the use of the least absolute shrinkage and selection operator (LASSO) method [145]. It is a regularization technique often used to reduce overfitting by discarding fewer important features. It does this by adding a penalty parameter, λ , to the minimization function that determines the number of features to shrink. Larger values of λ shrink some of the least essential feature coefficients to zero, reducing the feature space, while $\lambda = 0$ leaves the objective function unchanged.

$$Obj = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p w_j \quad (3.58)$$

where (w_j) is the coefficient or weight of the feature (j). The penalty feature λ is determined using 5-fold cross-validation (CV) by fitting a linear regression model.

3.1.12 Hyperparameter Tuning

The best-performing hyper-parameters for each ML model discussed above can be found during the hyper-parameter search. This can be done through either a grid search combination that combines all the possible hyper-parameters listed to find the best optimum parameter or a random search by randomly selecting hyper-parameters to find the best-performing ones. The former method is more robust because it searches through all possible combinations; however, it has a high computational cost. One can choose the search method depending on the level of accuracy desired and the available computational resources. During the hyper-parameter search, it is often necessary to perform five-fold cross-validation

to assess the model performance by searching for the best parameters. In this work, the entire hyperparameter search and tuning was done using algorithms implemented in Scikit-Learn [144].

3.2 Molecular Dynamics

3.2.1 Molecular Dynamics Simulation

Molecular dynamics (MD) is a simulation technique that relies on classical newton mechanics to propel atoms through time and space. Initially, the position (\mathbf{r}) and the velocity (\mathbf{v}) are defined for the atom (i) with a mass (m_i). To propagate the particle (i), we need to know its acceleration (\mathbf{a}_i), which can be calculated from the force (\mathbf{F}_i) acting on the particle in all directions. The force can be calculated from the potential energy (U_{Total}), which is solved analytically using the following equation:

$$U_{Total} = U_{Non-Bonded} + U_{Stretching} + U_{Bending} + U_{Torsion} \quad (3.59)$$

$$U_{NonBonded} = \sum_i \sum_j 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_i \sum_j \frac{q_i q_j e^2}{4\pi\epsilon_o r_{ij}} \quad (3.60)$$

$$U_{Stretching} = \sum_{Bonds} \frac{K_r}{2} (r - r_{eq})^2 \quad (3.61)$$

$$U_{Bending} = \sum_{Angles} \frac{K_\theta}{2} (\theta - \theta_{eq})^2 \quad (3.62)$$

$$U_{Dihedral} = \sum_{Dihedral} \frac{V_1}{2} [1 + \cos(\phi)] + \frac{V_2}{2} [1 - \cos(2\phi)] + \frac{V_3}{2} [1 + \cos(3\phi)] \quad (3.63)$$

The total potential energy (U_{Total}) takes into account bond stretching, bond bending, dihedral angle for the bonded interactions, electrostatics, and van der Waal's interaction for non-bonded interactions. The van der Waal's interaction is approximated using Lennard Jones (LJ) parameters σ_{ij} , ϵ_{ij} , r_{ij} which represents the

distance at which the energy is zero, the well depth, and the distance between atoms i and j . q is the partial atomic charge for electrostatic interactions. r , r_{eq} , θ , θ_{eq} are bond length, equilibrium bond length, bending angle, and equilibrium bending angle. K_r , K_θ are force constants for stretching and bending. V_i denotes the torsional parameters for dihedrals [146,147].

The equilibrium parameters, the charge of atom species, and LJ interactions are known as forcefield (FF) parameters that are often derived from experiments or quantum mechanical calculations by fitting to properties such as density or enthalpy of vaporization. There are several kinds of FF parameters one can choose from, such as united atom FF, which lumps a group of atoms into a single unit, making it computationally less expensive because of the fewer atoms that need to be simulated. The other popular one is all-atom FFs, which treats each atom as an interaction site, making it more computationally expensive but often required to calculate properties such as x-ray structure factor.

The force acting on each particle can be then calculated as follows:

$$F_i = -\frac{dU_i}{dr_i} \quad (3.64)$$

Once the force is known, one can use the newton law of motion to calculate the position and velocity of the particle as follows:

$$\frac{F_i}{m_i} = \frac{d^2 r_i}{dt^2} \quad (3.65)$$

$$v_i = \frac{dr_i}{dt} \quad (3.66)$$

$$a_i = \frac{dv_i}{dt} \quad (3.67)$$

As for calculating the position for the particle at time $t + 1$ time step given their current position and previous position. This can be calculated using the Verlet algorithm, which is one of the most widely used integration techniques [146, 147]. This can be derived using the central finite theorem for the second derivative.

$$y'' = \frac{y_{i+1} - 2y_i + y_{i-1}}{\Delta x^2} \quad (3.68)$$

$$x_{t+1} = 2x_t - x_{t-1} + \Delta t^2 \frac{F_i}{m_i} \quad (3.69)$$

To summarize the entire MD protocol:

1. First assign initial position and velocity
2. Compute the potential energy U
3. Calculate the force acting on each of the particles
4. Update the velocity and position of the particle using newton's law of motion
5. Repeat step 2 until an equilibrium condition is met
6. Property calculation

3.2.2 Self Diffusion Constant

Once the system reaches equilibrium, several essential properties can be calculated using the equilibrated trajectory. One of them is the mean square displacement (MSD) which measures the deviation of the particle from the initial position as a function of time. The slope of MSD can be used to calculate the diffusion constant of the molecules as follows:

$$D = \frac{1}{6} \lim_{t \rightarrow \infty} \frac{d}{dt} \left\langle \sum_{i=1}^N [\vec{r}_i(t) - \vec{r}_i(0)]^2 \right\rangle \quad (3.70)$$

where, $\vec{r}_i(t)$ is the position of an ion at any given time, $\langle \dots \rangle$ is the average of the ensemble, and D is the self-diffusion constant. To determine the linear region for the fit, the MSD values are divided into blocks of time frame for the final *NPT* production run. Over these blocks, the non-Gaussian parameter $\beta(t)$ is calculated to find the diffusive regime:

$$\beta(t) = \frac{d \ln \langle \Delta r^2(t) \rangle}{d \ln(t)}. \quad (3.71)$$

A linear regime is indicated as $\beta(t)$ approaches a value of unity which is the appropriate block for calculating diffusion coefficient.

3.2.3 Ionic Conductivity

Similarly, taking into account the charge of ions, one can calculate the ionic conductivity of the system. The ionic conductivity of ILs can be calculated using two different equations. The first method is called the Nernst-Einstein (NE) equation, shown in the equation 3.72, which uses the self-diffusion constant (D) motion of the ions calculated using the equation 3.71. N stands for the number of ion pairs in the system, V stands for the system volume, T is the temperature of the system, k_b

stands for Boltzmann constant, q_+ , q_- stands for the charge on cation and anion, $\vec{r}_i(t)$ is the position of an ion at any given time and $\langle \dots \rangle$ is the average of the ensemble. The assumption using NE is that the movement of ions is independent of the presence of other ions, meaning the NE equation does not take into account the correlated movement of ions.

$$\sigma_{NE} = \frac{N}{Vk_bT}(q_+^2 D_+ + q_-^2 D_-) \quad (3.72)$$

The second method employs the Einstein method (equation 3.73) to calculate ionic conductivity (σ) that includes an additional term to take into account the correlated movement of ions on each other. These correlated ion movements do not participate in ionic conductivity, leading to lower ionic conductivity calculated using the Einstein equation compared to the NE equation.

$$\sigma = \frac{1}{6k_bVT} \lim_{t \rightarrow \infty} \frac{d}{dt} \left\langle \sum_{i=1}^N \sum_{j=1}^N q_i q_j [\vec{r}_i(t) - \vec{r}_i(0)] \cdot [\vec{r}_j(t) - \vec{r}_j(0)] \right\rangle \quad (3.73)$$

The difference in the ratio ($\frac{\sigma}{\sigma_{NE}}$) is often known as ionicity, signifies the degree of uncorrelated motion, where a ratio of 1 indicates highly uncorrelated movements for systems often seen with high ionic conductivity, while a lower ratio close to 0 indicates highly correlated motion between the ion pairs.

3.2.4 Radial Distribution Function

MD simulation also offers the possibility of probing the structural level arrangement of molecules with respect to each other as a distance function. One of the most calculated quantities is the radial distribution function (RDF), which quantifies the probability of finding a given molecule at a fixed distance to the reference molecule. This is because the molecular arrangement might be different for local

density at a closer distance than the bulk density. Thus, RDF calculates the ratio of local density to bulk density.

$$g(r) = \frac{\rho}{\rho_{bulk}} = \frac{V}{N^2} \left\langle \sum_i \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle \quad (3.74)$$

when the local density is similar to the bulk density, the $g(r)$ takes a value of 1.

3.2.5 Coordination Number

The RDF calculation can be further used to calculate the number of particles surrounding the reference particle by computing the coordination number. This is done by taking the integration of $g(r)$ from a distance of zero to the first solvation shell (r).

$$N = 4\pi\rho_b \int_0^r g(r)r^2 dr \quad (3.75)$$

3.2.6 Hydrogen Bonding

Besides calculating thermophysical and structural properties, one can also quantify the interactions, such as hydrogen bond dynamics, from the trajectory of the equilibrated system. In our work, we used the TRAVIS visualization package [148] to calculate the hydrogen bond dynamics.

For hydrogen bonding interactions, it is often interesting to examine the length of time the atoms stay hydrogen bonded together. This can provide insights into whether there is any hydrogen bonding preferential interaction between particular ion pairs. In the case of ILs, the hydrogen bonding lifetime was calculated between cations and anions.

This is calculated using the intermittent hydrogen bonding autocorrelation shown as follows :

$$C(t) = \frac{\langle h(0)h(t) \rangle}{\langle h(0) \rangle} \quad (3.76)$$

where $h(t)$ takes a value of 1 if the hydrogen bond is persistent but assumes a value of 0 when the hydrogen bond is broken. The cutoff angle between donor-acceptor-hydrogen was set to 30° , the cutoff distance between the acceptor-donor was selected to the corresponding first solvation shell of the center of mass RDF between the cation-anion, and the cutoff distance for the acceptor-hydrogen atom was limited to the first solvation shell of the hydrogen bonding sites.

The hydrogen bonding lifetime (τ_{lt}) can be calculated from the autocorrelation function as follows:

$$\tau_{lt} = \int_0^\infty C(t)dt \quad (3.77)$$

The integral is evaluated analytically by fitting a stretched exponential equation:

$$C(t) = a_1e^{-t/b_1} + a_2e^{-t/b_2} + a_3e^{-t/b_3} \quad (3.78)$$

where, $a_3 = 1 - a_1 - a_2$ yielding

$$\tau_{lt} = a_1b_1 + a_2b_2 + a_3b_3 \quad (3.79)$$

3.3 Electrochemical Stability

The electrochemical stability is an essential property for an electrolyte as it characterizes the electrochemical window (ECW) at which it is neither susceptible to oxidation nor reduction. The ECW calculations using computational techniques are performed using: (a) Thermodynamic cycle (TDC) approach, (b) One electron approach, and (c) Highest occupied molecular orbital (HOMO)/ Lowest occupied molecular orbital (LUMO) approach. The TDC method involves the calculation of the Gibbs Free Energy of reduction and oxidation potential for the cation and anion. This method consists of a series of steps as one has to calculate the Gibbs free energy of (G_{cation} and $G_{cation+e^-}$) and (G_{anion} and $G_{anion-e^-}$) in the gas phase and solution phase. As such, the TDC approach is computationally expensive, but at the same time, the ECW values are found to be in excellent agreement with experimental measurements [42, 104]. In the TDC calculations for ILs, the cathodic limit is equivalent to the reduction of a cation and the anodic limit to the oxidation of an anion. Thus, the ECW is calculated by taking the difference in ($E_{ox} - E_{red}$).

The one electron approach is similar to TDC with the energy calculated for the reduction of cation and the oxidation of anions. As such, the cathodic limit and anodic limit is given by:

$$\Delta E(cathodic) = C - C^+ \quad (3.80)$$

$$\Delta E(anodic) = A - A^- \quad (3.81)$$

The vertical transition energy (E) here refers to sum of total energy and zero potential energy calculated from DFT calculations. The ECW can be calculated as follows with respect to Li/Li+ electrode:

$$\text{Cathodic potential} = -\frac{\Delta E_{\text{cathodic}}}{F} - 1.46 \quad (3.82)$$

$$\text{Anodic potential} = \frac{\Delta E_{\text{anodic}}}{F} - 1.46 \quad (3.83)$$

where F is the faraday's constant, the constant (-1.46) converts the potential to Li/Li+ reference electrode [149–151]. The one electron approach has been found to perform extremely well in the presence of a solvent rather than vacuum to calculate electrochemical window [128].

Lastly, the HOMO/LUMO approach is based on the Koopman's theorem [152] that approximates the negative of the HOMO energy with ionization energy of a molecule meaning the energy required to remove an electron, while the LUMO energy with the electron affinity of a molecule. Thus the HOMO energy is equal to the anodic limit, and the LUMO energy is equal to the cathodic limit. In contrast to TDC and one electron approach, the HOMO/LUMO approach is pretty straightforward as one has to obtain the HOMO/LUMO energy readily provided by DFT calculations.

$$V_{CL} = \frac{-\epsilon_{LUMO}}{e} \quad (3.84)$$

$$V_{AL} = \frac{-\epsilon_{HOMO}}{e} \quad (3.85)$$

$$ECW = V_{CL} - V_{AL} \quad (3.86)$$

where e denotes the charge of an electron.

Panidian et al. [128] did an extensive study on calculating the ECW of ILs using various methodologies. According to the authors, the Δ SCF Self Consistent Field

approach to calculating ECW in a vacuum had a deviation of 48–54.1% from experimental data, depending on the basis set. Changing the vacuum environment to an SMD solvent (dichloroethane) environment and using the one-electron addition approach reduced the deviation to 2.4% using the M06-L/6-311 + G(2d,p) basis set.

Building upon the work of Panidian and co-workers [128] we used a similar SMD approach with dichloroethane as the solvent, (M06-L) functional form, and (6-311 + G (2d,p)) as the basis set. The calculations are performed using Gaussian 09 package [153].

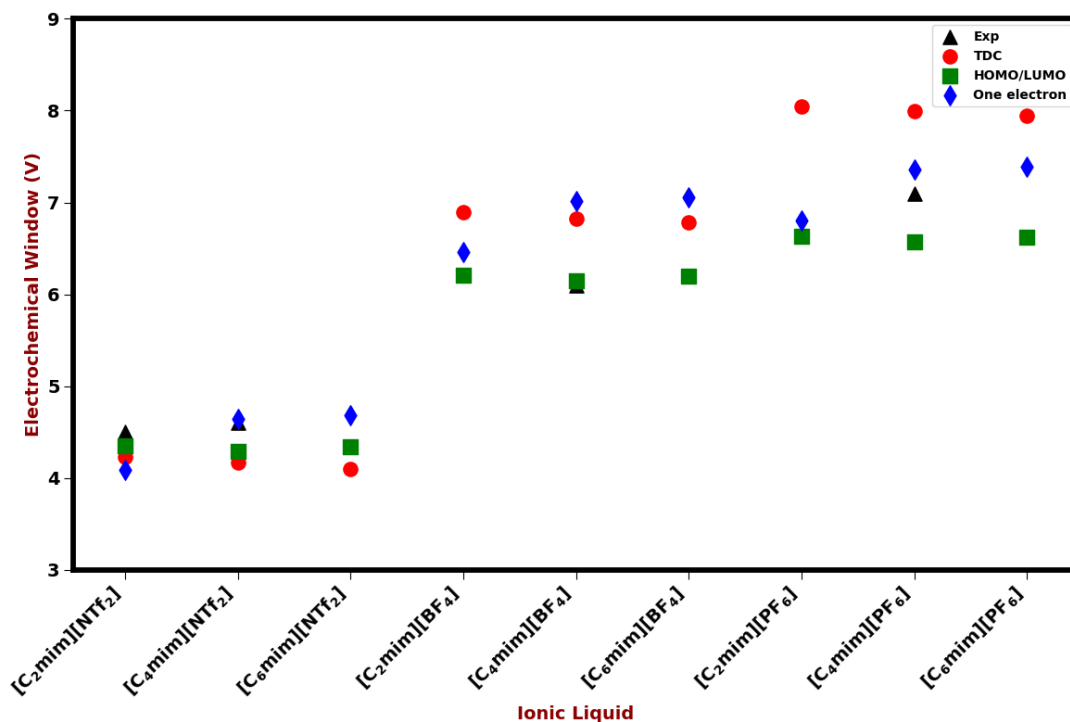


Figure 3.6: Comparison of ECW between TDC [42], HOMO/LUMO approach (this work), one electron approach (this work) and experimental data for various imidazolium ILs.

We compared the one-electron addition approach to the HOMO/LUMO approach for ECW compared to the TDC approach and experimental data. The data for TDC and experimental data were taken from Kazemiabnavi et al. [42] paper for im-

idazolium ILs. The comparison is shown in Figure 3.6 for all the three approaches and compared to experimental data where available. The HOMO/LUMO approach and one electron approach agree with the TDC method with RMSE of 0.88 V and 0.58 V, respectively.

Based on the RMSE error metrics, it is evident that the one-electron approach is much more accurate than the HOMO/LUMO approach. Besides accuracy, another limitation of the HOMO/LUMO approach is the lack of a way to determine the electrochemical stability with reference to an electrode, as that is very important for gauging the electrolyte stability towards the electrode [154]. This is taken into account in the one-electron approach with the addition of the constant term. However, as mentioned earlier, the one-electron approach requires twice as much calculation as the HOMO/LUMO approach, making it computationally expensive for many systems. Thus, to reduce the computational cost, we examined the possibility of finding a correlation between the HOMO/LUMO approach and one electron approach, as the latter has higher accuracy in predicting ECW. In contrast, the former approach involves twice as few calculations. Thus, correlating these two approaches could allow calculating the ECW based on one-electron approach using the HOMO/LUMO method with a minimal computational cost.

The correlation was done by calculating ECW using the HOMO/LUMO approach and one electron approach for twenty-one ILs that were found to have ionic conductivity greater than 2.0 S/m using the generalized machine learning model discussed further in chapter six. The ECW calculated using both approaches is shown in Figure 3.7 with a very high correlation coefficient.

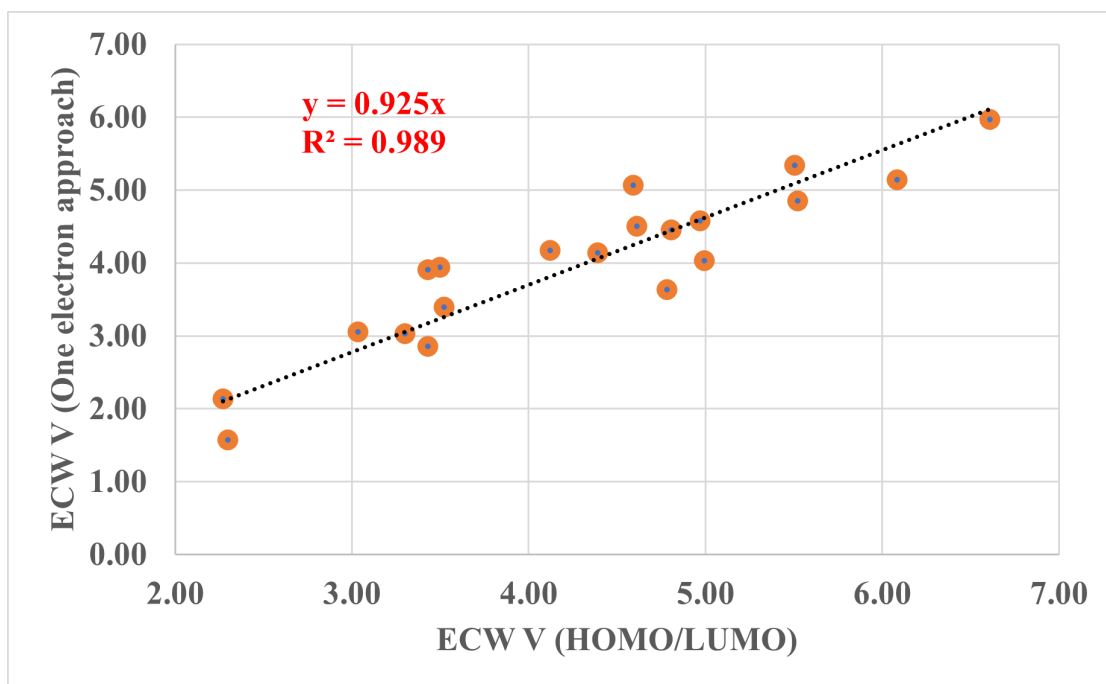


Figure 3.7: Correlation of ECW between HOMO/LUMO and one electron (Li/Li+) electrode approach.

CHAPTER 4

Developing Machine Learning Models for Ionic Conductivity of Imidazolium-Based Ionic Liquids

4.1 Abstract

In this study, we developed two machine learning models, support vector machine (SVM) and artificial neural network (ANN), to correlate ionic conductivity of pure ILs based on the imidazolium cations using the data acquired from the NIST ILThermo database. Both models were shown to successfully capture the entire range of ionic conductivity spanning six orders of magnitude over a temperature range of 275-475 K with relatively low statistical uncertainty. Due to slightly better performance, ANN was used to predict the ionic conductivity for 1102 ILs formed from every possible combination of 29 cations and 38 anions contained in the database. The procedure led to the generation of many ILs for which the ionic conductivity was estimated to be greater than 1 S/m. The ionic liquid dimethylimidazolium dicyanamide, not present in the original dataset, was identified to exhibit the ionic conductivity of 3.70 S/m, roughly 30% higher than the highest conductivity reported for any ionic liquid at 298 K in the database. The ANN model was also found to accurately predict the ionic conductivity for several ionic liquid-ionic liquid mixtures, for which experimental data are available. Encouraged by this result, we calculated ionic conductivity for all the possible binary ionic liquid-ionic liquid mixtures based on the cations and anions in the dataset. The model predictions revealed a large number of ionic liquid mixtures systems

exhibiting nonideal behavior where a maximum or minimum in the ionic conductivity was observed as a function of composition, similar to trends seen in binary ionic liquid mixture of water or conventional solvents with ILs.

4.2 Introduction

Room temperature ILs are a class of salts that are liquid at room temperature consisting exclusively of ions. They are currently one of the most studied solvents because of several unique properties such as negligible volatility, electrochemical stability, low melting point, and high thermal and chemical stability [155]. Because of all of these desirable properties, ILs are investigated for various industrial applications such as potential solvents to break minimum/maximum boiling azeotropes [156–159], extracting agent in LLE separations [160–163], electrolytes in electrochemical devices [11, 164–166], and solvent for gas-capture [167–170]. Despite the various favorable attributes inherent in an ionic liquid, high viscosity and low ionic conductivity of many ILs, especially at low temperatures, is a bottleneck for the application of ILs as electrolytes in batteries [171].

A widely adopted approach to mitigate potential drawbacks for using ILs is to tune the properties of an ionic liquid by altering functional group(s) attached to the cation, changing the cationic core (e.g. from aromatic to cyclic), and/or modifying the chemical composition of the anion. Developing new ILs this way requires considerable chemical intuition, expertise in synthesis, and subsequent measurements of properties. Given the breadth of the chemical space for cations and anions, it is practically impossible to study every possible combination of the cation and anion. The explosion in the chemical space is further exacerbated by the increasing popularity of exploiting ionic liquid-ionic liquid mixtures for tailoring properties of these solvents [51, 172, 173]. One estimate projects that there are as many as one billion ionic liquid systems [174]. Although daunting from an experimental or molecular simulation viewpoint, the vast chemical space of cations and anions also offers a unique opportunity to leverage machine learning and data

analytics-based techniques to search and design ILs with properties suited for a given application.

Indeed, several studies over the years have used artificial neural network (ANN) to model and predict ionic liquid properties such as density, [175] viscosity [176,177], melting point [178], toxicity [179], solubility of gases, such as CO₂ [180,181] and SO₂ [182] in ILs, surface tension [183], investigating ionic liquid-solvent mixtures, [184–187], and prediction of rate constants in ionic liquid-organic mixtures [188]. Additional examples involving the application of ANN for various properties for ILs can be found in a recent review article by Yusuf et al. [189] Recently, Beckner and Pfaendtner have demonstrated that it is possible to combine machine learning and genetic algorithm to develop new ILs with high thermal conductivity [190]. Some advances have also occurred for correlating ionic conductivity, an extremely useful property for selecting electrolytes in electrochemical applications and the topic of the present chapter. Krossing et al. used the concept of free volume and derived an empirical equation based on Cohen-Turnbull free volume theory to correlate transport properties such as ionic conductivity and viscosity for imidazolium based ILs that were in good agreement with experimental data at high temperature range, while some deviations were noted in the low temperature regime [191]. Passerini et al. found that the molar conductivity of pyrrolidinium based cations paired with sulfonylimide anions showed a high correlation ($R^2 = 0.9942$) with the sum of cation and anion volumes obtained from electronic structure calculations [110]. The observation suggested that the molar ionic conductivity decreased with an increase in the combined volume. However, no such monotonicity existed for imidazolium-based ILs, which is the focus of the present study. Beichel et al. used volume-based thermodynamics (VBT) approach to correlate ionic conductivity of ILs based on parameters such as molecular volume and surface area calculated using COnductor-like Screening MOdel (COSMO) [192]. The authors reported an overall root mean square error of 0.04-0.06 $\log(\sigma)$. Group contribution (GC) methods have also been found useful for developing a correlation between the ionic conductivity and various chemical features of ILs. For example,

Gharagheizi et al. employed a least-squared support vector machine GC method to estimate ionic conductivity consisting of a dataset with 54 different unique ILs with an absolute average relative deviation (AARD) of 3.3% [114]. Tochigi et al. developed a polynomial-based quantitative structure-property relationship (QSPR) to predict ionic conductivity for eight different cation families and sixteen different anions [109]. The authors reported an overall R^2 of 0.91 and standard deviation of 0.12 S/m for 139 data points. Coutinho et al. used a three-parameter GC method equation similar to Vogel-Tammann-Fulcher (VTF) for the estimation of ionic conductivity for pure ILs [193]. Wooley et al. applied a four-parameter GC-based approach to estimate ionic conductivity of ILs [113]. An attractive feature of GC methods is that chemically intuitive groups are usually selected as inputs to the model prior to optimizing model parameters. However, for billions of ILs with vastly different chemical functionalities, identifying and enumerating all the relevant groups can pose significant difficulties to eventual automated screening of ILs.

In this chapter, we explore a different approach rooted in the framework of machine learning techniques such as artificial neural network and support vector machine to correlate the ionic conductivity of pure ILs. We assess the performance of the two models and examine if the model can be extended to predict ionic conductivity of all possible combinations of unique cations and anions in the database and binary ionic liquid systems. As such the next section provides details on the data collection and processing, model formulation, and model validation. In the subsequent section, the models, trained with the ionic conductivity of pure ILs, are compared. The model with better accuracy is identified and is extended to predict the ionic conductivity for *in silico* ILs obtained by enumerating possible combinations of cations and anions contained in the dataset. We will demonstrate that such a procedure leads to the discovery of the ionic liquid with the highest conductivity, which matches with the experimental data at 298 K. The predictive capability of the model will be discussed in terms of the level of agreement for ionic conductivity for several binary ionic liquid systems. The possibility of obtaining enhancement in the ionic conductivity by formulating binary ionic liquid mixtures

will be presented followed by a summary of findings and the direction for future research.

4.3 Methodology

4.3.1 Data Collection and Processing

A total of 2895 ionic conductivity data for pure component imidazolium-based ILs were downloaded from the online ILThermo database maintained by NIST [73, 74] using the pyILT2 [142] utility. Majority of the imidazolium-based experimental data in the NIST ILThermo Database were measured using alternating current cell with electrodes [194, 195] while the rest were acquired using direct current cell with electrodes [17, 196], capillary cell, electrochemical (EC) cell [197], impedancemetry [198, 199] and conductivimeter [200]. The downloaded data were processed (see below) and formatted with an in-house Python script. The datapoints contained the ionic liquid name, temperature (K), pressure (kPa), reference from which the data was extracted, and the uncertainty in the measurement. Approximately 89% of the data represented ionic conductivity in the liquid state, while $\sim 10\%$ of the data for crystals, and a small fraction of the data with ionic conductivities for metastable liquids were discarded from the training set.

The next step involved a careful examination of the dataset. First, we eliminated any entries with missing values for the ionic conductivity or "NaN" in the dataset. To accomplish the removal of inconsistent data or typographical errors, we graphed ionic conductivity data as a function of temperature to identify outliers in the dataset. Some of the ionic conductivities were extremely low, in the range of 10^{-9} S/m belonging to ILs comprised of natural amino acids as the anions combined with 1-ethyl-3-methylimidazolium $[\text{C}_2\text{mim}]^+$ cation at 298.0 K [48]. We eliminated these points due to very low values of ionic conductivity and the fact that the model derived from an artificial neural network (ANN, see below) could not be extended to such small values. We also found that ionic conductivities for the pure 1-n-hexyl-3-methylimidazolium $[\text{C}_6\text{mim}]$ bromide Br and 1-n-octyl-

3-methylimidazolium [C₈mim]Br were reported to be 144.1 S/m and 116.4 S/m, respectively at 333.15 K [201]. These values are two orders of magnitude larger than those for many imidazolium-based ILs. For example, ILs with shorter alkyl chain length such as [C₂mim]Br and 1-n-butyl-3-methylimidazolium [C₄mim]Br have been reported to possess ionic conductivities of 1.06 S/m at 335.6 K [195] and 0.734 S/m at 373.1 K [13]. The visualization of the ionic conductivity as a function of the alkyl chain length also showed that the ionic conductivity decreases with the increase in the alkyl chain. Thus, the inconsistency led us to remove the seemingly high ionic conductivity datapoints. We pruned the dataset further by identifying duplicate ionic liquid fields (same cation, anion, temperature, and pressure) and keeping only the entry with the lowest uncertainty in the ionic conductivity measurements.

We further reduced the number of points for model development by visualizing the data to obtain a clue into the appropriate ranges for the ionic conductivity, temperature and pressure along with chemical identities of the ionic liquid in the database. We observed that a large fraction of the measurements have been conducted in the temperature range spanning 275-475 K (Figure A.1). Thus, we removed all the data points outside this temperature range. As there were only a limited number of points present at pressures other than 101 kPa, we decided to restrict the model development by fixing the pressure at 101 kPa. The resulting dataset contained a total of 1323 data points with ionic conductivities over six orders of magnitude from 4.1×10^{-5} S/m to 19.3 S/m as seen in supporting information. To assess the variability in the chemical identities of the cations and anions represented in the data set, we generated Figure 4.1 for every ionic liquid for which more than five data points were present; the size of the marker in the figure is proportional to the number of points reported for each of the ILs. It is clear that a large fraction of the ionic conductivity measurements cover the cations [C₂mim]⁺ and [C₄mim]⁺ paired with a broad variety of anions, while the remaining cations, on an average, are combined with two to three distinct anions. Overall, we found that the dataset contained 29 unique cations and 38 unique anions. There

were a total of 111 ILs, approximately 10% of the ILs that could be formed by combining cations and anions from the dataset.

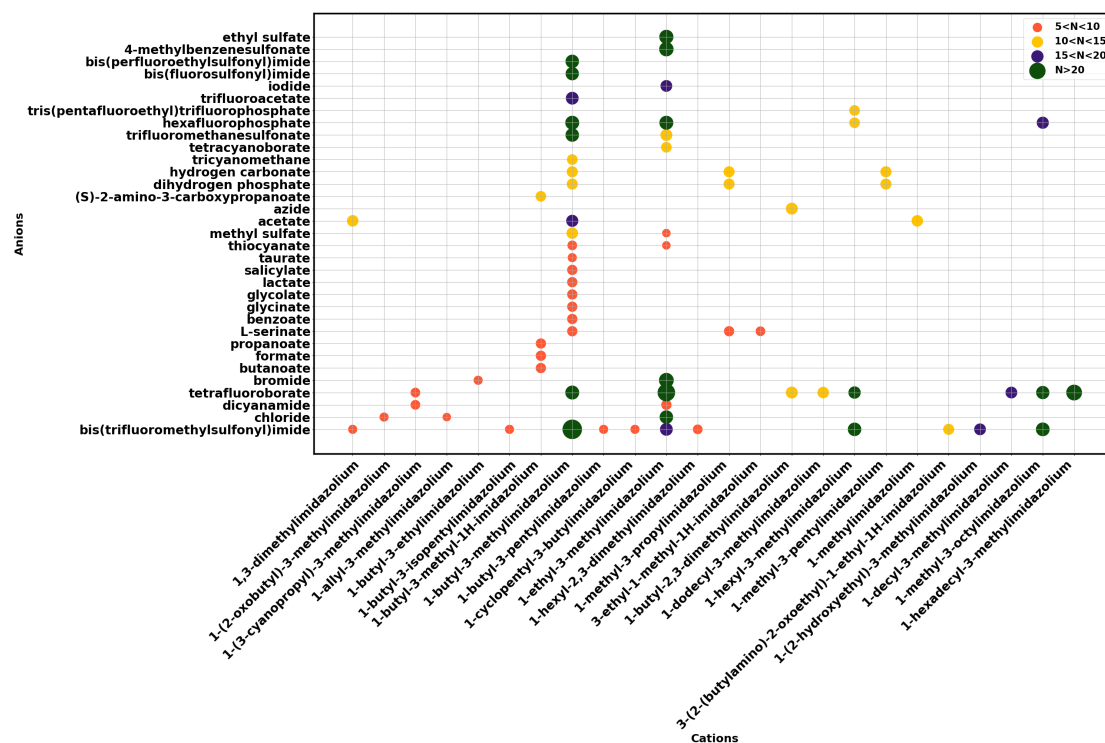


Figure 4.1: The number of ILs for a given cation-anion combination with more than five data points in the NIST ILThermo database is shown.

4.3.2 Feature Generation and Elimination

We translated the identities of the cations and anions to simplified molecular-input line-entry system or SMILES format using an open-source online website Open Parser for Systematic IUPAC Nomenclature OPSIN [10, 202]. One of the anions in the dataset [tetrakis(isothiocyano)cobaltate(2⁻)] could not be converted to SMILES format, therefore we removed the anion and corresponding ILs from further consideration. We used an open-source cheminformatics package RDKit [143] to generate descriptor features for the input to the models. RDKit produced a total of 196 descriptors for each of the cations and anions. A complete listing is available in the supporting information (Section 1.1). Prior to utilizing these features in the model development, we examined the correlation among features to reduce the dimensionality of the input and increase the speed of learning algorithms by

calculating cross correlation coefficients for every feature with every other feature. Comparing the correlation coefficients sequentially, we eliminated any feature that showed either a positive or negative correlation coefficient of greater than 0.9 with any of the previous features. This process brought the aggregate number of cation and anion features down to 38 and 59, respectively, for a total of 99 chemical features including temperature and pressure for a given ionic liquid. The final set of features used below for the model development is included in the supporting information (Sections 1.2 and 1.3).

4.3.3 Data Normalization

Data normalization is a standard technique in improving the model performance and minimizing biases in a multivariate regression with feature values varying over a wide range. For instance, the RDKit feature ‘hydrogen count’ would possess a considerably smaller range of values for the cations and anions in comparison to those for the ‘molecular weight’ feature, which will likely influence the corresponding weightage. On the output side, the ionic conductivity data varied over six orders of magnitude as pointed out earlier. Therefore, we decided to use Min-MaxScaler implemented in Scikit-learn [144] to normalize each input feature and the output by the difference in the maximum and minimum values, which led to any feature or output value to fall between 0 and 1. We preserved the scaling employed during the model generation for later use in the prediction.

4.3.4 Model Development

In this work, we used a total of 1323 experimental data points with a focus on cations exclusively from the imidazolium family to build machine learning model. The training set consisted of 90% of the total data, while the remaining 10% of the data was used as test case to evaluate the model’s performance. The model was constructed using two of the most popular machine learning methods, support vector machine for regression (SVR) and feed forward ANN (FFANN).

Support vector machine (SVM) is a supervised machine learning framework used for classification and regression problems [203–205]. The regression version of SVM is called SVR with the central objective of finding the best fit line in the hyperplane; the equation for the regressed line is derived by finding the maximum number of points located within a given tolerance. Hyper-parameter tuning of SVR parameters is extremely important to improve the model’s accuracy for regression analysis. Similarly, FFANN is also a supervised learning technique with a mapping function $y = f(\mathbf{x}; \theta)$ where θ is the parameter set that the model learns to provide the most optimal approximation of the function based on the input feature vector \mathbf{x} . The FFANN consists of three layers: an input layer, hidden layer and the output layer. The input layer consisted of chemical features along with the state points temperature T and pressure P.

Hyper parameters for both the models were tuned using GridSearchCV implemented in Scikit-learn [144]. GridSearchCV exhaustively searches all the hyper-parameter combination listed in the parameter search space to identify the best performing hyper-parameters. The search space for both the models along with the final hyper-parameters are provided in the supporting information (Sections 2.4 and 2.5). The GridSearchCV method is combined with 10 K-Fold cross validation to minimize any overfitting during the hyper-parameter search. The best performing model architecture with the highest accuracy during this hyper-parameter tuning process was selected as the final model with a further evaluation conducted on the test case set aside earlier. The workflow for cross-validation and testing of the model is depicted in Figure 4.2.

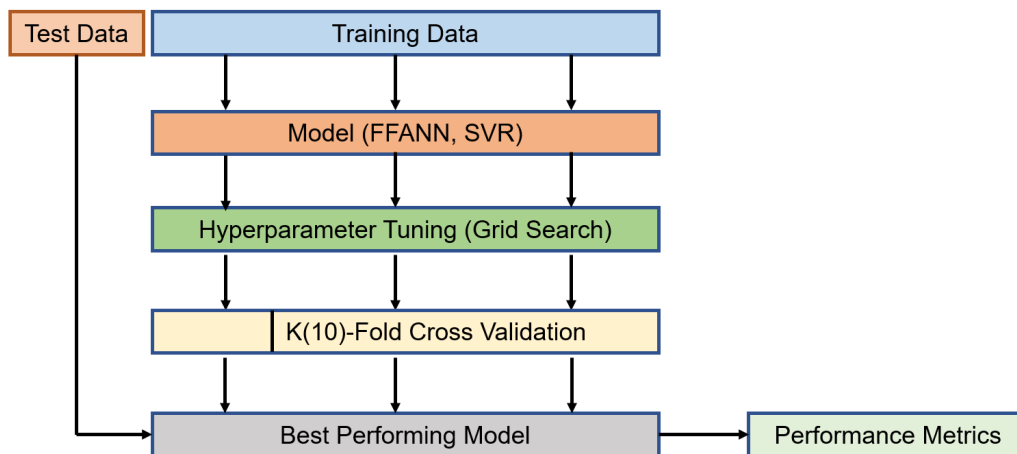


Figure 4.2: Description of model development followed in this study.

4.4 Results and Discussion

4.4.1 Model Validity

In this work, we developed machine learning models to predict ionic conductivity of imidazolium-based ILs using two different techniques. The best-performing model produced the lowest statistical uncertainty and captured the trends in the data such as the lowest ten and the top ten ionic conductivity measurements. During the model development, we observed that the models based on MinMax scaling as discussed earlier performed extremely well for predicting conductivity values in the higher magnitude range, but the predictive capability greatly diminished in the lower conductivity region. For instance, the experimental value of [C₆mim] tetrafluoroborate [BF₄] is 6.7×10^{-4} S/m at 298 K; however, the predicted value was 3.3×10^{-2} S/m - an error of two orders of magnitude. We noted this behavior for many other ionic liquid systems with ionic conductivity values on the lower end. The observation prompted us to convert the ionic conductivity values on a logarithm scale (base 10) before applying the MinMax scaling, which led to a dramatic improvement in the prediction of low ionic conductivity values. For example, the ionic conductivity prediction for the ionic liquid [C₆mim][BF₄] was 7.2×10^{-4} S/m in comparison to experimental measurement of 6.7×10^{-4} S/m.

Table 4.1: Comparison of the predictions results for FFANN and SVR for the training set, test set and the entire dataset. MSE is the mean squared error, MAE is the mean absolute error, RMSD is root mean square deviation and R^2 is the squared correlation between experiment and predicted data. \log_{10} scale refers to ionic conductivity scaled to \log_{10} .

Scale	Metric	Train		Test		Entire	
		SVR	FFANN	SVR	FFANN	SVR	FFANN
\log_{10} scale	R^2	0.995	0.993	0.976	0.991	0.993	0.994
	MSE	0.002	0.003	0.012	0.004	0.003	0.003
	MAE	0.014	0.036	0.038	0.044	0.017	0.037
	RMSD	0.047	0.057	0.111	0.071	0.057	0.059

Table 4.1 details the statistical assessment of SVR and FFANN for the training set, test dataset and the entire set. Both the models not only perform well for the training set, but they also have very high R^2 and low MSE, MAE and RMSD for the test set.

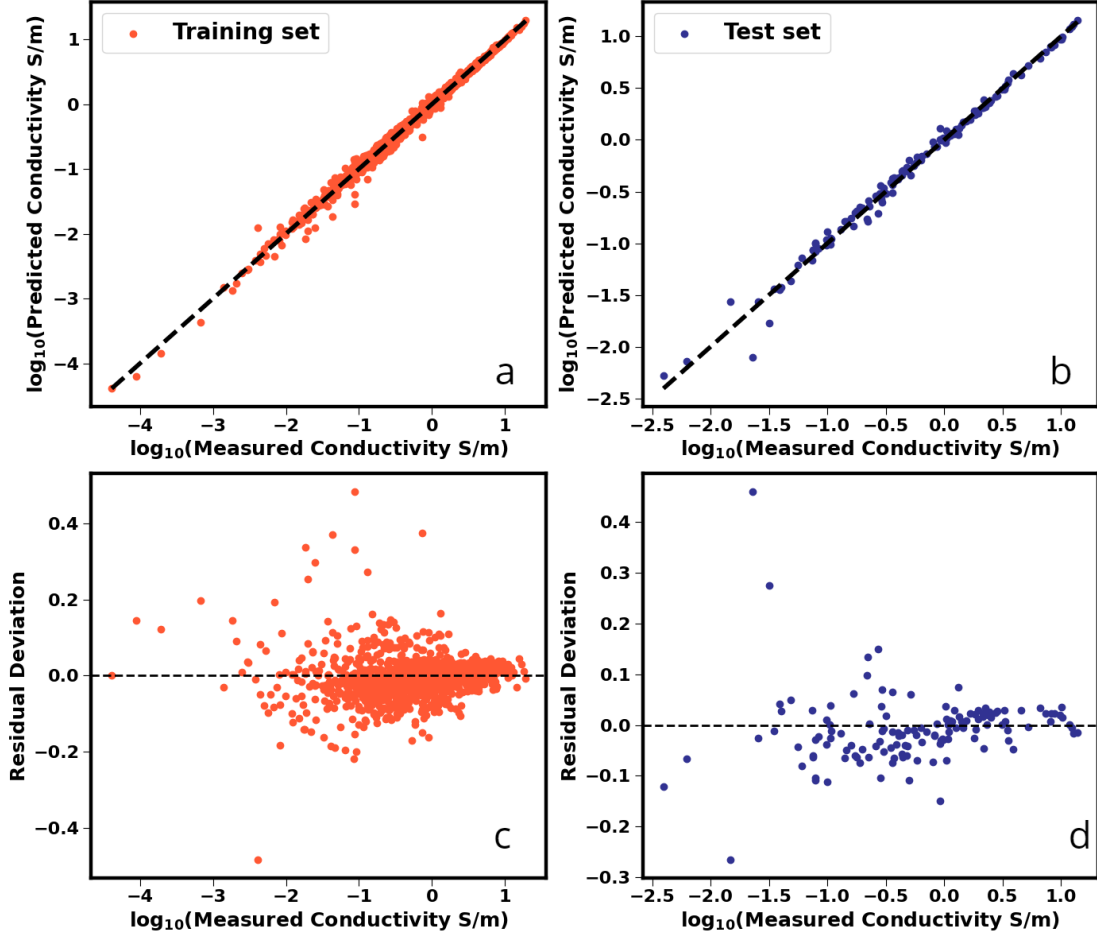


Figure 4.3: (a) Comparison of FFANN model predictions with the experimental data on a \log_{10} scale for the training set. A perfect prediction would fall on the $y = x$ dotted line; (b) comparison for the test set (c) Residual deviation on the log-10 scale calculated as $(\sigma_{\text{experiment}} - \sigma_{\text{prediction}})$ where σ refers to the ionic conductivity for the training set; (d) Residual deviation for the test set.

Figure 4.3(a) demonstrates that the FFANN model is able to capture the training data on the base-10 logarithmic scale spanning six orders of magnitude with a high accuracy in the low conductivity range. Figures 4.3(c) and (d) show the residual deviation calculated by taking the difference in experiment and predicted value vs the experimental data. It is important to note that the residual deviation stays within ± 0.5 log unit for the training set and the test set over the entire range of the experimental data.

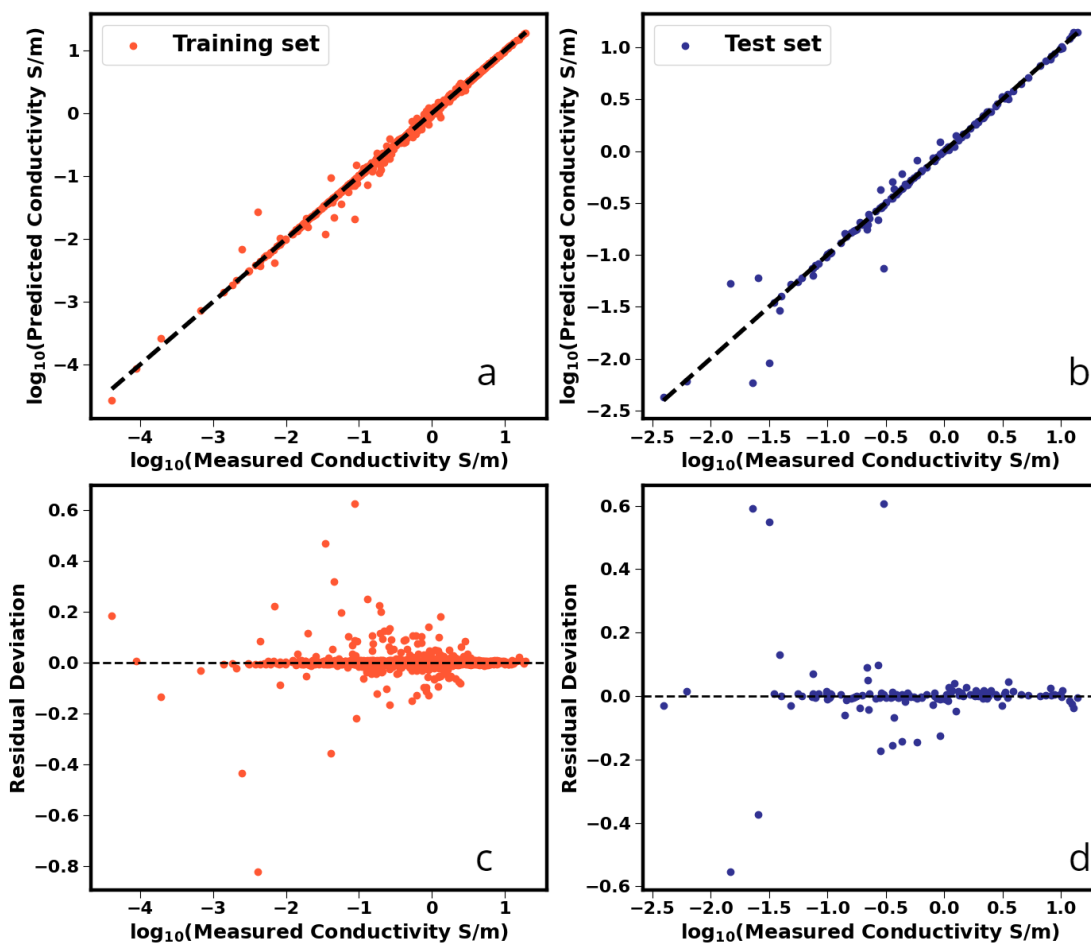


Figure 4.4: (a) Comparison of SVR model predictions with the experimental data on a log-10 scale for the training set. A perfect prediction would fall on the $y = x$ dotted line; (b) comparison for the test set (c) Residual deviation on the log-10 scale calculated as $(\sigma_{\text{experiment}} - \sigma_{\text{prediction}})$ where σ refers to the ionic conductivity for the training set; (d) Residual deviation for the test set.

Similarly, Figure 4.4(a) depicts that the ionic conductivity correlation using the SVR model for the training set and test set. In contrast to FFANN, SVR seems to have more deviation from the $y=x$ line for both sets at low ionic conductivity values. This is also reflected on the residual deviation plot Figure 4.4(c) and (d) where the maximum deviation reaches as high as ± 0.6 log unit for the training set. The normal scale ionic conductivity correlation using FFANN and SVR are provided in the supporting information as seen in Figure A.8 and Figure A.9.

The overall accuracy of both the model outputs is encouraging, especially when considered in the context of the ionic conductivity calculated from molecular simulations. Transport properties such as ionic conductivity are quite challenging to accurately predict from atomistic simulations requiring long simulation times and optimization of force field parameters. The problem is further exacerbated for sluggish ILs possessing extremely low ionic conductivity as probed here. In such scenarios, the simulation results of ionic conductivity can differ by a factor up to 10 (by 1 unit on log 10-scale) from the corresponding experimental observations. An added advantage of the proposed model is to provide guidance, at almost no computational cost, on the ionic conductivity values for *in silico* generated pure ILs and mixtures of binary ILs obtained from possible combinations of cations and anions studied here. However, we admit that the machine learning model cannot provide molecular-level insight that is inherent in molecular simulations. For the discussion below, we focus on FFAAN as the accuracy of the model is slightly better for the entire data set.

We also compared ionic conductivity data predicted using FFANN model against a recently published Group Contribution (GC) method [113] in comparison with experimental data. This was done for a total of 203 data points consisting only the $[C_{n=(2,4,6,8)}\text{mim}][\text{NTf}_2]$ subset series at various temperatures. The statistical uncertainty in predictions with the FFANN model and the GC model are similar compared to experimental data. The experiment data, GC predictions and FFANN predictions can be found in the supporting information along with the statistical uncertainty in predictions.

We also probed the accuracy with which the FFAAN model captured trends. For this, we chose the data at 298 K as there were only a few systems for which the data was available over the entire temperature window. In Figure A.10, we compare the predictions of the FFANN model for the ten lowest ionic conductivity values reported in the NIST ILThermo Database at 298 K. We observe that the model accurately predicts the ordering of the ILs while the ionic conductivities are

also in very good agreement. The plot also reveals that long alkyl chains or amino acid-based anions tend to produce low conductivity ILs. Similarly, Figure A.11 represents a comparison between the FFANN model predictions and experimental measurements for the ten largest conductivity values at 298 K. It is evident that the predictive capability of the model is excellent. It is also important to highlight that not only does the model capture the quantitative trend accurately, but it also performs correctly in terms of taking into account the cation and anion properties and behavior. For example, $[\text{BF}_4]^-$ when paired with a long alkyl chain cation $[\text{C}_{12}\text{mim}]^+$ yields one of the lowest ionic conductivity ILs, while its combination with $[\text{C}_2\text{mim}]^+$ generates an ionic liquid with five orders of magnitude higher ionic conductivity than that for $[\text{C}_{12}\text{mim}][\text{BF}_4]$. We also point out that the change in the identity of the anion can dramatically affect the ionic conductivity as exemplified by 1-allyl-3-methylimidazolium $[\text{AMIm}][\text{Benzoate}]$ and $[\text{AMIm}][\text{Formate}]$, the latter with the ionic conductivity four orders of magnitude higher than that for the former; the model successfully predicts the trend.

4.4.2 Unique Ionic Liquid Combination

Next we generated all the combinations of 29 unique cations and 38 anions present in the dataset, which resulted into 1102 pure ILs at 298 K for which we predicted ionic conductivity at 298 K. We first tested the accuracy for such predictions using FFANN and SVR model based on two test cases that were not part of the training set. The first system is $[\text{C}_2\text{mim}]$ bis(fluorosulfonyl)imide $[\text{FSI}]$. The database contained $[\text{C}_4\text{mim}][\text{FSI}]$ as the only ionic liquid containing $[\text{FSI}]^-$. The model prediction for the ionic conductivity using FFANN was found to be 1.60 S/m for $[\text{C}_2\text{mim}][\text{FSI}]$ at 298.15 K, in excellent agreement with the corresponding experimental measurements of 1.61 ± 0.02 S/m compared to significantly under predicted value of 0.189 S/m using SVR method [206]. The second system is represented by the ionic liquid $[\text{C}_1\text{mim}][\text{DCA}]$, which was predicted to possess the highest ionic conductivity of 3.70 S/m at 298 K using FFANN, which is roughly 30% higher than the highest ionic conductivity of 2.83 S/m for $[\text{C}_2\text{mim}][\text{DCA}]$. We found two experimental papers confirming that the ionic conductivity of $[\text{C}_1\text{mim}][\text{DCA}]$ at

298.0 K is around 3.60 S/m, [40, 207] once again in excellent agreement with the value obtained from our FFAN model. The SVR model, however, suggested the ionic conductivity to be 0.061 S/m for the same ionic liquid, a significant underprediction. These observations point to the fact that the FFANN model is well suited to estimate ionic conductivity for ILs as long as the constituent ions are present in the training set and the features for the ions generated are also present in the dataset. However that is not the case for SVR which seems to perform poorly for ILs beyond the training set. It is also worth pointing out that we are able to obtain the ionic conductivity for $[\text{C}_1\text{mim}][\text{DCA}]$ higher than the largest value of 2.83 S/m at 298 K because the model was fitted using the ionic conductivity data up to ~ 19 S/m (see Figure A.8(a).) The applicability of FFANN and SVR model to predict ionic liquid combination with either the cation or anion missing from the training set is further illustrated for several test cases as shown in the supporting information.

The high accuracy of the FFANN to model the experimental ionic conductivity data prompted us to generate ionic liquid predictions as seen in Figure 4.5 along with the experimental ionic conductivity values at 298 K. It is clear that a large fraction (87.3%) of the ILs exhibit ionic conductivity below 0.5 S/m. More interestingly, the procedure yielded a number of ILs (approximately 8.3%) with ionic conductivity between 0.5 S/m and 1.0 S/m while 47 ILs were predicted to possess ionic conductivity greater than 1.0 S/m. As a comparison, the original data contained a very few ILs crossing this threshold (five out of 73 data points). Cyano-based anions such as dicyanamide $[\text{N}(\text{CN})_2]^-$, tricyanomethanide $[\text{C}(\text{CN})_3]^-$, tetracyanoborate $[\text{B}(\text{CN})_4]^-$, and thiocyanate $[\text{SCN}]^-$ accounted for the two thirds of the ILs with ionic conductivity greater than 1 S/m. As for the cation, $[\text{C}_1\text{mim}]^+$, $[\text{C}_2\text{mim}]^+$, and $[\text{C}_3\text{mim}]^+$ were found in two thirds of the ILs for which the ionic conductivity is greater than 1 S/m.

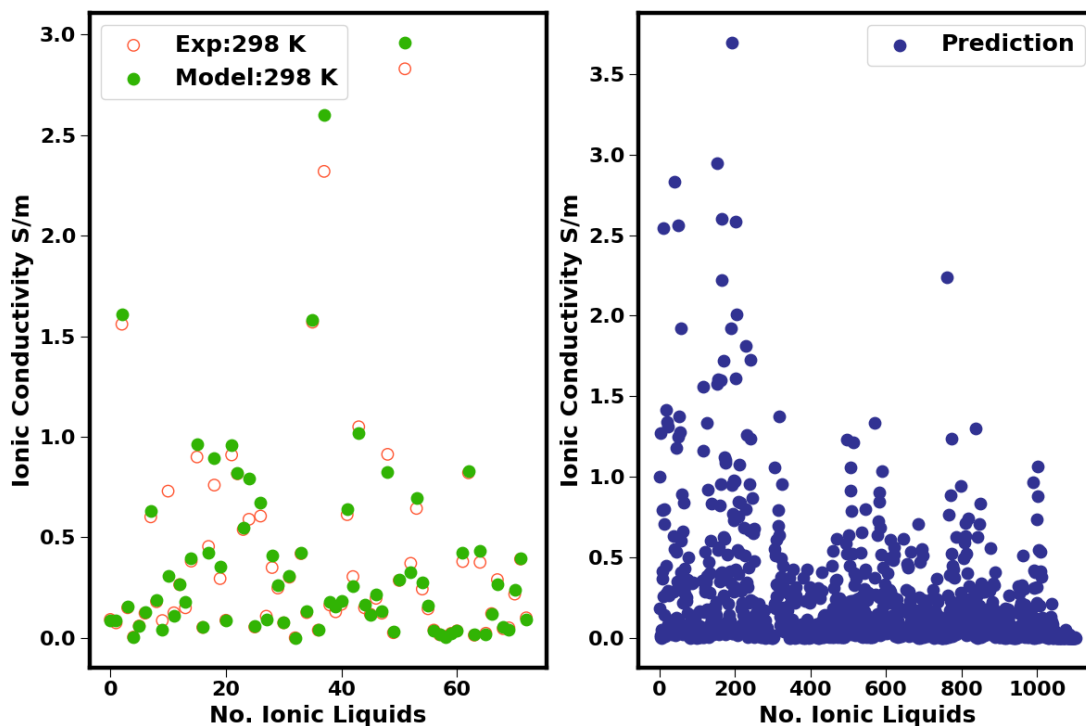


Figure 4.5: (a) Ionic conductivity comparison between experiment (open circle in red) and model prediction (green) using FFANN for all those data at 298 K. (b) Unique ionic liquid predictions using FFANN for 1102 ionic liquid obtained by combining 29 unique cations and 38 anions at 298 K. The ionic conductivity for these ILs at 350 and 400 K appear in the supporting information (Figure A.15.)

One potential issue with the generation of ILs by combining cations and anions is the lack of knowledge concerning whether such ILs would exist in the liquid state at the temperature of interest. For example, Martino et al. reported that the physical state of $[C_1mim][DCA]$ is a supercooled liquid at room temperature [208]. In lieu of experiments, some clues into the physical state of these ILs can be gleaned from conducting molecular simulations and analyzing the resulting radial distribution functions as performed by Beckner and Pfaendtner [190]. We decided not to pursue such an approach as our primary motivation here is to identify pure ILs, and binary ionic liquid mixtures bearing high ionic conductivity. In future studies, we plan to perform molecular simulations to offer insights into the molecular-level mechanism for high conductivity of the novel ILs suggested by our model. We also hope that the promising ionic liquid candidates emerging from our work will

enable the experimental community to focus its efforts in the discovery for high ionic conductivity ILs.

4.4.3 Binary Ionic Liquid Mixtures

In this section, we evaluate the performance of the FFANN model in predicting ionic conductivity of binary ionic liquid mixtures using transfer learning, where the idea is to solve new tasks by transferring knowledge gained from a closely related problem. In this work, the transfer learning takes the form of using pure-component ionic conductivity data to develop a model to predict the ionic conductivity data for binary ILs which the model has not encountered before. The utility of the approach stems from the fact that there is a significant increase in the number of binary ILs due not only to the combinatorics but also the fact that the concentration of the constituent ILs is now an additional independent variable. For example, if the number of unique cations is N_c and N_a is the number of anions, there are potentially $N_c * (N_c - 1) / 2 * N_a$ binary ILs with common anion (Binary_C systems) and $N_c * N_a * (N_a - 1) / 2$ binary ILs sharing the identical cation (Binary_A systems); the number of ILs is further amplified by the number of practically realizable formulations. With 29 unique cations and 38 anions, we enumerated 15,428 Binary_C and 20,387 Binary_A systems. For each of these mixtures, we probed 19 intermediate concentrations spaced at an interval of 0.05 mole fraction between the pure ILs leading to a total of $\sim 680,000$ binary ILs.

For estimating the ionic conductivity of a given binary mixture, we combined the input features of the constituent ionic liquid cations and anions on a mole fraction basis. For example, for a Binary_C system designated as $[C1]_x[C2]_{1-x}[A]$, we obtained the cation features as the mole fraction-weighted average of the features for [C1] and [C2]. As this is an illustration for a common anion, we retained the input features for the anion as derived in the model development. Analogously, for Binary_A systems represented as $[C][A1]_x[A2]_{1-x}$, we kept the cation features while the anion features were derived by scaling the individual anion features by respective mole fractions and adding the features thus calculated. To examine the

overall accuracy of such an approach, we compared the model predictions with experimental data reported for several binary ionic liquid mixtures in the NIST database and literature [209–214].

4.4.4 Comparison of Experimental and FFANN-predicted Ionic Conductivity of Binary ILs

Table 4.2 lists the thirteen binary ionic liquid mixtures for which experimental data for ionic conductivity are available along with the number of data points and temperature range. Also included in Table 4.2 are the FFANN predictions for these systems and corresponding RMSD values. It is remarkable that the RMSD is less than 0.1 S/m for many systems, implying the suitability of the model for estimating the ionic conductivity for binary ionic liquid systems.

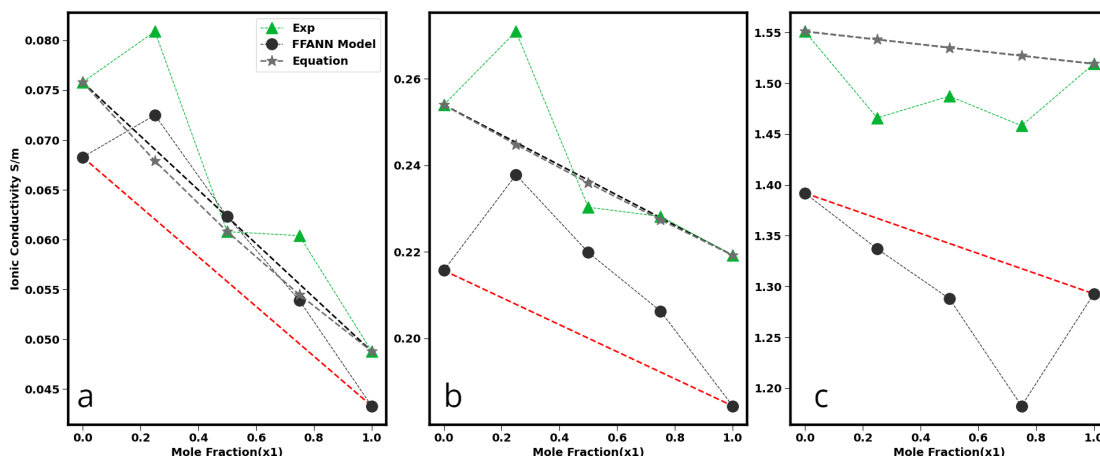


Figure 4.6: Comparison between experiment and FFANN model for $[\text{C}_4\text{mim}]^+[\text{Azide}]_{x_1} + [\text{C}_4\text{mim}]^+[\text{BF}_4]_{1-x_1}$ at (a) 303.15 K, (b) 323.15 K and (c) 368.15 K. The dashed lines connecting the pure end points are only guide to the eye. $[\text{C}_4\text{mim}]^+$ stands for 1-butyl-2,3-dimethylimidazolium cation [209]. The dashed line with \star is obtained by a logarithmic combining rule for ionic conductivity $\ln \sigma_{mix} = x_1 * \ln \sigma_1 + (1 - x_1) * \ln \sigma_2$, while the dashed line without symbol indicates estimates with a linear combining rule.

Out of the thirteen systems examined in Table 4.2, the binary system comprised of

Table 4.2: Root mean-squared deviation of the prediction and experimental data for binary ionic liquid mixtures. N.D stands for number of datapoints present in the dataset. [C₄mim]* stands for 1-butyl-2,3-dimethylimidazolium cation.

System	N.D	Temperature Range/K	RMSD S/m	Reference
[C ₂ mim][DCA] + [C ₂ mim][BF ₄]	9	298.15	0.46	[210]
[C ₂ mim][DCA] + [C ₂ mim][SCN]	30	298.15-323.15	0.36	[211]
[C ₄ mim][Cl] + [C ₄ mim][CF ₃ SO ₃]	5	298.0	0.05	[214]
[C ₄ mim][MeSO ₄] + [C ₄ mim][Me ₂ PO ₄]	4	298.0	0.17	[214]
[C ₄ mim][NTf ₂] + [C ₄ mim][CF ₃ SO ₃]	5	298.0	0.05	[214]
[C ₄ mim][NTf ₂] + [C ₄ mim][MeSO ₄]	5	298.0	0.10	[214]
[C ₄ mim][NTf ₂] + [C ₄ mim][Me ₂ PO ₄]	6	298.0	0.21	[214]
[C ₄ mim]*[Azide] + [C ₄ mim]*[BF ₄]	70	303.15-368.15	0.08	[209]
[C ₈ mim][Cl] + [C ₈ mim][BF ₄]	42	303.0-333.0	0.02	[213]
[C ₆ mim][Cl] + [C ₆ mim][BF ₄]	42	303.0-333.0	0.06	[213]
[C ₄ mim][NTf ₂] + [C ₄ mim][Acetate]	32	283.15-333.15	0.13	[212]
[C ₆ mim][Cl] + [C ₆ mim][PF ₆]	35	303.0-333.0	0.07	[213]
[C ₂ mim][BF ₄] + [C ₈ mim][BF ₄]	30	280.0-300.0 [†]	0.09	
Overall	315		0.167	

[†]Personal communication

[C₄mim]*[Azide] and [C₄mim]*[BF₄] exhibits non-ideal behavior, where the ionic conductivity achieves either a minimum or maximum at an intermediate mole fraction. Figure 4.6 shows that the ionic liquid passes through a maximum at lower temperatures (303 and 323 K), while a minimum is observed at higher temperature (368 K). The FFAAN model developed here seems to capture the trend with a reasonable agreement with experimental data. Furthermore, the model is accurate enough to identify the concentrations at which such extrema were measured in the experiment [209]. Overall this indicates that the model is robust enough to closely match both qualitative and quantitative trends; this is quite encouraging

given that the data for these binary ionic liquid systems were not part of the model development. We further tested the predictive capability of the model to reproduce such a non-ideal behavior reported by McFarlane et al. [50]. The authors measured the molar conductivity for the binary ionic system of [C₂mim][NTf₂] and [C₂mim][CF₃SO₃] and found a maximum at an intermediate mole fraction. Due to the lack of experimental data for molar volumes, a direct comparison was not possible; however, our model outputs (Figure A.16) indeed confirmed that the binary ionic liquid mixture would exhibit a maximum in ionic conductivity.

Encouraged by the success of the model in estimating ionic conductivity for several binary mixtures, we proceeded to examine if there are binary ionic liquid mixtures producing an extremum (either a maximum or minimum) in ionic conductivity as the mole fractions of the constituent ILs are varied. We discovered that there were a total of 5040 Binary_C systems, which yielded a maximum in the ionic conductivity. On the other hand, a total of 3771 Binary_A systems produced a maximum in the ionic conductivity at 298 K. Normalizing these systems by the corresponding number of possible binary ionic liquid systems, we calculated that approximately 32.6% of Binary_C and 18.4% of Binary_A systems could potentially be formed to obtain ionic conductivity higher than those of the two pure ILs forming the mixture. Two observations are worth pointing out: (a) binary ionic liquid systems offer a viable pathway for increasing ionic conductivity; (b) the likelihood for obtaining a maximum in ionic conductivity is higher when two different cations are mixed, particularly mixing cations with a large difference in the alkyl chain length.

In order to gain additional insight into the extent of enhancement in ionic conductivity, we calculated the percentage enhancement (E) using eq. 4.1 where σ_{\max} represents the maximum ionic conductivity for the mixture and $\sigma_{\max,\text{pure}}$ refers to the higher of the two pure ionic conductivities. Figure 4.7(a) and (b) present the binary enhancement factor for Binary_A and Binary_C systems, respectively.

$$E = \frac{\sigma_{\max} - \sigma_{\max,\text{pure}}}{\sigma_{\max,\text{pure}}} * 100 \quad (4.1)$$

It is apparent that the percentage enhancement is large for the ionic liquid mixtures systems with ionic conductivity values lower than 1 S/m and is partly attributable to the low conductivity values of the pure ILs appearing in the denominator of eq. 4.1. It is also interesting to observe that the Binary_C systems display a broader range of enhancement values in comparison to those found for Binary_A systems. The analysis suggests that there exists at least one ionic liquid mixture for each of the unique cations and anions exhibiting an enhancement. We also uncovered that the top three Binary_A mixtures for which maximum enhancement percentage was obtained contain $[\text{HSO}_4]^-$ and chloride as the anions. These mixtures are (i) $[\text{C}_6\text{mim}][\text{Cl}]_{0.75}[\text{HSO}_4]_{0.25}$ with a maximum pure value of 0.0021 S/m and enhanced maximum value of 0.0177 S/m leading to an enhancement of 715.4%, (ii) $[\text{C}_8\text{mim}][\text{Cl}]_{0.75}[\text{HSO}_4]_{0.25}$ with a maximum pure value of 0.0010 S/m and enhanced maximum value of 0.008 S/m with an enhancement of 675.4% and (iii) $[\text{C}_3\text{mim}][\text{Cl}]_{0.55}[\text{HSO}_4]_{0.45}$ with a maximum pure value of 0.006 S/m and enhanced maximum value of 0.047 S/m with an enhancement of 582.4%. As for the binary cation mixture seen in (b), 1-(1-cyanomethyl)-3-methylimidazolium_{0.6} 3-(2-(butylamino)-2-oxoethyl)-1-ethyl-1H-imidazolium_{0.4} $[\text{PF}_6]$ has a maximum pure value of 0.0196 S/m and the maximum value of 0.113 S/m leading to an increase of 485.2%.

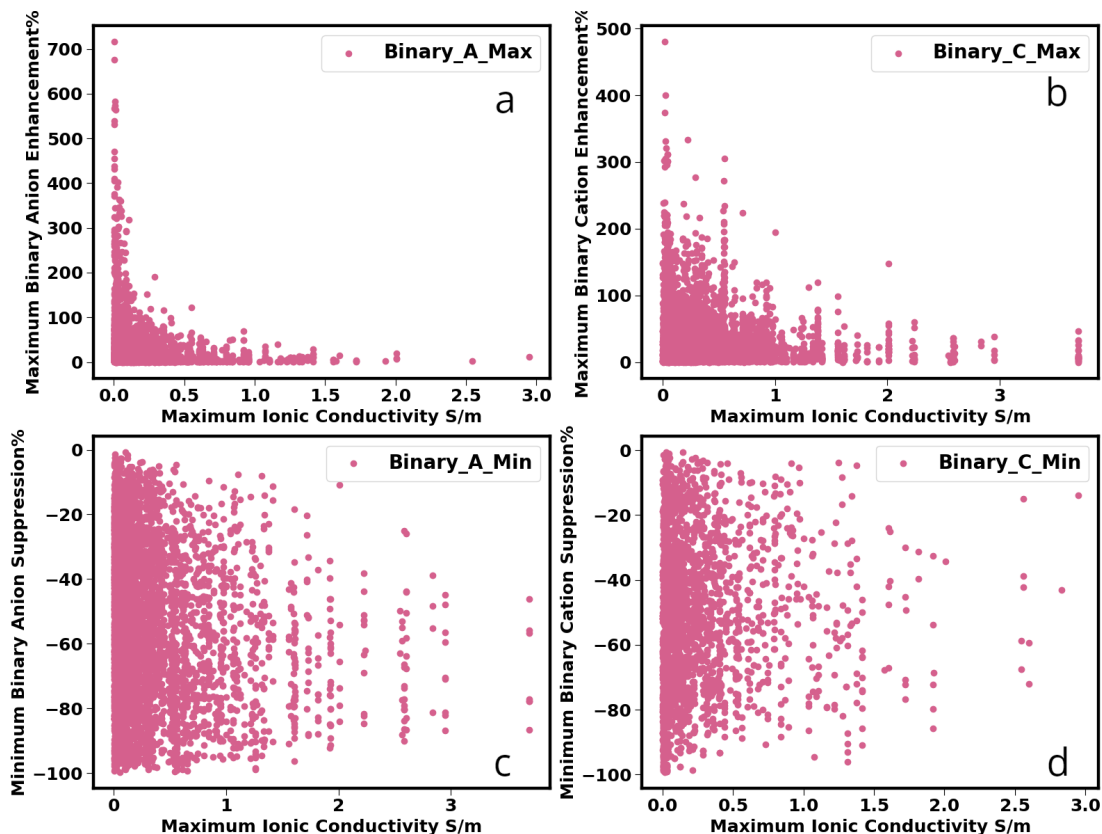


Figure 4.7: Percentage enhancement and suppression in ionic conductivity for binary ionic liquid mixtures at 298 K. Binary_A_Max stands for a binary mixture sharing a common cation showing maximum enhancement; Binary_C_Max stands for a binary cation mixture displaying maximum enhancement; Binary_A_Min denotes a binary anion mixture exhibiting minimum suppression, and Binary_C_Min represents a binary cation mixture producing minimum suppression.

Similarly, there were several binary ionic liquid systems which showed an opposite behavior, i.e., there is at least one binary ionic liquid composition at which the ionic conductivity is lower than those of the corresponding pure ILs. We uncovered 2305 Binary_C and 4284 Binary_A systems which showed a minimum in the ionic conductivity as a function of the ionic liquid composition at 298 K. To quantify the extent of lowering in the ionic conductivity, we calculated percentage suppression (S) using eq. 4.2 in which σ_{\min} denotes the minimum in ionic conductivity and $\sigma_{\max, \text{pure}}$ refers to the maximum of the two pure ionic conductivities. We elected

to measure the deviation from $\sigma_{\max,\text{pure}}$ to emphasize the reduction in the

$$S = \frac{\sigma_{\min} - \sigma_{\max,\text{pure}}}{\sigma_{\max,\text{pure}}} * 100 \quad (4.2)$$

ionic conductivity expected when an ionic liquid with lower conductivity is mixed with the one possessing high conductivity. Inherent in the definition in eq. 4.2 is the fact that the percentage lowering is capped at 100%. The extent of depression in the ionic conductivity depicted in Figure 4.7 confirms the expectation. It is noteworthy that the suppression in the ionic conductivity brought about by the mixture of anions is restricted to ILs with ionic conductivity below 1 S/m, while the depression in the ionic conductivity due to mixing of cations is predicted to cover the entire range of ionic conductivities. Furthermore, we identified the number for a given cation pair or anion responsible for elevating or depressing ionic conductivity for binary mixtures. The analysis is presented in the form of various heat maps (Figures A.17, A.18, A.19, and A.20). It is also interesting to note that the short chain alkyl cations such as 1-methylimidazolium and 1,3-dimethylimidazolium are promising cations for ionic conductivity enhancement when combined with other cations as seen in Figure A.19.

4.5 Conclusion

In this chapter, we made use of the NIST ILThermo Database to derive a FFAAN model and a SVR model for predicting ionic conductivity of pure imidazolium-based ILs. The ionic conductivity values ranged over six orders of magnitude and covered temperatures from 275 K to 475 K. The input features for the models were obtained using RDKit. The overall accuracy was found to be nearly identical for both the models. An examination of the predictions for the high ionic conductivity ILs suggested superior performance for FFAAN, which was then employed for subsequent predictions.

Using 29 unique cations and 38 unique anions in the database, the ionic conductivity for all the possible combinations (1102 in total) were predicted at 298 K. The procedure led to the identification of the ionic liquid [C₁mim][DCA] that is

not present in the training set with an ionic conductivity of 3.70 S/m - 30% higher than the highest ionic conductivity in the training set and the NIST ILThermo Database at 298 K. The prediction was confirmed with the experimental data available in the literature. A simple procedure for combining the features on a mole fraction-weighted basis was devised to evaluate the predictive capability of the model for ionic liquid-ionic liquid mixtures. The results obtained with the approach showed that model was able to accurately capture the ionic conductivity for several binary for which experimental data exist.

The present study suggests a large number of binary mixture with non-ideal behavior in terms of the ionic conductivity. We encourage the experimental and molecular simulation communities to test the predictions. Confirmation of such non-ideality will increase the confidence in such models, while any deviations of the measured or computed properties from the predictions will enable a further refinement of the model. In either case, it is expected that the concerted effort between the experimental, molecular simulation, and machine learning approaches will accelerate materials discovery in the ILS domain.

CHAPTER 5

Revealing Hydrogen Bond Dynamics between Ion Pairs in Binary and Reciprocal Ionic Liquid Mixtures

5.1 Abstract

Molecular dynamics simulations are performed to probe the molecular-level interactions between various ionic pairs in the reciprocal mixture consisting of equimolar amounts of the cations 1-ethyl-3-methylimidazolium $[\text{C}_2\text{mim}]^+$ and 1-n-hexyl-3-methylimidazolium $[\text{C}_6\text{mim}]^+$ and dicyanamide $[\text{DCA}]^-$ and bis(trifluoromethanesulfonyl)imide $[\text{NTf}_2]^-$ anions. Any enhancement or depletion in these interactions is compared with those existing in equimolar binary mixtures $[\text{C}_2\text{mim}][\text{C}_6\text{mim}][\text{DCA}]$, $[\text{C}_2\text{mim}][\text{C}_6\text{mim}][\text{NTf}_2]$, $[\text{C}_2\text{mim}][\text{DCA}][\text{NTf}_2]$, and $[\text{C}_6\text{mim}][\text{DCA}][\text{NTf}_2]$ and pure ILs $[\text{C}_2\text{mim}][\text{DCA}]$, $[\text{C}_6\text{mim}][\text{DCA}]$, $[\text{C}_2\text{mim}][\text{NTf}_2]$, and $[\text{C}_6\text{mim}][\text{NTf}_2]$. The simulation results indicate that the $[\text{C}_2\text{mim}]^+$ cation prefers to interact favorably with the strongly coordinating $[\text{DCA}]^-$ anion through enhanced hydrogen-bonding interactions, while showing no preferential interest towards the other anion in the reciprocal mixture. The average hydrogen bond lifetimes between $[\text{C}_2\text{mim}]^+$ cation and $[\text{DCA}]^-$ anion increases by a factor of two in the reciprocal mixture compared to that in the pure system. We find that the hydrogen bond lifetimes in the various systems are directly correlated to the first peak intensity in the center-of-mass radial distribution functions for corresponding ion pairs and have a direct bearing on the self-diffusion coefficients of the ions. Our results point to possibilities of tuning interactions between various species

in reciprocal ionic liquid mixture by appropriately changing the cation and anion combinations.

5.2 Introduction

For more than two decades now, ILs have gained widespread attraction because of several unique attributes such as low melting points despite ionic in nature, ability to exist as liquids at ambient conditions, low vapor pressures, and high thermal and chemical stability. In addition to these desirable properties, the ever-increasing attention paid to these ionic solvents is due to the fact that they offer the possibility of designing task-specific solvents with desired properties when an appropriate cation is paired with a suitable anion. Given that there exist seemingly an endless list of cations and a plethora of anions, it is no surprise that the chemical space for ILs is vast. The search for an ideal ionic liquid with requisite properties can be expanded further by considering mixtures of two or more ILs, which can potentially aid, for example, in improving sluggish dynamics encountered for many ILs.

A large number of experimental and molecular simulation studies have extensively examined the molecular level interactions and structural transitions in ILs mixtures and their relationship to deviations in macroscopic properties [51, 173, 210, 215]. Kapoor and Shah, in a recent study, showed that the difference in molar volumes of ILs and hydrogen bonding accepting ability between the anions played a driving force in determining whether a given binary ionic liquid mixture with a common cation would exhibit non-ideal behavior [52]. The authors conducted molecular dynamics (MD) simulation for 16 different binary anion mixtures and concluded that pure ILs with molar volume difference greater than $60 \text{ cm}^3/\text{mol}$ and hydrogen bonding basicity difference greater than 0.4 when mixed would most likely yield non-native structural transitions at microscopic level. Such guidelines could serve to screen potential ionic liquid mixtures to identify non-ideal behavior. Several other binary ionic liquid mixtures with two anions and a common cation have obtained similar results where the difference in hydrogen bonding ability of the anions

and the difference in anion size induces competition between the anions that causes the system to show non-ideal behavior. In such systems, the strongly coordinating anion prefers to occupy the locations proximal to the most acidic hydrogen in the imidazolium cation, which in turn displaces the less basic anion. [53, 54, 216, 217] .

As a consequence of the competition for the most acidic site in the imidazolium cation, binary anion mixtures also display hydrogen bond dynamics that differs from that in the pure ILs. For example, the MD study by Wang et al. reported an enhancement in the hydrogen bond dynamics between 1-ethyl-3-methylimidazolium [C₂mim] with tetrafluoroborate [BF₄] in the binary anion mixture containing [C₂mim]⁺, [BF₄]⁻ and bis(trifluoromethanesulfonyl)imide [NTf₂]⁻ [218]. At [C₂mim][BF₄]_{0.25}[NTf₂]_{0.75} mole fraction the lifetime of intermittent hydrogen bond dynamics between [C₂mim][BF₄] increases two-fold to 117 ps as compared to 57 ps for the pure [C₂mim][BF₄] ionic liquid. Interestingly the hydrogen bonding lifetime of [C₂mim][NTf₂] decreases as more [BF₄]⁻ anion is added to the mixture. Similarly, Gekhre et al. in an MD study, examined the molecular level interaction and hydrogen bond dynamics between 1-butyl-3-methylimidazolium [C₄mim]⁺ and Cl⁻ and trifluoromethanesulfonate [OTf]⁻ [219]. The authors observed an increase in the lifetime of hydrogen bond dynamics between [C₄mim] and Cl as Cl⁻ ion is added to the system, presumably due to the stronger hydrogen bonding ability of Cl⁻ than that for [OTf]⁻ anion.

ILs comprised of two cations and a common anion is another way to modulate properties of the resulting ionic liquid mixtures, primarily driven by the difference in the alkyl chain attached to the cations [220–222]. Shimizu et al. reported, from an MD study, that the equimolar mixture of [C₂mim]⁺ and 1-hexyl-3-methylimidazolium [C₆mim]⁺ with [NTf₂]⁻ obeyed a nearly ideal mixing law with a small negative excess molar volume [223]. In an experimental study, the Welton group found that the sign of the excess molar volume switches from negative to positive as the difference in the alkyl chain length between the imidazolium cation increases. The observation was attributed to the mismatch between the alkyl chain length

leading to a large number of voids in the mixture [224] as the longer alkyl chain is unable to incorporate the shorter chain into its non-polar domain, disrupting the continuous non-polar domain [225, 226]. Cosby et al., in a combined experimental MD simulation study, observed that the mesoscale morphologies in the mixtures of $[\text{C}_2\text{mim}]^+$ and 1-octyl-3-methylimidazolium $[\text{C}_8\text{mim}]^+$ with $[\text{BF}_4]^-$ could be tuned, which leads to an enhancement in the enhancement of static dielectric permittivity of the liquid by as much as 100% at equimolar concentrations as a result of change in the aggregation behavior of $[\text{C}_8\text{mim}]^+$ [227]. Such binary cation mixtures can also produce interesting hydrogen bonding dynamics as noted by Wang and co-workers in their MD study focusing on the mixtures of $[\text{C}_2\text{mim}]^+$ and $[\text{C}_4\text{mim}]^+$ cations with $[\text{BF}_4]^-$ [228]: the average lifetime for a hydrogen bond is greater for $[\text{C}_4\text{mim}]^+[\text{BF}_4]^-$ than that for $[\text{C}_2\text{mim}]^+[\text{BF}_4]^-$, which was reasoned to be partially responsible for the slow rotation of the $[\text{C}_4\text{mim}]^+$ cation.

A yet-another approach to tailoring the physicochemical properties of ILs is through reciprocal mixtures that contain at least two cations and two anions [214, 229]. These systems are interesting because they are usually formed by cations with varying alkyl chain lengths and anions that differ in their hydrogen bonding ability. By carefully selecting these parameters, which affect the non-polar and hydrogen bonding interactions, it may be possible to induce ionic associations that are not observable in neat or binary ionic liquid mixtures, leading to potentially non-ideal behavior. However, very few studies have probed molecular-level interactions in such mixtures. The experimental findings by Taige et al. indicate that the viscosity and ionic conductivity follow a non-ideal behavior for some reciprocal mixtures [230]. Experiments by Bharmoia et al. also suggest that a pronounced deviation from ideality is possible for various physicochemical properties compared to other binary mixtures [231]. An MD examination of reciprocal mixtures of 1-benzyl-3-methylimidazolium $[\text{Zmim}]^+$, $[\text{C}_4\text{mim}]^+$, $[\text{BF}_4]^-$, and $[\text{NTf}_2]^-$ revealed that the stronger hydrogen bonding ability anion $[\text{BF}_4]^-$ was seen to displace $[\text{NTf}_2]^-$ from the coordination shell of both the cations [232].

To the best of our knowledge, there exists no study in which hydrogen bond dynamics of the reciprocal is systematically studied, especially when there is a mismatch in the alkyl chain length in the cationic species combined with a considerable difference in the hydrogen bond accepting ability of the anions. In such systems, it would be important to investigate how the presence of a cation capable of forming a continuous non-polar domain perturbs the distribution of anions in coordination shell of the cation with shorter alkyl chain, specifically the hydrogen bonding environment and hydrogen bonding dynamics as they exert profound influence on the physicochemical properties of ILs. The focus of this work, therefore, is to employ an all-atom MD simulations to provide molecular-level insight into an equimolar reciprocal mixture that is composed of $[\text{C}_2\text{mim}]^+$, $[\text{C}_6\text{mim}]^+$, dicyanamide $[\text{DCA}]^-$ and $[\text{NTf}_2]^-$. Following Kapoor and Shah, the cations and anions are selected as they offer a significant difference in molar volumes and the hydrogen bonding ability. We systematically evaluate structure and dynamics in pure ILs, binary cation mixtures, binary anion mixtures, and the reciprocal mixture system.

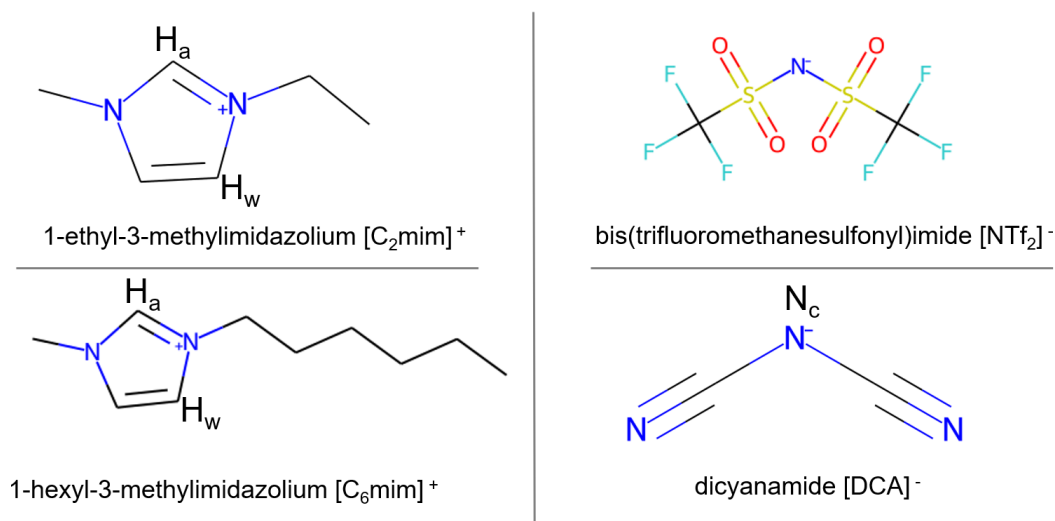


Figure 5.1: Chemical structures of the cations and anions used in this study.

5.3 Methodology

5.3.1 Force Fields

The all-atom version of Optimized Potentials for Liquid simulations (OPLS-AA) forcefield was used to model the ILs studied in this work. The forcefield for this study was taken from the work of Lopes and Paduá [233–235]. The charge scaling for electrostatic interactions was set to ± 0.8 to improve transport properties as demonstrated by other studies [236, 237]. We also conducted ± 1.0 charge scaling simulation to compare the difference in results for bulk density and radial distribution function. The results for ± 1.0 charge scaling are provided in the supporting information.

The OPLS-AA model treats each atom as an interaction site and the total potential energy U_{total} takes into account bond stretching $U_{\text{Stretching}}$, bond bending U_{Bending} , dihedral angle U_{Torsion} , Coulomb, and 12-6 Leonard Jones (LJ) terms for non-bonded interactions $U_{\text{Non-Bonded}}$. The functional form is shown below in Equations 1-5.

$$U_{\text{total}} = U_{\text{Non-Bonded}} + U_{\text{Stretching}} + U_{\text{Bending}} + U_{\text{Torsion}} \quad (5.1)$$

$$U_{\text{Non-Bonded}} = \sum_i \sum_j 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_i \sum_j \frac{q_i q_j e^2}{4\pi\epsilon_o r_{ij}} \quad (5.2)$$

$$U_{\text{Stretching}} = \sum_{\text{Bonds}} \frac{K_r}{2} (r - r_{eq})^2 \quad (5.3)$$

$$U_{\text{Bending}} = \sum_{\text{Angles}} \frac{K_\theta}{2} (\theta - \theta_{eq})^2 \quad (5.4)$$

$$U_{\text{Torsion}} = \sum_{\text{Torsion}} \frac{V_1}{2} [1 + \cos(\phi)] + \frac{V_2}{2} [1 - \cos(2\phi)] + \frac{V_3}{2} [1 + \cos(3\phi)] \quad (5.5)$$

where LJ parameters σ_{ij} and ϵ_{ij} represent the energy and size parameters, r_{ij} denotes the distance between atomic sites i and j , q is the partial atomic charge for electrostatic interactions. r , r_{eq} , θ , θ_{eq} are bond length, equilibrium bond length, bending angle and equilibrium bending angle, respectively. K_r , K_θ are force constant for stretching and bending, respectively; V_i is the torsional parameters

for dihedrals. LJ potential between unlike pairs was calculated using the geometric combining rule for σ and ϵ . The intramolecular non-bonded interactions between atoms separated by three bonds was evaluated using a scaling factor of 0.5 for both the LJ and electrostatic interactions.

5.4 Simulation Details

Molecular Dynamic (MD) simulations were performed for nine different systems at 323.0 K using Gromacs 2018 software package [238, 239] including four pure ionic liquid systems [C₂mim][DCA], [C₂mim][NTf₂], [C₆mim][DCA], [C₆mim][NTf₂], four binary ionic liquid systems with equimolar compositions [C₂mim][DCA]_{0.5}[NTf₂]_{0.5}, [C₆mim][DCA]_{0.5}[NTf₂]_{0.5}, [C₂mim]_{0.5}[C₆mim]_{0.5}[DCA], [C₂mim]_{0.5}[C₆mim]_{0.5}[NTf₂], and an equimolar reciprocal system [C₂mim]_{0.5}[C₆mim]_{0.5}[DCA]_{0.5}[NTf₂]_{0.5}. A cubic simulation box was employed for each of the ionic liquid systems. The length of the simulation box for pure systems was estimated from experimental density at 323 K, while the simulation box length for the binary mixtures was set assuming ideal mixing behavior. Similarly, the initial volume for the reciprocal system was calculated by considering the ideal mixing volume for the ILs [C₂mim][DCA], [C₂mim][NTf₂], [C₆mim][DCA], and [C₆mim][NTf₂] with mole fractions of each of the ILs set to 0.25. Packmol was used to prepare an initial configuration for all the systems for 500 ion pairs [240] with periodic boundary conditions applied in all directions.

The simulation protocol involved five steps: minimization, annealing, *NVT* equilibration, and *NPT* equilibration followed by two production runs in the *NPT* ensemble. Each system was first minimized using the steepest-descent algorithm for 10000 steps and subjected to a 1.5 ns annealing protocol. The temperature was slowly increased to 323 K during the first stage of annealing followed by relaxation of the system at 323 K. The temperature was further raised to 523 K to ensure additional relaxation of the system before bringing the system down to 323 K. An *NVT* equilibration for 10 ns was performed next, in which the temperature of the

system was controlled at 323 K using a Berendsen thermostat with a time constant of 0.4 ps. An *NPT* equilibration for 10 ns was carried out so that the density approached an equilibrium value; Berendsen thermostat and barostat were applied during this stage with time constants of 0.4 ps and 0.2 ps, respectively. The first *NPT* production run lasted 60 ns in which the Nosé-Hoover thermostat [241] and Parrinello-Rahman barostat [242] were activated at a coupling constant of $\tau_t = 0.4$ ps and $\tau_p = 2.0$ ps. Finally data collection and analysis were carried out on the final 20 ns of the *NPT* production run. Results for each system were averaged over three independent simulations with different starting initial coordinates to estimate the uncertainty in property predictions.

During the course of the entire simulation protocol, the bonds involving hydrogen atoms were constrained using the LINCS algorithm. Both LJ and electrostatic interactions were truncated at 12 Å, while long-range electrostatic interactions were handled using particle-mesh Ewald (PME) summation with a PME order of 4 and Fourier spacing of 0.1 nm. Long-range corrections were also applied for LJ energy and pressure.

5.5 Results and discussion

5.5.1 Liquid Densities

The densities of all the nine different systems studied in this paper are shown in Table 5.1 and visualized in Figure B.1. The simulated densities of the pure systems with both charge scaling of ± 1.0 and ± 0.8 are compared with experimental data. Density data for the ± 1.0 charge scaling are provided in the supporting information. The pure ionic liquid densities obtained with ± 1.0 scaling are in excellent agreement with experimental measurement as indicated by the average absolute relative deviation (AARD) of 1.46%. In contrast, the AARD increases to 5.2% for pure ionic liquid densities with ± 0.8 charge scaling [122, 123, 236]. The deviations in densities are similar in magnitude to Goloviznina et al. study with ± 0.8 scaling using CL&P forcefields [236].

Table 5.1: Detailed description of the system composition and equilibrium densities at 323 K

System	Ions				Density (g/cm ³)
	[C ₂ mim] ⁺	[C ₆ mim] ⁺	[DCA] ⁻	[NTf ₂] ⁻	This work
[C ₂ mim][DCA]	500		500		1.026±0.001
[C ₂ mim][NTf ₂]	500			500	1.486±0.009
[C ₆ mim][DCA]		500	500		0.974±0.001
[C ₆ mim][NTf ₂]		500		500	1.332±0.008
[C ₂ mim][C ₆ mim][DCA]	250	250	500		0.996±0.001
[C ₂ mim][C ₆ mim][NTf ₂]	250	250		500	1.390±0.009
[C ₂ mim][DCA][NTf ₂]	500		250	250	1.292±0.001
[C ₆ mim][DCA][NTf ₂]		500	250	250	1.184±0.001
[C ₂ mim][C ₆ mim][DCA][NTf ₂]	250	250	250	250	1.231±0.001

5.5.2 Radial Distribution Function (RDF)

To understand the molecular-level structure between the ions in the reciprocal mixture, we computed the center of mass (COM) radial distribution function (RDF) plot for all the cation-anion interactions in the system as depicted in Figure 5.2(a). It was observed that the characteristic of the first solvation shell as deduced from the intensity of the first peak and the distance at which the peak is located varied with the system. For example, the cation-anion correlations are strongest for the [C₂mim]-[DCA] as the intensity of the first peak in the RDF is highest amongst all the RDFs. In contrast, the first peak height is nearly identical for [C₆mim][DCA], [C₂mim][NTf₂], and [C₆mim][NTf₂]. The observation points to a strong preferential interaction between the [C₂mim]-[DCA]. This is also reflected by the greater number of [DCA]⁻ surrounding [C₂mim]⁺ in the coordination number plot in Figure 5.2(b). This is most likely due to the easier access to the hydrogen bonding

sites of $[\text{C}_2\text{mim}]^+$ by the stronger coordinating ability and smaller size of $[\text{DCA}]^-$ in comparison to those for $[\text{NTf}_2]^-$. The location of the first peak height is almost invariant at $\sim 5 \text{ \AA}$ for the cation-anion interactions in $[\text{C}_2\text{mim}][\text{DCA}]$, $[\text{C}_6\text{mim}][\text{DCA}]$, and $[\text{C}_2\text{mim}][\text{NTf}_2]$, while the bulky nature of both the ions in $[\text{C}_6\text{mim}][\text{NTf}_2]$ leads to the first peak height locating further by 0.5 \AA . The anion size difference also affects the location of the first minimum in the RDFs such that the first solvation shell of $[\text{DCA}]^-$ around the cations is more contracted than that of $[\text{NTf}_2]^-$. The location of the first minimum remains unaffected for a given anion, which suggests that the size of the first solvation shell is determined primarily by the interaction of the imidazolium ring with the anion. The type of cation-anion pair also exerts subtle changes in the second solvation shell: the peak intensity is nearly identical for the pairs involving $[\text{C}_2\text{mim}]^+$ and is higher than those for pairs containing $[\text{C}_6\text{mim}]^+$. Beyond the second shell, the RDFs are nearly identical in all respects.

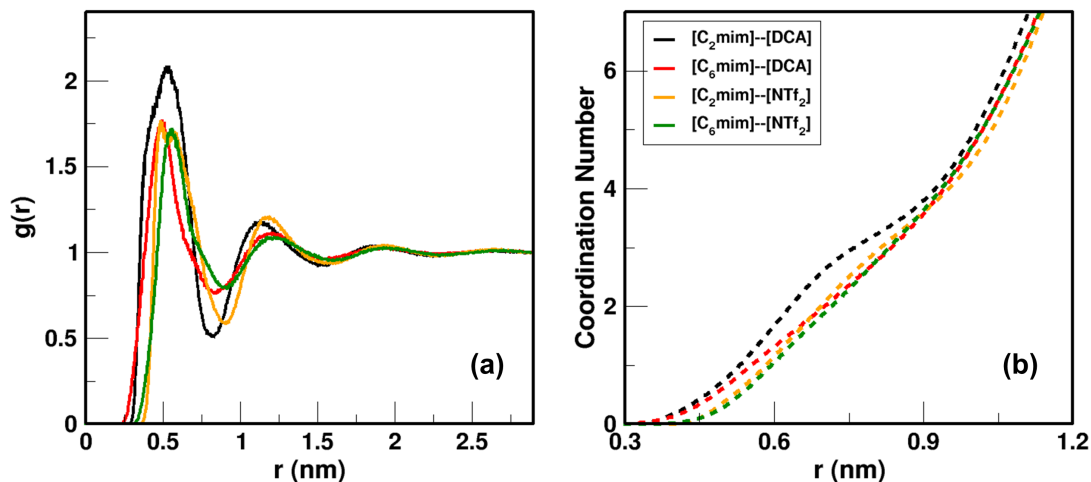


Figure 5.2: Center of mass (COM) radial distribution plot of all the cation-anion interactions in reciprocal mixture.

Next we determined the cation-anion COM radial distribution functions for $[\text{C}_2\text{mim}][\text{DCA}]$ in various ionic liquid mixtures to examine whether the preferential interaction of $[\text{DCA}]^-$ with $[\text{C}_2\text{mim}]^+$ is retained. Examination of these correlations in Figure 5.3 reveals that the first peak height follows the order $[\text{C}_2\text{mim}][\text{DCA}] < [\text{C}_2\text{mim}]_{0.5}[\text{C}_6\text{mim}]_{0.5}[\text{DCA}] < [\text{C}_2\text{mim}][\text{DCA}]_{0.5}[\text{NTf}_2]_{0.5} < [\text{C}_2\text{mim}]_{0.5}[\text{C}_6\text{mim}]_{0.5}[\text{DCA}]_{0.5}[\text{NTf}_2]_{0.5}$,

which clearly shows that the $[\text{C}_2\text{mim}][\text{DCA}]$ interactions become stronger in binary and reciprocal mixtures. The minor increase in the peak height when $[\text{C}_6\text{mim}]$ is added to $[\text{C}_2\text{mim}][\text{DCA}]$ is attributed to the formation of non-polar domains due to the presence of $[\text{C}_6\text{mim}]^+$, which leads to the confinement of the anion in the polar region and has an effect of pushing the anion closer to $[\text{C}_2\text{mim}]^+$. Shimizu et al. reported a similar increase in first peak height in the RDF between the carbon attached to H_a and oxygen atom in $[\text{NTf}_2]^-$ an equimolar mixture of $[\text{NTf}_2]^-$ and $[\text{C}_2\text{mim}][\text{NTf}_2]$ and $[\text{C}_6\text{mim}][\text{NTf}_2]$ [223]. A more pronounced influence on the first peak height is noted when a weakly coordinating anion $[\text{NTf}_2]^-$ is mixed with $[\text{C}_2\text{mim}][\text{DCA}]$, which is consistent with several reports in literature that the anion with stronger hydrogen bonding ability preferentially interacts with a cation in ionic liquid mixtures containing a common cation and two anions. Lastly, the reciprocal mixture shows the highest peak height compared to all the other systems exhibiting a combined influence of the larger cation and a weakly coordinating anion in driving the preferential interaction between $[\text{C}_2\text{mim}]^+$ and $[\text{DCA}]^-$. The trends seen for the reciprocal mixture are similar to a recent MD study by Sappidi et al. [232] where the authors found that the $[\text{BF}_4]^-$ anion tend to show a strong association for the two cations in the reciprocal mixture compared to $[\text{NTf}_2]^-$ anion.

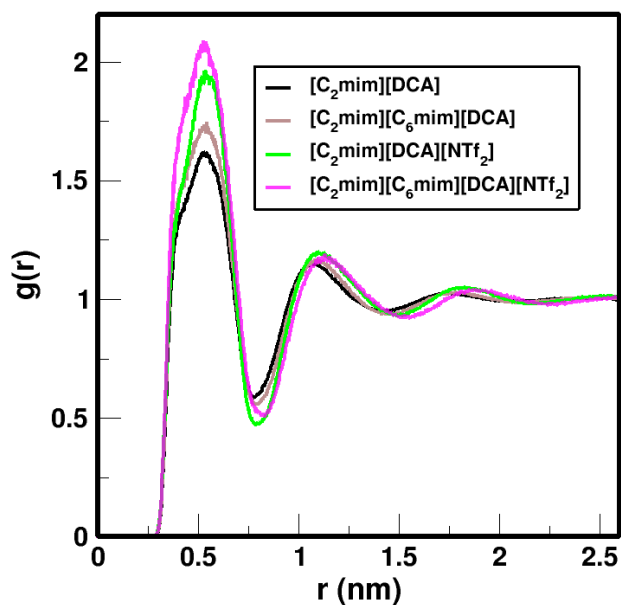


Figure 5.3: COM RDF plot for $[\text{C}_2\text{mim}]^+ \cdots [\text{DCA}]^-$ interactions in pure and various mixtures.

The origin of the preferential interaction between $[\text{C}_2\text{mim}]^+$ and $[\text{DCA}]^-$ in mixtures relative to the pure ionic liquid is examined further using the radial distribution functions around the hydrogen bonding sites in the imidazolium ring. Figure 5.4(a) depicts the interactions between H_a position in the cation with the N_c position in the $[\text{DCA}]^-$ anion as one of the dominant hydrogen bonding sites (please refer to Figure 5.1 for the nomenclature). The magnitude of the first peak intensity greater than 1 at short distances for all the systems indicate a presence of hydrogen bonding interactions. The RDF's exhibit split peaks in the first solvation shell, suggesting that $[\text{DCA}]^-$ interacts with the hydrogen bonding site through multiple conformations. It is interesting to note that the two peaks in the first solvation shell for all the systems occur at around the same distance, while a clear change in the peak heights is observed for pure and mixture systems. The neat ionic liquid $[\text{C}_2\text{mim}][\text{DCA}]$ exhibits the lowest peak height, which is enhanced upon the addition of the $[\text{C}_6\text{mim}]^+$ cation. The addition of $[\text{NTf}_2]^-$ induces a significant increase in the first peak height. Such a behavior, brought about the difference

in the hydrogen bonding ability of the two anions, is well documented in literature [52, 243]. Finally, the reciprocal mixture shows the largest increase in the peak height amongst all the mixture systems, demonstrating the combined effect of the difference in the hydrogen bonding ability of the anions and the formation of non-polar domains due to the $[\text{C}_6\text{mim}]^+$ cation. The RDF's for the H_w position in the cation and the $[\text{DCA}]^-$ anion are provided in Figure 5.4(b). Although the first peak intensities are lower than those found for the H_a site, the ranking of peak heights is similar to that obtained for the most acidic hydrogen bonding site, which clearly shows that enhanced interactions of $[\text{C}_2\text{mim}]^+$ with $[\text{DCA}]^-$ in ionic liquid mixture systems are due to an increased coordination of $[\text{DCA}]^-$ at all the hydrogen bonding sites. Overall, these findings confirm that the preferential interaction of the anion with stronger hydrogen bonding ability can be dramatically enhanced by forming ionic liquid mixtures. As hydrogen bonding interactions have been implicated in several physicochemical properties of ILs, it is expected that ILs properties can be tuned through a subtle manipulation of these interactions through various combinations of cations and anions in an ionic liquid mixture.

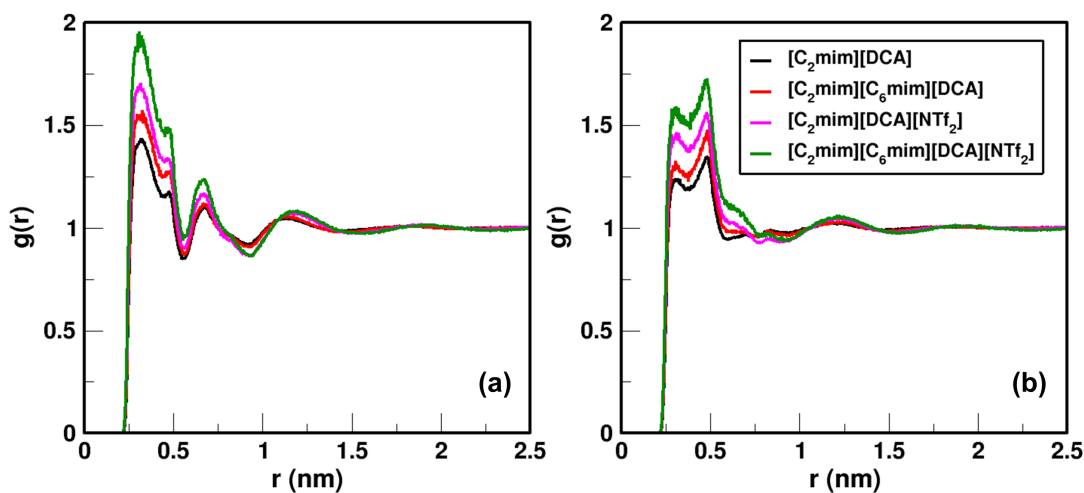


Figure 5.4: RDF plots for site-site interaction in pure and various mixtures. (a) $\text{H}_a - -\text{N}_c$ site-site interactions between $[\text{C}_2\text{mim}]$ and $[\text{DCA}]$. (b) $\text{H}_w - -\text{N}_c$ site-site interaction between $[\text{C}_2\text{mim}]$ and $[\text{DCA}]$.

We also analyzed the hydrogen bonding interactions between the most negatively charged oxygen atom in $[\text{NTf}_2]^-$ and the H_a and H_w sites in the $[\text{C}_2\text{mim}]^+$ cation. The radial distributions functions for the H_a site are shown in Figure 5.5, which reveal that the ordering of the first peak intensity for the various systems is dependent on the type of the mixture under consideration. As seen for the hydrogen bonding interaction of $[\text{DCA}]^-$ with H_a , there is a rise in the intensity when a longer alkyl chain cation is added to the system. However, the ionic liquid mixture in which $[\text{DCA}]^-$ is present, there is depression in the first peak intensity in the RDF of H_a with O relative to that for the pure system. This is in direct contrast to the behavior observed in the RDF of H_a with N_c of $[\text{DCA}]^-$, confirming the preferential interaction of $[\text{DCA}]^-$ with the hydrogen bonding site. The first peak height is elevated when the reciprocal mixture is analyzed; however, the magnitude is still smaller than that obtained in the pure system. The RDF's between H_w and O mirror the trends for the H_a and O RDF's with a decrease in the first peak height for respective systems.

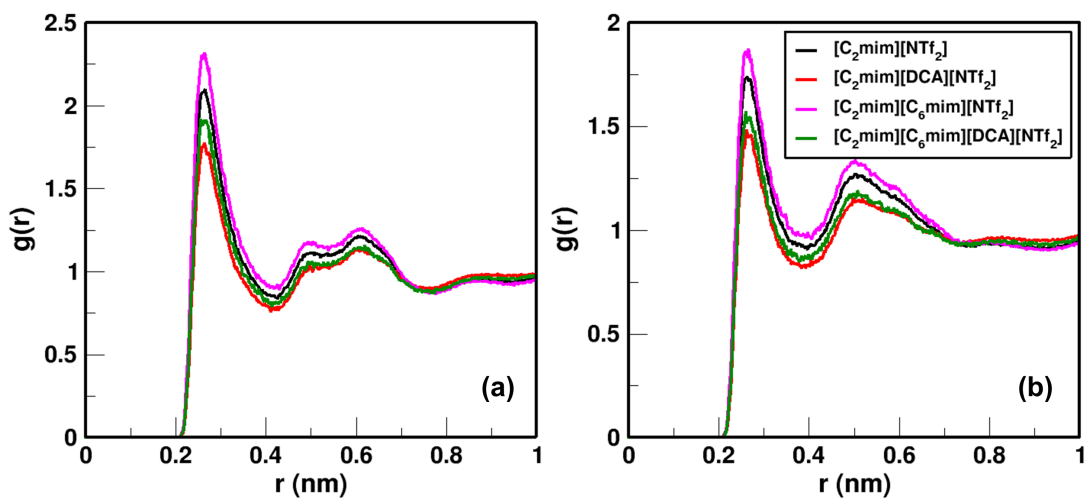


Figure 5.5: RDF plots for site-site interaction in pure and various mixtures. (a) H_a - $-\text{O}$ site-site interactions between $[\text{C}_2\text{mim}][\text{NTf}_2]$. (b) H_w - $-\text{O}$ site-site interaction between $[\text{C}_2\text{mim}][\text{NTf}_2]$.

5.6 Hydrogen Bonding Lifetimes

To understand how this enhancement in interaction for the hydrogen bonding sites affects the dynamics of the system, the lifetime of hydrogen bond dynamics between the ion pairs for all the systems were calculated using the intermittent hydrogen bonding autocorrelation function (Eq. 5.6) calculated over the first 1 ns of the second *NPT* production run using TRAVIS [244]

$$C(t) = \frac{\langle h(0)h(t) \rangle}{\langle h(0) \rangle} \quad (5.6)$$

where $h(t)$ takes a value of 1 if the hydrogen bond is persistent, but assumes a value of 0 when the hydrogen bond is broken. We defined the hydrogen bond following the criteria laid out in the work by Kirchner et al. [237,245]. The cutoff angle between donor-acceptor-hydrogen was set to 30° , the cutoff distance between the acceptor-donor was set at the distance corresponding to the first minimum in the center of mass RDF between the cation-anion, and the cut-off distance for the acceptor-hydrogen atom was limited to the first solvation shell of the hydrogen bonding sites. Further details are provided in the supporting information (Figure B.9, B.10, B.11, B.12). The autocorrelation was calculated for the $C_a-H_a-N_c$ and C_a-H_a-O interactions, where C_a and H_a atom belongs to $[C_2mim]^+/[C_6mim]^+$ cation, N_c is the central nitrogen atom in $[DCA]^-$ anion and O is the oxygen atom in the $[NTf_2]^-$ anion.

The average hydrogen bonding lifetime between a given pair of atoms was calculated by integrating the hydrogen bonding autocorrelation function

$$\tau_{lt} = \int_0^\infty C(t)dt \quad (5.7)$$

This integral was evaluated analytically by fitting a stretched exponential equation

$$C(t) = a_1e^{-t/b_1} + a_2e^{-t/b_2} + a_3e^{-t/b_3} \quad (5.8)$$

where, $a_3 = 1 - a_1 - a_2$ yielding

$$\tau_{lt} = a_1b_1 + a_2b_2 + a_3b_3. \quad (5.9)$$

Table 5.2: Average lifetime τ_{it} of hydrogen bonds. C_{a2} and H_{a2} atom are attached to the $[C_2mim]^+$ cation, C_{a6} and H_{a6} atom are attached to the $[C_6mim]^+$ cation, N_c is attached to the $[DCA]^-$ anion and O is attached to the $[NTf_2]^-$ anion.

System	$C_{a2} - -H_{a2} - -N_c$	$C_{a2} - -H_{a2} - -O$	$C_{a6} - -H_{a6} - -N_c$	$C_{a6} - -H_{a6} - -O$
$[C_2mim][DCA]$	42.2±5.7			
$[C_2mim][NTf_2]$		29.2±1.5		
$[C_6mim][DCA]$			88.7±6.1	
$[C_6mim][NTf_2]$				50.1±6.4
$[C_2mim][C_6mim][DCA]$	57.6±10.2		66.9±0.3	
$[C_2mim][C_6mim][NTf_2]$		38.8±7.7		37.3±0.9
$[C_2mim][DCA][NTf_2]$	55.9±5.3	26.5±3.5		
$[C_6mim][DCA][NTf_2]$			109.4±3.6	49.6±4.4
$[C_2mim][C_6mim][DCA][NTf_2]$	73.9±11.9	34.6±4.9	95.1±16.8	37.9±2.5

The hydrogen bonding lifetime values obtained from the analysis are included in Table 5.2. For the pure ILs, we observe an increase in the hydrogen bonding lifetime as the alkyl chain in the imidazolium cation is elongated for a given anion, attributable to the formation of the non-polar domain confining the movement of the anion to the polar domain. The trend is in similar agreement with the work of Kircher et al [245]. where the authors found the hydrogen bonding lifetime for the same hydrogen bonding sites of $[C_4mim][NTf_2]$ at 323.0 K using CL&P forcefields to be 35.4 ps which is in between the lifetime of $[C_2mim][NTf_2]$ and $[C_6mim][NTf_2]$ as seen from the table. Reddy et al. saw a similar trend in the hydrogen bonding lifetime where it increased by a factor of 14 as the alkyl length of the primary ammonium cation paired with formate anion increased from methyl chain to propyl chain [246].

The increase in the lifetime for hydrogen bonding interaction is primarily dominated by the identity of the anion. For instance. the ratio of the average lifetime for $[C_6mim][DCA]$ is ~ 2 times the lifetime of $[C_2mim][DCA]$, while the ratio drops to ~ 1.7 in going from $[C_2mim][NTf_2]$ to $[C_6mim][NTf_2]$. The hydrogen bonding lifetime ratio is also similar to the change in viscosity for a given ionic liquid series. Experimental viscosity of $[C_2mim][DCA]$ [247] at 323.15 K is found to 8.51 cp com-

pared to 18.72 cp for $[C_6mim][DCA]$ [248] which is roughly 2.20 times the increase in viscosity as the alkyl chain length of the cation is increased. Similarly, the viscosity of $[C_2mim][NTf_2]$ [249] at 323.15 K which is 15.61 cp that only increases to 26.25 cp for $[C_6mim][NTf_2]$ [250] which is roughly a factor of 1.68 almost identical to the ratio of the hydrogen bonding lifetime for these systems. Our results are consistent with the findings by Zhang and Maginn that the transport properties such as self-diffusion coefficients and the ionic conductivity are inversely related to the ion-pair correlation timescales [251]. For mixtures, we notice that the hydrogen bonding lifetimes increase in accordance to the increase in the height of the RDF's between H_a and N_c for $[C_2mim][DCA]$ and H_a and O for $[C_2mim][NTf_2]$ as illustrated in Figure 5.6.

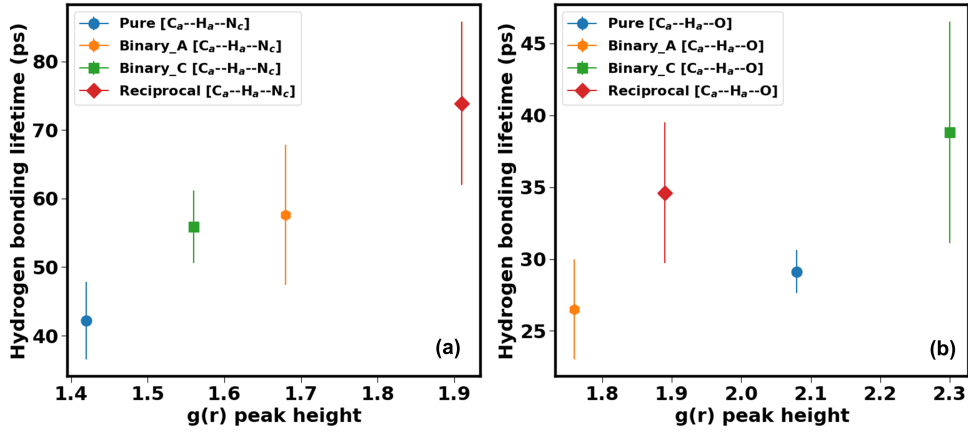


Figure 5.6: Correlation between hydrogen bonding lifetime and RDF peak height of the first solvation shell of hydrogen bonding sites. (a) $[C_2mim][DCA]$. (b) $[C_2mim][NTf_2]$. Binary_A refers to $[C_2mim][DCA][NTf_2]$ system, Binary_C refers to $[C_2mim][C_6mim][DCA]$ or $[C_2mim][C_6mim][NTf_2]$ system and Reciprocal refers to $[C_2mim][C_6mim][DCA][NTf_2]$ system.

5.6.1 Self Diffusion Constant

The self-diffusion constant of the ions was obtained by fitting the linear region of the mean square displacement (MSD) using the following equation.

$$D = \frac{1}{6} \lim_{t \rightarrow \infty} \frac{d}{dt} \left\langle \sum_{i=1}^N [\vec{r}_i(t) - \vec{r}_i(0)]^2 \right\rangle \quad (5.10)$$

where, $\vec{r}_i(t)$ is the position of an ion at any given time, $\langle \dots \rangle$ is the average of the ensemble, and D is the self-diffusion constant. To determine the linear region for the fit, the MSD values were divided into three blocks spanning 0-10 ns, 10-15 ns and 15-20 ns for the final *NPT* production run. Over these blocks, we computed the non-Gaussian parameter $\beta(t)$:

$$\beta(t) = \frac{d \ln \langle \Delta r^2(t) \rangle}{d \ln(t)}. \quad (5.11)$$

A linear regime is indicated as $\beta(t)$ approaches a value of unity, which, for most systems occurred over the 10-15 ns block, as assessed from the slope of the log-log plot MSD versus time (please refer to Figures B.13 - B.16) The reported values of self-diffusion coefficients were calculated by tracking the MSD over this timeframe.

Table 5.3 summarizes the average self-diffusion constant of ions for pure, binary, and reciprocal mixtures at 323.0 K obtained from this study along with a comparison of self-diffusion constant data obtained from MD simulation in literature and NMR diffusion experiments. The self-diffusion constant for the cation and anion in pure [C₂mim][DCA] show an interesting trend where the self-diffusion coefficient of the anion is greater than that of the cation, a trend that is opposite for many imidazolium-based ILs [252]. This is primarily because of the smaller size and planar shape of the [DCA]⁻ anion that allows it to easily diffuse through the ionic network. Goloviznina et al. reported a similar trend for [C₂mim][DCA] [236]. The authors employed ± 0.8 charge scaling as in this work, which results in a considerable speeding up of the dynamics and an improvement in the self-diffusion coefficients compared to ± 1.0 scaling for the electrostatic interactions. The self-diffusion coefficients for both the cation and anion in [C₆mim][DCA] drop by a

Table 5.3: Self Diffusion constant ($D \times 10^{-7}$ cm²/sec) of ions for pure, binary and reciprocal mixtures at 323 K.

System	Self Diffusion constant			
	[C ₂ mim] ⁺	[C ₆ mim] ⁺	[DCA] ⁻	[NTf ₂] ⁻
[C ₂ mim][DCA]	10.64±0.67[12.8] ^a		12.29±0.14[14.6] ^a	
[C ₂ mim][NTf ₂]	7.85±1.61[10.6] ^c			6.22±1.56[6.91] ^c
[C ₆ mim][DCA]		3.75±0.32[8.4] ^b	4.92±0.30[9.6] ^b	
[C ₆ mim][NTf ₂]		4.48±1.20[4.64] ^c		3.90±0.93[4.31] ^c
[C ₂ mim][C ₆ mim][DCA]	7.84±0.54	5.38±0.02	7.72±0.41	
[C ₂ mim][C ₆ mim][NTf ₂]	6.77±1.55	5.3±1.48		4.59±1.08
[C ₂ mim][DCA][NTf ₂]	8.63±0.57		10.24±1.02	6.89±0.77
[C ₆ mim][DCA][NTf ₂]		3.28±0.19	4.07±0.46	2.67±0.12
[C ₂ mim][C ₆ mim][DCA][NTf ₂]	6.14±0.60	4.83±0.29	6.37±0.55	4.10±0.15

^a D data obtained using MD simulation at 303 K. [236]

^b D data obtained using MD simulation at 333 K. [236]

^c D VFT equation from Tokuda et al. [252]

factor of 2.4 compared to the self-diffusion coefficients of ions in [C₂mim][DCA]. The reduction in the self-diffusion coefficient is similar to the ratio of the viscosity of [C₂mim][DCA] to that of [C₆mim][DCA] and is in line with change in the hydrogen bonding lifetimes for these ILs. Switching the anion from [DCA]⁻ to [NTf₂]⁻ leads to a decrease in ionic self-diffusion coefficients despite the hydrogen bonding lifetime being smaller than that for [C₂mim][DCA], which suggests that additional factors such as the bulkier size of the anion and a propensity for anion- π interactions also contribute to determining dynamics in [C₂mim][NTf₂]. Consistent with findings in the literature, results from our simulations indicate that the cation diffuses faster than the anion in the pure ionic liquid systems containing [NTf₂]⁻ as the anion. Similar to the [DCA]⁻ systems, substituting [C₂mim]⁺ cation with [C₆mim]⁺ causes slowing down of the dynamics for both the ions. The calculated self-diffusion coefficients for the ions in [C₂mim][NTf₂] and [C₆mim][NTf₂] agree reasonably well with those computed from the parameters of the VFT equation derived by Tokuda et al. [252] For binary and reciprocal mixtures, we see that the species exhibiting higher self-diffusion coefficient slows down, while the self-

diffusion coefficients of the slower moving ions either remains nearly identical or increases with respect to their respective values for the pure ILs, e.g., the self-diffusion coefficients of $[\text{C}_2\text{mim}]^+$ and $[\text{DCA}]^-$ are lower, while that of $[\text{C}_6\text{mim}]^+$ enhances in $[\text{C}_2\text{mim}][\text{C}_6\text{mim}][\text{DCA}]$ mixture relative to those in pure systems. A similar observation can be made for the $[\text{C}_2\text{mim}][\text{DCA}][\text{NTf}_2]$ system. We also calculated the overall diffusion coefficients for the pure, binary, and reciprocal mixtures using the Eqs. 5.12–5.15 to quantify the dynamics:

$$D = \frac{D_+ + D_-}{2} \quad (5.12)$$

$$D = \frac{D_+ + x * (D_{1-} + D_{2-})}{2} \quad (5.13)$$

$$D = \frac{x * (D_{1+} + D_{2+}) + D_-}{2} \quad (5.14)$$

$$D = \frac{x * (D_{1+} + D_{2+} + D_{1-} + D_{2-})}{2} \quad (5.15)$$

The results for the overall diffusion constant of the system are displayed in Figure 5.7. As expected from the individual ions self-diffusion coefficients, $[\text{C}_2\text{mim}][\text{DCA}]$ has the fastest dynamics compared to all the systems. This is again because of the pairing between short alkyl chain cation and a smaller size anion that does not hinder the movement of ions leading to faster diffusion. Changing the $[\text{DCA}]^-$ anion to bulkier $[\text{NTf}_2]^-$ however reduces the overall diffusion of the system significantly. The effect of replacing the $[\text{C}_2\text{mim}]^+$ cation with $[\text{C}_6\text{mim}]^+$ has a more pronounced effect on the overall dynamics of the ionic liquid. For the binary ionic liquid mixtures, the collective dynamics falls in between the dynamics of the constituent ILs.

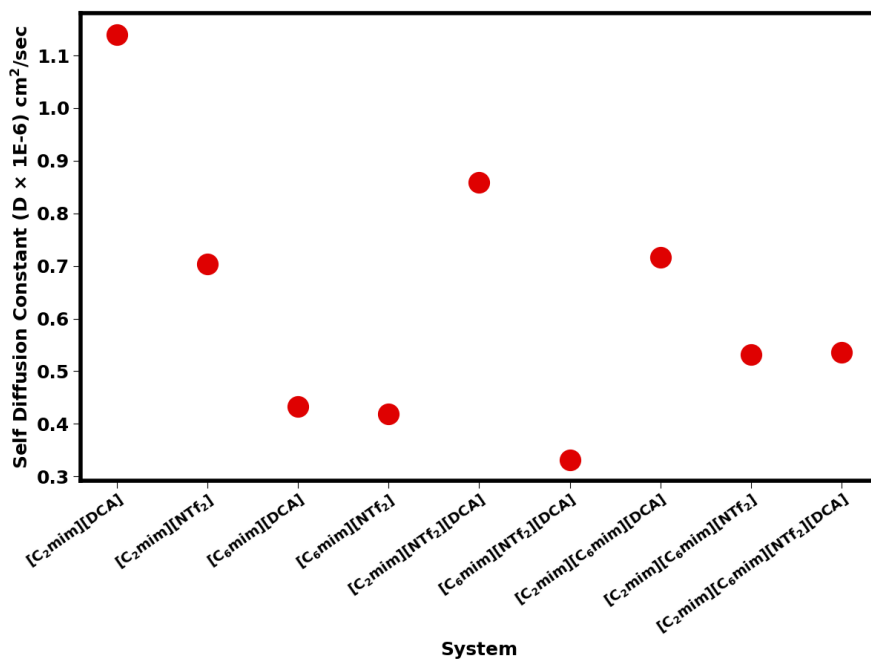


Figure 5.7: Self Diffusion constant of all the systems calculated using equation 5.12, 5.14, 5.13 and 5.15.

5.7 Conclusion

In this work, we performed MD simulations on various systems of pure ILs [C₂mim][DCA], [C₆mim][DCA], [C₂mim][NTf₂], [C₆mim][NTf₂], and equimolar binary and reciprocal mixtures resulting from them. We characterized these systems in terms of the ionic center-of-mass radial distribution functions (RDF), hydrogen bonding interactions and corresponding hydrogen bonding lifetimes, and self-diffusion coefficients. We observed that the RDF between [DCA]⁻ with [C₂mim]⁺ displayed the highest first peak intensity in the reciprocal mixture followed by that in the binary anion mixture, binary cation mixture, and pure ionic liquid system, indicating that the association of the [DCA]⁻ with [C₂mim]⁺ could be tuned through the addition of a cation, anion, or both to [C₂mim][DCA]. Further, evaluation of the hydrogen bonding site-site RDF interaction exhibited a similar enhancement in the first peak intensity, revealing the role of hydrogen bonding in preferential interaction. The first peak intensity in the center-of-mass RDF of [C₂mim]⁻[NTf₂]⁺ in the reciprocal mixture, however, was found to be less than that for the pure ionic

liquid $[\text{C}_2\text{mim}][\text{NTf}_2]$, providing additional evidence of the preferential interaction of $[\text{C}_2\text{mim}]^+$ with the anion possessing higher hydrogen bonding ability.

We also demonstrated that the preference for $[\text{DCA}]^-$ to associate with $[\text{C}_2\text{mim}]^+$ translates into an increase in the hydrogen bonding lifetimes in binary and reciprocal mixtures. In fact, the hydrogen bonding lifetime in the reciprocal mixture is almost twice that for the pure ionic liquid system. Although the hydrogen bonding lifetime for $[\text{C}_2\text{mim}][\text{NTf}_2]$ also increases when mixtures are considered, the enhancement is non-monotonic. The hydrogen bonding lifetime for the $[\text{C}_2\text{mim}][\text{NTf}_2]$ interaction stayed relatively the same or reduced in the presence of $[\text{DCA}]^-$ anion, suggesting that this preferential interaction only occurs for the short alkyl chain cation and the strongly coordinating anion. The overall dynamics of the various systems, as analyzed from the diffusion coefficients, revealed trends similar to the hydrogen bonding lifetimes, with the exception of pure $[\text{C}_2\text{mim}][\text{DCA}]$ and $[\text{C}_2\text{mim}][\text{NTf}_2]$ systems, which was explained in terms of the ability of $[\text{NTf}_2]^-$ to participate in additional modes of interaction with the cation.

Our study provides molecular-level details on the emergence of preferential ionic interactions and their implications on the hydrogen bonding between ion pairs in the binary and reciprocal mixtures, opening up possibilities for inducing non-idealities in such systems. From a practical point of view, as hydrogen bonding interactions are intimately connected to physicochemical properties of ILs, the present work exemplifies a pathway for designing task-specific ionic liquid mixtures by combining cations with differing in the length of the alkyl chain length and/or anions possessing widely different hydrogen bonding abilities.

CHAPTER 6

A Generalized Machine Learning Model for Predicting Ionic Conductivity for Ionic Liquids'

6.1 Abstract

Ionic Liquids (ILs) are currently being considered as potential electrolyte candidates for next-generation batteries and energy storage devices due to their high thermal and chemical stability. However, high viscosity and low conductivity at lower temperatures have severely hampered their commercial applications. To overcome these challenges, it is necessary to develop structure-property models for ionic liquid transport properties to guide the ionic liquid design. This work expands our previous effort in developing a machine learning model on imidazolium-based ILs to now include ten different cation families, representing structural and chemical diversity. The model dataset contains 2869 ionic conductivity values over a temperature range of 238-472 K collected from the NIST ILThermo database and literature values for 397 unique ILs. The database covers 214 unique cations and 68 unique anions. Three machine learning models, multiple linear regression, random forest, and extreme gradient boosting, are applied to correlate the ionic liquid conductivity data with cation and anion features. Shapely additive analysis is performed to glean insights into cation and anion features with significant impact on ionic conductivity. Finally, the extreme gradient boosting model is used to predict ionic conductivity of ILs from all the possible combinations of unique cations and anions to identify ILs crossing the ionic conductivity threshold of 2.0 S/m.

6.2 Introduction

The asymmetric cationic structures and articulated nature of anion are responsible for charge delocalization and frustrated crystal packing for a large number of ILs leading many to exist as liquid at ambient conditions. In contrast to conventional solvents, ILs offer several unique and desirable properties such as negligible vapor pressure, low melting point and nonflammability. These attributes are primary reasons ILs are studied extensively for various industrial applications such as solvents in chemical separation/purification [157,253], as catalysts [254,255], use in CO₂ capture [256,257] and potential electrolytes for battery application [11,165].

The use of ILs for battery applications and energy storage medium is primarily due to their high thermal [258] and chemical stability [259] to address tremendous safety concern associated with the current state-of-the-art electrolytes found in Li-ion batteries [260–262]. For example, current electrolytes powering Li-ion batteries are carbonate-based electrolytes mixed with salts such as lithium hexafluorophosphate LiPF₆, which are very volatile, flammable, and potentially hazardous during thermal runaway reactions or short-circuit [263,264]. Kalhoff et al. [6] carried out an extensive study on the performance and safety of electrolytes based on organic carbonates (OC) and ILs among others. The authors noted the superiority of OC electrolytes in terms of ionic conductivity; however, the performance of OC was poor for electrochemical and thermal stability. Additionally, these solvents posed safety concerns. On the other hand, ILs received a high rating for electrochemical and thermal stability, and safety consideration, but only medium for ionic conductivity, and suffered from poor low-temperature performance. Thus, for ILs to be considered potential electrolyte candidates, an improvement in low ionic conductivity performance at sub-ambient conditions is needed in the next-generation of ILs.

As is common for almost all applications involving ILs, a systematic improvement in the transport properties of ions can be accomplished by selecting an op-

timal cation-anion combination using chemical intuition. The approach, however, is likely to be slow and time consuming due to the staggering number of such possible combinations [174] in the range of 10^{14} . The presence of a myriad of interactions such as electrostatic, hydrogen bonding, π - π stacking, anion- π , and van der Waals further complicates choosing cation-anion pairing to deliver anticipated property enhancement. For example, the attempt to alter the hydrogen bonding interactions through alkyl substitution of the most acidic hydrogen site in the imidazolium cation led to an increase in the viscosity of the resulting ionic liquid - a counterintuitive result [13]. An experimental high throughput screening approach may also not be feasible due to the requirement for ensuring the purity of the synthesized ILs. Similarly, molecular simulation-based techniques such as molecular dynamics and Monte Carlo simulations can, in principle, accelerate the search of ILs with desired properties; however, describing interactions between ionic liquid components continues to be a nontrivial task. Given these challenges and the availability of the ionic liquid property database - ILThermo - maintained by NIST, machine learning-based methods are gaining attention as a pre-screening tool to correlate ionic liquid properties with attributes that describe cations and anions [105, 265–267]. Genetic mutation and generative-based models also allow the accelerated discovery of ILs with properties within desired range [118, 119].

In our previous proof-of-concept article, [267] we focused on modeling ionic conductivity using an artificial neural network and support vector regression models for imidazolium-based ILs as these ILs are generally less viscous and possess high ionic conductivity at room temperature - key properties for battery electrolytes [11]. Additionally, a large amount of data is available for imidazolium-based ILs enabling machine learning model development. One of the difficulties of using imidazolium-based ILs is that the electrochemical stability of imidazolium cations is rather low - less than 4.0 V (vs. Li/Li⁺) - which is not suitable for high voltage battery application [268]. The primary reason for this behavior is the susceptibility of the cation to reduction at the most acidic proton at the C₂ position. Protecting this position by substituting various functional groups improves the stability but leads

to slower dynamics [13] in comparison to that for the parent ionic liquid. The next closest relative to imidazolium cations are the pyridinium-based cations that are more sluggish with high viscosity and low ionic conductivity, which is why there is a limited amount of study done on exploring its application as electrolytes for battery application [14–16]. Beyond the aromatic cations, cyclic cations such as pyrrolidinium and piperidinium cations have generated tremendous interest as they have a high biodegradability rate and low toxicity [18,19]. The pyrrolidinium cation also offers low viscosity and high ionic conductivity, and unlike imidazolium cations, are more electrochemically stable, with a majority of them exhibiting electrochemical window reaching above 4.5 - 5.0 V [11]. Along with faster dynamics, pyrrolidinium cations also have better stability towards lithium metal, making them an ideal candidate for battery application as potential electrolytes [20,21].

Modifying the five-ring pyrrolidinium structure to a six-ring structure gives rise to piperidinium cations. Similar to pyridinium cations, piperidinium cations have slower dynamics than pyrrolidinium cations because of the bulky nature of the cation. As such, there are relatively few studies that have explored the possibility of piperidinium cations as electrolytes for battery application [22–24]. Besides cyclic and aromatic cations, other central atom-based cations such as tetralkylphosphonium, tetraalkylammonium, and trialkylsulfonium, are also extensively studied for various applications [25–27]. The ammonium-based cations are characterized by a high electrochemical window compared to imidazolium but suffer from high viscosity and low ionic conductivity [28]. An alternative to nitrogen-based cations is phosphonium-based cations that have similar properties as ammonium ILs, with some of them outperforming ammonium cations [26,29]. The other common cation type is sulfonium-based ILs which have favorable properties compared to phosphonium-based cations because of the small volume occupied by the core sulfur atom leading to lower viscosity and high ionic conductivity [32–34]. In addition to the commonly studied cations, there are several other cation types such as morpholinium [23,24,269], pyrazolium [270,271], oxazolidinium [272] which might offer desirable properties for battery application but there is very limited informa-

tion on the physicochemical properties of these ILs in the literature.

Given the availability of ionic conductivity data for ILs belonging to a large variety of different cation types, it is conceivable to find an ionic liquid with high ionic conductivity, if an accurate structure-property relationship is uncovered. With this objective, the present article focuses on developing machine learning models capable of predicting ionic conductivity covering various cation families and anions with high accuracy. Additionally, important features contributing to the ionic conductivity have been identified using shapely additive (SHAP) analysis technique. The insight is used to develop a classification model to categorize cations that are likely to yield ILs, with a given anion, into high/low ionic conductivity. Lastly, ionic conductivity for all possible pairings of the cation and anion are predicted to identify ILs possessing high ionic conductivity.

6.3 Methodology

6.3.1 Data collection and processing

In this study, we developed machine learning models trained on experimental ionic conductivity data primarily obtained from NIST ILThermo Database [73, 74]. We supplemented the data extracted from the ILThermo database with data collected from various sources found in literature [36, 40, 75–94, 273]. This led to a total of 4786 data points covering ten different cation types as seen in Figure C.1. Data download, data cleaning, duplicate removal, and conversion of chemical structures to SMILES convention followed a similar approach outlined in our previous study [267]. The state property filter was set between 95–110 kPa, eliminating some of the very high-pressure data, while no restrictions were imposed on the temperature. The temperatures and pressure were selected considering that ionic liquid-based batteries would be operated over a wide range of temperatures and close to atmospheric pressure. The final dataset contained 2869 data points, 397 unique ILs, 214 unique cations, and 68 unique anions ranging from 238 K to 472 K covering ionic conductivity from 10^{-5} S/m to 19.3 S/m, spanning six orders of

magnitude.

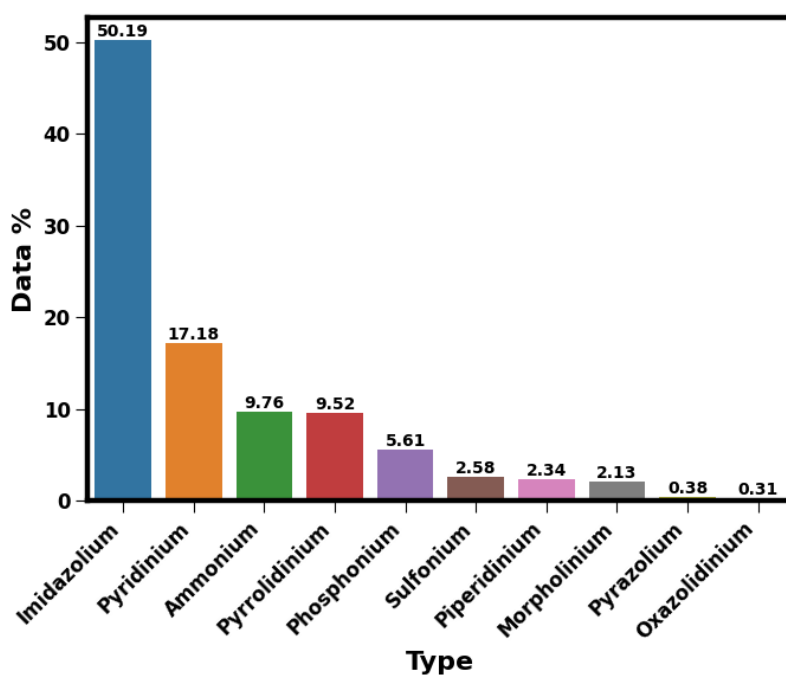


Figure 6.1: Experimental ionic conductivity data distribution % by cation type for the model development data set after data cleaning. The percentage for each of the cation family is calculated using the number of data points collected in Table C.1.

The percentage distribution of the individual cation type in the model development data set is depicted in Figure 6.1. As expected majority of the data belongs to the imidazolium family as it is one of the widely studied cations. Additional 40% of the data points are contributed by pyridinium, ammonium, pyrrolidinium, and phosphonium cations. One concern with such skewed data distribution would be the bias in prediction towards the imidazolium data set due to the relative abundance of the ionic conductivity data for this cation family. In a later section, we discuss our approach to systematically evaluate the model's prediction by cation type to evaluate such bias in prediction.

6.3.2 Feature generation and processing

Features for the cations and anions are generated using open-source cheminformatics RDKit package [143] that produced 196 unique features each for cation and anion. Temperature and pressure were included as additional features which led to a total count of features to 394. Some of the features, however, were not essential for model development as they were assigned a value of zero for all the cations and anions. Besides, a high-dimensional feature space could lead to overfitting of the data, resulting in a poor performance of the model for test data [274]. To avoid such issues, we first reduced the number of features by eliminating features exhibiting high correlations. A further reduction in the dimensionality of the feature space was achieved through the Least absolute shrinkage and selection operator (LASSO) [176,275] algorithm. Lasso is a regularization technique that is used to shrink the dimensionality of the feature space by adding a penalty parameter λ to the minimization function that denotes the amount of feature shrinkage (eq. 6.1). Larger values of λ parameter lead to the coefficients of the features that are of less importance to zero, thereby reducing the number of features necessary for a model, while the minimization function is recovered for $\lambda = 0$.

$$\text{Obj} = \sum_{i=1}^n (y_i - \sum_j x_{ij} w_j)^2 + \lambda \sum_{j=1}^p w_j \quad (6.1)$$

The hyperparameter λ was determined using 5-fold cross-validation (CV) technique by fitting a linear regression model with a $\log \lambda$ in the range of [-6, 50]. Based on the CV, the optimum value ($\log \lambda = -5$) helped reduce the number of features to 51 cations, 47 anion features leading to a total of 100 features including temperature and pressure.

6.3.3 Model Development

For the model development, the data set was split into 90% training set, while the remaining 10% of the data was set aside as a test case. The input features and the ionic liquid conductivity data were normalized to fall within the range of [0,1] using MinMax scaling implemented in Scikit-learn [144]. The ionic conductivity

values were represented on a log₁₀ scale before scaling as the values spanned six orders of magnitude. Three different models (Figure 6.2) were developed to correlate the ionic conductivity data: multiple linear regression (MLR), random forest (RF), and extreme gradient boosting (XGBoost). The rationale for choosing these models compared to the widely popular neural network was to offer insights into the importance of individual features.

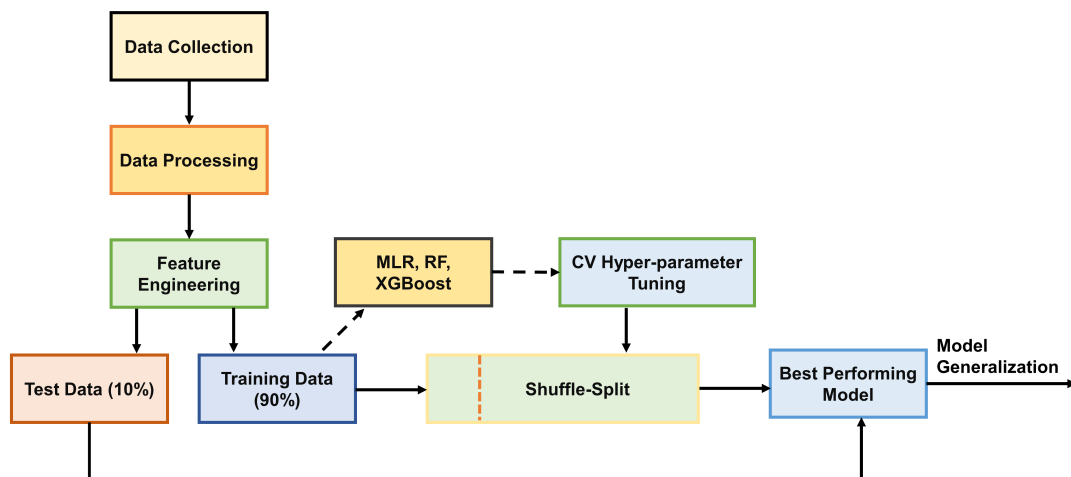


Figure 6.2: Workflow for developing machine learning models for correlating ionic conductivity of ILs.

6.3.4 Multiple Linear Regression

Correlation of ionic conductivity is first attempted using the multiple linear regression (MLR) model as it is the simplest form of regression method. In an MLR model, the structure-property relationship is expressed as a linear combination of features x_i (eq. 6.2)

$$y_p = b + w_1x_1 + w_2x_2 \dots + w_nx_p. \quad (6.2)$$

where b is the bias in the model, and w_i corresponds to the weight of feature x_i , which are determined by minimizing the least square error between the model prediction and the labels. Note that the model contains p features.

6.3.5 Random Forest

Random forest (RF) is a supervised machine learning method based on ensemble learning technique similar to decision-tree (DT) method [144]. However, unlike the DT method, for which outputs are generated using a single tree, RF methodology consists of multiple decision trees, which are generated in parallel, in an effort to reduce the possibility of overfitting and minimizing any bias towards feature selection. A sample from the training set is drawn at random with replacement to initiate a given tree. The final prediction is the ensemble average of the outputs predicted by individual trees. The number of trees or estimators and the depth of the trees are hyper-parameters of the model, which were determined using randomized cross-validation (RandomizedCV) [144]. In this approach, a random combination of hyperparameters is evaluated using 5-fold CV. Figure C.2 provides the grids used for the number of trees and estimators.

6.3.6 XGBoost

Extreme gradient boosting (XGBoost) is a decision tree-based ensemble method similar to RF that uses a gradient boosting algorithm [276]. However, unlike RF, where individual trees are formed in parallel, XGBoost consists of a series of trees built iteratively. The model starts with weak learners that are intentionally added to make a significant error which gets added to the loss function of the subsequent tree using a gradient descent algorithm. The objective of the XGBoost function is to minimize the loss as each tree is added until the accuracy no longer improves. The hyperparameters for XGBoost are determined using RandomizedCV using 5-fold CV [144]. The final set of hyperparameters for the XGBoost model is listed in the supporting information (Figure C.3).

6.3.7 Cross Validation and Model Evaluation

In most cases, performance evaluation of a model on the test data is enough to assess the ability of the model to generalize on out-of-bag samples. However, it is not always guaranteed that the models will generalize, especially when there is an over-

abundance of one or more types of data. For instance, in this study, approximately half of the data is represented by the imidazolium family (Figure 6.1). Even with a proper random shuffle of train and test split, most of the data in the training set and test set could belong to the imidazolium cation, leading to high accuracy on the training set and test set, which might not reflect the model’s inherent ability to generalize beyond the imidazolium-based ILs. Another challenge with the present data is that there may be ILs for which there is data at multiple temperatures, while for others, the ionic liquid conductivity is reported at only one temperature; the cation 1-ethyl-3-methylimidazolium or the anion bis(trifluoromethylsulfonyl)imide anion are such examples as they are usually studied over a wide range of temperatures. Despite the train/test split, most of the data could be for the same ionic liquid, simply reflecting the ability of the model to scale ionic conductivity as a function of temperature, not truly reflecting the underlying structure-property relationship.

For this work, we initially started with a 90:10 train/test split, where the test set data was never exposed during the model development. Next, we employed the shuffle-split technique by dividing the training set data into 90:10 for model development and validation data, respectively, repeated 100 times. The extensive sampling is expected to minimize any bias in the split for training and validation data and is likely to provide uncertainty in the prediction based on the data split. Performance metrics for the training and validation set were evaluated at each instance. The model with the best performance on the validation set was selected as the model for test set evaluation.

The predictive capability of the model was also determined for each of the cation types to examine overfitting of the model towards a certain class of the cation type. Finally, the performance of the model was evaluated on a separate external test set consisting of 30 ILs gathered from a literature review [36, 76, 82, 92, 94, 277–279]. The external test set contained unique ionic liquid combinations that are not present in the model data set. Here, unique ionic liquid combination refers to ILs with given cation and anion pairs that the model did not encounter during

the training phase. However, these cations and anions were present in the dataset paired with a different anion/cation. Furthermore, a few of the ILs in this test case are novel cation family types that did not belong to any of the ten cation families for which models were developed. As the chemical structures of the cations in these ILs resemble those in the model dataset, such performance evaluation would be informative to understand the extent to which these models can be generalized to cations families beyond those studied here.

6.4 Results and Discussion

6.4.1 Model Performance Metrics

In this work, we evaluate the correlation of ionic conductivity for ten different cation types, 214 unique cations, 68 unique anions with ionic conductivity data over a temperature range of 238-472 K using linear and non-linear machine learning methods. The accuracy and robustness of these models are evaluated using three different performance metrics: correlation coefficient (R^2), root-mean-squared-error (RMSE) and mean absolute error (MAE). The average and standard deviation of the performance metrics for the 100 shuffle-split is reported in Table 6.1. The model with the best performance on the validation becomes the model of choice for test set evaluation.

Among the three models, MLR has the lowest accuracy in correlating ionic conductivity compared to the other two models with low R^2 and high RMSE/MAE for training, validation, and test set. Increasing the complexity of the model that takes into account non-linear behavior in the model drastically increases the model performance as seen with the RF method, implying a non-linear correlation between features and ionic conductivity. A further improvement in the performance metrics can be observed with the XGBoost model, as the model is designed to iteratively learn and correct the errors incurred in the previous steps.

Correlations plots for MLR, RF, XGBoost are provided in the supporting information (Figure C.4, C.5 and C.6). Based on the trends in the figures, it is readily

Table 6.1: Average and standard deviation of the performance metrics for the training and validation set using MLR, RF, XGBoost model. RMSE is the root mean squared error, MAE is the mean absolute error, and R^2 is the correlation coefficient between experiment and predicted data. Shuffle-Split indicates random data shuffle into 100 different training/validation splits. The model with the best performance on the validation set during shuffle-split becomes the final choice of model for test set evaluation. Note: The RMSE and MAE values are for ionic conductivity in the \log_{10} scale.

Model	Data Set	Shuffle-Split			Best Performing		
		R^2	RMSE	MAE	R^2	RMSE	MAE
MLR	Training	0.877±0.002	0.260±0.003	0.173±0.002	0.873	0.265	0.175
	Validation	0.867±0.023	0.268±0.029	0.180±0.011	0.914	0.213	0.159
	Test	–	–	–	0.853	0.322	0.204
RF	Training	0.994±0.000	0.059±0.001	0.031±0.000	0.994	0.060	0.031
	Validation	0.956±0.013	0.152±0.026	0.083±0.007	0.977	0.116	0.074
	Test	–	–	–	0.963	0.161	0.087
XGBoost	Training	0.999±0.000	0.020±0.001	0.012±0.000	0.999	0.021	0.012
	Validation	0.977±0.011	0.109±0.026	0.050±0.006	0.993	0.061	0.039
	Test	–	–	–	0.987	0.094	0.047

apparent that the non-linear models significantly outperform the MLR model, similar to other studies that have examined correlation of ionic liquid properties using linear and non-linear approaches [280, 281].

We also examined the overall correlation coefficient (R^2) for the XGBoost and RF model for each of the cation families, the results of which are presented in Figure 6.3. It can be observed that the performance of the XGBoost model is somewhat independent of the type of the cation family, while the RF model is more sensitive to the type of cation and the corresponding number of data points.

For example, the R^2 value drops to 0.55 for morpholinium cation, despite being present in the training set, while the R^2 predicted using the XGBoost method is 0.9. Similarly, the R^2 value obtained with the RF model for oxazolidinium cations drops below 0.90, whereas the XGBoost model again yields R^2 values ~ 0.9 . For all other cations types, the correlation coefficients are nearly perfect as deduced from the XGBoost method, while those calculated from the RF model are lower, which shows that the XGBoost model can be accurate even when the data is limited (Figure 6.1). Based on the performance metrics by cation type, the XGBoost model is chosen as the choice of model for further prediction as it outperforms the RF model for individual cation types, which is essential for unique ionic liquid prediction discussed in the later section.

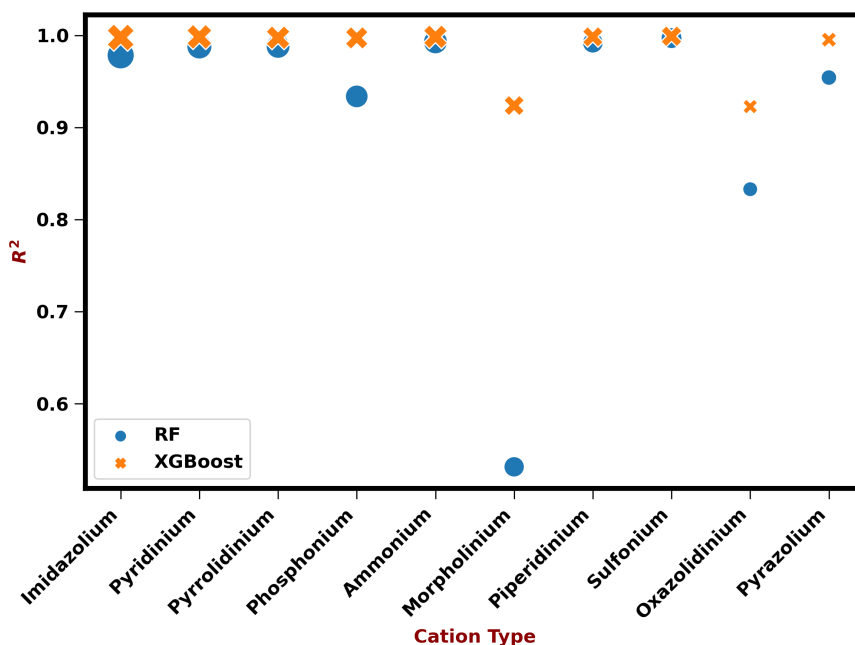


Figure 6.3: Correlation coefficient (R^2) for RF (circles) and XGBoost (crosses) models for the entire data set as a function of cation family. Size of the marker indicates the relative proportion of data points present for the given cation type.

We further tested the predictive capability of the XGBoost model for the external data, which were neither part of the training data or test data, consisting of 30 data points collected from the literature. This data set included 27 unique ILs with a temperature range between 293-323.15 K and ionic conductivity range of 0.06-1.68

S/m. The XGBoost model obtained an R^2 of 0.80, RMSE of 0.20 S/m, and MAE of 0.14 S/m for this external test set compared to experimental data. The entire prediction on this external test case using XGBoost along with experimental data, the source, along with schematics of the cations and anions are provided in Table C.2 (unique ionic liquid combinations), Table C.3 (cations structurally similar to those on which the model was trained) and Table C.4 (substituted imidazolium-based cations).

6.4.2 Model Interpretation

A significant advantage of ensemble-based models over 'black box' models such as neural network is the easy interpretability of the feature importance. This insight can be valuable for developing design heuristics for the search and development of new cations with high ionic conductivity. In this work, we employed the Shapley additive explanations (SHAP) [282] method, which provides a reliable way to explain the importance of features and the model decision making [283, 284]. As shown in Figure 6.4, the SHAP analysis ranks the features in terms of their importance, while the SHAP value indicates how varying a certain feature is likely to affect the output, ionic conductivity in the present case. A negative SHAP value suggests a lowering of conductivity, while a positive value implies an increase in conductivity.

Figure 6.4 presents nine features deemed most important for ionic liquid predictions. Out of these nine features, six features correspond to the cations, while two features are linked to the anions. Temperature was identified as the most dominant feature among these features. This is not surprising as the dynamics of ILs are accelerated with an increase in temperature. The second most important feature is the IPC descriptor for the cation that takes into account the information content of the molecule, such as the number of atoms through a graph representation [285]. Based on the SHAP value, it is clear that a high value of IPC (denoted by the red color) negatively impacts the output. As the IPC descriptor is related to the content of the molecule, higher values of the feature delineates bulky cations,

for example those containing long alkyl chains, slowing down the dynamics of the system [286, 287]. Furthermore, Figure C.10 (a) also shows the relation between the IPC descriptor of the cation and experimental ionic conductivity at 298.15 K, demonstrating the effect of the descriptor on the ionic conductivity; there is a general trend of decreasing ionic conductivity with IPC as revealed by the SHAP analysis. The next feature Chi0, that contributes to the ionic conductivity is also related to cations, capturing the nature of molecular connectivity [288]. The influence of Chi0 is similar to that identified for IPC in that it is negatively correlated to the ionic conductivity as confirmed in Figure C.10 (b). Besides these two descriptors, four more cation descriptors show very high importance in the model output based on SHAP value. These descriptors pertain to the topological connectivity (Balabanj, BertzCT), electronegativity (MaxAbsEStateindex), and the van der Waals surface area (VSA_EState8) of the cations (Figures C.10 (c)-(f)).

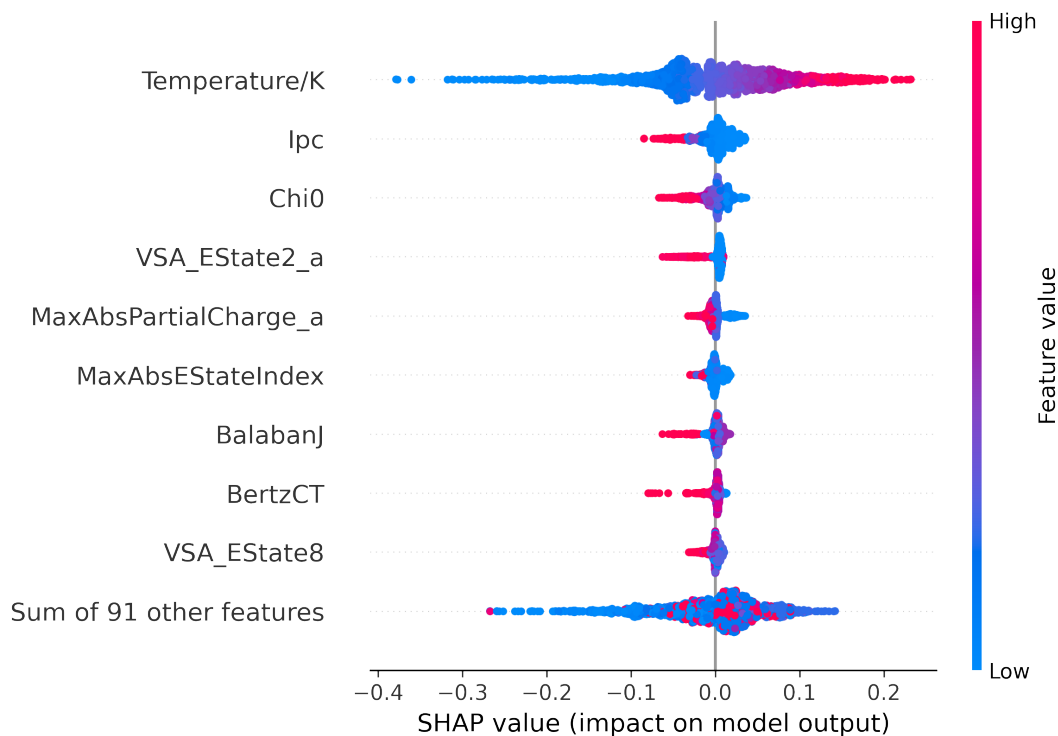


Figure 6.4: SHAP feature importance for the training set data. Features ending with 'a' indicates features for anions.

As for the two anions descriptors seen from Figure 6.4, VSA_EState2_a is the sum

of the electrotopological state index of an atom with van der Waals surface area between 4.78 - 5.0 [143]. The electrotopological descriptor encodes both the electronic and topological state of the anion [289]. The electronic state here refers to the electron distribution of the atoms in a molecule. Next, MaxAbsPartialCharge_a descriptor stands for the maximum absolute partial charge of the molecule calculated using the Gasteiger partial charge method based on electronegativity of the atoms in the molecule [290]. In the experimental data set, the highest value for this descriptor is for the halogen anions with a maximum partial charge of 1.0, followed by anions based on the phosphorous atom, oxygen-based anions, and cyano-based anions. The descriptor relation to ionic conductivity can be explained through the SHAP feature importance insight as the cyano group has the lowest MaxAbsPartialCharge_a, which results in higher ionic conductivity. In contrast, the halogen, phosphorous, and oxygen-based anions have the highest MaxAbsPartialCharge_a reducing ionic conductivity. Relation between these descriptors and ionic conductivity can be seen from Figure C.10 (g) and (h).

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. To examine how important these features are with respect to ionic conductivity, we also attempted to build a small-scale 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 by using a few selected features as inputs for the model's development. Further details on the classification model development and results can be found in the supporting information. The classification model is able to classify ILs in the high/low ionic conductivity categories with 98% accuracy for the training set, 92% for the test set, and 63% for the external test set. The accuracy is very high considering that the model is built only with six descriptors.

6.4.3 Unique ILs

As the XGBoost model is rigorously cross-validated for the test case and an external test case, we proceeded to combine all the unique cations (214) and anions

(68), significantly expanding the pool of ILs from mere 337 ILs to a staggering 14,552 unique ILs. Although impressive, it is important to note that not all the ILs generated from the combination may exist in a liquid state at 298.15 K, necessitating a separate model to estimate the melting point of these ILs. The XGBoost predictions for 14,552 unique ILs and experimental measurement for 337 ILs at 298 K are provided in Figure 6.5 for different cation families. As can be observed from the figure, the pyrrolidinium cation type has the highest experimental ionic conductivity followed by imidazolium, ammonium, pyrazolium, and pyridinium. The model predictions accurately capture this trend.

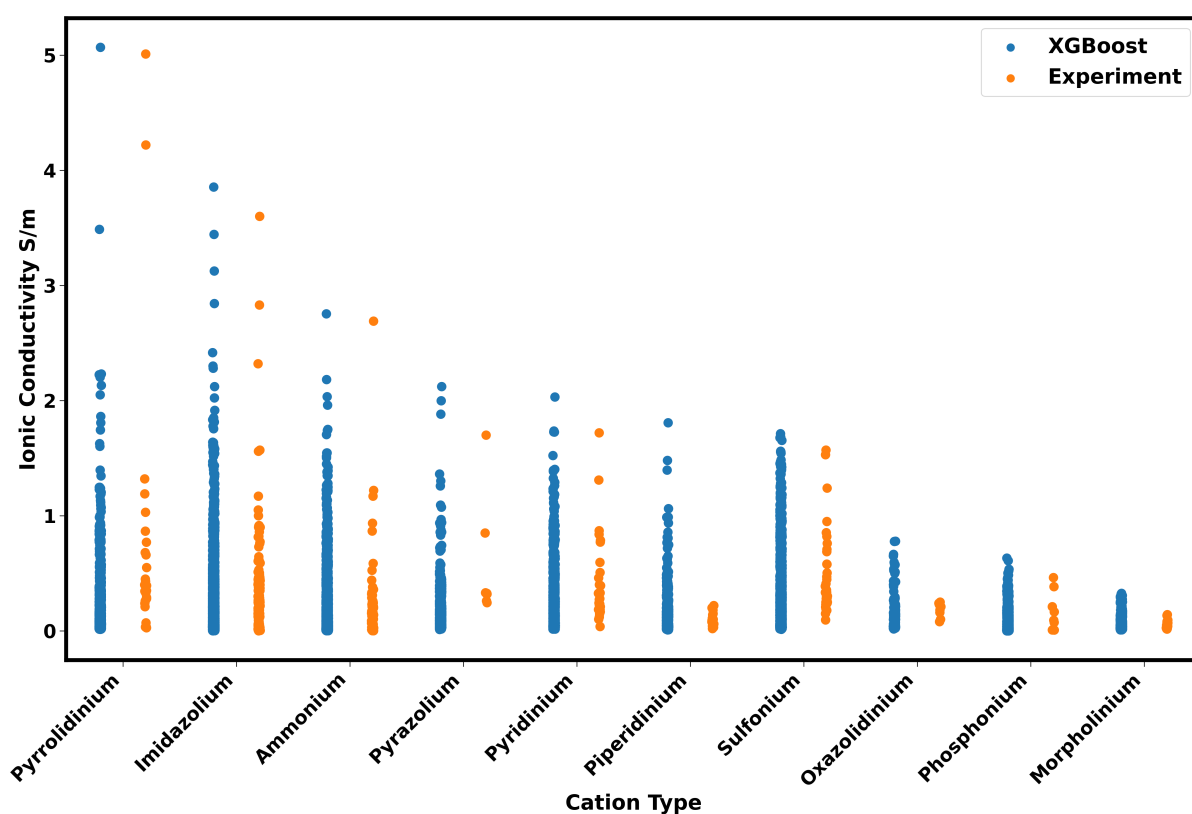


Figure 6.5: Categorical data of unique ILs based on cation family type at 298.15 K.

The current conventional electrolyte LP30 found in Li-ion batteries consists of 1 M LiPF_6 in 1:1 ethyl carbonate (EC) and dimethylcarbonate (DMC) mixture that is known to have an ionic conductivity of 1.26 at 298.15 K [71,291]. Thus for ILs to be considered as a potential electrolyte additive to replace LP30, the ionic conductivity

ity target should at least be close to 2.0 S/m as the addition of Li salts dramatically reduces the ionic conductivity and increases viscosity by 30-50% [96, 278, 292]. In our model development database, there are only five ILs with ionic conductivity greater than 2.0 S/m, which is now expanded to 21 ILs using the unique ionic liquid combination. This is possible as some of the cations for the ILs with experimental data higher than 2.0 S/m when combined with other cyano-based anions present in the data set lead to more ILs with high ionic conductivity. Breaking the unique ionic liquid combination analysis by cation type, there are just two cations beyond 2.0 S/m present in the experimental data set for the pyrrolidinium cations. Using the model, this region of space is now expanded to seven unique pyrrolidinium ILs. Similarly, the number of ILs beyond 2.0 S/m has grown from four to nine ILs for imidazolium cations. For the piperidinium experimentally, there is no data beyond 0.5 S/m at 298.15 K. That now has expanded to a large number of them crossing the 1.0 S/m as the piperidinium cations are paired with some other anions, mainly cyano based anions, as they can push ILs to have ionic conductivity. Based on the cations and anion combinations resulting from available experimental data, our analysis suggests that there are no high ionic conductivity ILs that can be formed using oxazolidinium, phosphonium, or morpholinium cations.

6.5 Conclusion

In search of ILs with high ionic conductivity for battery application, we developed three different machine learning models to correlate ionic conductivity of ten different cation types covering a temperature range of 238-472 K. It was found the multiple linear regression model was least accurate, while the non-linear model XGBoost performed the best. Although the accuracy of the model developed using RF methodology was similar to that for the XGBoost model, a degradation in its predictive capability was noted for cation families that represented a very small portion of the overall data set. On the other hand, the XGBoost model retained its high accuracy across all the cation families.

Feature importance based on SHAP analysis showed temperature, six cation features, and two anions features to have the most influence on ionic conductivity output. The insight gained from the SHAP analysis was used to develop a decision tree-based model containing only six cation features to classify ILs containing $[\text{NTf}_2]^-$ anion into two categories: high ionic conductivity and low ionic conductivity. The model showed a high accuracy, successfully classifying 92% of the ILs from the test set, demonstrating the usefulness of the SHAP analysis.

Lastly, all the unique cations and anions in the database were combined to dramatically expand the chemical space of ILs as demonstrated by the increase in the number of ILs from 337 to 14,552 unique ILs. The model predictions hint at 21 ILs possessing ionic conductivity greater than 2.0 S/m at 298.15 K. We envision that the large database of ionic liquid conductivity predictions can serve as a roadmap for future computational and experimental efforts in search for ILs with very high ionic conductivity suitable for battery application as electrolytes.

CHAPTER 7

Accelerated Discovery of Novel Ionic Liquid Cations using a Continuous Latent Space Representation of Chemical Space

7.1 Abstract

Ionic liquids have generated tremendous interest in the research community over the years due to several unique and desirable properties compared to conventional solvents. However, its slow transport properties and high material costs have limited its commercial application development. Due to the vastness of the ionic liquid chemical space, narrowing down the potential candidates using experiment or atomistic simulation with faster dynamics, high stability, and low material cost has proven difficult. Machine learning (ML) algorithms, on the other hand, could serve as an alternative tool for accelerated material discovery with desirable properties. This paper employs a generative-based deep machine learning method to discover new cations in the ionic liquid chemical space, thereby increasing the number of potential cation candidates for battery applications. According to our findings, some of the newly discovered cations belong to known cation families, whereas most cyclic and aromatic cations occupy a new region of chemical space. To assess their suitability as potential electrolytes, we calculated the electrochemical stability of these cations using DFT calculations. A few newly discovered cations have very high electrochemical stability, making them ideal for battery applications.

7.2 Introduction

Ionic liquids (ILs) are a class of salts composed entirely of ions and are often found to be liquid at a temperature below 100 °C. They have gained widespread interest because of their unique properties, such as negligible vapor pressure, low flammability, high thermal and chemical stability. Growing concerns over the toxicity and emission of volatile organic contaminants have made ILs a potential candidate for various industrial applications such as solvents for extraction [293, 294], CO₂ capture [56, 295], catalysis [296, 297], electrolytes for battery application [11, 298]. In addition, it also offers the possibility of designing and tuning the properties by combining the appropriate cation and anion combinations to form task-specific ILs. There is a large list of available cations and anions to choose from that can be further modified by changing the functional group and attachment position.

Among all the known cation types, imidazolium cations are one of the most widely studied cations with desirable transport properties such as low viscosity and high ionic conductivity, making them suitable as solvents for various industrial applications [299, 300]. However, the presence of the proton at the C2 position in the imidazolium cation makes them easily susceptible to reduction with a small electrochemical window [301]. Any modification to the C2 position by replacing the proton with an alkyl chain reduces ionic conductivity and increases viscosity dramatically [13]. In addition, an increase in alkyl chain length attached to the side of the cation also significantly slows down the dynamics of the ionic liquid, prompting researchers to search for cations beyond the imidazolium family [302, 303].

Alternatives such as quaternary ammonium-based ILs also appear extensively in the literature due to their inexpensive nature, high electrochemical stability, and less toxicity compared to imidazolium-based cations [304]. However, they have very high viscosity and low ionic conductivity in comparison to their imidazolium counterparts [305]. That has shifted the attention towards quaternary phosphonium cations with similar thermal/electrochemical stability but twice as low viscosity

as the quaternary ammonium cations [29, 31]. Similarly, sulfonium-based cations are another central atom-based cation with low viscosity and relatively higher ionic conductivity [27, 33]. However, they have a similar electrochemical window to imidazolium-based cations and low thermal stability compared to other cation types [125, 301].

Recently, pyrrolidinium-based ILs have also emerged as the next choice of interest because of favorable properties such as high electrochemical window, low viscosity, and high ionic conductivity [20, 21, 306]. The modification of the five-membered ring of the pyrrolidinium cation to the six-membered ring gives rise to the piperidinium cations. They exist in liquid form for a wide temperature range, making them suitable as solvents for various applications with low toxicity and high electrochemical stability [307]. Pyridinium cations are another cation type that is relatively less studied compared to the ones mentioned above, as they are more viscous than their aromatic counterpart [308, 309].

Despite the large variety of cation types, they either have favorable transport properties but low electrochemical stability or the opposite [11]. For instance, one of the most commonly studied ILs: 1-ethyl-methylimidazolium dicyanamide has an ionic conductivity of 2.82 S/m at room temperature, but it has an electrochemical window of less than 3.20 V [310]. Imidazolium cations paired with bis(trifluoromethylsulfonyl) imide anions, on the other hand, have an electrochemical window of 4.50 V [11, 311], with a maximum ionic conductivity of 0.91 S/m at room temperature [312]. Changing the imidazolium cation to pyrrolidinium cation significantly increases the electrochemical window, with most of them reaching above 4.50 V. However, the ionic conductivity of pyrrolidinium cations is relatively lower compared to imidazolium ILs [301, 313]. The other potential candidate is the sulfonium cations that offer low viscosity and high ionic conductivity compared to ammonium and phosphonium cations [33] and are found to be more suitable for battery application compared to pyrrolidinium and phosphonium cations [27, 34]. However, beyond these three cation types, the other commonly studied cation

types have very sluggish dynamics that are ideally not preferred for battery application [33,314].

Thus, for ILs to be considered electrolytes for battery applications, there is a definite need for more cations that could offer similar characteristics as the one mentioned above, increasing the list of potential candidates. This might also reduce the overall material cost as more candidates are explored and available [315]. Expanding the candidate list by increasing the alkyl chain length attached to the cation would not be a viable option as it slows the dynamics significantly [11]. Another route would be to mutate cations by adding functional groups such as ether to the alkyl chain, which is known to reduce viscosity and increase ionic conductivity with relatively little impact on the electrochemical window [36,37]. Besides that, the other option would be to discover an entirely new family of cations with desired properties.

Using experimental techniques to find additional new cation families would be challenging and time-consuming as there are estimates that the number of possible ILs could be in the range of billions [174]. Material search and design using atomistic simulation in such a vast ionic liquid chemical space would be expensive and unfeasible as well. Thus, this provides a unique opportunity to deploy machine learning tools to virtually explore the vast chemical space and accelerate the discovery of new cations with possibly desirable properties.

Machine learning tools have already attracted tremendous interest in the ionic liquid community, such as regressing properties and searching the ionic liquid chemical space for top-performing ILs beyond the ones already known in the literature [105, 119, 267]. Traditional regression-based methods using neural networks and support vector machines have become a go-to choice of method to regress data and expand the property database for a given property [118, 176, 178, 181, 267]. However, these methods are limited only for regressing properties as the design of new ILs can be challenging using traditional descriptor-based methods because of

the difficulty of going from chemical descriptors back to SMILES.

Instead, deep learning methods such as a generative-based model that uses the concept of transfer learning to generate new data by learning patterns from the input data have been on the rise to design, discover, and predict properties of novel materials [316–319]. Such methods offer accelerated material discovery by representing chemical structure information in a continuous form of latent space rather than a discrete form such as the traditional chemical descriptors. Another advantage of the generative-based method over the traditional descriptor is that the model decides the information and attributes it requires to represent the chemical information accurately in a low-dimensional space, removing the painstaking process of calculating descriptors and deciding the importance of each descriptor for property prediction.

Bombarelli et al. developed a generative-based model using Variational AutoEncoder (VAE) to generate a low-dimension representation of the chemical information combined with a regression-based model that takes this low-dimension data of chemical structures as input and predicts properties. Thus, allowing the search and discovery of materials with property prediction in the same instance [318]. Lim et al. [317] developed a conditional variational autoencoder model (CVAE), a modification of the VAE model that takes desired properties and the SMILE structure as input data. The trained model serves as a medium for an inverse design that outputs chemical structures with properties within that specified property range. Beckner et al. [119] combined the VAE model with a QSPR-based model to generate and predict properties for new ILs by using a transfer learning approach. Initially, the model learns the chemical space from an extensive list of organic molecule SMILES, followed by a smaller set of ionic liquid SMILES, to generate new ionic liquid structures with the desired property range.

Based on a similar approach to discovering new materials using a generative-based neural network, this work focuses on searching for new cation structures beyond the

existing ones present in the literature to increase the chemical space for the search and design of ILs with desirable properties for battery application. We employ a Variational AutoEncoder (VAE) machine learning model trained on known cation structures represented in the form of SMILES. The hypothesis is that, as these known cations are represented in a continuous low-dimensional vector space, the region of chemical space between known cations should be home to undiscovered cations representing new cation family types with possibly desirable properties suited for battery application.

In the next section, we elaborate on the concept of VAE, followed by data generation and representation for the model development. Next, the post-processing steps for new cation data generated using VAE, such as SMILE validity and stability, are discussed. As the interest is in evaluating cations for battery application, we also describe the methodology involved in this study to calculate the electrochemical window of the cations to examine their electrochemical stability towards battery application. The results and discussion section describes the model validation, data generation procedure, latent space arrangement, cation discovery, classification of cations by family, and electrochemical stability of all the novel cation types.

7.3 Methodology

7.3.1 Variational AutoEncoder (VAE)

VAE is a generative-based machine learning model that represents the information mapping between input and output data using a probability distribution. It consists of two deep neural networks (encoder and decoder): an encoder network learns to encode the input information in a compressed form in a low-dimensional latent space, learning only the specific attributes of the input data. The decoder network is trained to decompress that latent space information back to reconstruct the original input data. Combining the encoder and decoder allows the compression of information from higher-dimensional space into a low-dimensional space

that contains all the necessary information about the input data. The low dimensional latent space serves as a reservoir to generate new data, which is passed through the decoder to reconstruct it back in the form of the original input.

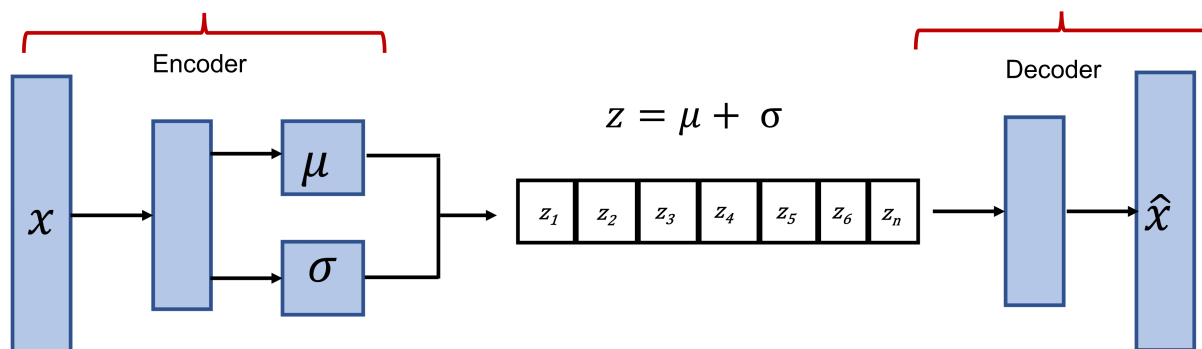


Figure 7.1: Variational Autoencoder schematic representation. The encoder takes in the input information and compresses it to low dimensional space on the latent space (z). The decoder takes the compressed information from the latent space and reconstructs it back to the form of the original input. The latent space is represented in the form of a Gaussian distribution with a mean (μ) and variance (σ). $z_1, z_2..z_n$ represents the attributes of the latent space with (n) dimensions.

7.3.2 Data Generation

The VAE model is trained on SMILES data based on various commonly studied cation families, including imidazolium, ammonium, phosphonium, piperidinium, pyridinium, sulfonium, and pyrrolidinium, with only alkyl functional groups attached to them. The length of the alkyl group varies from 1 to 10, including isomers. The individual percentage of each cation type is depicted in Figure 7.2 (a). Further details on the data generation are provided in the supporting information.

7.3.3 Model Development

The SMILES structures are converted to a one-hot encoding vector [320] format as depicted in Figure 7.2 (b). The one-hot encoding approach represents the SMILES string in "1" and "0" indicating whether a given character is present in the SMILE structure. The SMILE string vocabulary is the collection of all the unique characters present in the model dataset that form the column of the one-hot encoded vector, which in our case is a total of 21 unique characters and a row length of 100. The rationale for choosing this particular row length of 100 is to avoid the VAE model from generating extremely long alkyl isomer chain cations, which are often accompanied by extremely sluggish transport properties [321]. Lastly, every SMILE string is modified to have 'E' at the terminal, indicating the end of the string to make the conversion process back to SMILES from one-hot encoding efficient as there is no need to translate bits beyond the 'E' character [317, 322]. The one-hot encoding vector is a multidimensional (100,21) array reshaped to a 1D array of (1,2100) array size as the input layer for the VAE model.

In total, there are 18,141 SMILE data points, where 10% of the data is kept aside as a test set, and the remaining data is divided into training and validation sets (90:10) for model development and hyper-parameter tuning. Further details on the hyper-parameter tuning for the final VAE architecture are also provided in the supporting information.

7.3.4 Post Processing

New cation data generated from the VAE model is processed for SMILE validation, duplicate check, and charge validation. The validity of the generated SMILES is visualized using the OPSIN web application [10] and tested using the RDKit package [323]. The duplicate check involves two parts : (a) a comparison between the newly generated SMILES and the training dataset to ensure that all the smiles generated are unique, and (b) an internal duplicate check within the VAE gener-

ated dataset to ensure none of the SMILES are repeated. Both tests employ the Tanimoto similarity index score to check for duplicates implemented in the RDKit package [323]. The SMILE string is also checked for the presence of a positive charge to ensure no neutral organic molecules are present in the final generated set.

As there are no inbuilt valence rule heuristics in the VAE model, some of the new SMILE structures generated by the model take the form of radical cations. Instead of discarding such cations, we manually added hydrogen to such cations where necessary to convert such cations into protic cations. Besides valence rules, the VAE also has no inbuilt heuristics to determine whether a given cation structure exists in a stable form. In order to examine the stability of these generated cations, we ran Quantum mechanical (QM) calculations on all of them for a stability test. The result and discussion section have information about the stability test in detail. The three-dimensional coordinates of the cations for QM calculations are generated using the Openbabel package [324]. Lastly, we also attempted to build a classification model based on Random Forest (RF) [144] to classify the VAE generated cations using chemical descriptors generated through RDKit [323]. This eliminates the need to examine the novel cations and classify them manually. More details about the classification model development are provided in the supporting information.

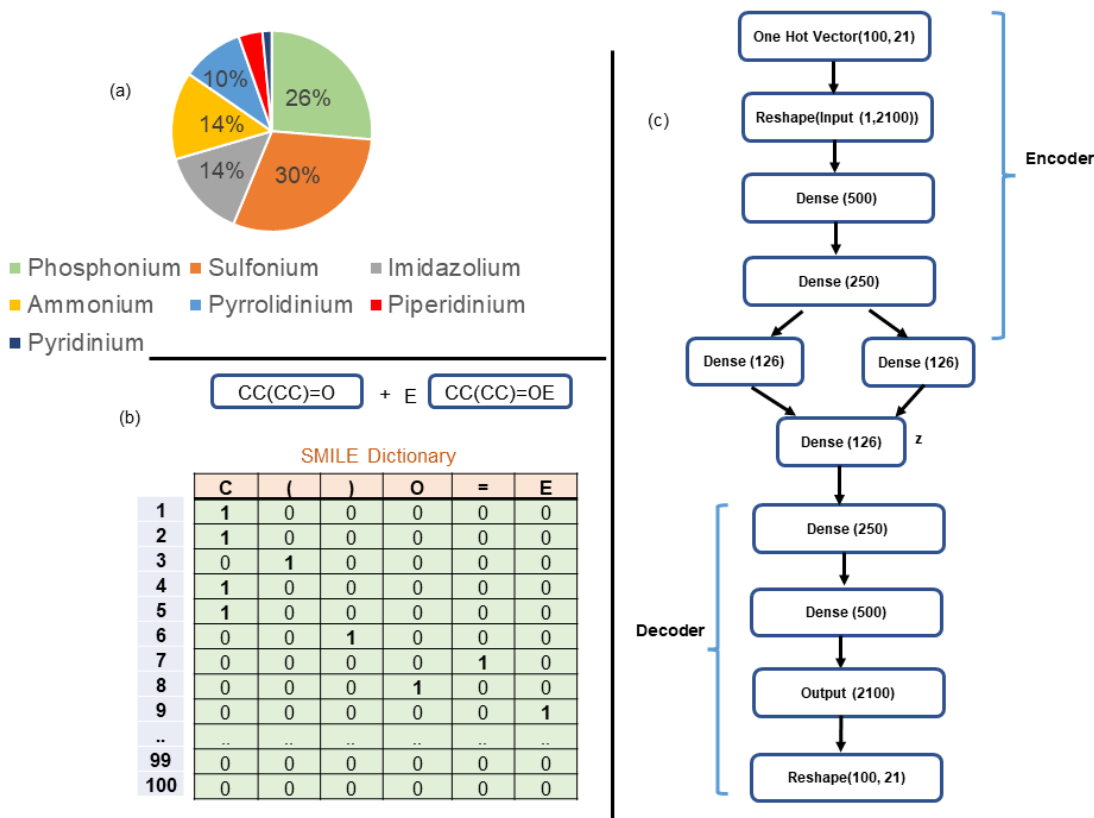


Figure 7.2: (a) % Representation of different cation type in the VAE model development. (b) One hot encoding vector of a sample organic molecule with a vocabulary size of 6 and length of 100. 'E' denotes the end of SMILE string. (c) Neural network representation of the encoder and decoder for the VAE model.

7.3.5 Electrochemical Window

As this work is focused on finding new cations suitable for battery applications, we evaluated the electrochemical stability of the VAE generated cations by calculating their electrochemical window (ECW). ECW is a critical property of an electrolyte as it characterizes the range of potential over which it is stable against oxidation and reduction. As the ionic liquid space expands, measuring ECW for each ionic liquid using experimental techniques can be challenging and expensive. Numerous studies have employed computational tools to calculate the ECW of ILs. [42, 104, 126, 128, 325, 326] This is done by approximating the cathodic limit (CL) with cation reduction and the anodic limit (AL) to the oxidation of anions.

The two common ways to calculate ECW using computational tools are through the thermodynamic cycle (TDC) approach and the highest occupied molecular orbital (HOMO)/lowest occupied molecular orbital (LUMO) method. The TDC method involves calculating the Gibbs free energy of oxidation and the reduction potential of the cations and anions. On the other hand, the HOMO/LUMO method is based on Koopmann’s theory that approximates the electron affinity with the LUMO energy and the ionization potential with the HOMO energy [152]. Thus, for ECW calculations, the LUMO energy is equal to the cathodic limit and the HOMO energy to the anodic limit.

$$V_{CL} = \frac{-\epsilon_{LUMO}}{e} \quad (7.1)$$

$$V_{AL} = \frac{-\epsilon_{HOMO}}{e} \quad (7.2)$$

$$ECW = V_{CL} - V_{AL} \quad (7.3)$$

Where e denotes the charge of an electron.

Panidian et al. [128] did an extensive study on calculating the ECW of ILs using various methodologies. The Δ SCF Self Consistent Field approach to calculating ECW in vacuum had a deviation of 48–54.1% from experimental data, depending on the basis set, according to the authors. Changing the vacuum environment to an SMD solvent (dichloroethane) environment and using the one-electron addition approach reduced the deviation to 2.4% using M06-L/6-311 + G(2d,p) basis set. The one-electron addition approach calculates the ECW using the reduction potential of the cations and the oxidation potential of the anions. The reduction potential is the difference in Gibbs free energy between V_C and V_{C+e^-} . Similarly, the oxidation potential is the Gibbs free energy difference of V_A and V_{A-e^-} .

Building upon the work of Panidian and co-workers [128] we used a similar SMD approach with dichloroethane as the solvent, (M06-L) functional form, and (6-311

+ G (2d,p)) as the basis set. The calculations are performed using Gaussian 09 package [153]. However, instead of the one-electron addition approach, we use the HOMO/LUMO approach using equation 7.1, 7.2 and 7.3 to calculate ECW. Unfortunately, the authors do not provide ECW data to benchmark the calculations. Instead, we compare our HOMO/LUMO calculations to the work of Asha et al. [104] for pyrrolidinium ILs and Kazemiabnavi et al. [42] for imidazolium ILs. The result and discussion section elaborate more on the comparison.

7.4 Results and Discussion

7.4.1 Model Validation

Type	Average reconstruction accuracy %
Training Set	97.67±1.45
Validation Set	95.95±1.65
Test Set	95.57±1.76

Table 7.1: Reconstruction accuracy % of the SMILE data using the VAE model for different data set called upon 100 different instances. Reconstruction accuracy % is calculated by taking the total SMILE generated by the VAE model that matches the input SMILE to the total number of input SMILE.

Table 7.1 shows the reconstruction accuracy % of the training, validation, and test set. As the VAE model is a probabilistic model, it is crucial to ensure that the model reconstruction accuracy of the SMILE string stays relatively the same every time the model gets called. Thus we reported the mean and standard deviation of the three sets called 100 different times. Given that all of the data points are unique SMILE strings with varying cation chain length and complexity in branching, the test set reconstruction accuracy is approximately 95.57±1.76%, indicating that the VAE model is well equipped to construct and deconstruct information to valid cation SMILES.

Figure 7.3 depicts the projection of the training set latent space to a two-dimensional axis using principal component analysis (PCA) labeled accordingly by the cation type. Each subplot denotes the cation type location in the latent space. It is evident from the subplots that the VAE self arranges the cations into individual clusters by cation type despite having no information about the cation label. Beckner et al. [119] reported similar self-arrangement trends for the latent space with clusters of individual cations.

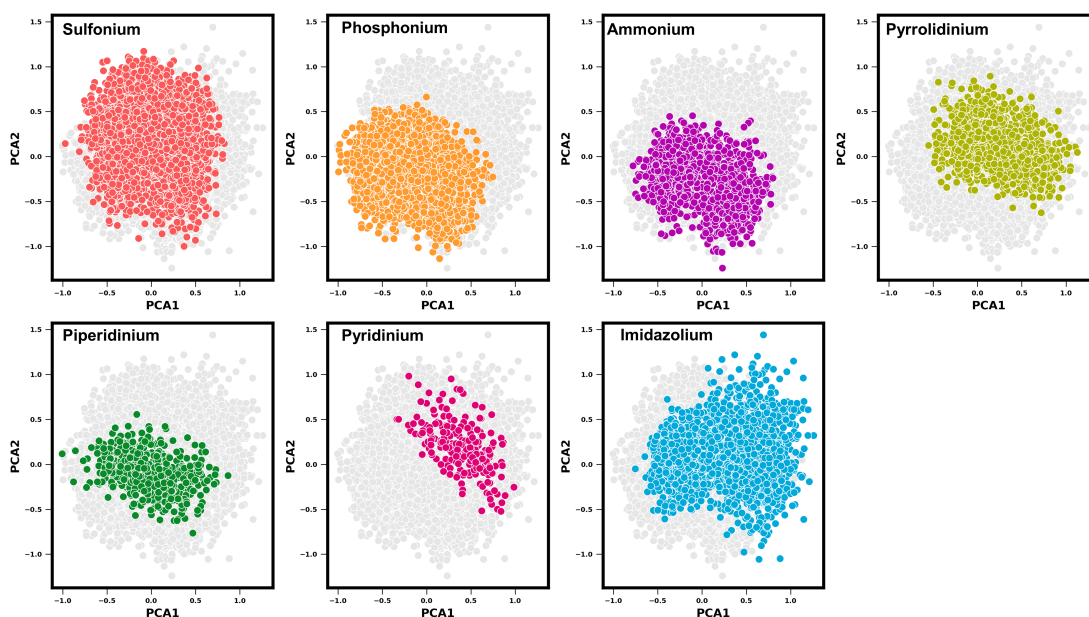


Figure 7.3: Projection of latent space into two dimensional axis using the first two principal components. Each of the subplot indicates the region for each of the cation type on the latent space.

The self-arrangement of the latent space is even more remarkable considering that these cation clusters are not random arrangements but instead follow a pattern based on similarity. For instance, the non-cyclic/aromatic cations (sulfonium, ammonium, phosphonium) are arranged close to each other as they have a common feature, i.e., a central-based atom. The location of ammonium cations is fascinating as it serves the transition to cyclic/aromatic cations, starting with pyrrolidinium, followed by piperidinium, pyridinium, and finally, the imidazolium cations that

occupy the entire right side of the cluster. The supporting information (Figure D.5) also depicts the third and fourth principal component analyses along with the first and second components.

7.4.2 Data Generation

With the model validated on the test set, we focused our attention on discovering new cation structures by sampling the latent space and passing it through the decoder to extract the output in the form of a one-hot encoding vector converted to SMILES. Theoretically, it would be possible to sample every possible region in the latent space and generate an enormous number of new cation structures ranging in the millions. However, processing them might be challenging and prone to errors during smile validation, duplicate removal, and family identification.

Based on the kernel density estimate of all the latent space attribute (Figure D.6), we limited the latent space sampling to a $\mu = 0$ and σ ranging from 0 to 1 with an increment of 0.01. This was done for 300 data points at a time, leading to 30,000 sampling instances. Applying the post-processing steps discussed earlier that included SMILE validation, duplicate removal, and neutral molecule removal led to 756 valid unique SMILE structures.

7.4.3 Stability Test

Often, with a generative-based model such as VAE, it is challenging to provide insights such as the stability of the generated SMILE structure, as the model has no inbuilt heuristics to know whether a given chemical species could exist in a stable form. In this section, we examined the number of cations that could exist in a stable form by running a QM stability test.

The stability evaluation of the cation geometry and frequency optimization is done using the HF/STO-3G theory implemented in the Gaussian 09 package [153]. The stability criteria of the structures included threefold: (a) converged geometry and frequency optimization, (b) no imaginary frequencies, (c) change in the bond dis-

tance less than 10%. The last test ensured that all the cations preserved their bond structure during geometry optimization.

Out of 756 SMILE structures, 425 passed the three stability test criteria. All the failed systems had convergence issues. Examining some of the unstable structures through visual inspection showed that some of these cations had unusual structures, such as two nitrogen atoms attached or nitrogen and phosphorus atoms next to one another, which might have made the cation unstable.

7.4.4 Non-Aromatic/Non-Cyclic Cations

There were 323 cations out of 425 that belonged to the non-aromatic/non-cyclic family. Figure 7.4 (a) shows the classification of these cation type using the classification model. The highest number of new cations belonged to the sulfonium group, followed by phosphonium and ammonium, leaving only eleven cations as "Unknown" types. Of the eleven "Unknown" cations, six of them are primary ammonium cations with a long branched alkyl chain length that are not present in the VAE training set. Despite the model being trained entirely on aprotic cations, the model generated numerous radical cations, as discussed earlier. Thus, we added hydrogen atoms to the charged atom during the post-processing step to overcome this shortcoming. Krenn et al. [327] in a recent study, implemented a new way of representing molecules through a technique called self-referencing embedded strings (SELFIES) that aims to solve some of the issues associated with the traditional SMILE structures.

Next, we examined the chemical space of the cations represented using a two-dimensional PCA projection based on cation features generated using RDKit. Figure 7.4 (b) shows the chemical space of known cation types used during classification model development. This explains why the classification model can classify each cation type with such high accuracy, as each cation family tends to have its own unique chemical space. The alkylated imidazolium (1-alkyl-2-alkyl-3-alkyl imidazolium) cation is the small region of the imidazolium cation separated from

the big block of imidazolium. Although the VAE training data does not have 1-alkyl-2-alkyl-3-alkyl imidazolium cations as they tend to induce sluggish transport properties [13], this was added to the classification training data to ensure that the model knows to identify them as imidazolium instead of labeling them as "Unknown" if the VAE generates them during sampling.

Figure 7.4 (c) shows the PCA projection of the new generated non-aromatic or non-cyclic cations along with the existing cations used for the classification model development. Notice the overlap around the ammonium cation region, as those are the primary ammonium cations that the model misclassified as "Unknown" cation types. The remaining four cations are from the sulfonium and phosphonium families with nitrogen atoms on the alkyl chain length, which the model classified as "Unknown" as no such cations were present in the training set. Bates and co-workers report the synthesis of such amine/amino-based cation structures in the imidazolium alkyl chain specific for CO₂ adsorption [256]. Zhang and co-workers also demonstrated the possibility of synthesizing amino-based phosphonium cations paired with amino-based anions for CO₂ adsorption that showed promising results [328].

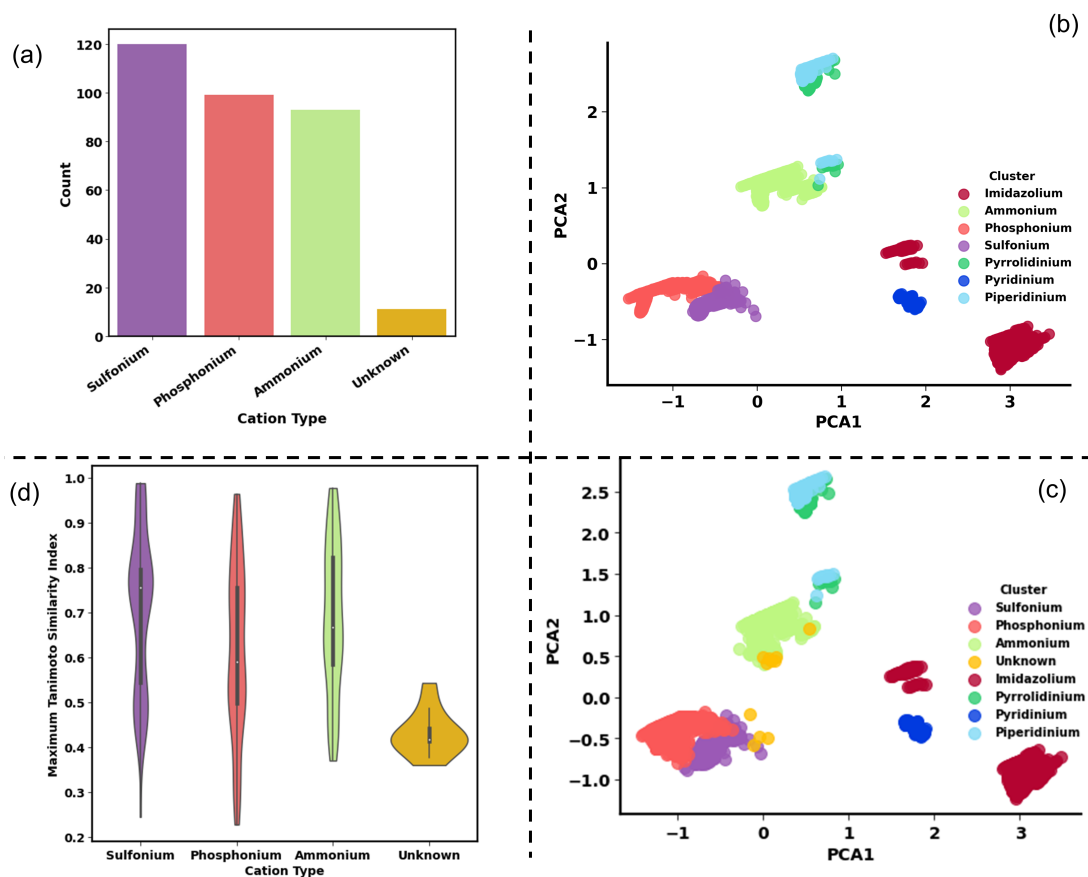


Figure 7.4: (a) Bar chart of the cation type generated using the VAE model for non-aromatic/non-cyclic based cations. (b) Two-dimensional PCA plot of the known cation chemical space with their respective label generated using features from RDKit. (c) A two-dimensional PCA plot of the new VAE generated non-cyclic/non-aromatic cations on the cation space in addition to the classification training data. The features for these cations are generated using RDKit. Here "Unknown" refers to cations that do not belong to known cation families. (d) Maximum Tanimoto Similarity index for the "Unknown" cation type compared to the rest of the known cation type.

We also examined the novelty of these cations by calculating the Tanimoto Similarity index for each of the cations. This is done by calculating the index with respect to all the known cations in the training set and taking the maximum of each index, as we are interested in knowing the closest similarity obtained for the new cations. Figure 7.4 (d) shows the violin plot distribution of the three known

cation types and unknown types. For the sulfonium cations, the closest similarity index varies from 0.99 to 0.25, with the most similar cations having a long alkyl chain with a large number of branches along the way. However, the least similar cations have double bonds on the alkyl chain attached to the sulfonium cation, as such cations were not present in the VAE development. The model places the double structure on the sulfonium cations by learning from the aromatic cations that had the double structure present in them. Again, this shows the transfer learning ability to generate new structures from known structures and functional groups. We see similar trends for the phosphonium and ammonium cations where the closest similarity cations are low branched cations, while the furthest similar cations are alkyl chains with a double bond, as seen in Figure D.7.

Based on this finding, it is safe to say that the level of cation complexity generated by the VAE model can increase exponentially with the addition of new functional groups such as ether, hydroxyl, and carboxyl, as there are more ways to mutate these functional groups.

7.4.5 Aromatic/Cyclic Cations

Similar to non-aromatic/non-cyclic cations, we evaluated the cation clustering for aromatic/cyclic-based cations using the method discussed above. Figure 7.5 (a) shows the bar plot of the frequency of each of the cation types, where around 76 out of 102 cations do not belong to the known family type, followed by the piperidinium cation type, ammonium, and imidazolium. Compared to the PCA plot between Figure 7.5 (b) and 7.5 (c), almost all the "Unknown" cation fill up the new region around ammonium, pyrrolidinium, and piperidinium cations.

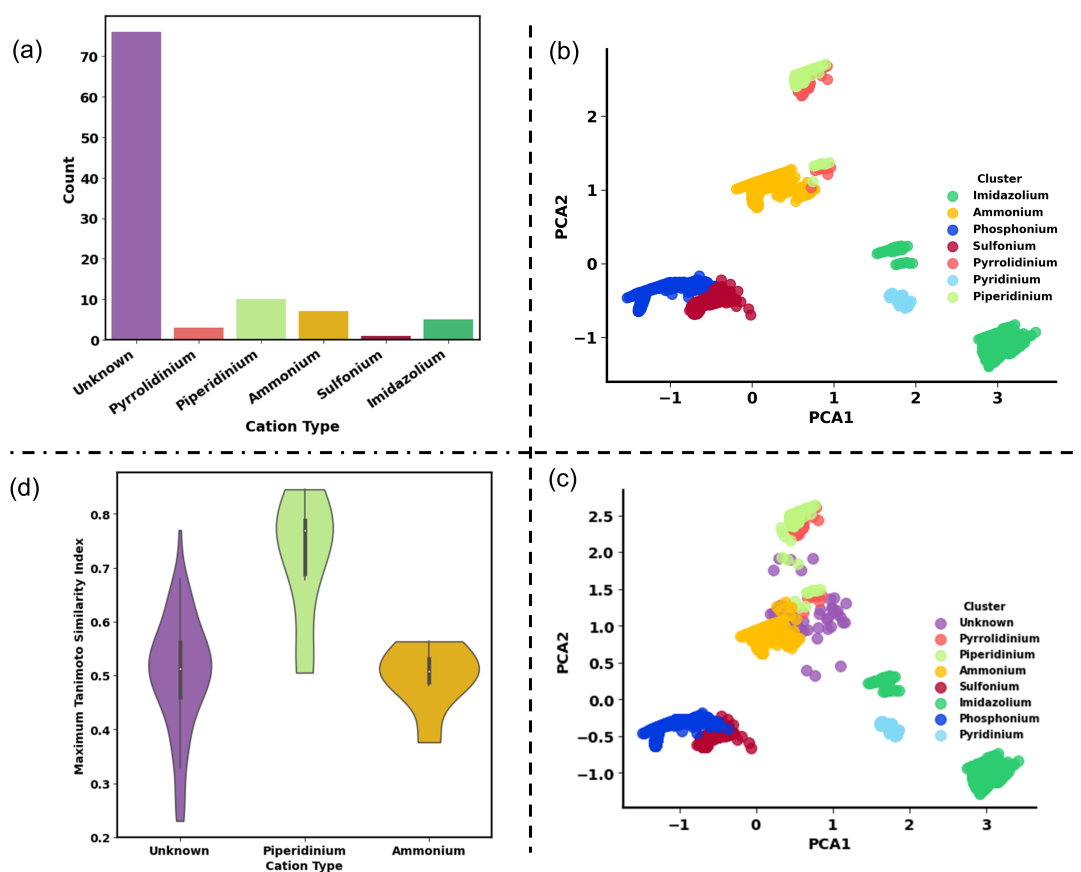


Figure 7.5: (a) Bar chart of the cation type generated using the VAE model for aromatic/cyclic-based cations. (b) Two-dimensional PCA plot of the known cation chemical space generated using features from RDKit with their respective label. (c) A two-dimensional PCA plot of the new VAE generated cyclic/aromatic cations on the cation space in addition to the classification training data. Here "Unknown" refers to cations that do not belong to known cation families. (d) Maximum Tanimoto Similarity index for the "Unknown" cation type compared to the rest of the known cation type

Figure 7.5 (d) depicts the Tanimoto similarity coefficient for the "Unknown" label cations where some of the cations have a maximum Tanimoto similarity index close to 0.20, which suggests a very minimal similarity with the known cation type. To put that in context, one can imagine the two cation types, 1-methyl-1-propylpyrrolidinium, and 1-methyl-1-propylpyridinium, which are very distinct cations with varied physical structures and properties that have a Tanimoto simi-

larity index of roughly about 0.22.

Figure D.8 shows some of the least similar cations with their respective similarity index. The first cation on that list has a structure similar to the imidazolium cation but with only one nitrogen atom present in it. The second seven-member structure is the modification of the six-member piperidinium cation. The diversity and uniqueness of these cation structures again highlight the vastness of the cation chemical space and the number of new structures that are yet to be discovered with possibly unique properties.

7.4.6 Electrochemical Window

To evaluate the applicability of these cations as electrolytes for battery applications, we calculated the (ECW) of the cyclic/aromatic "Unknown" cations in this section. Before calculating the ECW of them, we benchmarked our method with a few different studies. Figure D.9 (a) compares the ECW calculated in this work to the TDC approach for Li⁺/Li reference electrode electrochemical and experimental data [11, 42]. The ECW of [C_{n=2,3,4,5,6}mim][NTf₂] using the HOMO/LUMO approach are in excellent agreement with the TDC method and few experimental data. Similarly, the ECW of [C_{n=2,4,6}mim][BF₄] shows similar quantitative agreement as to the TDC method, with one of the experimental data, sets closer to the HOMO/LUMO approach as compared to TDC. Lastly, for the [C_{n=2,4,6}mim][PF₆] series, the TDC method predicts the ECW to be close to 8.0 V, while the HOMO/LUMO method has values close to 6.5 V. The experimental data for [C₄mim][PF₆] is found to be between the two computational methods. Overall, the Average Absolute Relative Deviation (AARD) between our approach and the TDC method is about 9.03%.

Similarly, Figure D.9 (b) compares the ECW of [Py14]⁺ cation paired with three different anions. In all three of them, the HOMO/LUMO approach has AARD of 10.0% compared to the TDC for Li⁺/Li reference electrode and 9.52% against experimental data [22, 104, 301]. The benchmark data is provided in the supporting

information.

With the ECW method in excellent agreement with TDC and experimental data, we proceeded to compute the HOMO/LUMO of all the 74 cyclic/aromatic cations labeled "Unknown" in the earlier section. Out of the 76 cations we discovered earlier using the VAE, two of the cations repeatedly failed to converge during geometry optimization. As such, the two cations are discarded from the calculations, leaving 74 cations. The ECW of these cations are calculated in combination with four commonly studied anions (dicyanamide $[DCA]^-$, bis(fluoromethylsulfonyl)imide $[NTf_2]^-$, bis(fluorosulfonyl)imide $[FSI]^-$ and tetrafluoroborate $[BF_4]^-$) by simply taking the difference in the anion $HOMO$ energy and the cation $LUMO$ energy using equation 7.1, 7.2 and 7.3.

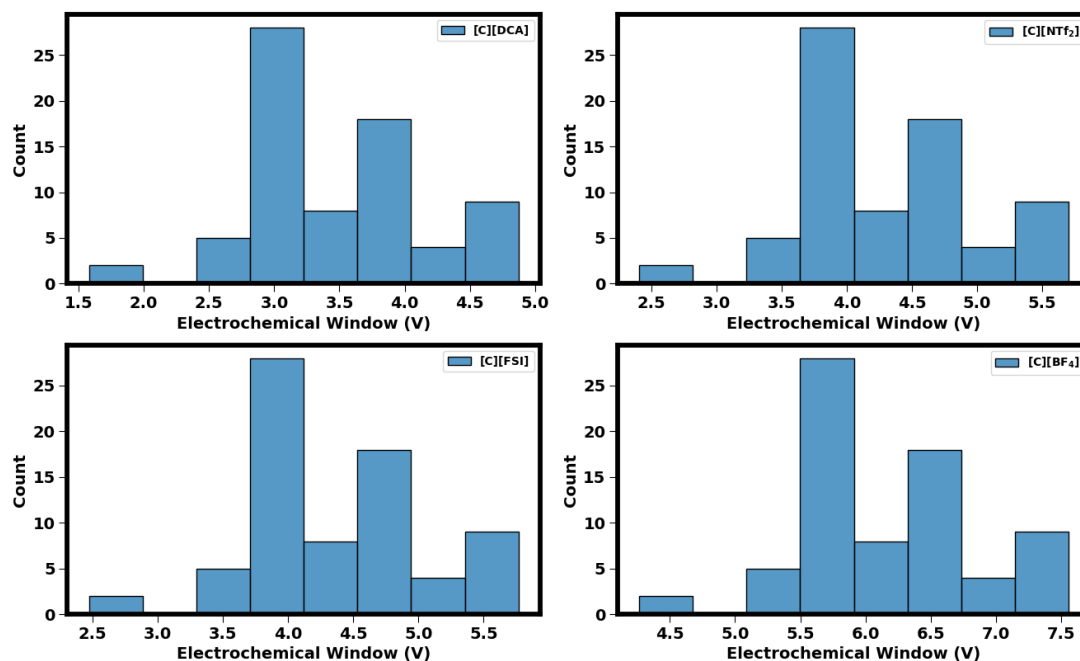


Figure 7.6: Electrochemical window values of all the "Unknown" cyclic/aromatic cations paired with (a) $[DCA]^-$, (b) $[NTf_2]^-$, (c) $[FSI]^-$ and (d) $[BF_4]^-$ anion.

Figure 7.6 (a) depicts the ECW of all the 74 cations paired with the $[DCA]^-$ anion. As expected, the ECW paired with $[DCA]^-$ anion has a very low electrochemical window, with most of it concentrated at around 3.0 V. This is because the $[DCA]^-$

anion easily gets oxidized as compared to other anions. Asha et al. [104] reported the HOMO of the $[\text{DCA}]^-$ to be concentrated around the nitrogen atom, which could explain why it gets oxidized easily. A few of the cations cross 4.0 V, with the maximum ECW reaching a value of 4.75 V.

Next, Figure 7.6 (b) and (c) show the ECW distribution of the cations paired with the $[\text{NTf}_2]^-$ and $[\text{FSI}]^-$ anion. The two anions are relatively more stable to oxidation than the $[\text{DCA}]^-$ anion as studies have shown the HOMO to be located on the oxygen atom of the anion that has higher electronegativity compared to the nitrogen atom [11, 42, 104]. Majority of the cations now have ECW at around 4.0 V with a maximum reaching around 5.5 V. Lastly, Figure 7.6 (d) shows the ECW of $[\text{BF}_4]^-$ anion with a majority of the ILs with ECW concentrated around the 5.5-6.0 V region. Kazemiabnavi et al. [42] reported similar trends for the $[\text{BF}_4]^-$ anion as the authors alluded to the location of HOMO that is found only on the fluorine atoms as it has very high electronegativity values, making it difficult to oxidize.

7.4.7 Conclusion

In summary, we used a Variational Autoencoder (VAE) based machine learning method to discover new cation structures from the known cation family. The VAE latent space seems to self-arrange into clusters of cation families based on chemical similarity despite having seen no such information about the cations. Sampling the latent space generated 756 new unique cations, of which 425 formed stable structures based on the QM stability test. These cations are classified into respective families using a classification model. The model classified 76 new aromatic/cyclic cations as "Unknown" type. A two-dimensional PCA projection of the cation chemical space revealed new regions of chemical space occupied by the newly discovered cations labeled "Unknown". The Tanimoto similarity index of the cations provided further confirmation of the novelty of the new cation family.

We also calculated the electrochemical window of these "Unknown" aromatic/cyclic

cations using the HOMO/LUMO approach to assess their electrochemical stability for battery applications. The majority of the cations paired with [DCA]⁻ had ECW close to 3-3.5 V, the ones paired with [FSI]⁻ and [NTf₂]⁻ had an electrochemical window greater than 4.0 V. In contrast, cations paired with [BF₄]⁻ anion had the highest electrochemical window, with some of them reaching 7.5 V. Future work would be to add more functional groups to the VAE cation training set and increase the sampling fold to generate more possible cation structures with mixed functional groups that could potentially occupy new niches in the ionic liquid cation chemical space, offering desirable properties.

CHAPTER 8

Mapping the Frontier Orbital Energies of Imidazolium-Based Cations Using Machine Learning

8.1 Abstract

This paper investigates the use of a machine learning model based on extreme gradient boosting (XGBoost) to correlate frontier orbital energies calculated from density functional theory (DFT). The system of interest is the imidazolium cation with an alkyl chain length ranging from 1 to 10, including all possible isomers. Enumerating the entire imidazolium isomer space yields 315,872 distinct cation structures. Calculating frontier orbital energies for each would be computationally very expensive and time-consuming using DFT. Instead we develop a machine learning model to correlate these energies. Compared to DFT calculations, the model can predict the external test set with MAE of less than 0.4 eV for HOMO and 0.2 eV for LUMO energies. Aside from accurate energy prediction, the entire energy calculation using the model for 317,382 cations is completed in about four hours using a personal computer. This drastically reduces computation time for the entire chemical space using DFT, which otherwise might have required more than a million hours using high-performance computing.

8.2 Introduction

A tremendous amount of research has been devoted to finding the next generation of electrolytes found in Li-ion batteries to solve some of the safety concerns

associated with organic carbonate-based electrolytes that are highly volatile and flammable during thermal runaway or short-circuit [329, 330]. Several promising alternative candidates, such as solid-state electrolytes, polymer-based electrolytes, and ILs, are being considered because of their high thermal and chemical stability [6, 331].

Among them, ILs are being considered a promising candidate as they offer several exciting features, such as negligible vapor pressure, high thermal/chemical stability, and design flexibility. Furthermore, there is an extensive list of cations, anions, and functional groups that one can choose from to design task-specific applications for a given application [332, 333]. In addition to pure ILs, it is also possible to form binary mixtures by mixing ILs to improve some of their weaknesses, such as sluggish transport properties [51, 334, 335]. However, several essential electrolyte properties, such as ionic conductivity and electrochemical stability, must be adequately understood before using them for battery applications. In our earlier work we were focused on developing a model to generate high ionic conductivity data using a machine learning model applicable for a large set of diverse cations and anions [267, 336]. Thus, the focus of this work is geared towards electrochemical stability of ILs.

Knowing the electrochemical stability of an electrolyte is critical for determining the operating condition of the battery with minimum safety concerns. Electrochemical stability is quantified using an electrochemical window (EW), the operating range of voltage where a species does not undergo reduction or oxidation. In the case of ILs, that would be the voltage range where the cation does not undergo reduction and the anion is immune to oxidation. Experimentally, EW is calculated using a linear sweep or cyclic voltammetry techniques with reference to a working electrode [125]. However, the sheer volume of ILs that are possible using the available cation-anion combination would make the experimental process of measuring EW for each of them expensive and time-consuming, as the number of ILs possible is estimated to be in the range of billions [174].

Instead, several studies have employed the use of molecular dynamics (MD) combined with DFT or direct DFT calculations to compute the electrochemical window of ILs [42,104,126,337]. This involves the thermodynamic cycle approach that calculates the Gibbs free energy of reduction and oxidation of cations/anions using thermodynamic cycle. The other approach would be the calculation of the highest occupied molecular orbital (HOMO) and lowest occupied molecular orbital orbitals (LUMO). During ECW calculations, the LUMO energy is assumed as the cathodic limit or cation reduction, while the HOMO energy is assumed as the anodic limit or the oxidation of an anion [337].

Liang et al. [126] used DFT calculations to compute the HOMO/LUMO energies of 42 cations and 42 anions combined to calculate the EW of 1764 ILs. The authors found the fluoro-based anions to have the highest electrochemical stability compared to other anions. Saeed et al. [42] calculated the EW of imidazolium-based ILs with varying lengths and anions using the thermodynamic cycle and HOMO/LUMO methods. The authors found the EW calculated using the HOMO and LUMO method to have a considerable deviation compared to experimental data, while the thermodynamic cycle method results were in excellent agreement with experimental data. Ilawe et al. [127] in a computational study, showed that the choice of DFT functional to calculate HOMO/LUMO greatly impacts the accuracy of EW compared to experimental data.

Besides battery application, several other studies have shown the importance of HOMO/LUMO as chemical descriptors to predict ionic liquid properties such as toxicity, biodegradability, and solubility [338–340]. Numerous studies have also investigated ILs as corrosion inhibitors for various metals by examining the HOMO/LUMO parameters to determine the corrosion efficiency of the ionic liquid for a given metal. Yousefi et al. [341] examined the corrosion efficiency of six imidazolium ILs on mild steel in HCl solution using experimental and quantum mechanical calculations. The authors found the linear equation and non-linear

equation consisting of just the two parameters, HOMO and LUMO, as descriptors, were sufficient to correlate inhibition efficiency accurately. Similarly, Tian et al. [342] found that the HOMO/LUMO and other electrochemical descriptors accurately correlated the corrosion inhibition efficiency of imidazolium hydrogen sulfate ILs on copper.

The HOMO/LUMO energy itself can provide valuable insights into the stability and hardness of a molecule. Based on Koopman's theory, HOMO energy is approximately equal to the ionization potential energy that quantifies the energy required to withdraw an electron from a molecule. On the other hand, LUMO is approximately equal to electron affinity energy, the energy released when an electron is added to the system. Several studies have found the change in alkyl chain length to have minimal impact on the reduction potential of the imidazolium cation as the LUMO is located on the imidazolium cation ring [42, 127, 343]. Any changes to the imidazolium ring, such as the substitution of a functional group at the (C₂) position on the ring with other functional groups, have been found to improve electrochemical stability [344].

Besides linear alkyl chain substitutes, various studies have also demonstrated the usefulness of branched alkyl chains as cation attachments. Erdmenger et al. [345] reported the branched alkyl chains imidazolium cations to have higher decomposition temperature than linear alkyl chain imidazolium cations paired with two different anions. Similarly, Corvo et al. [346] in an experimental study, reported the branched alkyl chain imidazolium cations exhibited higher CO₂ solubility as compared to the linear alkyl chain imidazolium cations at high pressure. Besides physical properties, branched alkyl chain imidazolium cations are also known to have lower toxicity as compared to linear alkyl chain imidazolium cations [347, 348]. Despite some of the unique advantages associated with branched alkyl chain imidazolium cations over linear alkyl chain imidazolium cations, very little is known about its energy gap and electrochemical stability that are some of the critical properties required for electrolytes characterization for battery application.

One potential reason why there are very few studies on the branch alkyl chain functional group is the enormous number of cation enumerations possible with varying lengths, attachment positions, and isomers. Performing experimental measurements or computational calculations would be expensive and time-consuming, even using high-performance computing. Furthermore, this becomes even more challenging for long alkyl chains with multiple branching levels attached to both sides of the cation, as the DFT calculations may run into a convergence issue.

Thus, this presents a unique opportunity to use machine learning methods to correlate frontier orbital energies of imidazolium cations from density functional theory (DFT) calculations. This work attempts to build a machine learning model trained on a small set of DFT cation data that is assumed to represent the imidazolium isomer chemical space. The trained model is then utilized to map out the entire space, thus reducing the computational cost of running expensive calculations for a large set. The cation of interest for this study is the 1-alkyl-3-alkylimidazolium ($[C_nC_m\text{mim}]^+$) cation, which is widely studied for various applications due to its extremely low viscosity and high ionic conductivity property as compared to other cation types [11].

The following section discusses the SMILE data generation procedure, DFT calculations, and the machine learning methodology. We evaluate the model’s performance on the test set based on three different error metrics following model development. In addition to the test set, we compare the model’s prediction to DFT calculations for an external test set (cations that are not related in alkyl chain length to the training set cations) to further support the model’s predictive capability. Next, we employ the SHAP feature analysis technique to determine some of the most influential features for the HOMO/LUMO energy model. Lastly, with the model cross-validated, we map out the entire isomer space using the model.

8.3 Methodology

8.3.1 Data generation

The cation of interest for this study is imidazolium cation with an alkyl functional group attached to positions 1 and 3 (1-alkyl-3-alkylimidazolium $[C_nC_m\text{mim}]^+$). The alkyl chain is varied from 1 to 10 in length, with all possible isomers leading to 317,382 unique imidazolium cations, as seen in Figure 8.1. For some of the long alkyl chain cations, the number of possible isomers is in $5 \log_{10}$ units, indicating the vastness of the isomer chemical space. The enormous cation chemical space makes it difficult and expensive to perform DFT calculations on all of them as mentioned earlier.

Cations denoted by (*) in Figure 8.1 indicate the cation for which we performed DFT calculations to compute its HOMO/LUMO energy for model development. Starting with $[C_{n=1,10}C_1\text{mim}]^+$ as this would include all the possible isomers from alkyl chain 1 to 10, thus ensuring the model has seen all the possible isomer variations. Next, we add $[C_{n=2,8}C_2\text{mim}]^+$ cations to the training database to take into account the change in HOMO/LUMO energy as the alkyl chain on one side of the cations gets extended from methyl to ethyl chain. Similarly, we added $[C_{n=3,6}C_3\text{mim}]^+$, $[C_{n=4,6}C_4\text{mim}]^+$, $[C_5C_5\text{mim}]^+$ and $[C_6C_6\text{mim}]^+$. Adding these cations would ensure that the model learns the HOMO/LUMO energy change as the alkyl chain length varies on both sides of the cation and when both sides have the same chain length. This leads to 1399 cation data points from these alkyl combinations for model development.

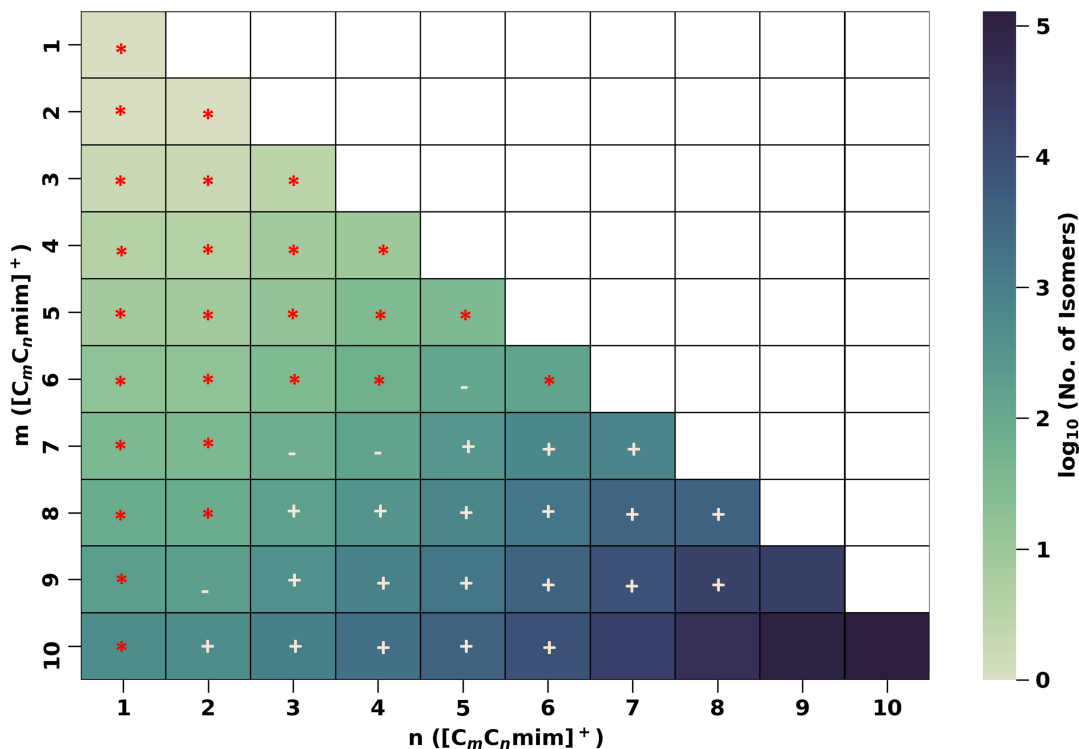


Figure 8.1: Number of isomers for $[C_n C_m \text{mim}]^+$ cation in log10 scale. Cations denoted by (*) indicates data used for model development. Cations denoted by (-) and (+) indicate data used for external test set validation.

We also generated external test set data to evaluate the model’s generalization ability compared to DFT calculations. Note that this data set is not part of the training set. Cations denoted by the (-) symbol as seen in Figure 8.1 indicates four of the cation sets on which we ran DFT calculations in their entirety and extracted its HOMO/LUMO energy. This includes $[C_2 C_9 \text{mim}]^+$ with 211 cation structures, $[C_3 C_7 \text{mim}]^+$ has 78 cations, $[C_5 C_6 \text{mim}]^+$ has 136 cations and $[C_4 C_7 \text{mim}]^+$ 155 cations lead to a total of 580 data points. For the remaining cation set, generating all the possible isomers for test set evaluation would be expensive and time-consuming, so we generated HOMO/LUMO energy for 20 random data points for each cation, as indicated by the (+) symbol. In total, the external test set contained 940 unique cation structures.

8.3.2 DFT calculations

The HOMO/LUMO energies are computed from DFT calculations based on m06 level theory and 6-311++g(d,p) basis set using Gaussian 09 package [349]. The input files for the calculations are prepared using openbabel package along with the generation of initial three-dimensional coordinates [324]. To ensure that the system is optimized, we evaluated the convergence criteria for geometry optimization along with making sure there are no imaginary frequencies for the system.

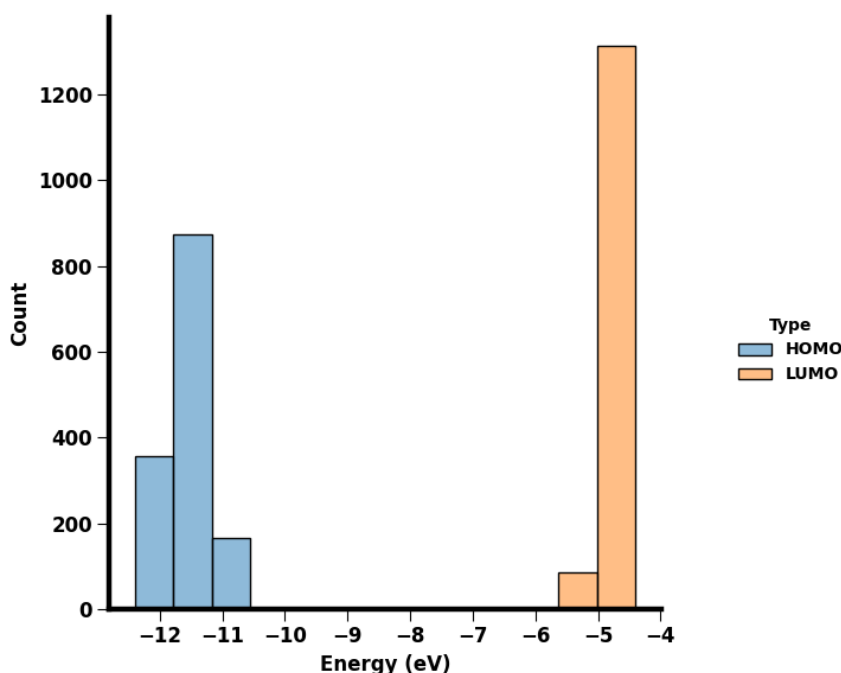


Figure 8.2: HOMO and LUMO energy distribution of cations present in the model development data set calculated using DFT.

Figure 8.2 depicts the DFT-calculated distribution of HOMO/LUMO energy. The HOMO energy is distributed between -12.5 and -10.5 eV, while almost all the LUMO energy is concentrated between -5.0 and -4.0 eV. As mentioned earlier, this is because the LUMO is primarily found on the ring of the imidazolium cations which is not impacted by the change in alkyl chain length or branching.

8.3.3 Machine learning

Features for the cations are generated with the open-source cheminformatics RD-Kit package, which can generate 196 two-dimensional features [323]. The data is divided into 90% training set and the remaining 10% as test data for model evaluation. The training set data is further divided into 90% for model development and 10% for hyper-parameter evaluation referred as validation data. The input and output data are normalized and scaled using MinMax scaling [0-1] implemented in Scikit-learn [144]. To reduce the number of features for the HOMO and LUMO models, we employed the Least absolute shrinkage and selection operator (LASSO) technique [144]. The best LASSO penalty parameter is determined using a 5-fold cross-validation (CV) search. The algorithm reduces the number of features for the HOMO model to 33 and 31 for the LUMO model.

For the model development, we employed the XGBoost ensemble-based method as we found it to perform better than the random forest (RF) method in our previous study [336]. Another reason for choosing an ensemble-based method over a 'black-box' model such as a neural network is to get further insights on the importance of the feature for HOMO/LUMO energy prediction.

The hyper-parameters for the XGBoost model are determined using a randomized grid search method with a 5-fold CV implemented in Scikit-learn [144]. This includes the number of trees for the XGBoost model (`n_estimators`), the maximum depth of each tree (`max_depth`), the learning rate of the gradient descent algorithm (`learning_rate`), the fraction of data used to build individual tree (`subsample`), the fraction of features to choose from to build individual tree (`colsample_bytree`) and the fraction of features used to train individual node in a tree (`colsample_bylevel`). The entire hyper-parameter space and the final optimized parameters are shown in Figure E.2 and E.3.

8.4 Results and discussion

8.4.1 Model validation

The performance metrics for HOMO and LUMO energies predicted using the XGBoost model are shown in Table 8.1. The correlation coefficient (R^2) stays above 0.90 for all three sets for both energies. The RMSE and MAE for both the HOMO and LUMO models are less than 0.1 eV for all three sets, suggesting that it is possible to correlate DFT properties with high accuracy using two-dimensional descriptors. This is even more impressive because all the cation data points are unique cation structures with varying alkyl chain lengths, branching differences, and attachment positions. The correlation plots for both the HOMO/LUMO energy can be seen from Figure E.4 and E.5. The maximum relative deviation stays below 3.0% for both of the models.

Property	Data	R^2	RMSE (eV)	MAE (eV)
HOMO	Train	0.998	0.015	0.011
	Validation	0.957	0.064	0.048
	Test	0.914	0.081	0.061
LUMO	Train	0.996	0.006	0.004
	Validation	0.945	0.028	0.022
	Test	0.906	0.037	0.026

Table 8.1: Performance metrics of the XGBoost Model for HOMO (eV) and LUMO (eV) energies. (R^2) is the correlation coefficient, (RMSE) is root mean squared error and (MAE) is mean absolute error.

8.4.2 External test validation

Next, we evaluated the model’s performance on all the external test sets highlighted by the (-) and (+) symbols seen in Figure 8.3 compared to DFT calculations. As mentioned earlier, cations denoted by (-) indicate all the smile structures for that

particular cation are utilized for evaluation, while (+) indicates the selection of 20 data points from the entire list of possible smile structures. The goal of the external test is to evaluate the generalization ability beyond the model cation dataset compared to DFT calculations.

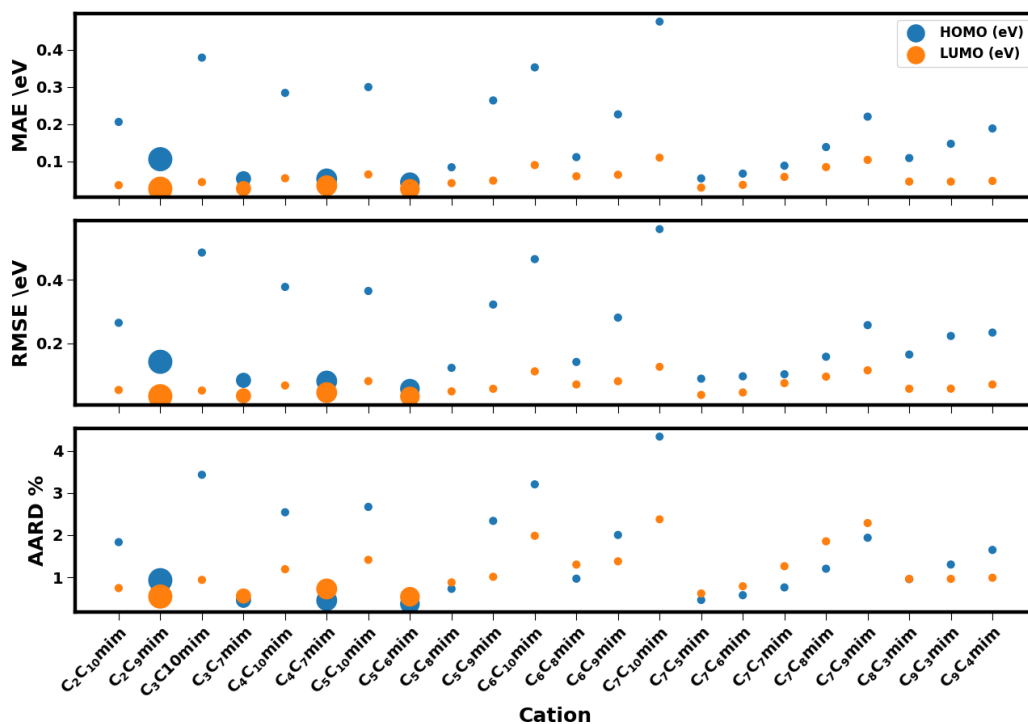


Figure 8.3: Performance metrics for the external set using HOMO/LUMO energy model. The cations listed here are not present in the model development data. The size of the marker varies based on the amount of branched isomer data present for the individual cation set.

Overall, the LUMO model has a much lower RMSE/MAE/AARD for all the cation types than the HOMO model. One reason for this is that the LUMO energy is primarily dominated by the cation ring, as there is very little change in energy with the increase in alkyl chain length or branching [42]. However, HOMO energy for shorter alkyl chain cations is also located on the imidazolium ring, which shifts to the alkyl chain and branching as the length, and the complexity of the chain increases [350].

Compared to the test set as seen in Table 8.1, the MAE/RMSE for the external test reaches a maximum of 0.4 eV for HOMO energies and 0.2 eV for LUMO energies for the external test set, as seen from Figure 8.3 . The training/test data contains cations of the same alkyl chain length and position, with the only difference being the variation in isomer branching. That could explain why the model has MAE/RMSE lower than 0.1 eV for the test set. However, for this external test set, none of the cations with this alkyl chain length and branching attached to both sides are present in the model development, thus reflecting the actual generalization ability. Nevertheless, MAE/RMSE less than 0.4 eV and AARD less than 4% can still be considered very high accuracy compared to DFT calculations, as this removes the computational need of generating data for the rest of the cations.

8.4.3 Feature importance

A major advantage of the ensemble-based method over neural networks is the easy interpretability of the model decision-making and the feature importance of the output. In this work, we employ Shapely Additive Explanation (SHAP) [282] similar to other studies [283, 284, 351] to gain insights on some of the most important features influencing HOMO/LUMO energies. Figure 8.4 (a) depicts the SHAP analysis of feature importance for the HOMO model. The first on the list is the Balabanj feature, where a lower value of this feature (denoted by the blue color) lowers the model output. The Balabanj feature is a topological descriptor that considers the structural complexity [352]. As the training set consists of cations with various alkyl chain lengths and levels of chain branching for the isomers, it is understandable why the Balabanj feature is the most important as it takes into account all of that structure complexity concerning HOMO energy.

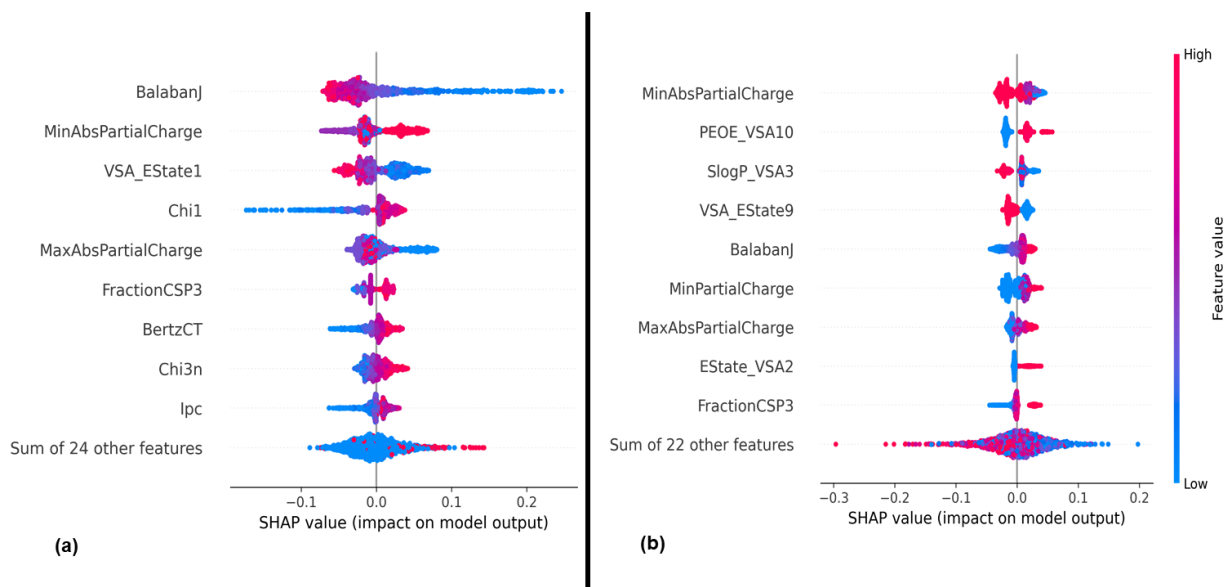


Figure 8.4: SHAP feature importance for (a) HOMO energy model and (b) LUMO energy model.

The Min/MaxAbsPartialCharge descriptors are the absolute minimum and maximum of the partial charge calculated using the Gasteiger method based on the electronegativity of the atoms [290]. These descriptors consider the electron-donating and withdrawing capability of a molecule directly related to HOMO/LUMO energies. The third important descriptor is the VSA_EState1 descriptor, the electrotopological state for the atoms with van der Waal's surface area of less than 4.78 [289, 323]. Electrotopological descriptor considers the topological and the electronic state of the molecules, where electronic state refers to electron distribution on the molecule. The Chi1 and Chi3n descriptors take into account the cations' connectivity index. As there are various levels of branching for the isomers attached to the cations, the influence of these two descriptors makes sense. The FractionCSP3 descriptor considers the number of carbon atoms that are sp³ hybridized.

The BertzCT descriptor is a topological descriptor that relates to the complexity of the molecule through graph theory [353]. The descriptor is the sum of two quantities: the complexity of the atoms and the complexity of the connectivity.

Thus, it makes sense why this descriptor is influential for the HOMO energy as for isomers, the complexity of the atom term remains the same since they have the same number and type of atoms. This leaves the second connectivity term to consider the subtle difference between the two isomers for the effect of branching concerning HOMO energy. Lastly, the IPC descriptor accounts for the information content of molecules, such as the number of atoms, the bonds, and the branching complexity [285]. The importance of this descriptor ties with the complexity of the branching on the alkyl chain with HOMO energy.

As for the LUMO model seen from Figure 8.4 (b), the most important descriptor is the MinAbsPartialCharge that calculates the minimum of the Gasteiger partial charge based on the electronegativity of the atoms in the molecule. The following three influential descriptors are based on van der Waal's surface area (VSA), where the PEOE_VSA10 descriptor is the distribution of partial charge based on van der Waal's surface area for a partial charge between 0.10 and 0.15. SlogP_VSA3 descriptor is the octanol-water partition coefficient for atoms with a specific van der Waal's surface area. It is unclear why this descriptor is one of the most influential for LUMO energy. The remaining descriptors are similar to the HOMO model, such as Balabanj, Min/Max Partial charge, and FractionCSP3.

8.4.4 Entire isomer chemical space mapping

As the external set is validated against DFT prediction with excellent accuracy, we proceeded to calculate the HOMO and LUMO energies using the model for the entire chemical space shown earlier in Figure 8.1. This includes 317,382 unique imidazolium cations with an alkyl chain varying from 1 to 10 with all possible isomers. Figure 8.5 (a) shows the energy gap calculated by taking the difference in LUMO and HOMO energy for the training set using DFT calculations. The cations are divided into three categories based on chain length. The energy gap for 1-methyl-3-alkylimidazolium $[C_1C_{n>=2}mim]^+$ cations vary the most, from 7.25 eV to 5.50 eV. The lowest energy gap is for the long alkyl chain cations with HOMO located at the tail of the alkyl chain. Several other studies have seen a similar

trend with the increase in alkyl chain length attached to the cation [343, 350]. The energy gap for the next cation category ($[C_{m>=2}C_{n>2}mim]^+$) does not vary as much as the previous category, while the third category ($[C_{m>=1}C_{m=>1}mim]^+$) has the least change in energy gap, indicating high stability compared to the other two categories.

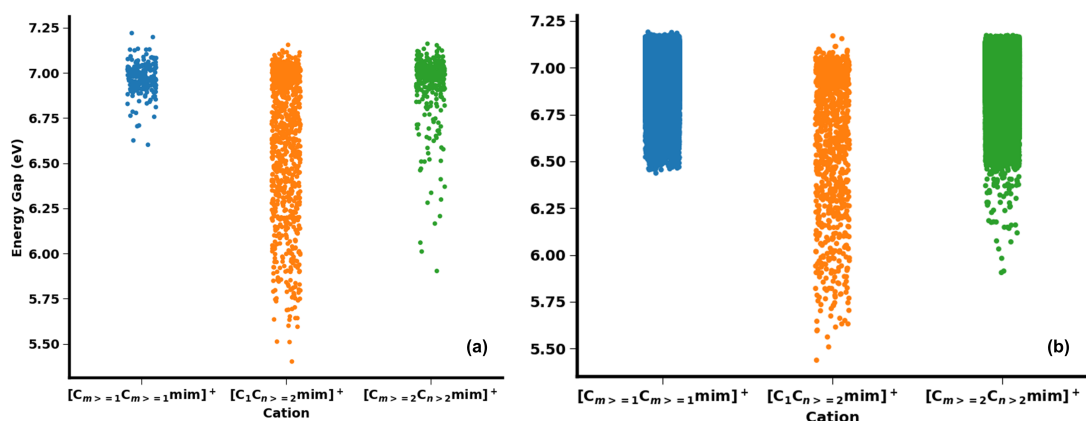


Figure 8.5: Energy gap of the entire $[C_n C_m mim]^+$ cations using (a) DFT for the 1399 data set for model development and (b) XGBoost model for 317,382 cations.

The energy gap for the entire isomer space is mapped using the XGBoost model for HOMO and LUMO energy by taking the difference in energies as depicted in Figure 8.5 (b). The trends observed from the model are in excellent agreement with DFT calculations. Calculating the entire energy gap for 317,382 cations using a personal computer with an Intel (R) Core (TM) i7-10750H CPU 2.60GHz and 16 GB of RAM took approximately four hours. Performing the same HOMO/LUMO calculations using DFT for a single cation takes about three to four hours, depending on the alkyl chain length, using high-performance computing (HPC) with Intel Xeon Gold 6130 CPUs using the 32-core specification. This drastically cuts down the computation time required for the entire chemical space using DFT calculations, which otherwise might have required more than a million hours using HPC.

8.5 Conclusion

This work attempts to build a machine learning model to correlate frontier orbital energies for imidazolium-based cations with alkyl chains varying from 1 to 10, including isomers. Enumerating the entire cation space leads to 317,382 unique cations, for which running DFT calculations on each of them would be unfeasible due to high computational cost. Instead, we utilize a machine learning approach using the XGBoost method to correlate HOMO and LUMO energies for a small sample set of cations assumed to represent the entire isomer space. The trained model can accurately predict the energies for the external test set, which was not part of the training data, with MAE of less than 0.4 eV for HOMO and 0.2 eV for LUMO energy compared to DFT calculations. Besides accuracy, the model can also map the entire HOMO/LUMO energy space for 317,382 cations within four hours on a personal computer. As demonstrated in this study, the machine learning method could be used in situations where DFT calculations are not easily accessible or are too expensive to run on a large-scale system.

CHAPTER 9

CONCLUSION

9.0.1 Conclusions

In conclusion, the work done in this dissertation utilizes a computational-based approach toward the search and design of ionic liquids as electrolytes for battery applications. The motivation for this work began with the background on ionic liquids and the limited amount of data present for ionic liquids with high ionic conductivity. In addition, the ones with high ionic conductivity, however, have extremely low electrochemical stability, making them undesirable for battery applications.

The first part of the work attempts to build a machine learning model for expanding ionic conductivity data for imidazolium ionic liquids, as it is one of the most widely studied ionic liquids cations with low viscosity and high ionic conductivity. The model of choice for this study is a feed-forward artificial neural network and support vector machine, as they are well suited for large data and can capture complex non-linear behavior. The ionic conductivity values of the imidazolium experimental data set ranged over six orders of magnitude and covered temperatures from 275 K to 475 K. The neural network model performed slightly better than the support vector machine on the test set and much better on the external test for unique ionic liquid combination. Utilizing the available cations and anions in the database, we combined them to form 1102 unique ionic liquids. The procedure

led to the identification of the ionic liquid [C₁mim][DCA] that is not present in the training set with an ionic conductivity of 3.70 S/m, which is 30% higher than the highest ionic conductivity in the training set and the NIST database. Furthermore, combining cations and anions as a function of mole fraction leads to a large number of binary mixtures. Comparison between experimental and model mixture data showed excellent agreement with low root mean square deviation. Several binary mixtures also showed non-ideal behavior, such as enhancement or suppression in ionic conductivity as a function of mole fraction with values greater or lower than the pure ionic liquid values.

The second part of the work is a molecular dynamics study to understand the molecular-level interactions of ionic liquid mixtures to explain the macroscopic non-ideal deviation in properties such as ionic conductivity. Radial distribution function (RDF) analysis showed preferential interaction between the short alkyl chain cation and strongly coordinating anion compared to the longer alkyl chain cation and weakly coordinating anion present in this reciprocal mixture. This preferential interaction is highest for the reciprocal mixture, followed by the binary anion, the binary cation mixture, and the pure system. Hydrogen bond dynamic analysis follows a similar trend as the RDF peak height with the same particular cation-anion exhibiting an extended hydrogen bond lifetime in the reciprocal mixture.

As mentioned earlier, imidazolium ionic liquids are the most commonly studied cations because of their low viscosity and high conductivity, but they have very low electrochemical stability compared to other cations types. Thus, a tremendous amount of research is being undertaken to find new cations with desirable properties. In the third part of the work, we expand the machine learning model to cover ten available cation types, including imidazolium cations. For the model development, we employed ensemble-based methods such as Random Forest (RF) and Extreme Gradient Boosting (XGBoost) that are robust, less prone to overfitting, and can provide insights into the model's decision-making. The XGBoost model per-

forms consistently better for all the cation types regardless of the amount of data present, unlike the RF model, which is sensitive to the amount of data present for individual cation types. The feature analysis technique showed six cation features, two anion features, and temperature as the most influential features concerning ionic conductivity.

The motivation behind the fourth part of the work began with the work done in the previous chapter. Despite expanding the ionic conductivity data to 14,000+ ionic liquids, fewer than twenty-one ionic liquids have ionic conductivity comparable to conventional organic electrolytes. This work utilizes a generative-based machine learning model to accelerate the discovery of cations in search of ionic liquids with desirable properties for battery application. The generative model is based on a variational autoencoder (VAE) that compresses input information to a low-dimensional latent space that serves as a reservoir to sample and generate new data similar to the input data. The VAE model is trained on commonly studied cation SMILE strings based on imidazolium, pyrrolidinium, piperidinium, pyridinium, sulfonium, ammonium, and phosphonium cations. The cation data generated from the VAE undergoes post-processing steps that involve duplicate check, charge validation, and stability tests using QM calculations to ensure the cations exhibit a stable structure. Sampling the latent space generated 756 new unique cations, of which 425 formed stable structures based on the QM stability test. These cations are classified into respective families using a classification model. The model classified 76 new aromatic and cyclic cations as "Unknown" types. A two-dimensional principal component analysis projection of the cation chemical space revealed new regions of chemical space occupied by the newly discovered cations labeled "Unknown". The Tanimoto similarity index of the cations provided further confirmation of the novelty of the new cation family. Electrochemical window calculations for these "Unknown" cations showed some of the cations to exhibit very high electrochemical stability suitable for battery application.

Some of the cations, when paired with an anion such as tetrafluoroborate anion,

exhibit an electrochemical window up to 7.0 V, which is exceptionally high considering the electrochemical window of most conventional ionic liquids that are found in the range of 4.0 V [11]. Despite such a high electrochemical window most of the commercially studied cathode materials are known to have degradation issues at high voltage (> 4.5 V vs Li/Li⁺), reducing the life cycle capacity of the electrode [354, 355]. Nevertheless, it is still desirable to have ionic liquids with such a high electrochemical window as this dissertation aims to expand the list of potential candidates with desirable electrolyte properties.

The final part of the work aims to develop a machine learning model to correlate DFT-derived properties such as highest occupied molecular orbital (HOMO) and lowest occupied molecular orbital (LUMO) energy, which is extremely important to determine the stability of the molecule, toxicity, and biodegradability. The motivation for this begins with the need to calculate the HOMO/LUMO energy of cations for electrochemical stability calculations. Running expensive DFT calculations on each one would be extremely costly, time-consuming, and almost unfeasible on a large scale. Thus, this last work aims to build a machine learning model based on two-dimensional descriptors to correlate HOMO/LUMO energy. The machine learning model can accurately predict the energies for the external test set with an MAE of less than 0.4 eV for HOMO and 0.2 eV for LUMO energy compared to DFT calculations.

9.0.2 Future Work

Besides generating new ionic conductivity data using a machine learning model, it would also be desirable to validate and gain insights at a molecular level using molecular dynamics (MD). However, the current forcefields for ionic liquids severely underpredict ionic conductivity, making it challenging to rely on data generated from MD simulations. Developing forcefields to improve ionic conductivity prediction and expanding them to a large number of ionic liquids would help expand the generation of new data and increase the ionic conductivity space in search of potential candidates with a desirable range.

The variational autoencoder (VAE) method in this study is limited to generating novel cations without any property prediction. However, several studies have shown the possibility of combining a property predictor and a VAE model to simultaneously generate new data and property known as conditional variational autoencoder. This could be useful to estimate the property range for the newly discovered materials without additional calculation.

Several studies have shown that the accuracy of the electrochemical window calculation using DFT depends on the choice of the theory and basis set [127, 128]. Thus, it would be desirable to have a machine learning model to predict energy for multiple theories and basis sets as required for a given application with the desired level of accuracy. However, this would require the development of models trained on multiple theories and a basis set. Running DFT calculations on each instance to generate a large amount of training data would not be feasible due to the extremely high computational cost of running such a large set of calculations. Using a transfer learning approach by retraining a base model with a small amount of data for other theories and basis sets might help solve this issue.

For any machine learning model development, the first step includes data collection and processing. In the case of ionic liquids, NIST has maintained a repository of experimental data for various properties that can be used to download experimental data [356]. The entire downloading process can be automated with the use of a publicly available tool (pyilt) [142]. However, besides the NIST IL thermo database, there is no other major data repository of experimental data for ionic liquids. In such a case, the data gathering procedure has to be done manually for the property of interest through literature, making the protocol time-consuming and prone to error. Thus it would be desirable to have a natural language processing (NLP) script that scans over journals published online for experimental data for a given property of interest, similar to the work done in other material science domains [357, 358].

CHAPTER A

Appendix: Developing Machine Learning Models for Ionic Conductivity of Imidazolium-Based Ionic Liquids

Performance Metrics

Average Absolute Relative Deviation

$$AARD\% = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i^{pred} - y_i^{exp}}{y_i^{exp}} \right| \quad (A.1)$$

Root Mean Square Deviation

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

Mean Absolute Error

$$MAE = \frac{\sum_{i=1}^N |y_i^{pred} - y_i^{exp}|}{N} \quad (A.3)$$

Mean Squared Error

$$MSE = \frac{\sum_{i=1}^N (y_i^{pred} - y_i^{exp})^2}{N} \quad (A.4)$$

Relative Deviation

$$RD\% = \frac{(y_{exp} - y_{pred})}{y_{exp}} * 100 \quad (A.5)$$

Mean Absolute Percentage Error

$$MAPE = \frac{100}{N} \frac{\sum_{i=1}^N |y_i^{pred} - y_i^{exp}|}{y^{exp}} \quad (A.6)$$

Where:

N is the number of data points

y^{exp} is experimental data

y^{pred} is predicted data using model

Temperature Distribution Profile

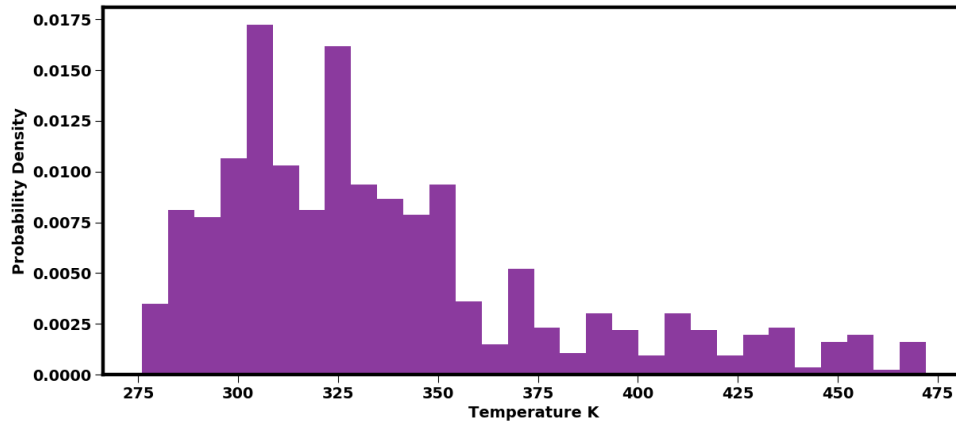


Figure A.1: Temperature distribution profile of all the experimental data points used in the model development.

Experimental Ionic Conductivity Distribution Profile

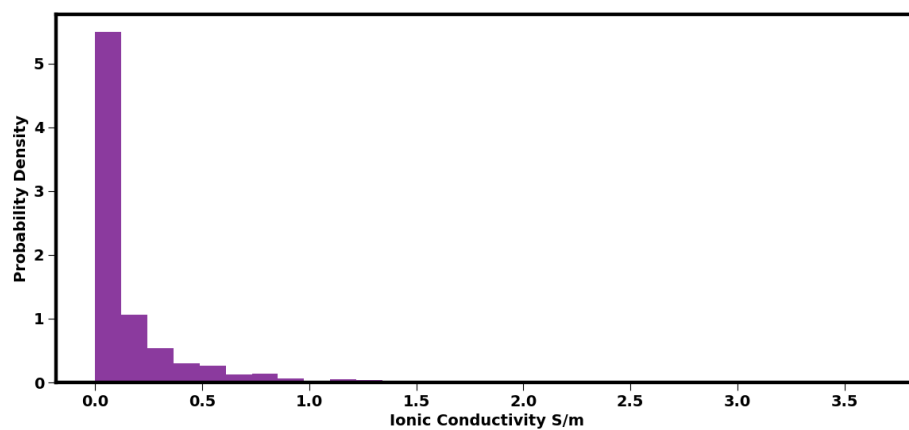


Figure A.2: Ionic conductivity distribution profile of all the experimental data points used in the model development.

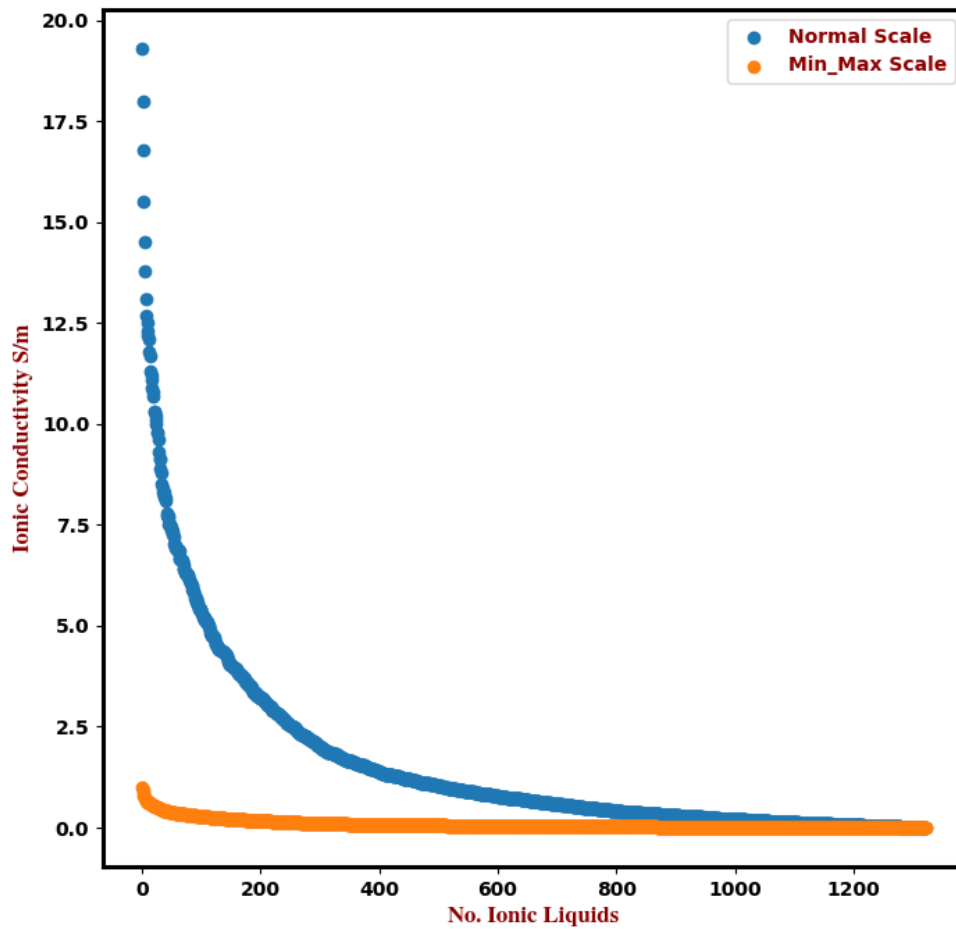


Figure A.3: Comparison between experimental data in normal scale and transformed data using Min Max Scaler to scale the data between 0-1 before model development.

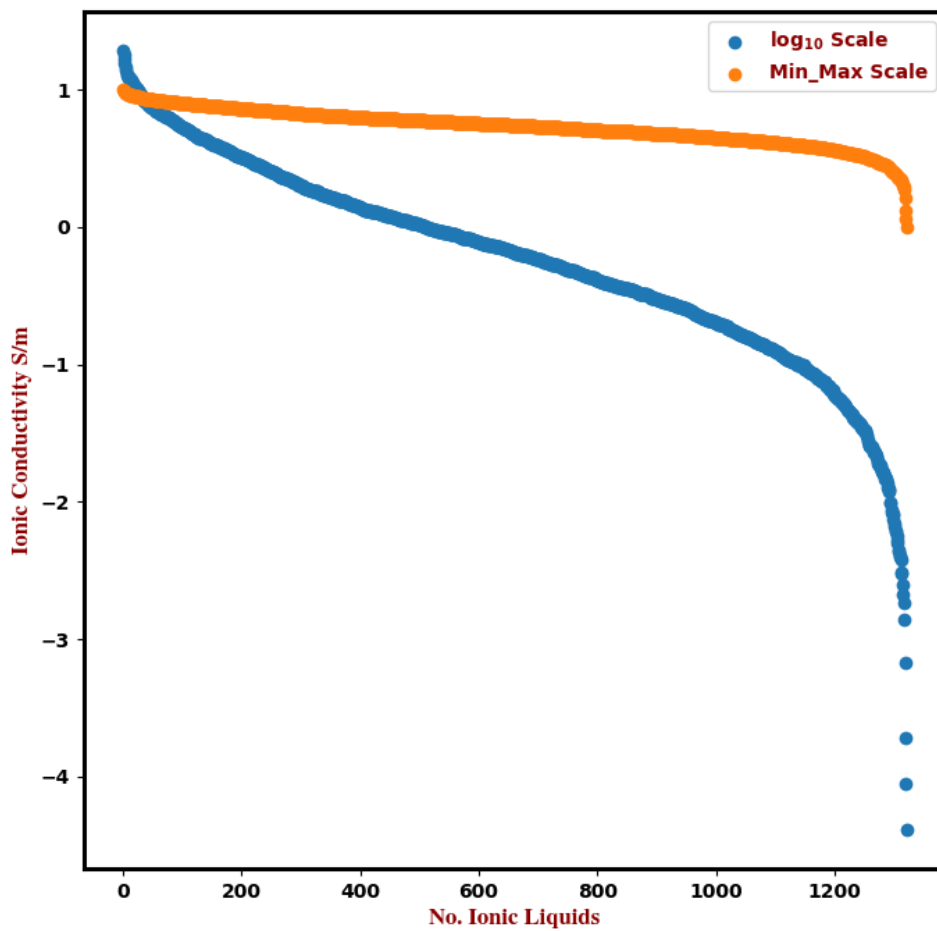


Figure A.4: Comparison between experimental data in \log_{10} scale and transformed data using Min Max Scaler to scale the data between 0-1 before model development.

Correlation HeatMap

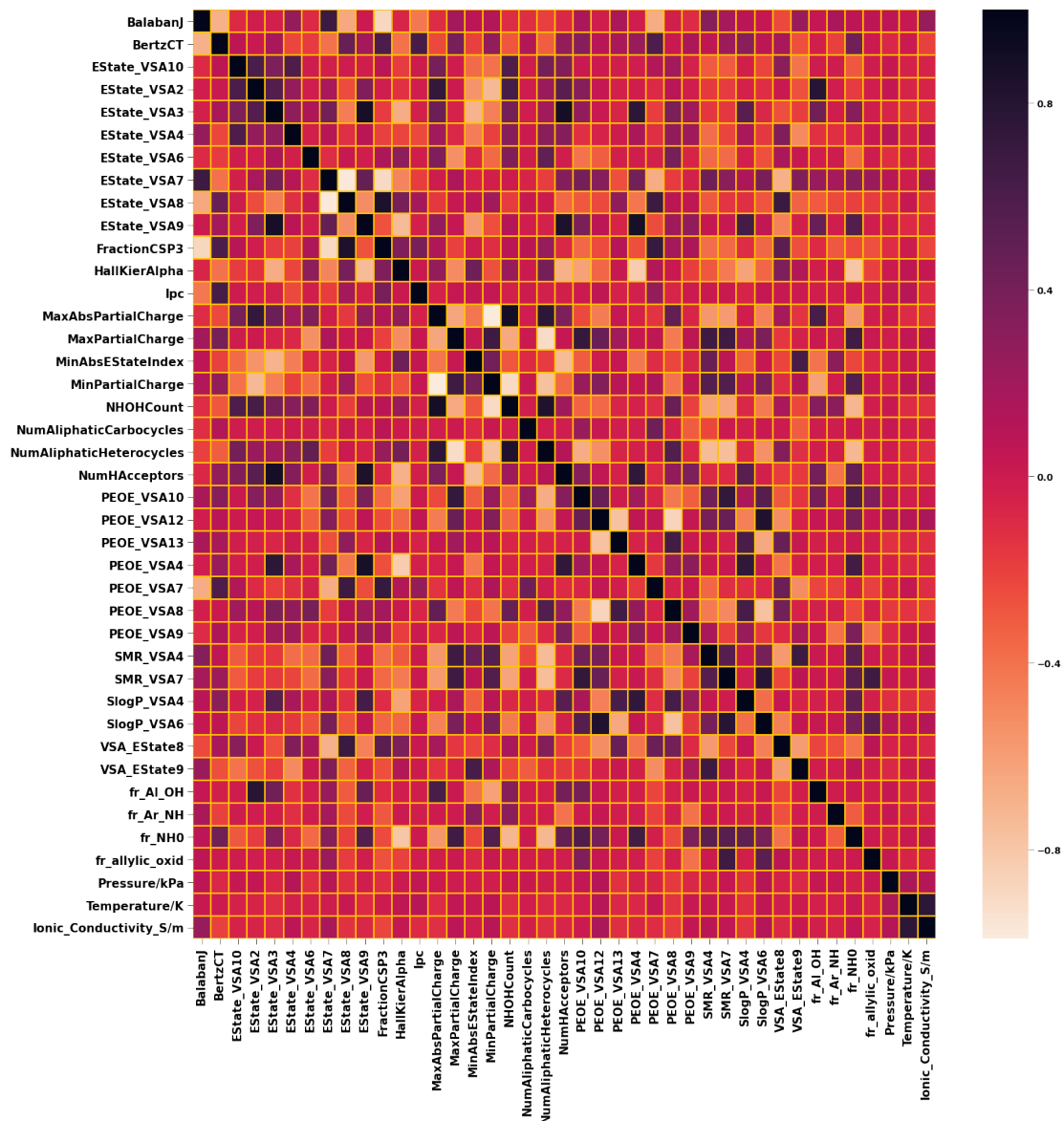


Figure A.5: Heat map correlation of cation features with experimental ionic conductivity data.

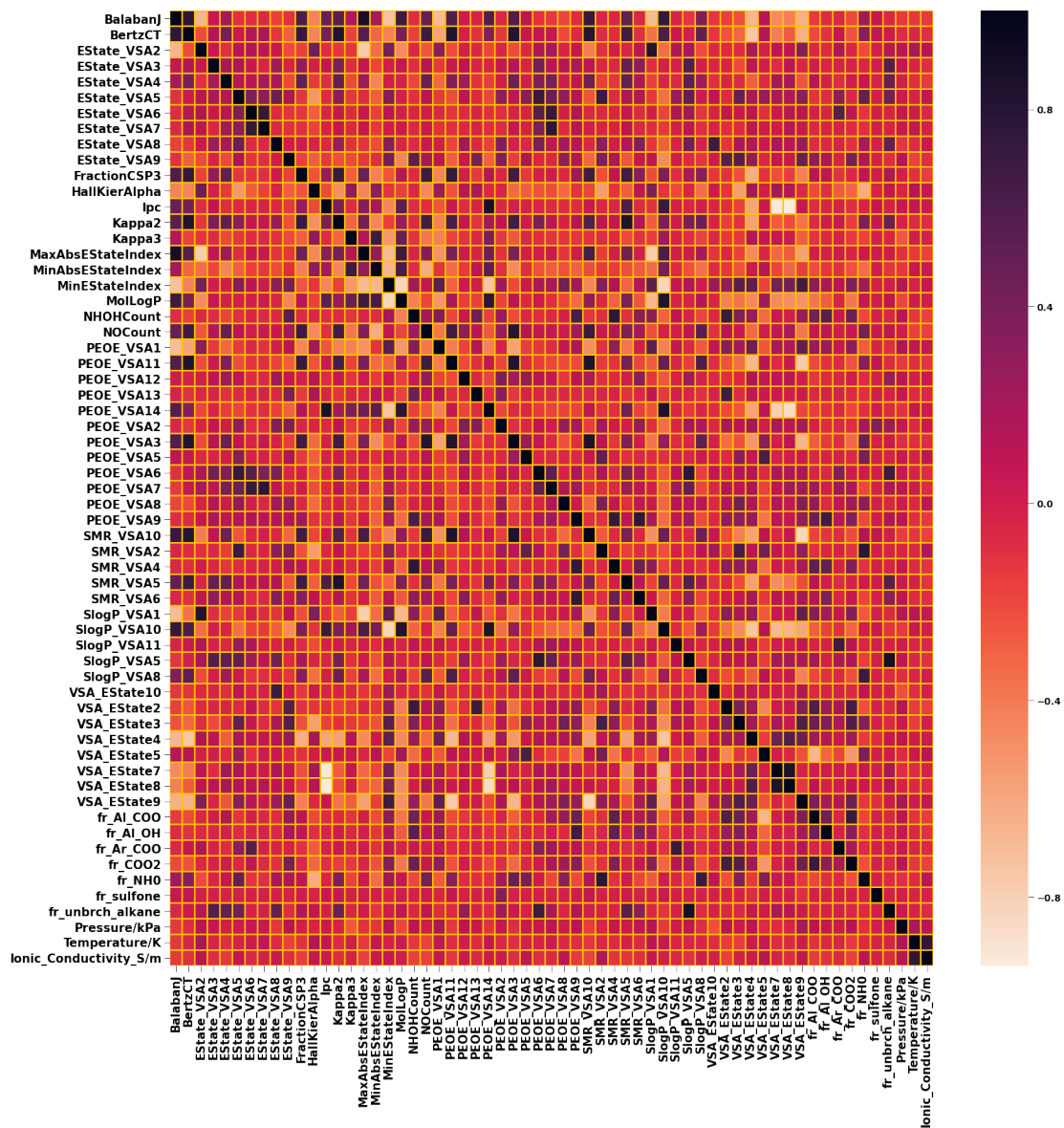


Figure A.6: Heat map correlation of anion features with experimental ionic conductivity data.

FFANN Hyper-parameter Search Space

Number of nodes in the first hidden layer was set to the number of input features
: 99

Search Space:

Learning Rate = [0.001,0.002,0.003,0.004,0.005,0.006,0.1]

Neurons=[15,20,21,22,23,24,25]

epochs = [50,150,250,350,450,550]

batch size = [100,300,500,700]

Default Parameters:

loss='mse'

metrics=['mae','acc','mse']

Final Hyper-parameter:

Learning Rate : 0.002

Number of Hidden layers : 2

Number of neurons in Hidden Layer1 : 99

Number of neurons in Hidden Layer2: 24

Activation Function : relu

epochs : 450

batch size : 100

SVR Hyper-parameter Search Space

Search Space:

```
estimator=SVR(kernel='rbf'),
```

Search Space:

```
'C': [0.1, 1,10,25,50,100,110,120,130,140,150,  
160,170,180,190,200,250,350,500]
```

```
'epsilon': [0.0001,0.0002,0.0003,0.0004,0.0005,  
0.001,0.002,0.003,0.004,0.005]
```

```
'gamma': [0.0001,0.0002,0.0003,0.0004,0.0005,  
0.0006,0.0007,0.001, 0.005,0.006,0.007,0.008,  
0.1,0.2,0.3,0.4,1, 3, 5]
```

Final Hyper-parameter:

```
'C': 250
```

```
'epsilon': 0.001
```

```
'gamma': 3
```

Final FFANN Topology

Model: "sequential_2"

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 99)	9900
dense_7 (Dense)	(None, 24)	2400
dense_8 (Dense)	(None, 1)	25

=====
Total params: 12,325
Trainable params: 12,325
Non-trainable params: 0

Figure A.7: FFANN model topology. Total number of data points in the training set : 1190. Ratio of weight and bias to the total number of data points in the training set : 10.35

Test Train Partition

The random state of the test/train split is also known to affect the model performance. Thus we looked at how the FFANN and SVR model’s true performance varies with different random state. This was done by changing the random state over 90 different points. The average and standard deviation of this evaluation is reported below.

Table A.1: Comparison of the predictions results for FFANN and SVR for the training set and test set by varying the partition of the test/train split. This was done by changing the random state parameter in sklearn. MSE is the mean squared error, MAE is the mean absolute error, RMSD is root mean square deviation and R^2 is the squared correlation between experiment and predicted data. \log_{10} scale refers to ionic conductivity scaled to \log_{10} , while normal scale refers to ionic conductivity data without any scaling in S/m.

Scale	Metric	Train		Test	
		SVR	FFANN	SVR	FFANN
\log_{10} scale	R^2	0.995±0.001	0.991±0.004	0.982±0.012	0.974±0.024
	MSE	0.002±0.001	0.005±0.002	0.009±0.007	0.014±0.014
	MAE	0.015±0.001	0.050±0.015	0.029±0.007	0.062±0.017
	RMSD	0.050±0.002	0.070±0.014	0.091±0.007	0.111±0.045
Normal scale	R^2	0.992±0.007	0.977±0.036	0.997±0.002	0.971±0.045
	MSE	0.004±0.002	0.130±0.204	0.015±0.013	0.166±0.254
	MAE	0.028±0.008	0.151±0.074	0.048±0.012	0.169±0.084
	RMSD	0.068±0.004	0.321±0.166	0.118±0.043	0.357±0.197

FFANN final model validation

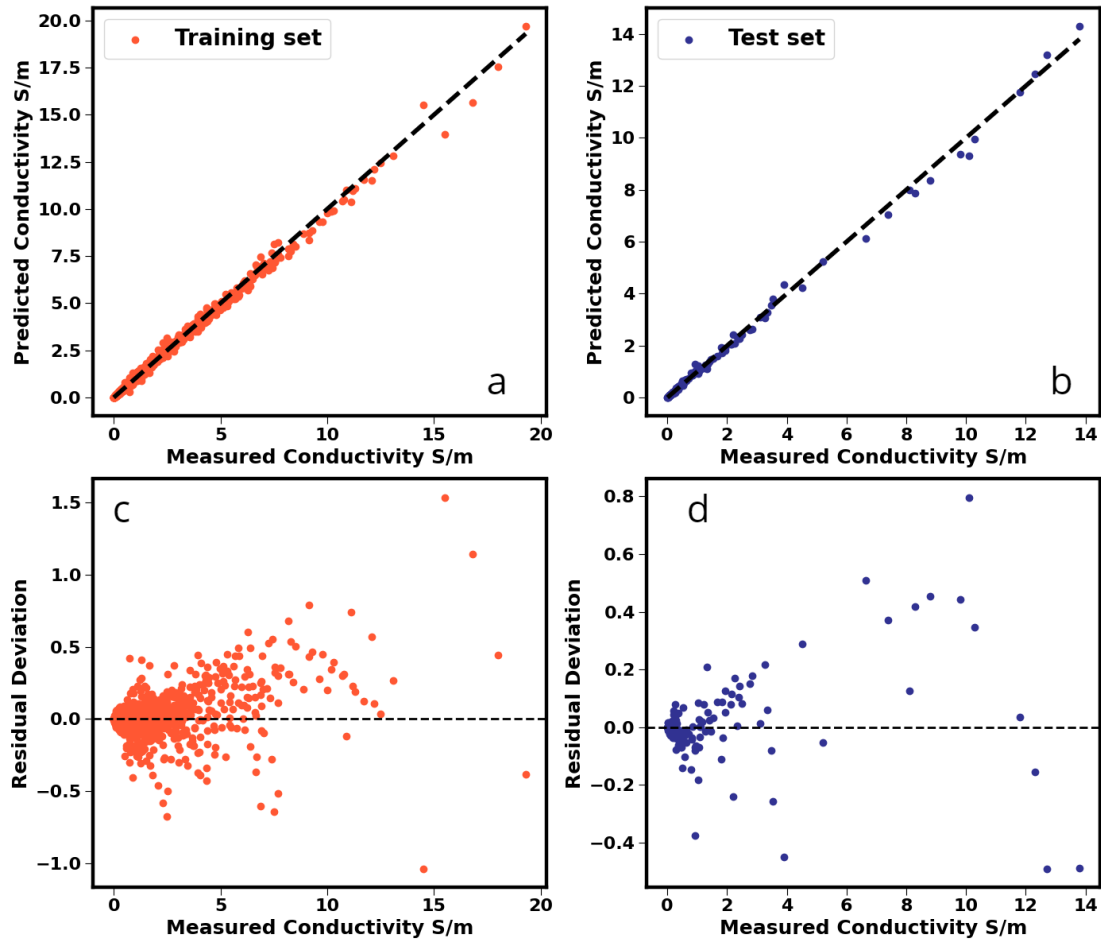


Figure A.8: (a) Comparison of SVR model predictions with the experimental data on normal scale for the training set. A perfect prediction would fall on the $y = x$ dotted line; (b) comparison for the test set (c) Residual deviation on the normal scale calculated as $(\sigma_{\text{experiment}} - \sigma_{\text{prediction}})$ where σ refers to the ionic conductivity for the training set; (d) Residual deviation for the test set.

SVR final model validation

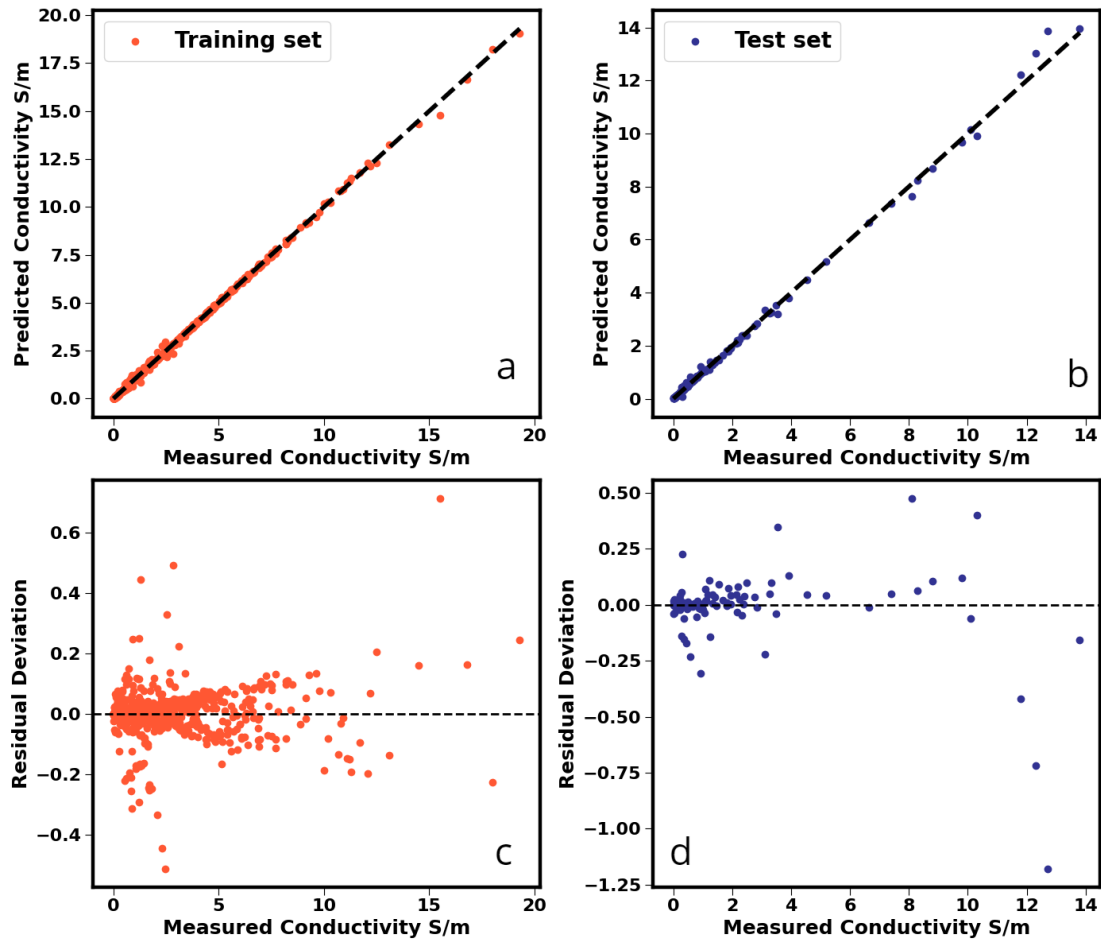


Figure A.9: (a) Comparison of SVR model predictions with the experimental data on normal scale for the training set. A perfect prediction would fall on the $y = x$ dotted line; (b) comparison for the test set (c) Residual deviation on the normal scale calculated as $(\sigma_{\text{experiment}} - \sigma_{\text{prediction}})$ where σ refers to the ionic conductivity for the training set; (d) Residual deviation for the test set.

Table A.2: Comparison of the predictions results for FFANN and SVR for the training set, test set and the entire dataset. MSE is the mean squared error, MAE is the mean absolute error, RMSD is root mean square deviation and R^2 is the squared correlation between experiment and predicted data. MAPE is mean absolute percentage error in %.

Scale	Metric	Train		Test		Entire	
		SVR	FFANN	SVR	FFANN	SVR	FFANN
Normal scale	R^2	0.999	0.995	0.997	0.995	0.999	0.996
	MSE	0.003	0.021	0.023	0.024	0.005	0.022
	MAE	0.027	0.079	0.059	0.086	0.031	0.081
	RMSD	0.061	0.148	0.153	0.160	0.075	0.149
	MAPE	3.58	8.50	9.03	10.01	4.14	8.65

Ionic Conductivity Comparison

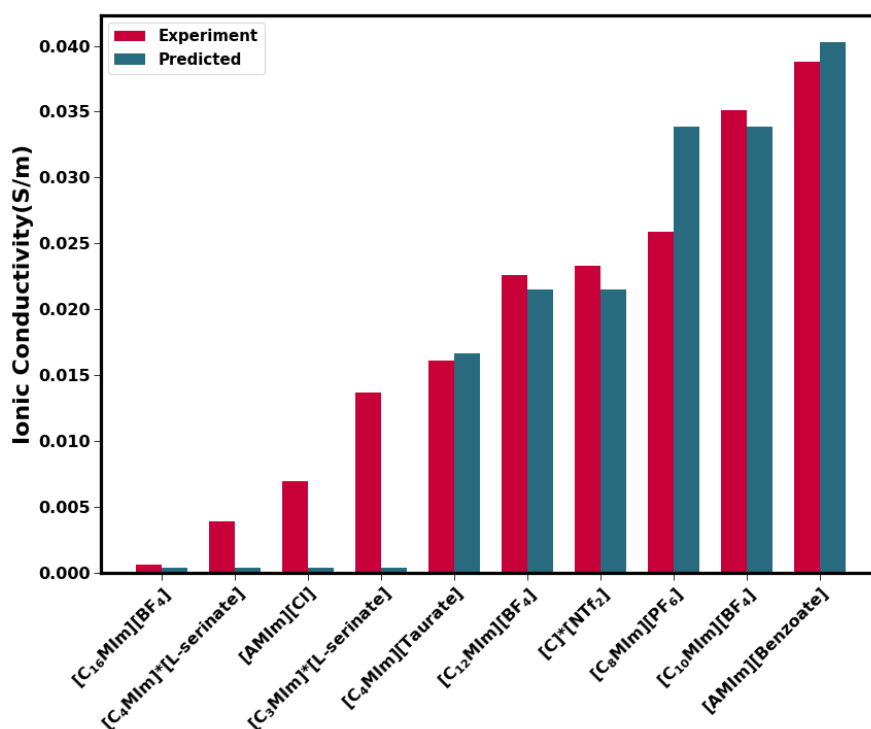


Figure A.10: Comparison between experiment and model for ten ionic liquids at the bottom of the ionic conductivity values at 298 K using FFANN model. [C₄mim]^{*}: 1-butyl-3-methyl-1H-imidazolium; [C₃mim]^{*}: 1-propyl-3-methyl-1H-imidazolium; [C]^{*}: 3-(2-(butylamino)-2-oxoethyl)-1-ethyl-1H-imidazolium; [Amim] stands for 1-allyl-3-methylimidazolium.

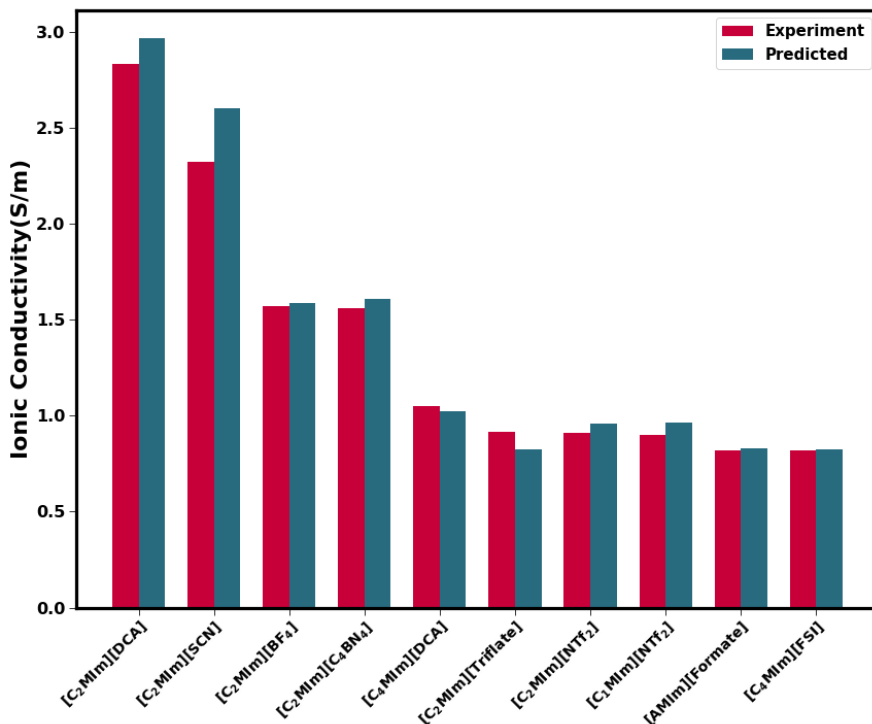


Figure A.11: Comparison between experiment and model for top 10 ionic liquids with the highest ionic conductivity at 298.0 K using FFANN model. [AMIm] stands for 1-allyl-3-methylimidazolium.

Model Comparison

Table A.3: Comparison with experimental data for the GC method and FFANN model for the $[C_{n=(2,4,6,8)}\text{mim}][\text{NTf}_2]$ subset series at various temperatures. The experimental data and GC data for this predictions were taken from the supporting information of Chen et al. work [113].

Method	R^2	MAE	RMSD	MSE
GC	0.985	0.035	0.074	0.005
FFANN	0.986	0.053	0.072	0.005

Alkyl Chain length Dependency

As a further demonstration of the predictive power of the proposed model, we provide additional alkyl chain length-dependent behavior of several ionic liquids for which ionic conductivity data are available in the literature. Zhou et al. reported several physiochemical properties including ionic conductivity for alkyltrifluoroborate and alkenyltrifluoroborate anions paired with the $[\text{C}_2\text{mim}]^+$ cation [359]; none of the anions except the nonalkylated analogue $[\text{BF}_4]^-$ is present in the training dataset. We calculated the ionic conductivity for these ionic liquids the two models model and compared the values with the experimental data (Figure A.12). Remarkably, the ionic conductivity predictions are in reasonable agreement with those reported from experiments with a MSE of 0.138 using FFAAN and 0.340 for SVR method. In addition, the model captures the trend of decreasing ionic conductivity with the alkyl chain length, the trend reported for the effect of increasing alkyl chain length in the cation on ionic conductivity. The comparison provides a further evidence that the FFANN model is flexible enough to handle combination of functional groups if they are present in the original dataset. It can also be inferred that the tuning of the hyper-parameter has not led to over-fitting of the training data.

Table A.4: Comparison between experiment and model for ionic conductivity of alkyl and alkenyltrifluoroborate anions paired with $[\text{C}_2\text{mim}]^+$ cation at 298.15 K using FFANN model and SVR model. Experiment data were taken from ref [359].

No.	Ionic Liquid	T/K	Exp(S/m)	FFANN	SVR
1	$[\text{C}_2\text{mim}][\text{CH}_3\text{BF}_3]$	298.15	0.9	1.49	0.06
2	$[\text{C}_2\text{mim}][\text{C}_2\text{H}_5\text{BF}_3]$	298.15	0.63	1.19	0.06
3	$[\text{C}_2\text{mim}][\text{C}_3\text{H}_7\text{BF}_3]$	298.15	0.57	1.01	0.06
4	$[\text{C}_2\text{mim}][\text{C}_4\text{H}_9\text{BF}_3]$	298.15	0.32	0.24	0.06
5	$[\text{C}_2\text{mim}][\text{C}_5\text{H}_{11}\text{BF}_3]$	298.15	0.27	0.09	0.06
6	$[\text{C}_2\text{mim}][\text{CH}_2\text{CHBF}_3]$	298.15	1.05	1.22	0.07
7	$[\text{C}_2\text{mim}][\text{BF}_4]$	298.15	1.36	1.58	1.54

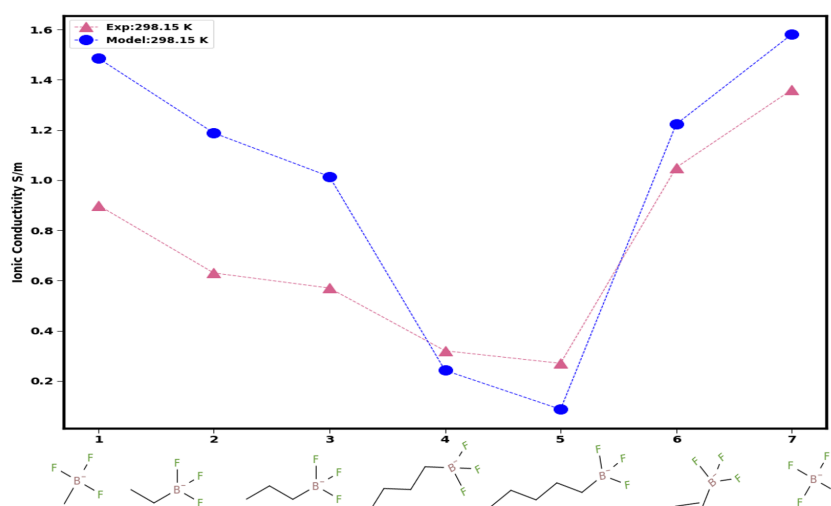


Figure A.12: Comparison between experiment and model for ionic conductivity of alkyl and alkenyltrifluoroborate anions paired with $[\text{C}_2\text{mim}]^+$ cation at 298.15 K using FFANN model. Experiment data were taken from ref [359].

We carried out another test for the model by computing the ionic conductivity for the homologous series of the alkylsulfate $[\text{C}_n\text{SO}_4]^-$ ($n = 1, 2, 4, 6,$ and 8) and hydrogen sulfate $[\text{HSO}_4]^-$ anion paired with the $[\text{C}_2\text{mim}]^+$ cation at 298.15 and 323.15 K in order to compare the calculations to the experimental data reported in the work of Garbal et al. [360]. Note that the anions $[\text{C}_4\text{SO}_4]^-$ and $[\text{C}_6\text{SO}_4]^-$

were not present in the training dataset and hence are test cases for the model. On the other hand, ionic liquids $[\text{C}_2\text{mim}][\text{HSO}_4]$ and $[\text{C}_2\text{mim}][\text{C}_1\text{SO}_4]$ are examples of unique ionic liquid combinations. A comparison presented in Figure A.13 clearly demonstrates that the model performs remarkably well in capturing the nonlinear trend of an increase in the ionic conductivity in going from $[\text{HSO}_4]^-$ to methylsulfate $[\text{C}_1\text{SO}_4]^-$ and a decrease in the ionic conductivity as the alkyl chain length increases. We also observe that the model predictions are close to the experimental data at both temperatures with MSE of 0.032 for FFANN and 0.037 for SVR.

Table A.5: Comparison between experimental data and model outputs for ionic conductivity of hydrogen sulfate and alkylsulfates $[\text{C}_n\text{SO}_4]^-$ anions ($n = 1, 2, 4, 6,$ and 8) paired with $[\text{C}_2\text{mim}]^+$ cation at 298.15 K and 323.15 K using FFANN model and SVR model. Experiment data were taken from ref [360].

No.	Ionic Liquid	T/K	Exp(S/m)	FFANN	SVR
1	$[\text{C}_2\text{mim}][\text{HSO}_4]$	298.15	0.09	0.02	0.01
2	$[\text{C}_2\text{mim}][\text{CH}_3\text{SO}_4]$	298.15	0.76	0.63	0.60
3	$[\text{C}_2\text{mim}][\text{C}_2\text{H}_5\text{SO}_4]$	298.15	0.38	0.40	0.38
4	$[\text{C}_2\text{mim}][\text{C}_4\text{H}_9\text{SO}_4]$	298.15	0.17	0.05	0.06
5	$[\text{C}_2\text{mim}][\text{C}_6\text{H}_{13}\text{SO}_4]$	298.35	0.08	0.05	0.06
6	$[\text{C}_2\text{mim}][\text{C}_8\text{H}_{17}\text{SO}_4]$	298.15	0.05	0.06	0.05
7	$[\text{C}_2\text{mim}][\text{HSO}_4]$	323.15	0.35	0.17	0.01
8	$[\text{C}_2\text{mim}][\text{CH}_3\text{SO}_4]$	323.15	1.69	1.28	1.41
9	$[\text{C}_2\text{mim}][\text{C}_2\text{H}_5\text{SO}_4]$	323.15	1.01	0.99	1.03
10	$[\text{C}_2\text{mim}][\text{C}_4\text{H}_9\text{SO}_4]$	323.15	0.49	0.14	0.06
11	$[\text{C}_2\text{mim}][\text{C}_8\text{H}_{17}\text{SO}_4]$	323.15	0.15	0.17	0.16

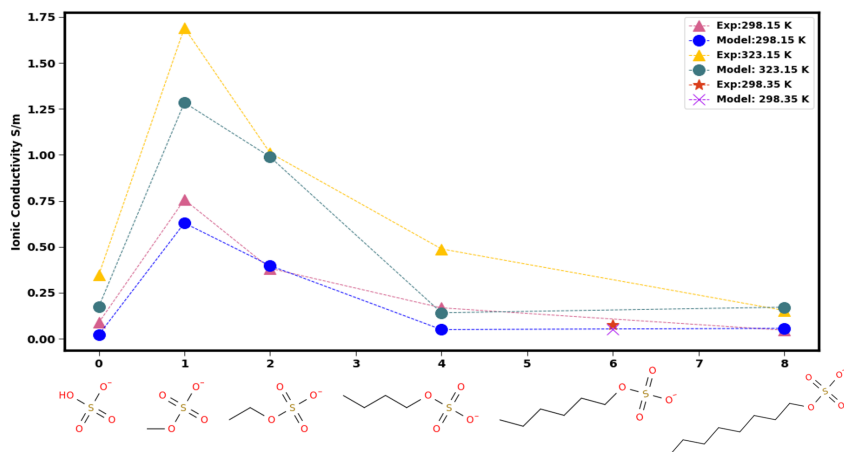


Figure A.13: Comparison between experimental data and model outputs for ionic conductivity of hydrogen sulfate and alkylsulfates $[C_nSO_4]^-$ anions ($n = 1, 2, 4, 6,$ and 8) paired with $[C_2mim]^+$ cation at 298.15 K and 323.15 K using FFANN model. Experiment data were taken from ref [360].

Next, we examined the model prediction capability for functionalized imidazolium cations by substituting various functional groups at R1, R2, and R3 positions as seen in Table A.6. The cations were paired with bis(trifluoromethanesulfonyl)imide $[NTf_2]^-$ anion and compared with experimental data taken from the work of Kakibe et al. [277]. As seen from the table the model does extremely well in predicting ionic conductivity for IL1 to IL5 as indicated by the low relative deviation% using FFANN compared to SVR. We believe that the behavior can be ascribed to the fact that the training data set contained several cations in which the R2 position was substituted with the methyl group. The model also captures the tendency that the substitution of the -H at the R2 position by methyl moiety leads to a decrease in the ionic conductivity which is often seen with imidazolium-based ionic liquids [13,361]. For the other substituted imidazolium-based, the percentage relative deviation increases significantly partly due to the fact that the model was not exposed to the ether-based functional groups such methoxyethyl and ethoxyethyl. Another reason could be related to the low ionic conductivity of these ionic liquids - the range over which the model accuracy for the training data is somewhat lower as compared to that for the higher ionic conductivity regime. This high-

lights the limitation of the model in terms of accuracy for ionic liquids with functional group substitution beyond the training set. Beckner et al. noted similar results when the imidazolium-based models were used to predict the viscosity of non-imidazolium based ionic liquids derived from phosphonium, pyridinium, and pyrrolidinium classes, where the overall MSE on the validation set increased from 0.006 ± 0.01 Pa s to 0.08 ± 0.01 Pa s [176].

Lastly, in Figure A.14, we provide a plot of ionic conductivity as a function of alkyl chain length in the 1-alkyl-3-methylimidazolium cations paired with 38 unique anions predicted using the FFANN model. Two key features are worth highlighting: (a) the model correctly identified the trend of decreasing ionic conductivity with an increase in the alkyl chain length; (b) the influence of the anion on the ionic conductivity diminishes as the alkyl chain is elongated, which is most likely related to the formation of non-polar domains in the imidazolium-based ionic liquid beyond the hexyl chain length creating a confinement effect for the cation and anion to freely diffuse [362].

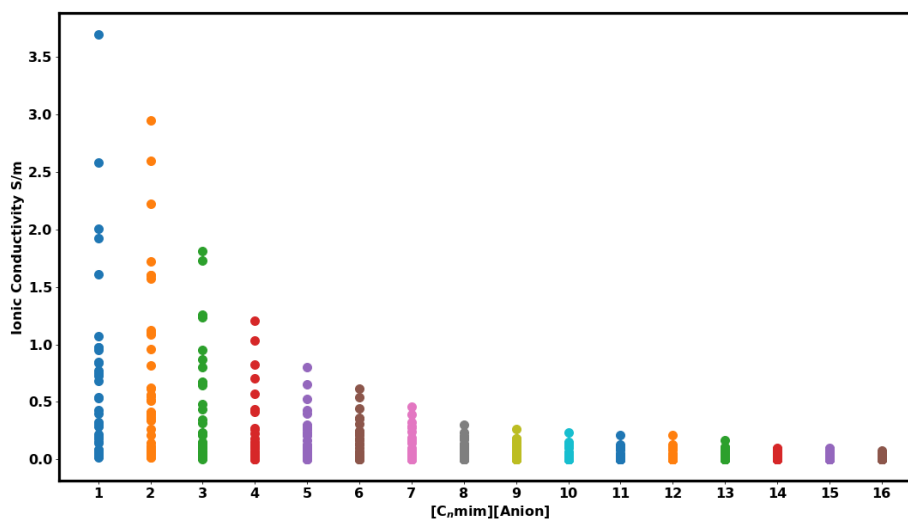


Figure A.14: Predicted ionic conductivity as a function of alkyl chain for the $[C_n\text{mim}]$ cation series for all the available anions at 298.0 K using FFANN model.

Predicted Ionic conductivity as a function of temperature

Figure A.15 shows the ionic conductivity for all the unique IL as a function of temperature using FFANN model. It is clear from the three sub-plots that conductivity increases as a function of temperature. However, the increase in conductivity is not similar for all the ionic liquids at different temperatures. Some of the ionic liquids that have very high conductivity at 298.0 and 350.0 K seems to fall off at 400.0 K indicating that the model does not scale the temperature linearly for all ILs like VFT equation would do, instead it scales the conductivity based on what cation and anion combination is present.

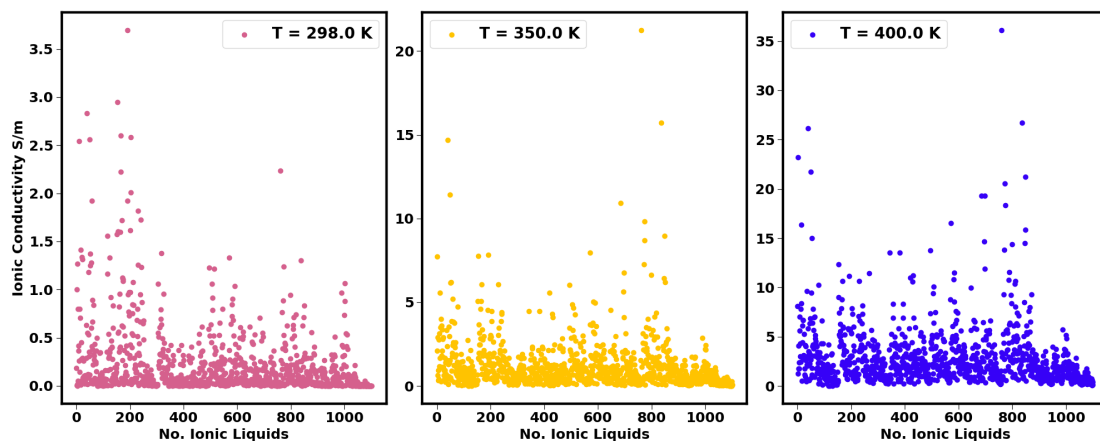


Figure A.15: Predicted ionic conductivity for all the unique ionic liquid combination(1102) as a function of temperature using FFANN model. Note:The numbering of ILs in x-axis is the same for all three sub-plots

Binary Liquid Mixture Enhancement

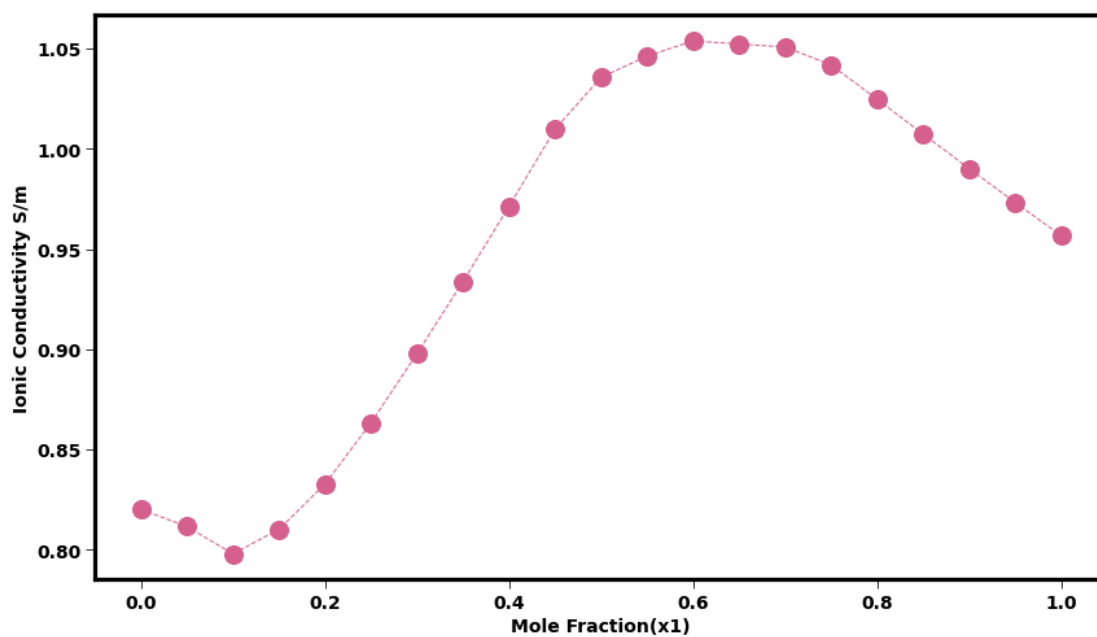


Figure A.16: Binary Mixture of $[C_2mim][NTf_2]_{x_1} + [C_2mim][CF_3SO_3]_{1-x_1}$ at 298.0 K.

Binary ionic liquids heat map visualization

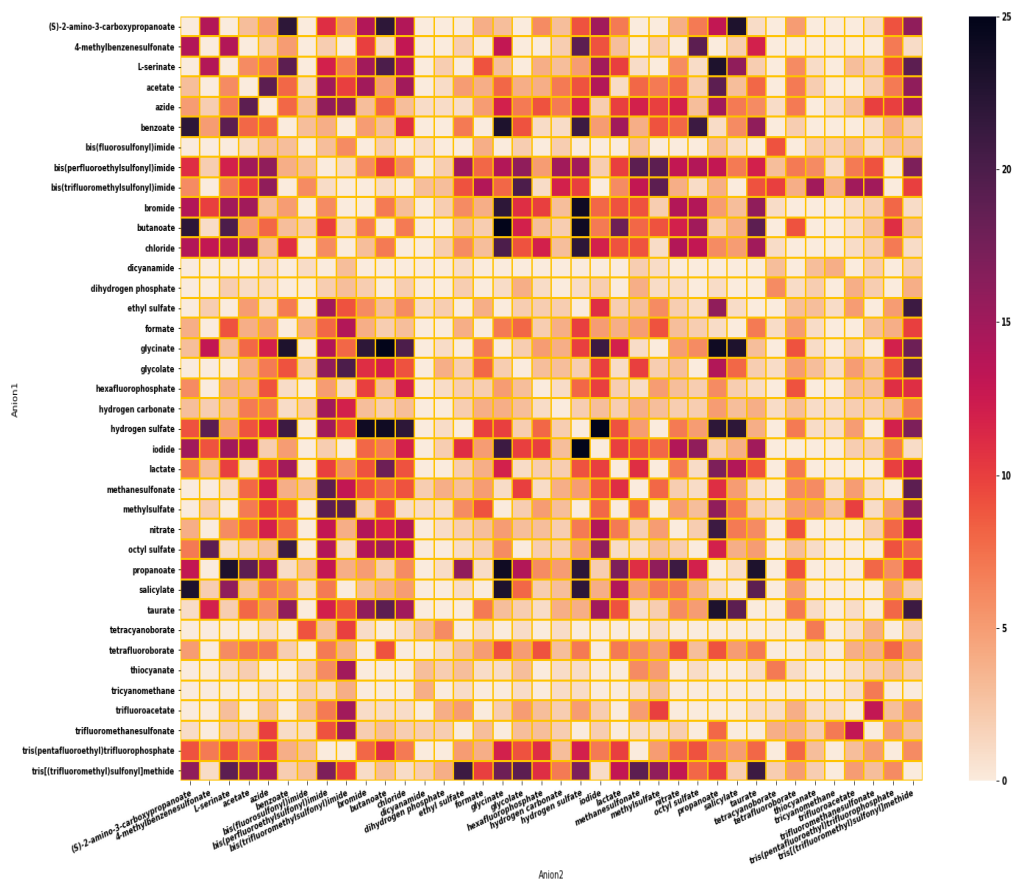


Figure A.17: Heat map of anions based on frequency of occurrence in binary anion common cation causing maximum enhancement at 298.0 K.

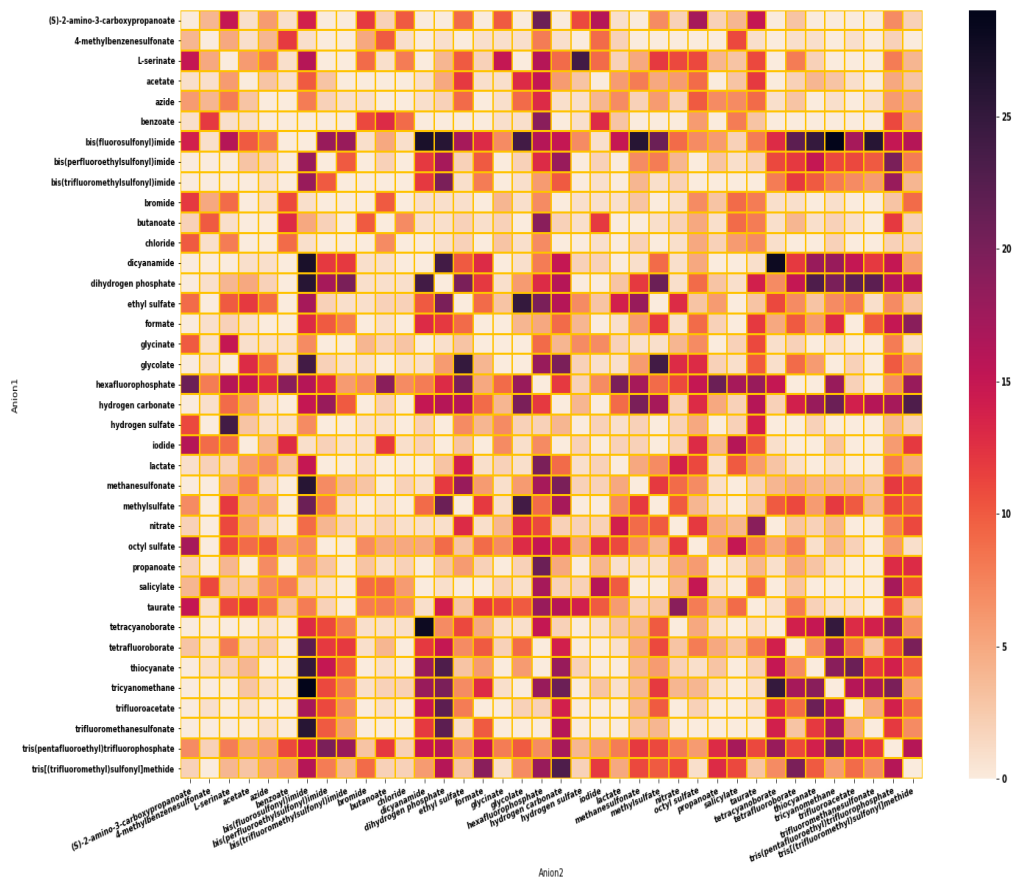


Figure A.18: Heat map of anions based on frequency of occurrence in binary anion common cation causing negative suppression at 298.0 K.

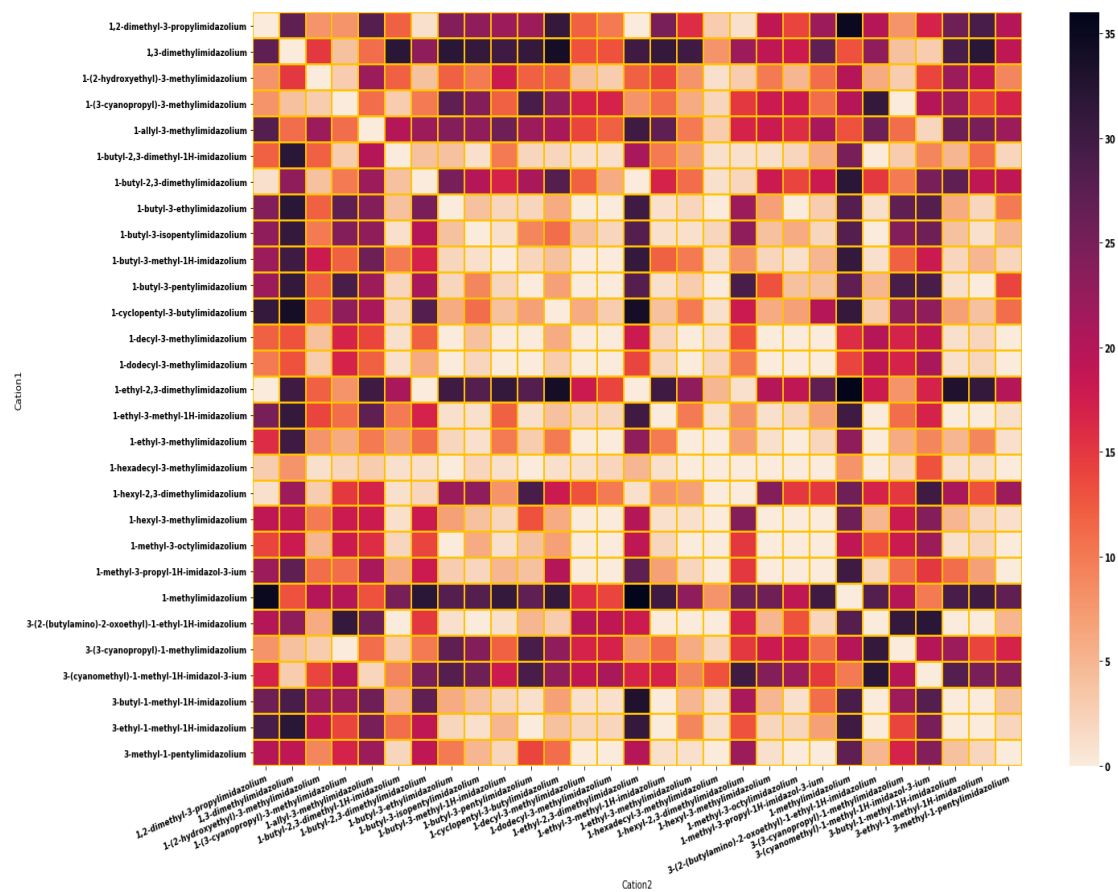


Figure A.19: Heat map of cations based on frequency of occurrence in binary cation common anion causing maximum enhancement at 298.0 K.

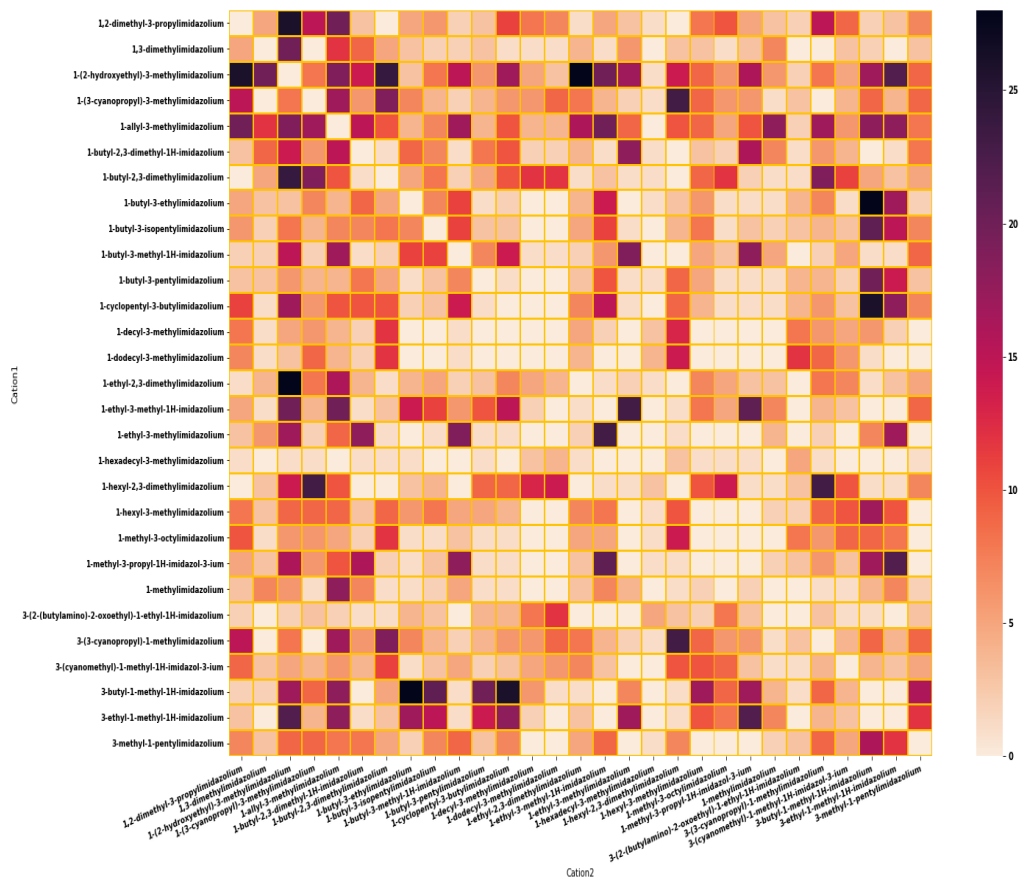
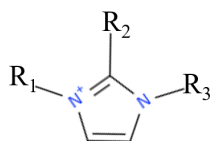


Figure A.20: Heat map of cations based on frequency of occurrence in binary cation common anion causing negative suppression at 298.0 K.

Table A.6: Comparison of ionic conductivity for $[R_1R_2R_3im][NTf_2]$ ionic liquids at 298.15 K between experiment and models. RD stands for Relative Deviation. Experiment data taken from ref [277].



No.	R ₁	R ₂	R ₃	Exp(S/m)	FFANN(S/m)	RD%	SVR	RD%
1	Methyl	H	Ethyl	0.96	0.96	-0.14	0.93	3.56
2	Methyl	Methyl	Ethyl	0.47	0.23	50.41	0.07	85.34
3	Methyl	Methyl	n-Propyl	0.36	0.20	44.75	0.08	78.41
4	Methyl	Methyl	n-Butyl	0.26	0.16	37.36	0.09	66.25
5	Methyl	Methyl	n-Pentyl	0.2	0.14	32.24	0.11	46.39
6	Methyl	Methyl	Allyl	0.52	0.10	81.17	0.06	88.51
7	Methyl	Methyl	Methoxyethyl	0.35	0.15	56.75	0.06	82.93
8	Methyl	Methyl	Ethoxyethyl	0.32	0.05	85.10	0.06	81.33
9	Allyl	Methyl	Ethyl	0.48	0.09	81.44	0.06	87.55
10	Allyl	Methyl	n-Propyl	0.4	0.07	83.05	0.06	85.07
11	Allyl	Methyl	n-Butyl	0.2	0.05	77.13	0.06	70.13
12	Allyl	Methyl	n-Pentyl	0.19	0.03	85.29	0.06	68.56
13	Allyl	Methyl	Allyl	0.44	0.19	57.85	0.06	86.42
14	Allyl	Methyl	Methoxyethyl	0.35	0.11	69.82	0.06	82.93
15	Allyl	Methyl	Ethoxyethyl	0.32	0.02	92.87	0.06	81.33

CHAPTER B

Appendix: Revealing Hydrogen Bond Dynamics between Ion Pairs in Binary and Reciprocal Ionic Liquid Mixtures

Potential Mean Force (PMF)

$$PMF = -k_b \cdot T \cdot \ln(g(r)) \quad (\text{B.1})$$

Liquid Densities

Table B.1: Comparison of liquid densities for pure ionic liquids against experimental data using two different charge scaling at 323 K.

System	Density[±0.8] (g/cm ³)	Density[±1.0] (g/cm ³)	Experimental (g/cm ³)	Reference
[C ₂ mim][DCA]	1.026±0.001	1.065	1.084	[211]
[C ₂ mim][NTf ₂]	1.486±0.009	1.513	1.493	[363]
[C ₆ mim][DCA]	0.874±0.0001	1.000	1.013	[364]
[C ₆ mim][NTf ₂]	1.332±0.008	1.370	1.350	[365]

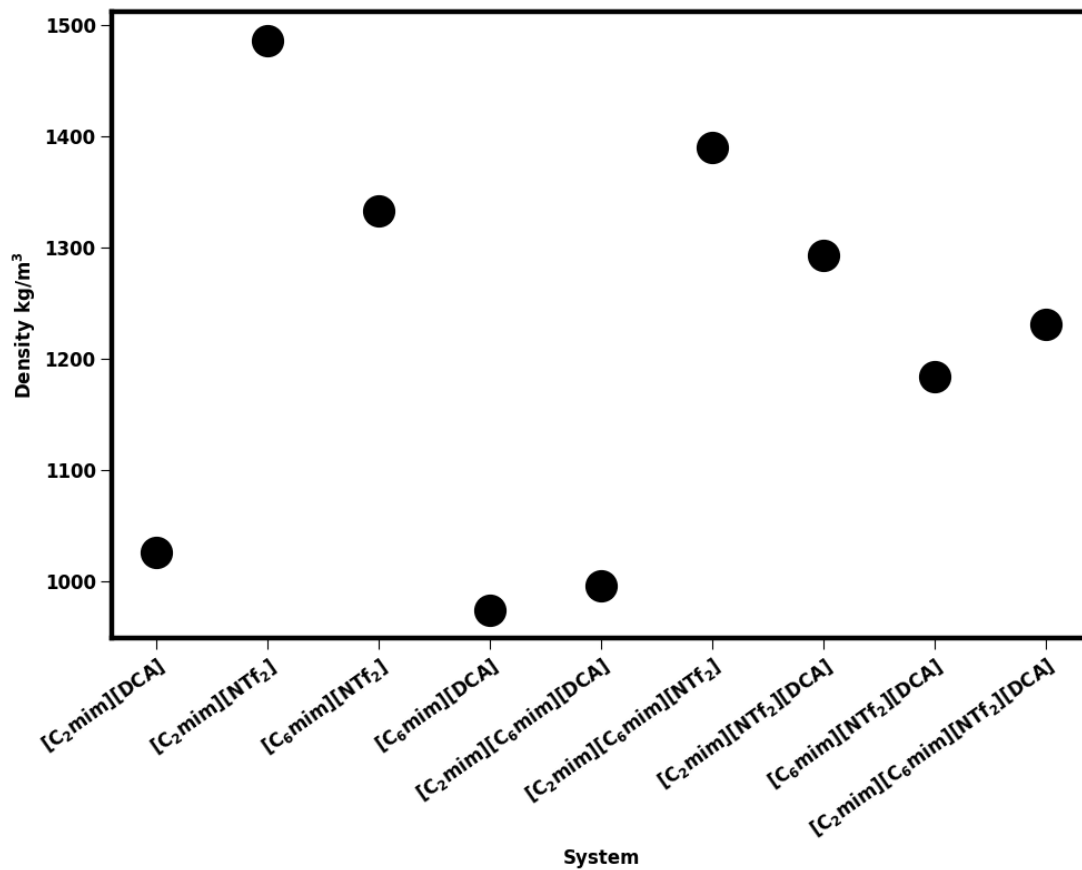


Figure B.1: Liquid phase density of all the system studied in this paper at 323.0 K using ± 0.8 charge scaling.

Radial Distribution Function

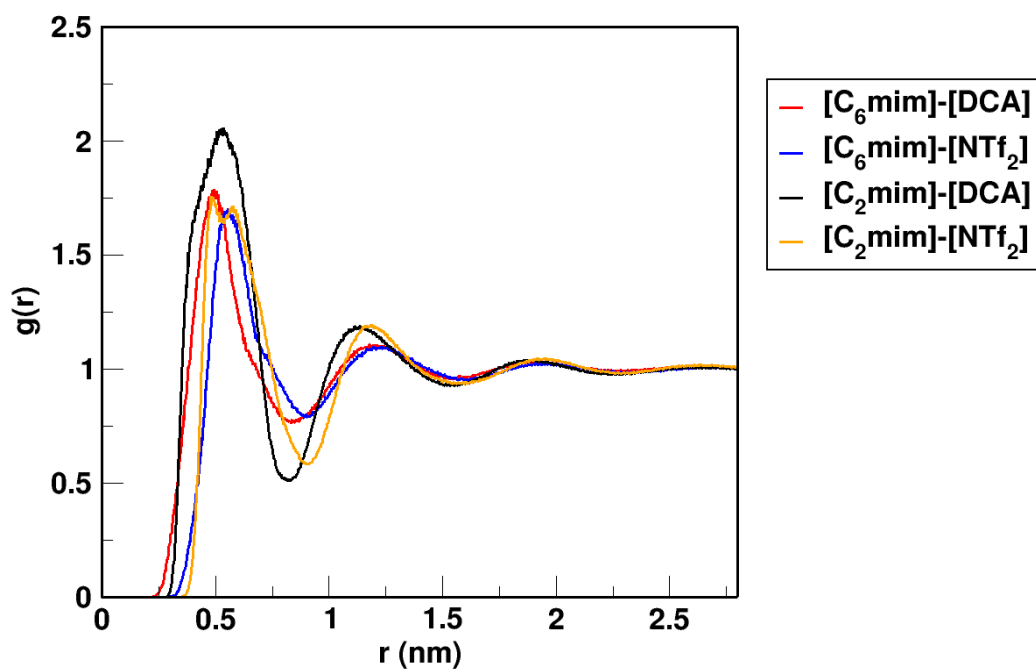


Figure B.2: Center of mass (COM) radial distribution plot of the cation ring and anion interactions in reciprocal mixture.

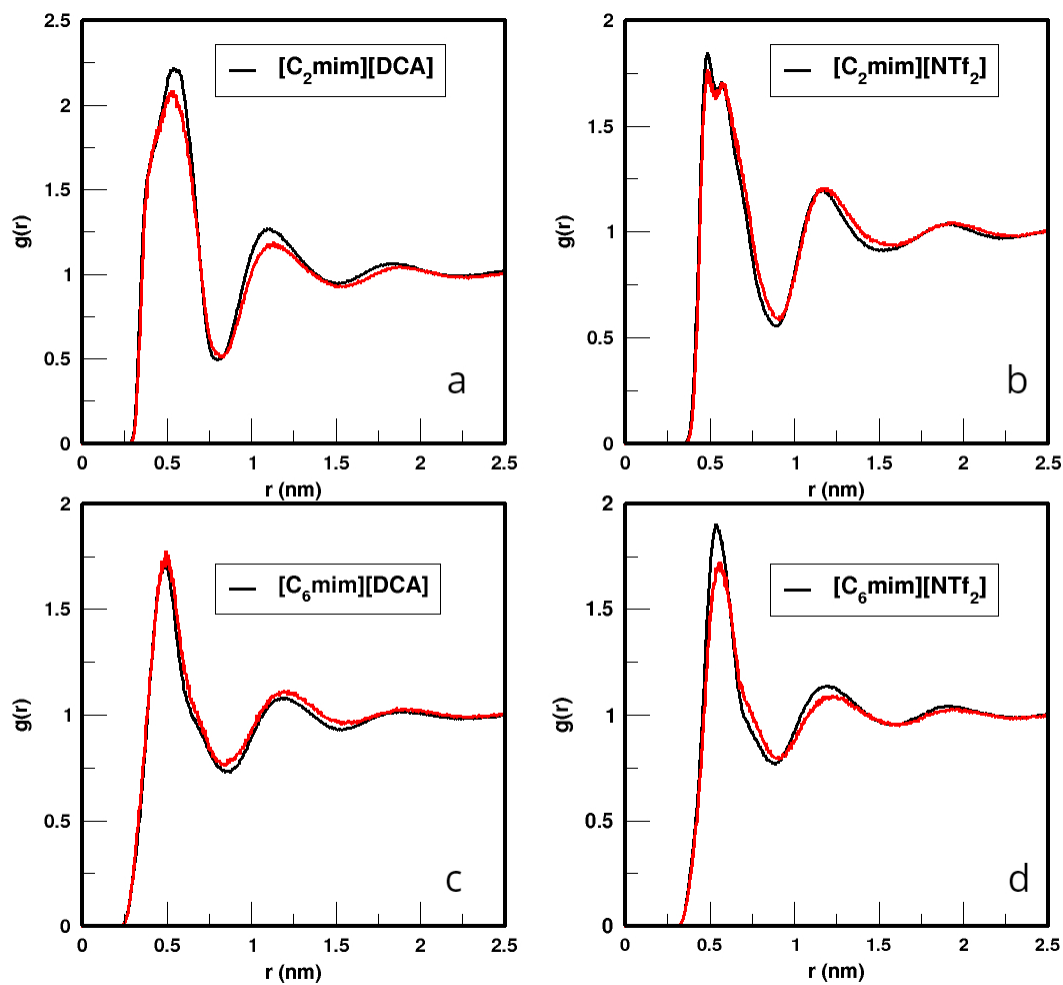


Figure B.3: Comparison of center of mass (COM) radial distribution plot of all the cation-anion interactions in reciprocal mixtures between two charge scaling. ± 1.0 charge scaling denoted by black color and ± 0.8 charge scaling denoted by red color.

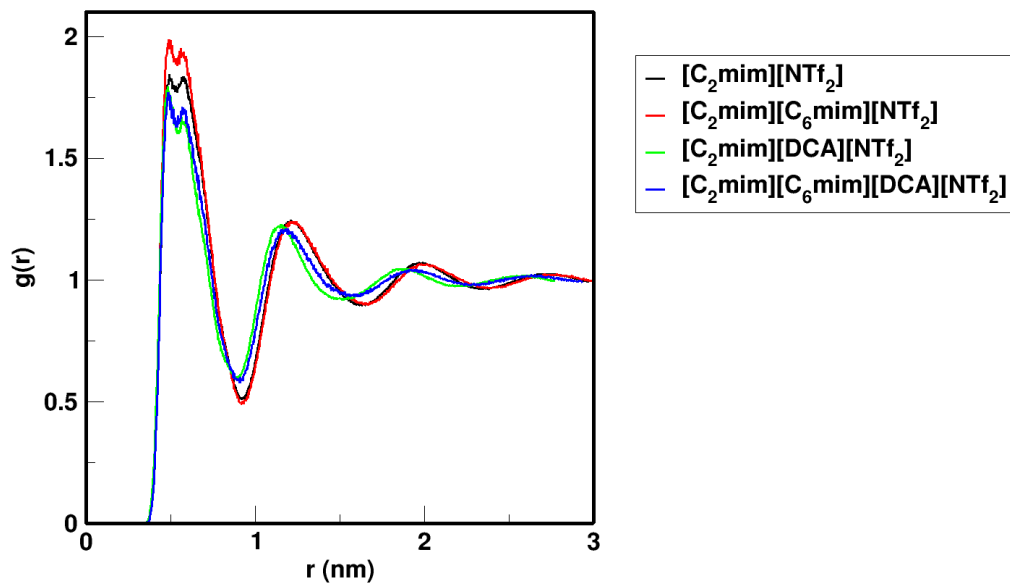


Figure B.4: COM RDF plot for $[\text{C}_2\text{mim}]-[\text{NTf}_2]$ interactions in various system containing $[\text{C}_2\text{mim}][\text{NTf}_2]$.

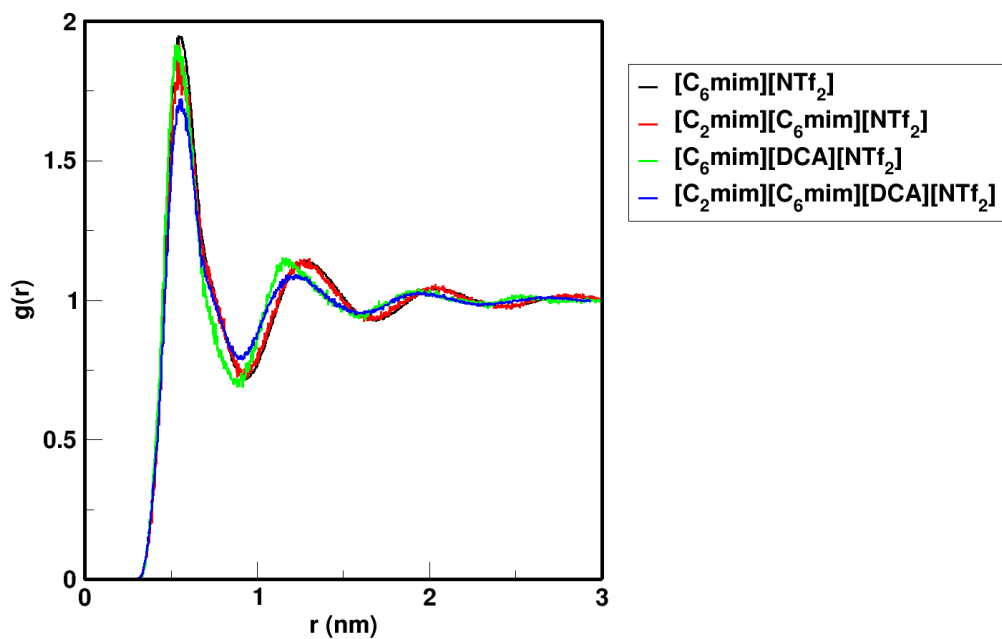


Figure B.5: COM RDF plot for $[\text{C}_6\text{mim}]-[\text{NTf}_2]$ interactions in various system containing $[\text{C}_6\text{mim}][\text{NTf}_2]$.

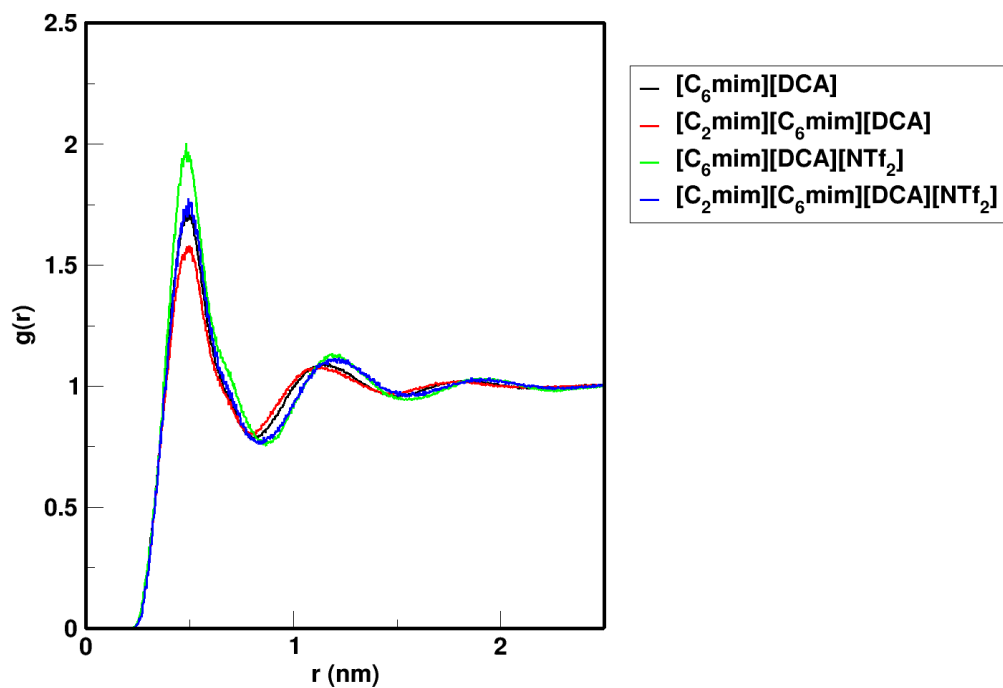


Figure B.6: COM RDF plot for $[\text{C}_6\text{mim}]-[\text{DCA}]$ interactions in various system containing $[\text{C}_6\text{mim}][\text{DCA}]$.

Spatial Distribution Function

Figure B.7 illustrates the spatial distribution function of anions around the $[\text{C}_2\text{mim}]^+$ cation for $[\text{C}_2\text{mim}][\text{C}_6\text{mim}][\text{DCA}][\text{NTf}_2]$ reciprocal mixture.

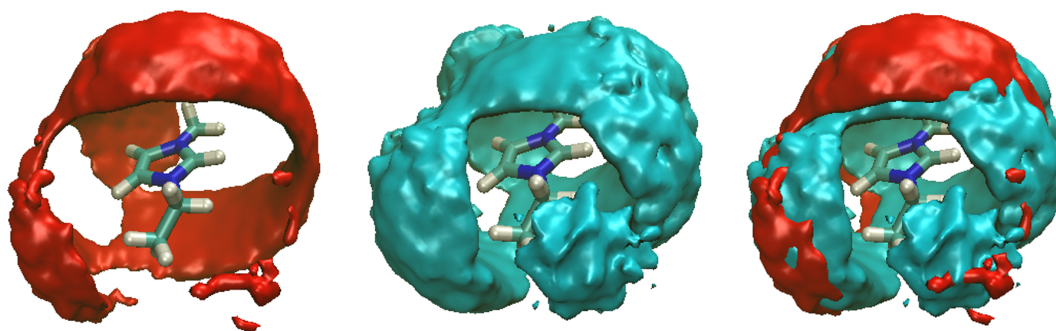


Figure B.7: Spatial distribution function (SDF's) of anions around the $[\text{C}_2\text{mim}]^+$ cation for $[\text{C}_2\text{mim}][\text{C}_6\text{mim}][\text{DCA}][\text{NTf}_2]$ reciprocal mixture. Red represents $[\text{NTf}_2]^-$ anion, cyan represents $[\text{DCA}]^-$ anion. Isosurface density was set to 2.2 times the bulk density.

Hydrogen Bonding Auto Correlation

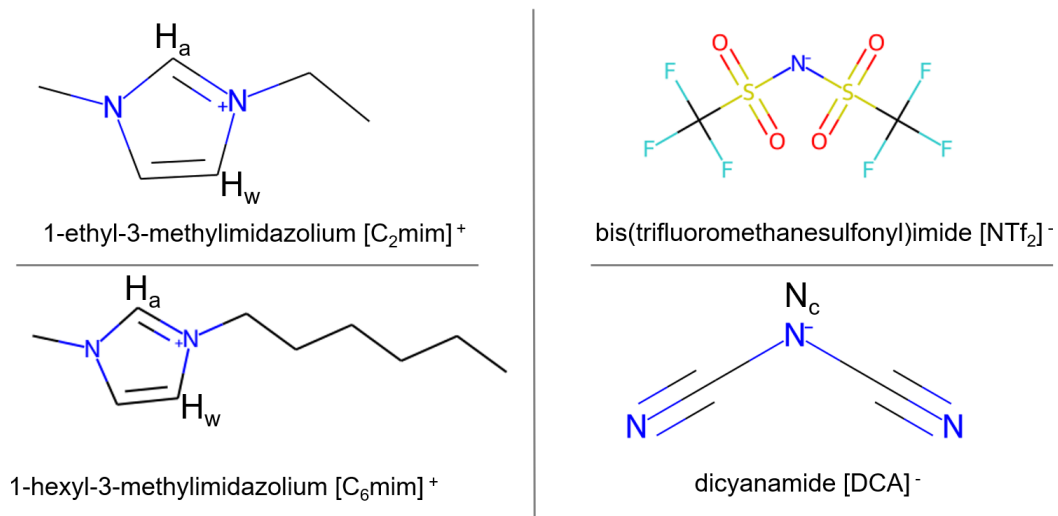


Figure B.8: Hydrogen bonding interaction sites for the cation and anion.

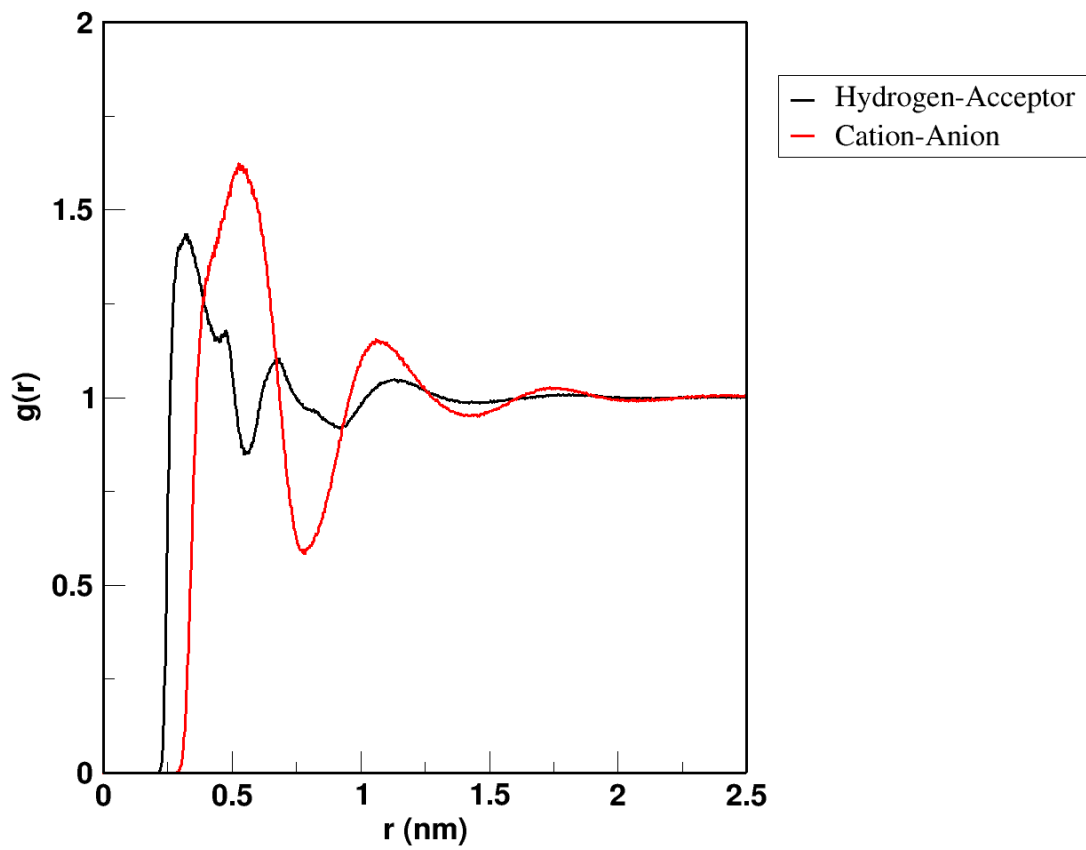


Figure B.9: RDF of site-site interaction for hydrogen (H_a) - acceptor (N_c) and com interaction between the cation and anion in $[C_2mim][DCA]$.

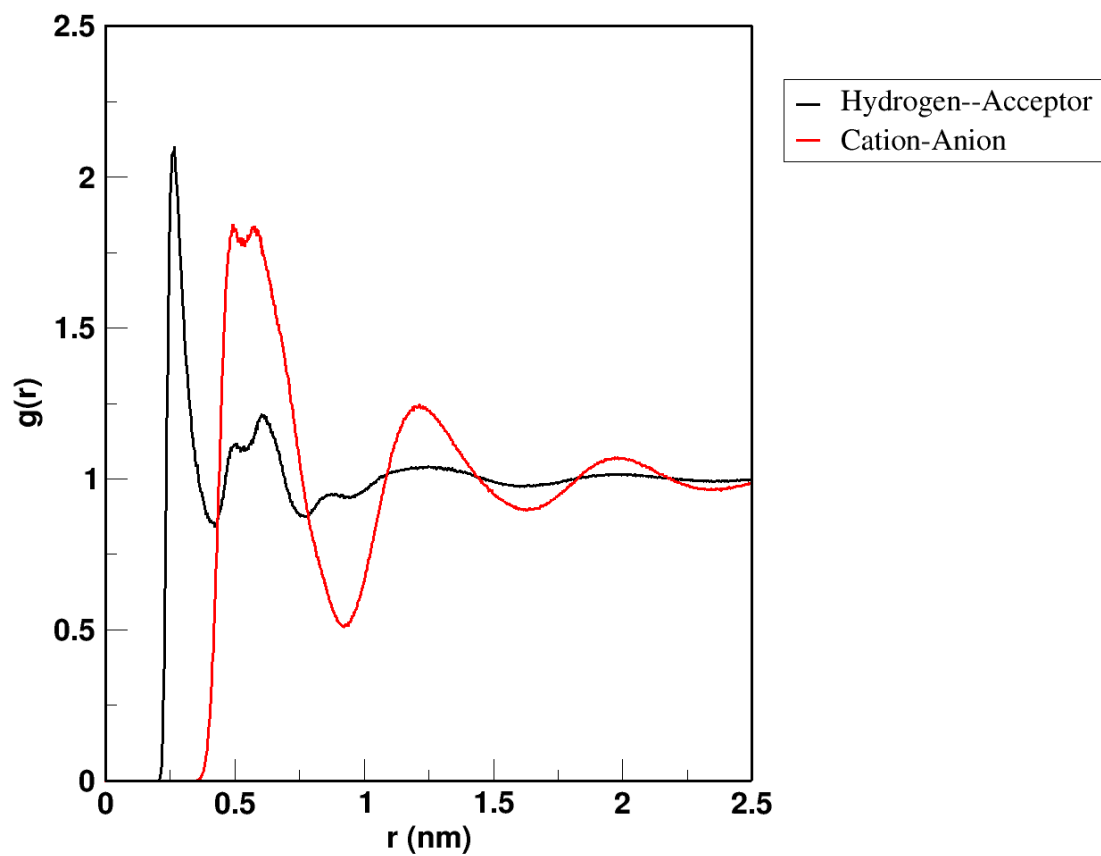


Figure B.10: RDF of site-site interaction for hydrogen (H_a) - acceptor (O) and com interaction between the cation and anion in $[C_2mim][NTf_2]$.

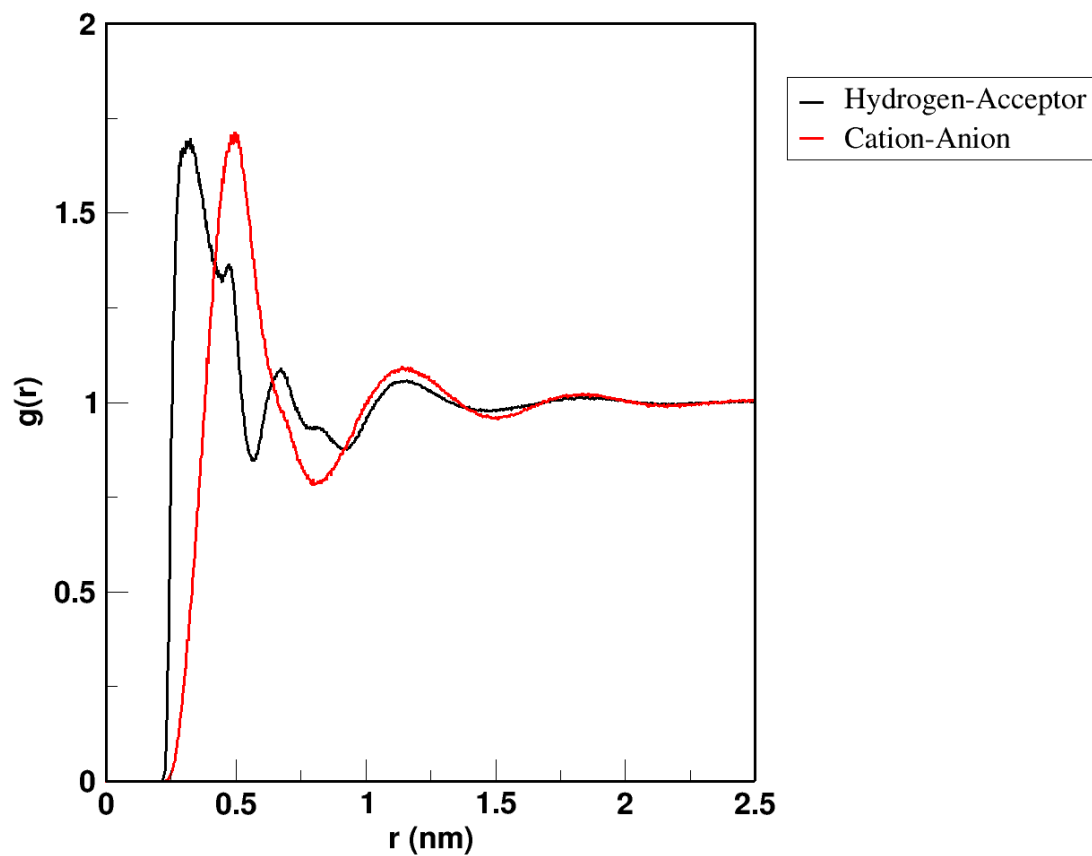


Figure B.11: RDF of site-site interaction for hydrogen (H_a) - acceptor (N_c) and cation-anion interaction between the cation and anion in $[C_6mim][DCA]$.

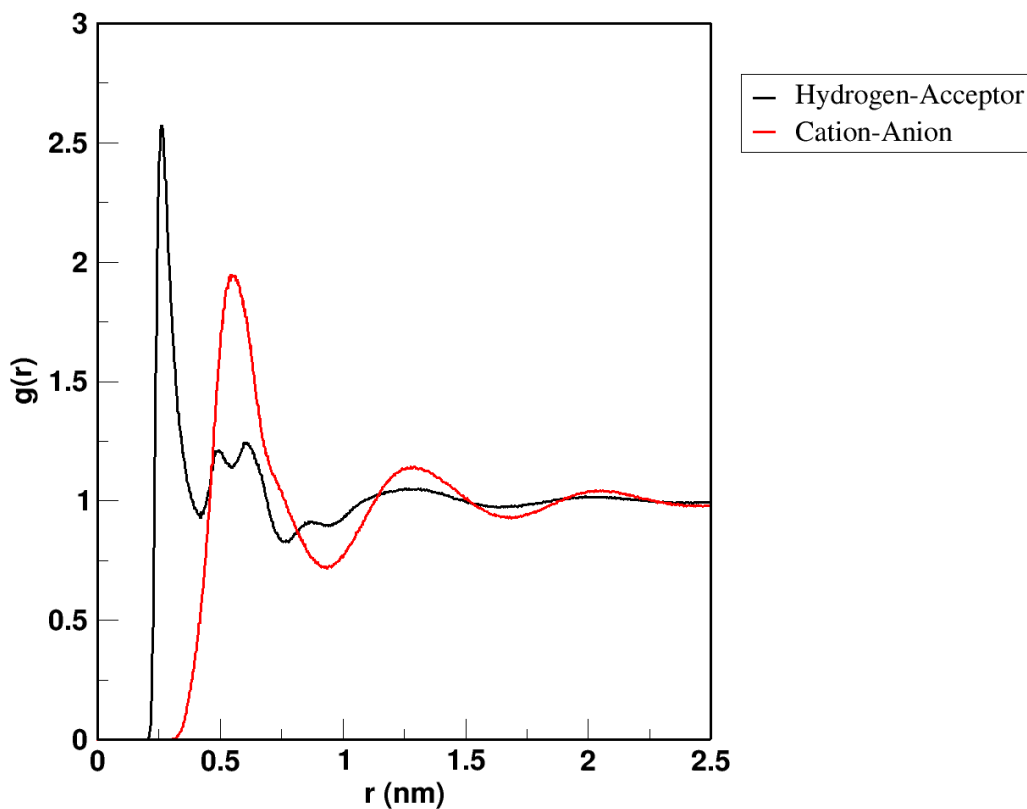


Figure B.12: RDF of site-site interaction for hydrogen (H_a) - acceptor (O) and cation-anion interaction between the cation and anion in $[C_6mim][NTf_2]$.

Table B.2: Cut-off distances (nm) used in the hydrogen bonding average lifetime calculations. The hydrogen-acceptor cut-off was set to the first solvation shell denoted by H_a-N_c/O . The distance for the donor-acceptor was limited to the first solvation shell of the center of mass (cation)-center of mass (anion) denoted by Cation-Anion.

System	H_a-N_c/O	Cation-Anion
$[C_2mim][DCA]$	0.554	0.768
$[C_2mim][NTf_2]$	0.428	0.910
$[C_6mim][DCA]$	0.559	0.800
$[C_6mim][NTf_2]$	0.408	0.920

Mean Square Displacement

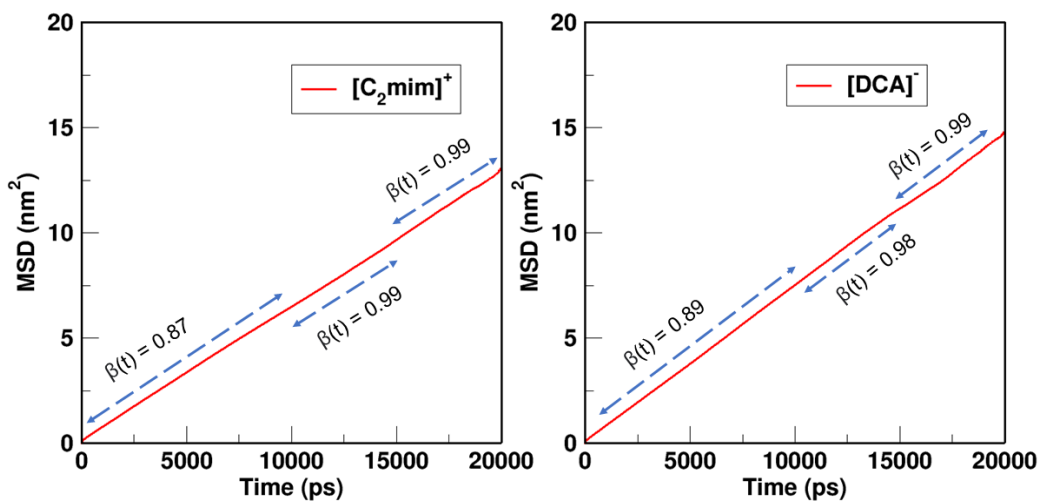


Figure B.13: Mean Square Displacement of [C₂mim]⁺ cation and [DCA]⁻ anion for pure [C₂mim][DCA] system.

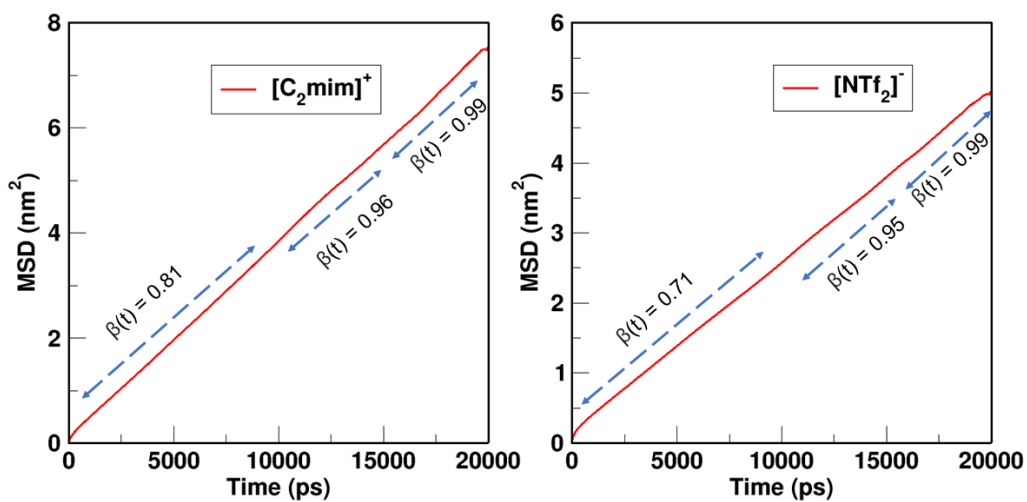


Figure B.14: Mean Square Displacement of [C₂mim]⁺ cation and [NTf₂]⁻ anion for pure [C₂mim][NTf₂] system.

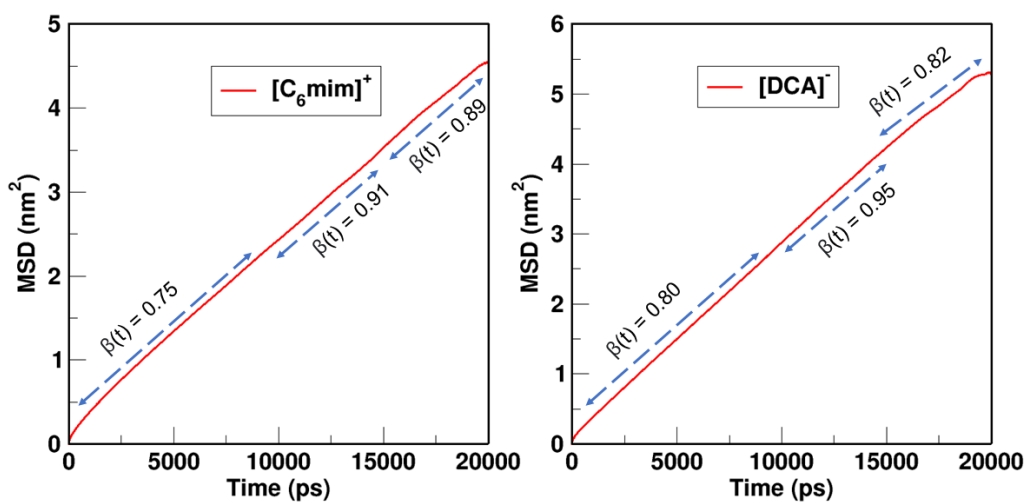


Figure B.15: Mean Square Displacement of $[\text{C}_6\text{mim}]^+$ cation and $[\text{DCA}]^-$ anion for pure $[\text{C}_6\text{mim}][\text{DCA}]$ system.

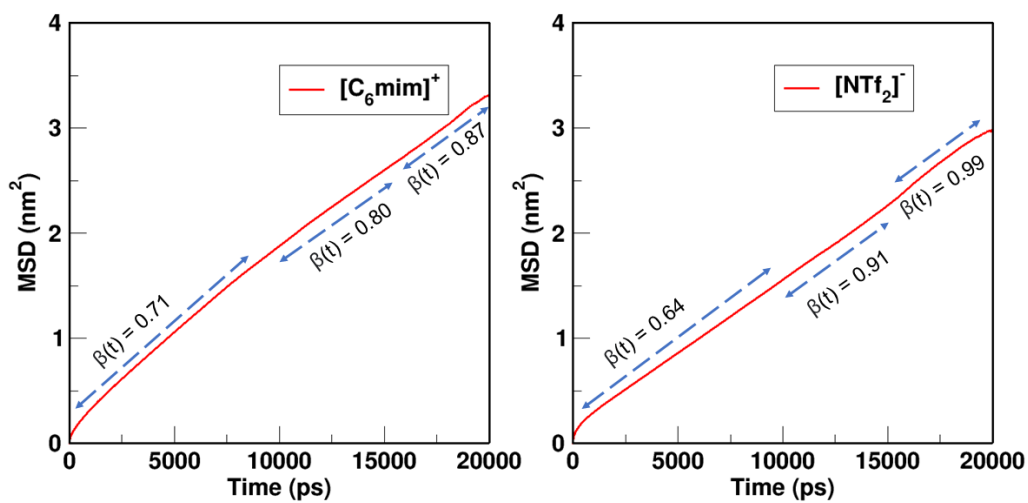


Figure B.16: Mean Square Displacement of $[\text{C}_6\text{mim}]^+$ cation and $[\text{NTf}_2]^-$ anion for pure $[\text{C}_6\text{mim}][\text{NTf}_2]$ system.

CHAPTER C

Appendix: A Generalized Machine Learning Model for Predicting Ionic Conductivity for Ionic Liquids'

Equations

Mean Absolute Error

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

Root Mean Squared Error

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

Residual Deviation

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

Where:

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

Cation Type

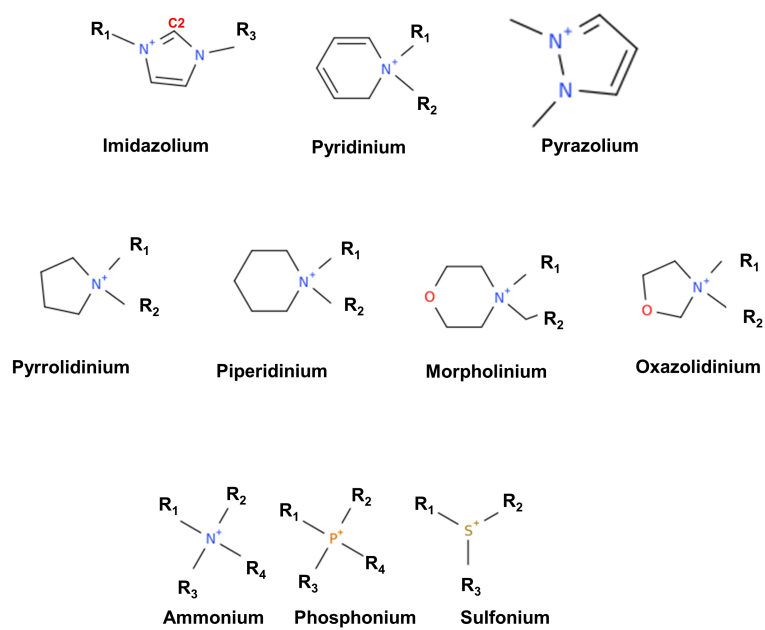


Figure C.1: Schematics of cation type used in this study for model development. R₁, R₂, R₃, and R₄ refer to substituents in the cation. Images were drawn using OPSIN package [10].

Table C.1: 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 C.2: 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 C.3: 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

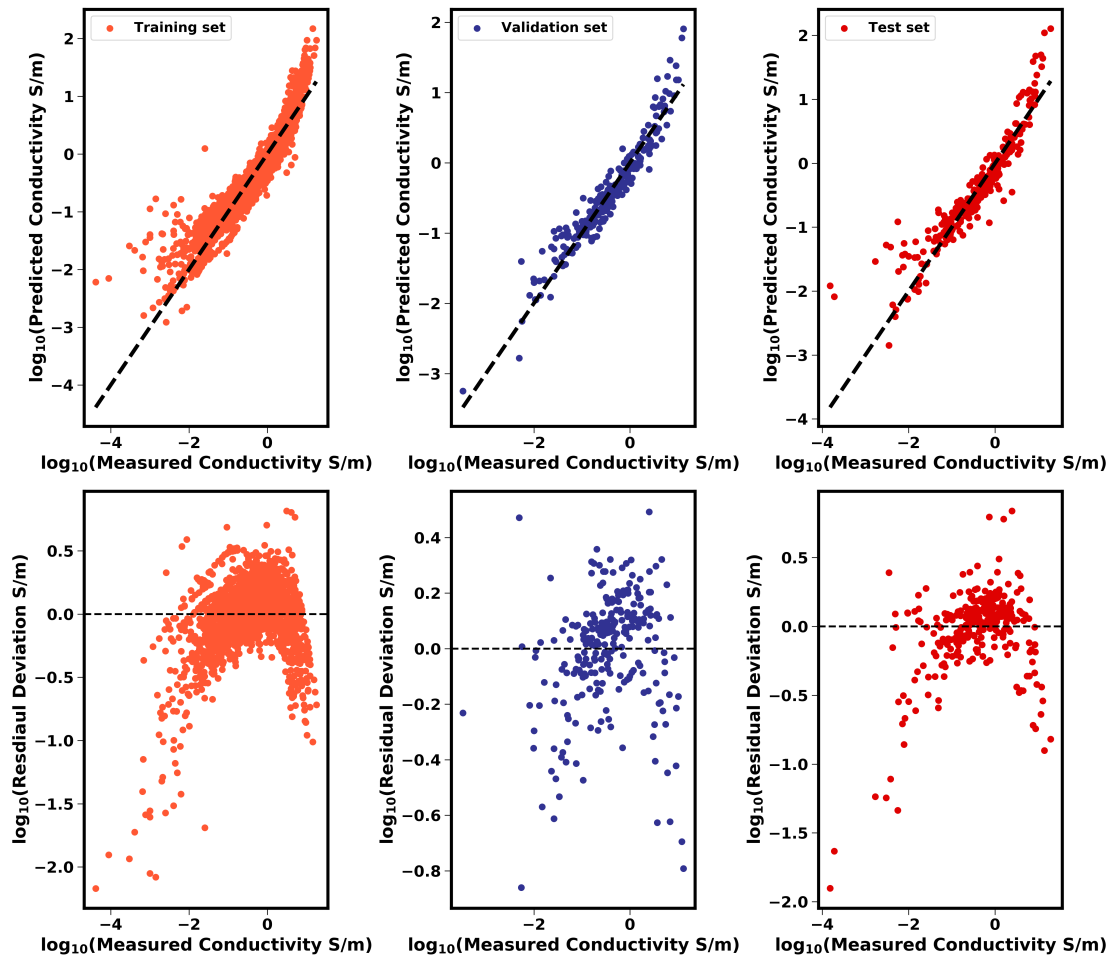


Figure C.4: Comparison of model predictions with the experimental data on \log_{10} scale for the training set (left pane), validation set (middle pane) and test set (right pane) using Multiple Linear Regression model.

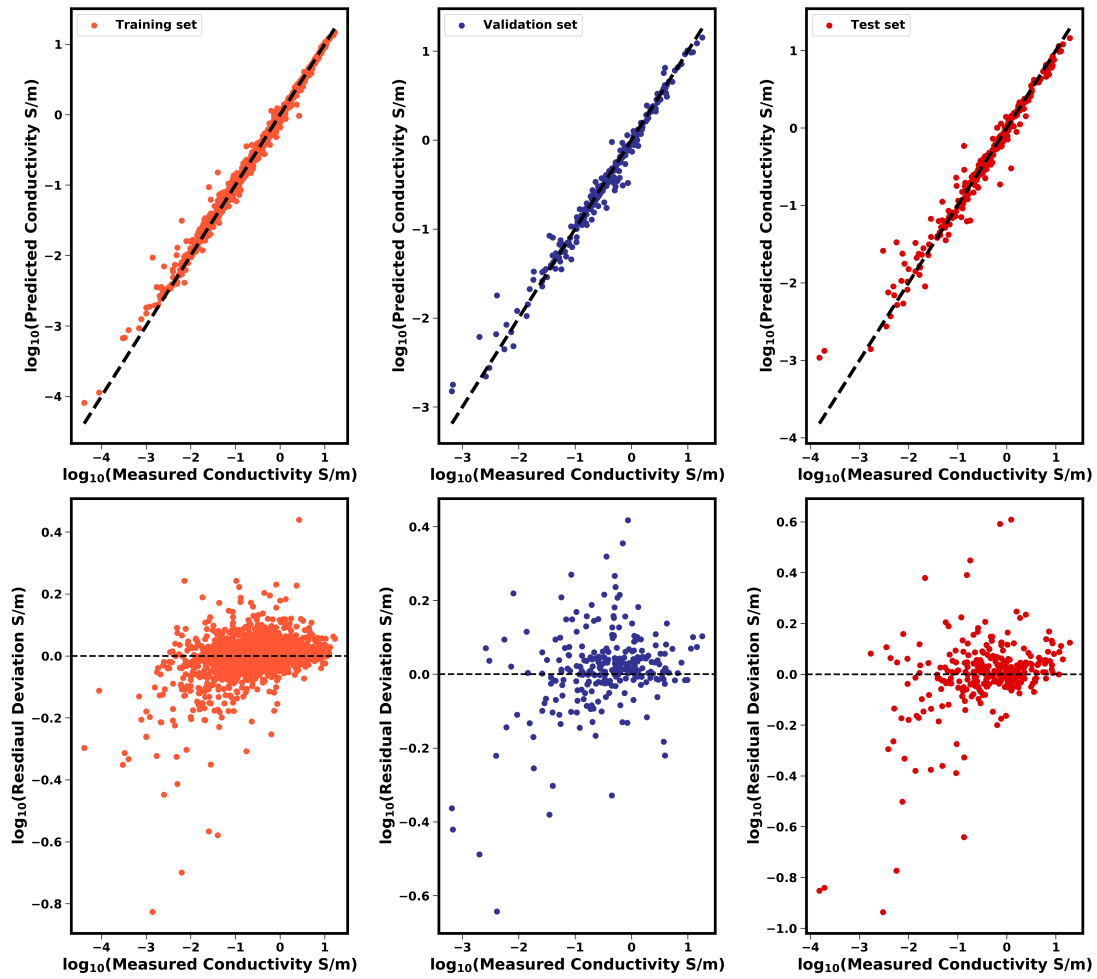


Figure C.5: Comparison of model predictions with the experimental data on \log_{10} scale for the training set (left pane), validation set (middle pane) and test set (right pane) using Random Forest model.

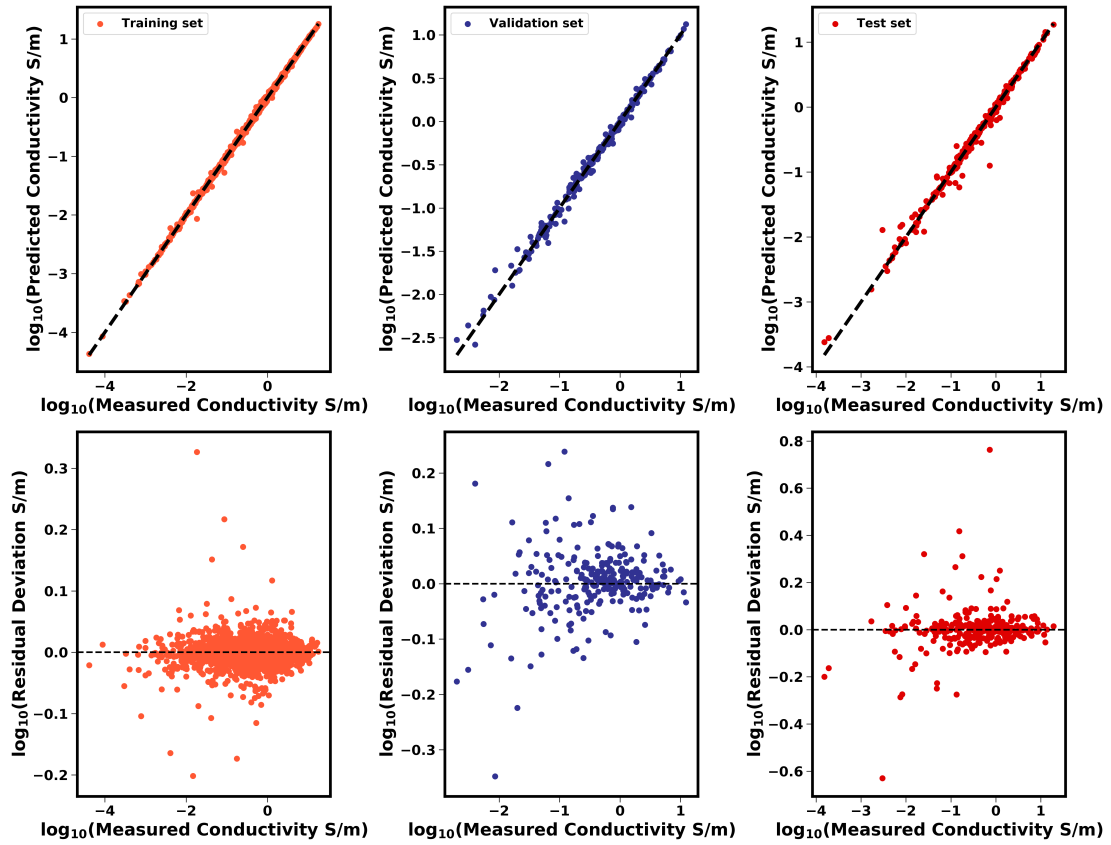


Figure C.6: Comparison of model predictions with the experimental data on \log_{10} scale for the training set (left pane), validation set (middle pane) and test set (right pane) using XGBoost model.

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 C.7 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.

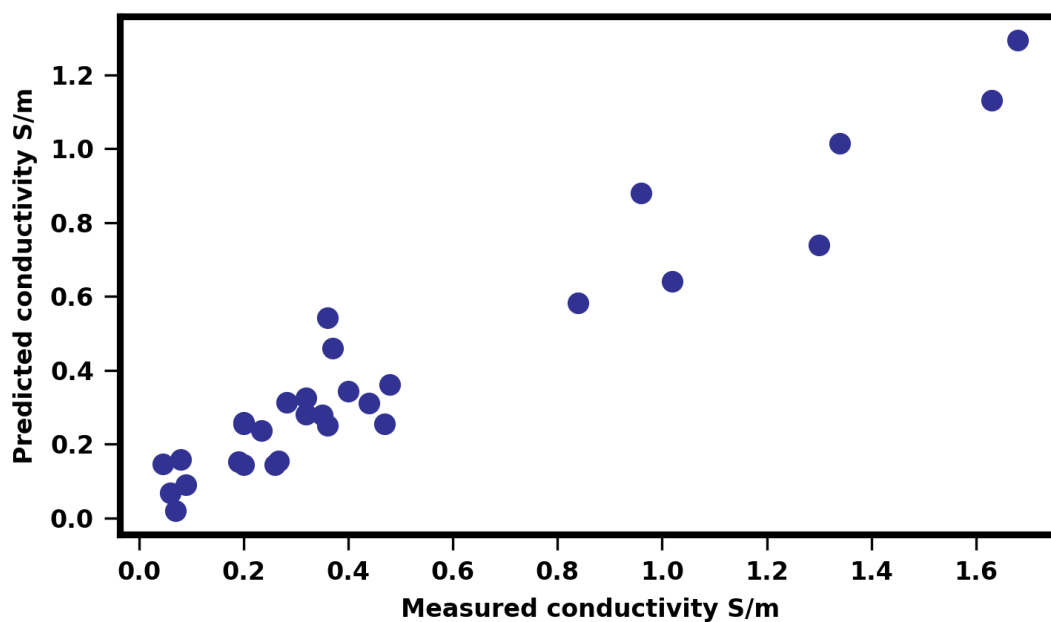
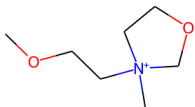
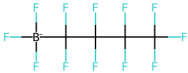
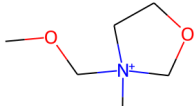
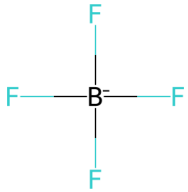
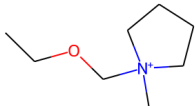



Figure C.7: Comparison of experimental external data set and XGBoost model prediction. The ionic conductivity values are reported in S/m.

Table C.2 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 C.2.

Table C.2: 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	[94]
2			0.08[298.15]	0.16	[94]
3			0.37[298.15]	0.46	[94]

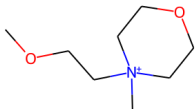

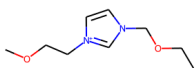

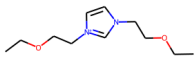

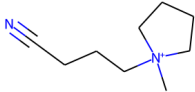

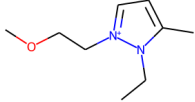
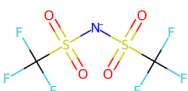
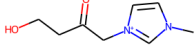
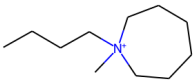

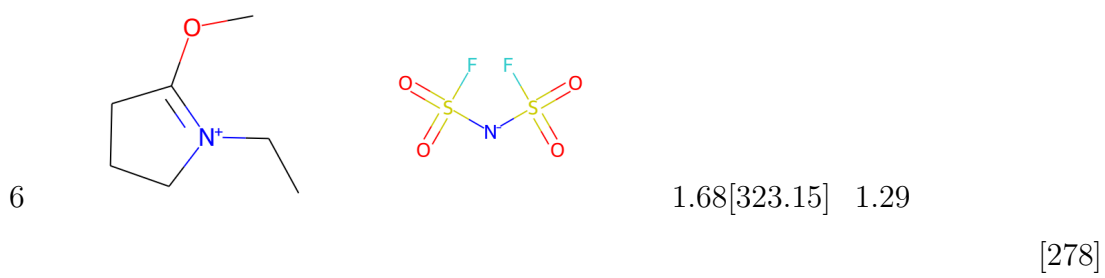
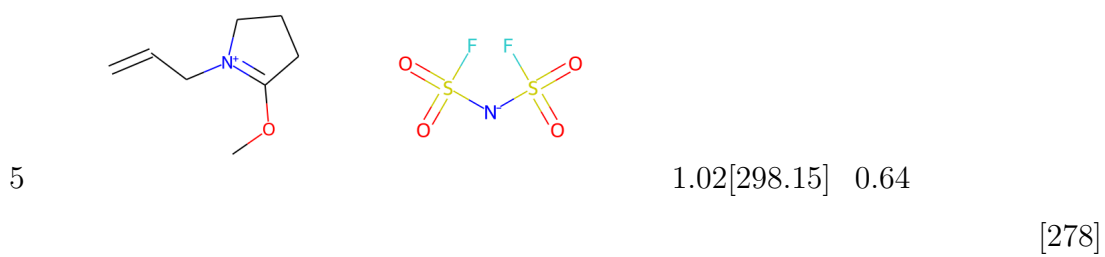
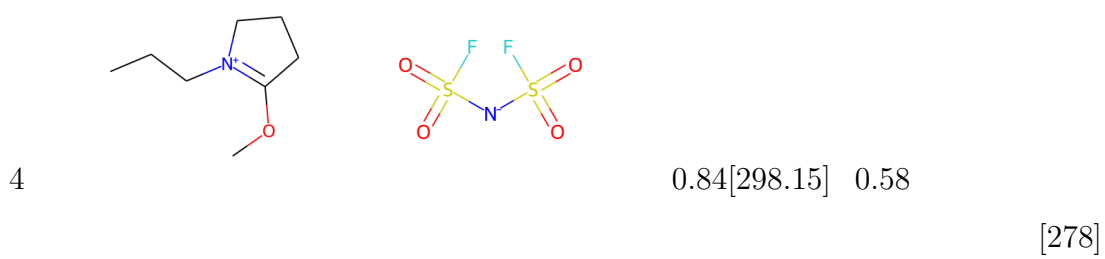
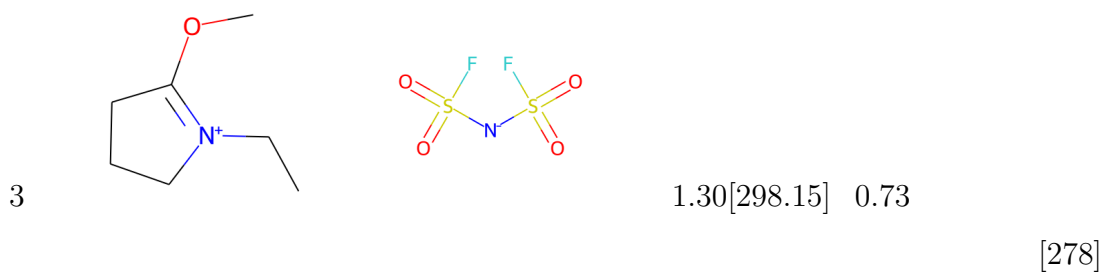
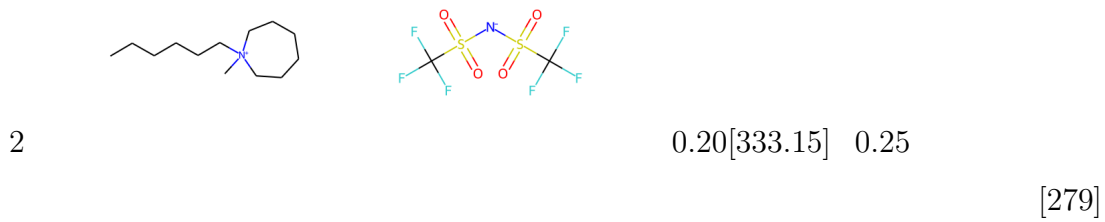
4			0.06[298.15]	0.07	[94]
5			0.28[298.15]	0.31	[36]
6			0.27[298.15]	0.15	[36]
7			0.05[298.15]	0.15	[76]
8			0.24[298.15]	0.24	[82]
9		Cl ⁻	0.07[293.15]	0.02	[92]

Table C.3 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. [278] 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 C.3: 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	[279]



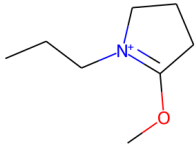
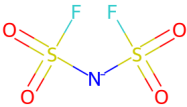
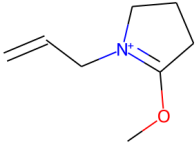
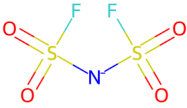
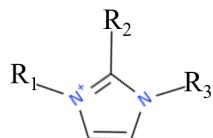
7			1.34[323.15] 1.01	[278]
8			1.63[323.15] 1.13	

Table C.4 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.



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

Table C.4: Comparison of ionic conductivity at 298 K between experiment [277] and XGBoost prediction for cations paired with bis(trifluoromethylsulfonyl)imide [NTf₂]⁻ anion.

Feature Importance

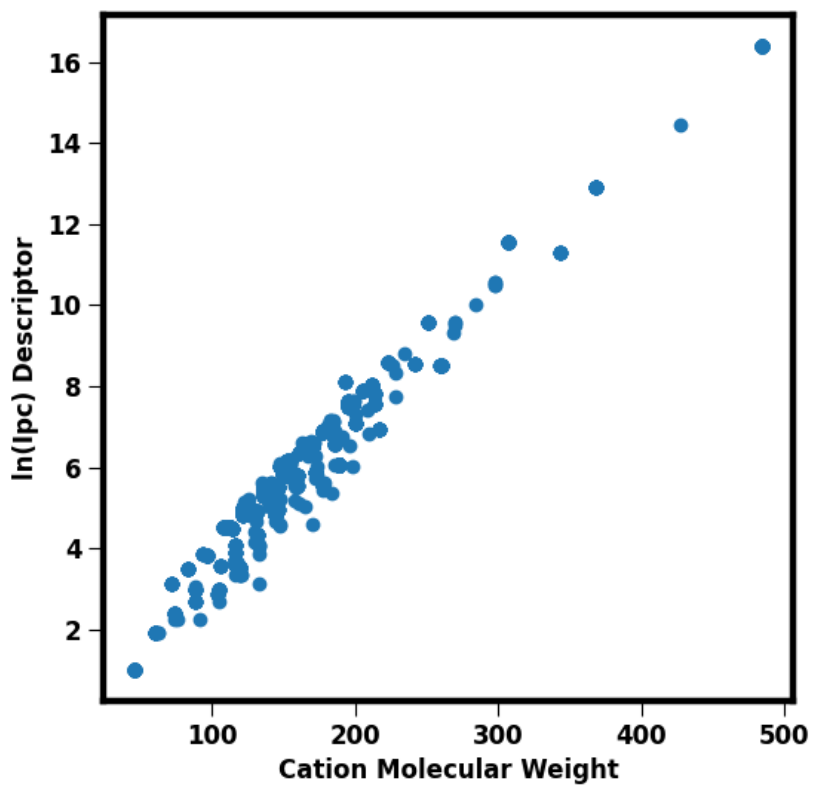


Figure C.8: Cation feature $\ln(I_{pc})$ vs molecular weights of the cations for which experimental data were included in the model development.

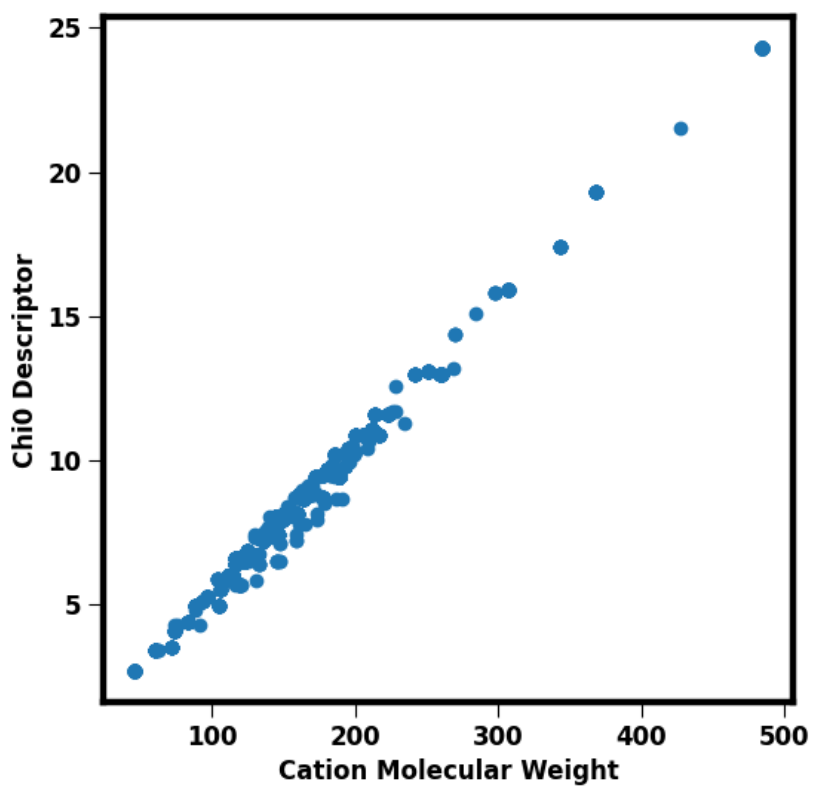


Figure C.9: Cation feature Chi0 vs molecular weight of the cations for which experimental data were used in the model development.

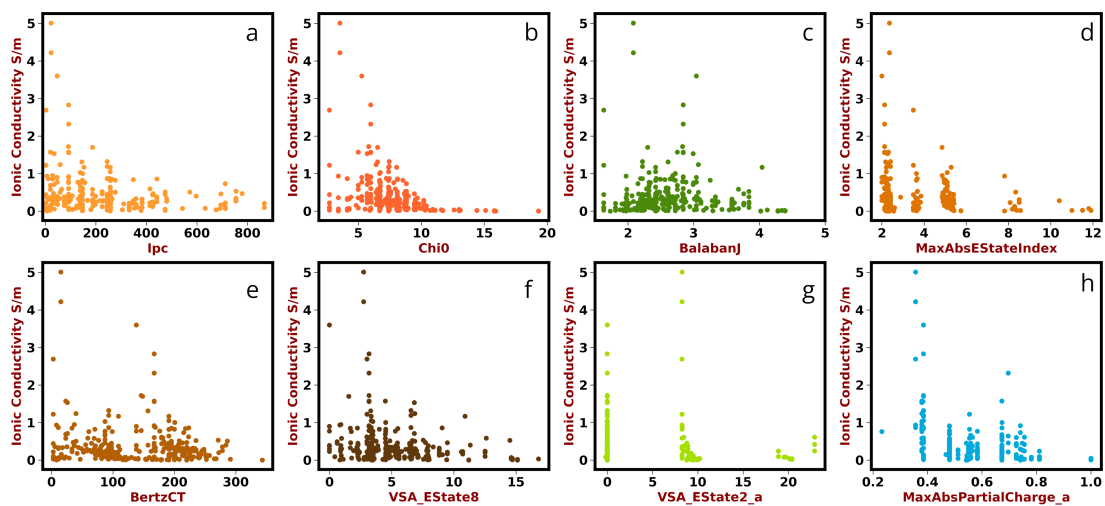


Figure C.10: Ionic conductivity vs. cation and anion features deemed to impact the ionic conductivity the most in the XGBoost model. The ionic conductivity data are plotted for the experimental data at 298.15 K. Features ending with '_a' indicates features for anions.

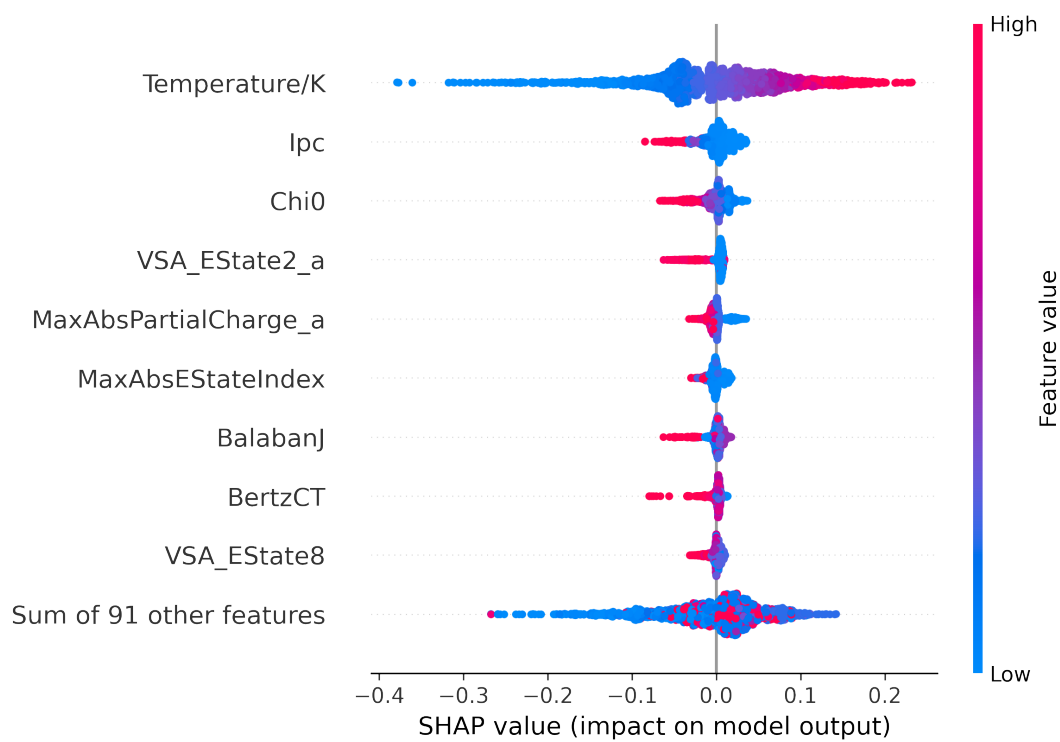


Figure C.11: SHAP feature importance for the training set data. Features ending with '_a' indicates features for anions.

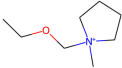
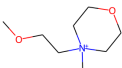
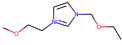
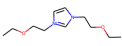
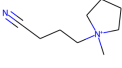
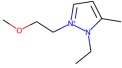
External Set Classification

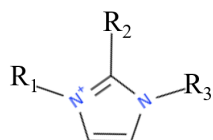
Based on the SHAP analysis, it is clear that some of the features have a very high influence on the ionic conductivity compared to the rest. In this section, we attempt to build a decision tree-based classification model to leverage insights generated from the SHAP analysis. The primary objective here 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-learn [144] 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 C.11). 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 C.5 and C.6). 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 C.5: 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	[94]
2		0.06	Low	Low	[94]
3		0.28	High	High	[36]
4		0.27	High	Low	[36]
5		0.05	Low	Low	[76]
6		0.24	Low	High	[82]



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

Table C.6: 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 ref [277].

Unique Ionic Liquids

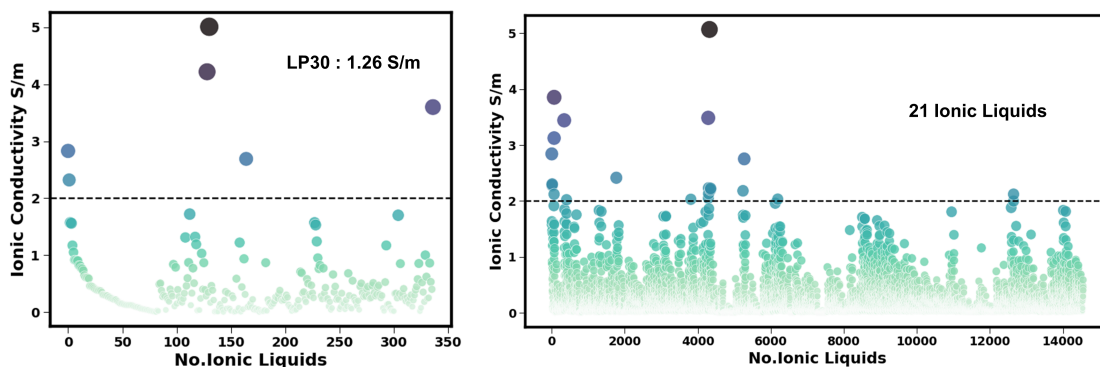


Figure C.12: (a) Experimental data at 298.15 K. (b) Unique ionic liquid predictions at 298.15 K using XGBoost model. LP30 here refers to the commonly used Li-ion electrolyte with an high ionic conductivity of 1.26 S/m at 298.15 K [71, 291]. Addition of Li⁺ salts to ionic liquids is known to reduce ionic conductivity by 30-50% [278].

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

Table C.7: 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.

CHAPTER D

Appendix: Accelerated Discovery of Novel Ionic Liquid Cations using a Continuous Latent Space Representation of Chemical Space

Data Generation

The VAE model is trained on SMILES data based on various commonly studied cation families, including imidazolium, ammonium, phosphonium, piperidinium, pyridinium, sulfonium, and sulfonium pyrrolidinium with an alkyl functional group attached to them. The length of the alkyl group is varied from 1 to 10, including isomers. Theoretically, it is possible to generate millions of such cation structures by attaching all the possible isomers at various positions. However, in this work, we are interested in finding the minimum cation data required to build a VAE model that accurately reconstructs SMILES for the test data and generates new SMILE strings.

Initially, we started with 10 different data set sizes: [2015, 4032, 6046, 8063, 10078, 12093, 14110, 16124, 18141, 20,151] that includes 1-alkyl-3-alkyl imidazolium, 1-alkyl-1-alkylpyrrolidinium, 1-alkyl-1-alkylpiperidinium, 1-alkyl pyridinium, tertiary sulfonium, quaternary phosphonium, and quaternary ammonium cations. The individual percentage of each cation seen in Figure D.1 (a) is kept the same for all the ten data sets. The rationale for choosing the particular cation ratio is to ensure that the ratio of non-nitrogen-based cation to nitrogen-based cation stays around 50%. For each data set, 10% of the data set is kept aside as test data

that remains untouched throughout the model development, while the remaining data is divided into training and validation set (90:10) for model development and hyper-parameter tuning.

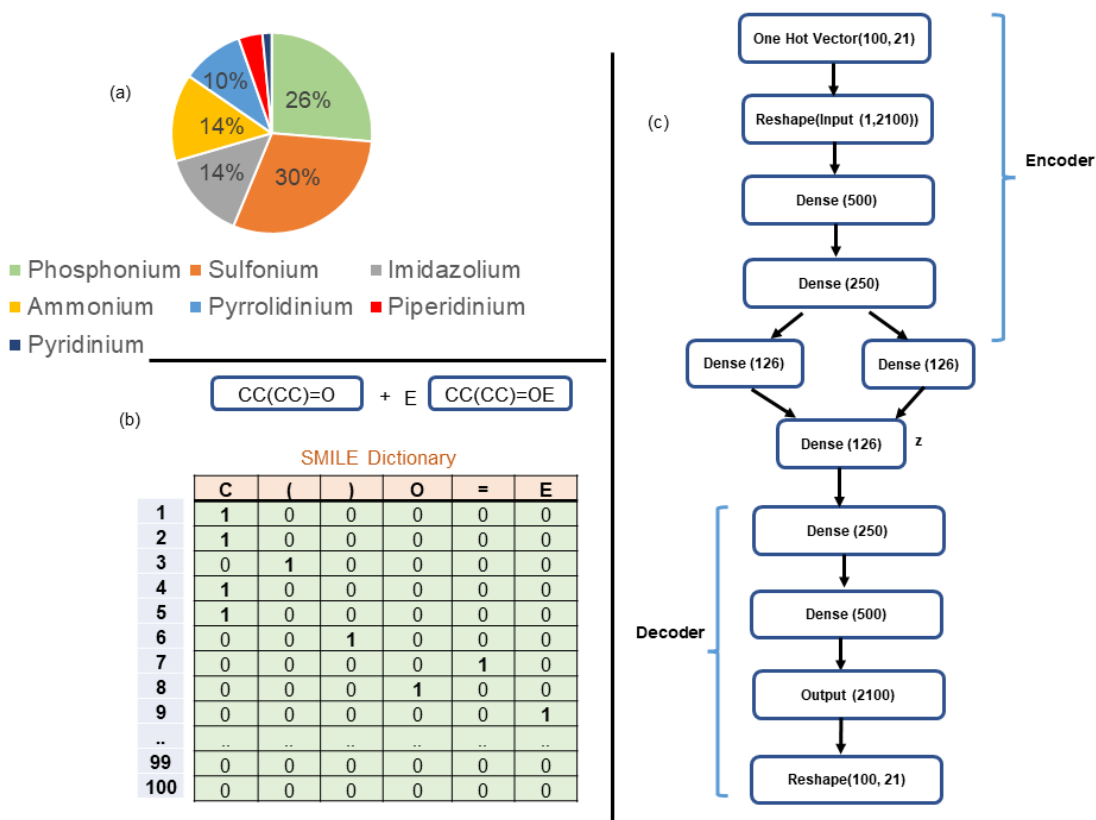


Figure D.1: (a) % Representation of different cation type in the VAE model development. (b) One hot encoding vector of a sample organic molecule with a vocabulary size of 6 and length of 100. 'E' denotes the end of SMILE string. (c) Neural Network representation of the encoder and decoder for the VAE model.

VAE Architecture and Hyper parameters

The one-hot encoding vector is a multidimensional (100,21) array reshaped to a 1D array of (1,2100) array size as the input layer. For the VAE architecture, three major hyper-parameters needed to be optimized : (a) the total number of training data points, (b) the dimension of the latent space, and (c) the number of layers and neurons for the encoder and decoder. (Note: total data points do not refer

to it as all of them were used for hyper-parameter search, out of the total data points, 10% of the data is not involved in any of the hyper-parameter searches, while the remaining 90% of the data is further divided into training and validation set). Rather than optimizing all these hyper-parameters at once, we systematically evaluate them by first optimizing the number of data points and dimensions of the latent space while keeping the number of layers and neurons fixed.

The VAE architecture is built following Keras tutorial on variational autoencoder [366]. Initially, the number of layers and neurons is fixed to 500 neurons in the first and second layer for the encoder and decoder. Next, we perform a grid search cross-validation to find the optimum number of dimensions for the latent space [21, 42, 63, 84, 105, 126, 147, 168, 189] and the total data points required to yield the highest validation set accuracy. Figure D.2 and D.3 show the variation in reconstruction accuracy as the total number of data points dimensions of the latent space is varied. Reconstruction accuracy % is calculated by taking the total SMILE generated by the VAE model that matches the input SMILE to the total number of input SMILE.

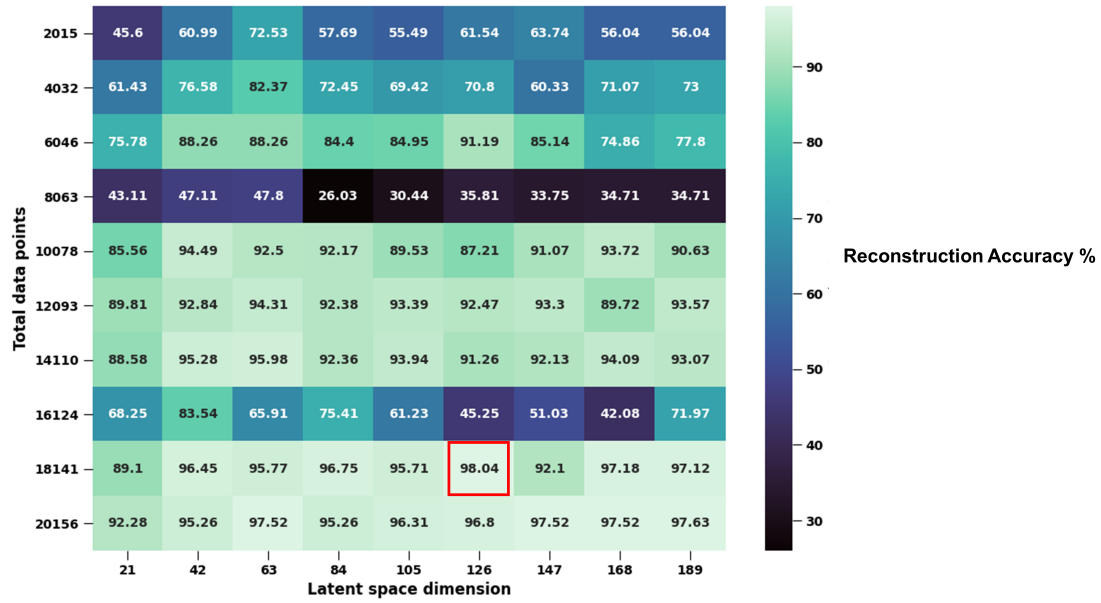


Figure D.2: Heat map of the the reconstruction accuracy % for the validation set during hyper-parameter search. X-axis indicates the latent space dimension. Y-axis indicates the total number of data points out of which 10% is kept aside as test data, the remaining 90% is further divided into training purposes and validation (90:10). Red square denotes the highest validation set reconstruction accuracy %.

Training Set

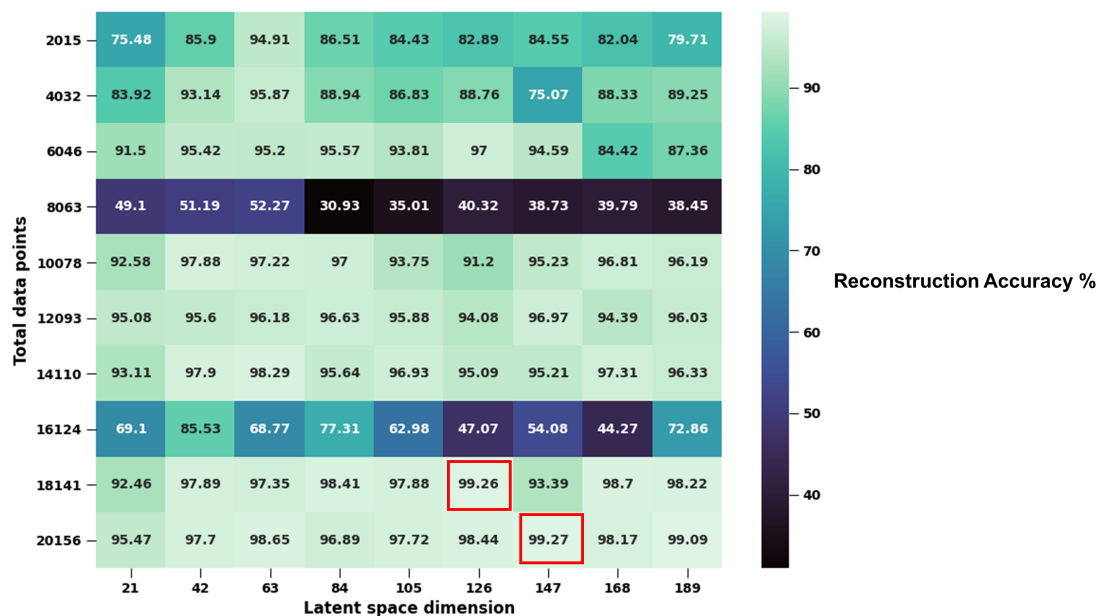


Figure D.3: Heat map of the reconstruction accuracy % for the training set during the hyper-parameter search. The X-axis indicates the latent space dimension. Y-axis indicates the total number of data points, of which 10% is kept aside as test data. The remaining 90% is further divided into training purposes and validation (90:10). Red squares denote the highest training set reconstruction accuracy %.

The highest reconstruction accuracy for the validation set peaks at 18,141 data points and 126 latent dimension space chosen as the optimal parameters. With the total data required for model development and latent space dimension fixed, we optimized the number of layers and number of neurons for the encoder and decoder. Based on the grid search combination, the validation set accuracy stayed relatively the same after reaching 500 neurons on the first layer and 250 on the second layer became the final architecture.

Figure D.1 (c) depicts the final optimized VAE architecture. The first hidden layer has 500 neurons with a dropout regularization of 0.3 to avoid overfitting. The second hidden layer has 250 neurons followed by a lambda layer of 126 neurons for μ

and σ to form the latent space with 126 dimensions. The decoder is represented exactly opposite of the encoder layer as it receives input from latent space and passes it to two dense layers of 250 and 500. The output of the decoder is in the shape of (1,2100), similar to input data. Finally, the one-hot encoded vector output is reshaped to (100,21) and converted to the SMILE string. The VAE is trained for 300 epochs with a batch size of 128 using the Adam optimizer as implemented in the Keras package [367].

Classification Model

Besides generating new cations, it is also important to classify them by cation family type to get a sense of the family lineage of these cations. Instead of manually examining them individually and classifying them by cation family type, we attempted to build a classification model that classifies cation into the respective family. However, the issue would be to classify cations that do not belong to the known cation groups and label them as "Unknown" as the model has never seen the structures for the label "Unknown" type. Thus, to solve this issue, we developed a classification model based on the cations SMILE data with the respective cation type family label and neutral organic molecule label as "Unknown". This idea was to train the model to classify anything besides the known cation family type as "Unknown" indicating that these cations do not belong to the known cation type and are new molecule structures. The neutral organic molecule database contained 9000 organic molecules from the ChEMBL database [368].

Chemical descriptors generated using RDKit [323] became input features for the classification model, normalized using MinMax Scaling [144]. Labels for the cations are individual family names, while the organic molecules are labeled "Unknown". The data is split into train/test split with a ratio of 90:10 to validate the model's accuracy. We employ Random Forest (RF) model implemented in Scikit-learn to classify the molecules [144].

Remarkably, the RF model with default parameters can classify all the cations and molecules on the training set and the test set with 100% accuracy, suggesting that these individual cation family types have a unique signature easily distinguishable from one another. We later discuss this unique signature of individual cation type with the visualization of the chemical features on a two-dimensional projection using Principal Component Analysis (PCA) that clearly shows the individual region of chemical space for each cation family type.

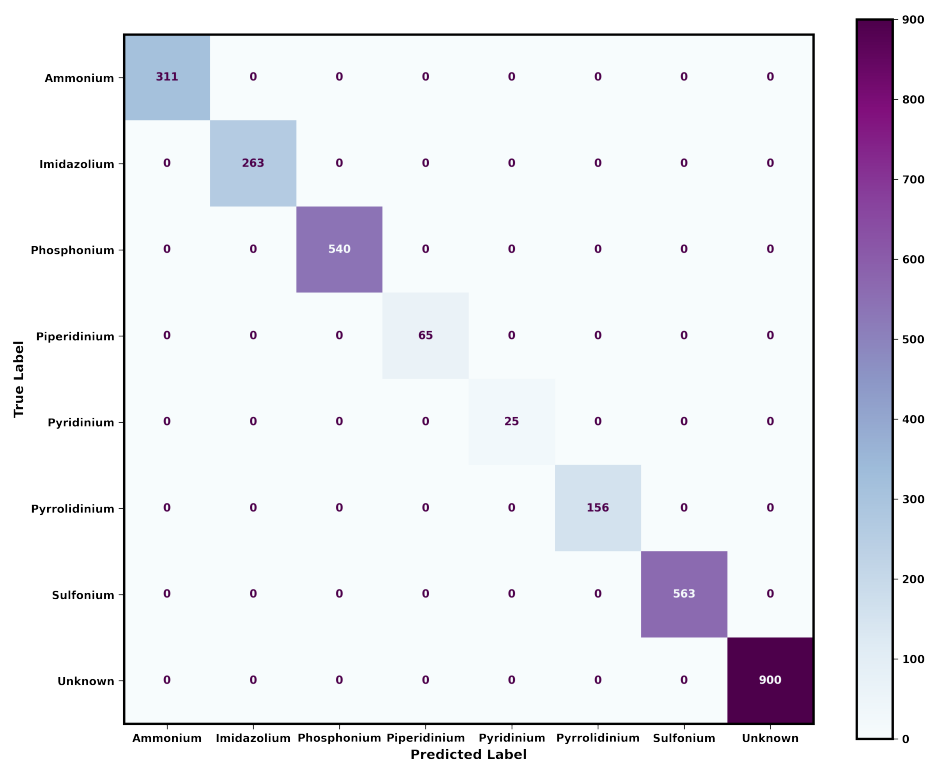


Figure D.4: Confusion matrix prediction of the test set using the classification model.

Latent Space Representation

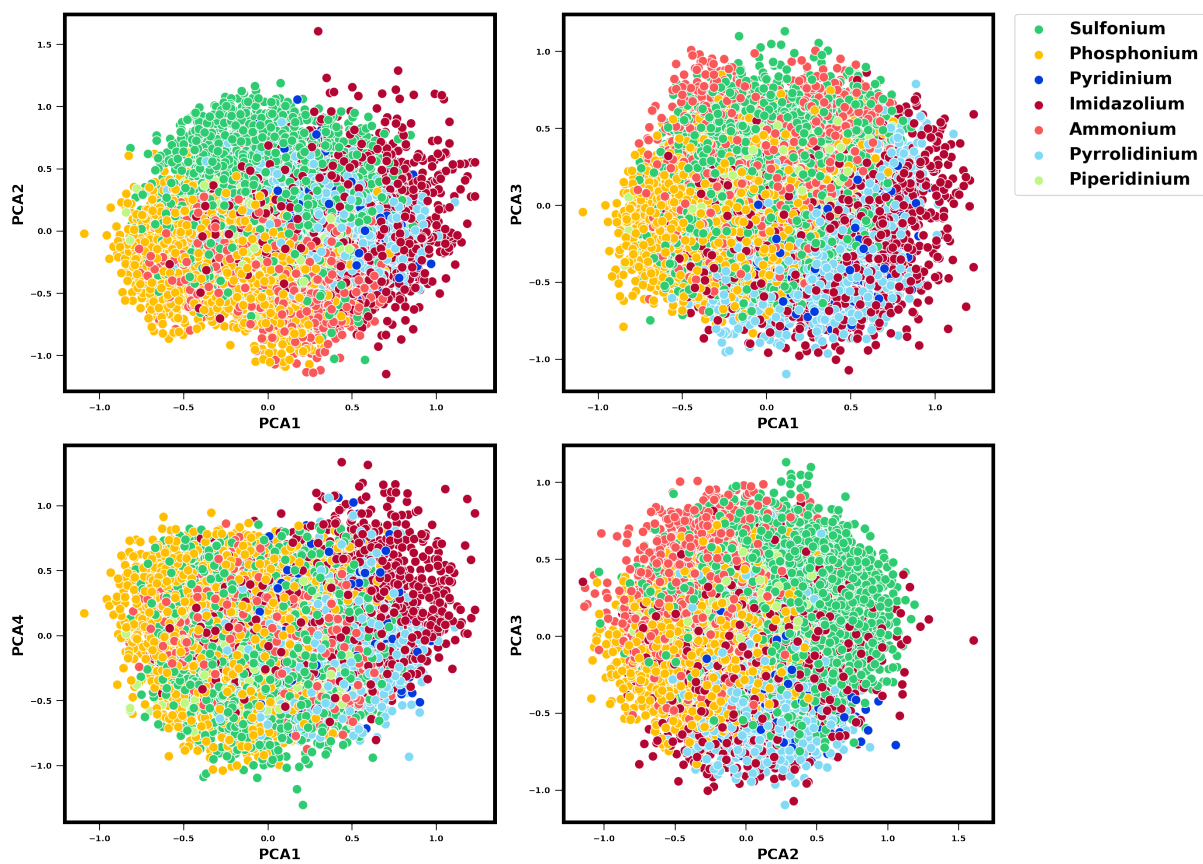


Figure D.5: Principal component analysis (PCA) of the latent space for the training data labeled by cation type.

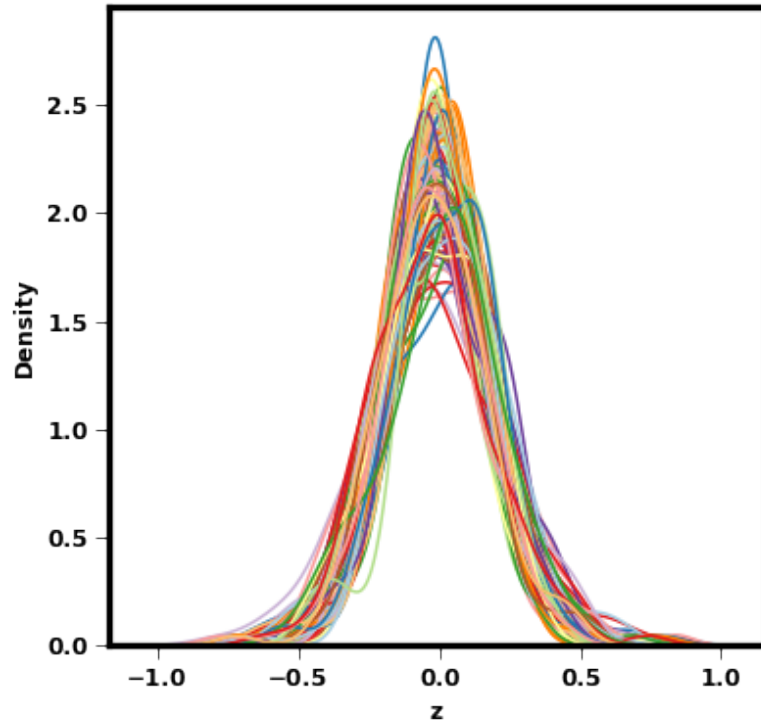


Figure D.6: Kernel Density Estimation (KDE) of each of the latent space attribute.

Non-Aromatic/Non-Cyclic cations

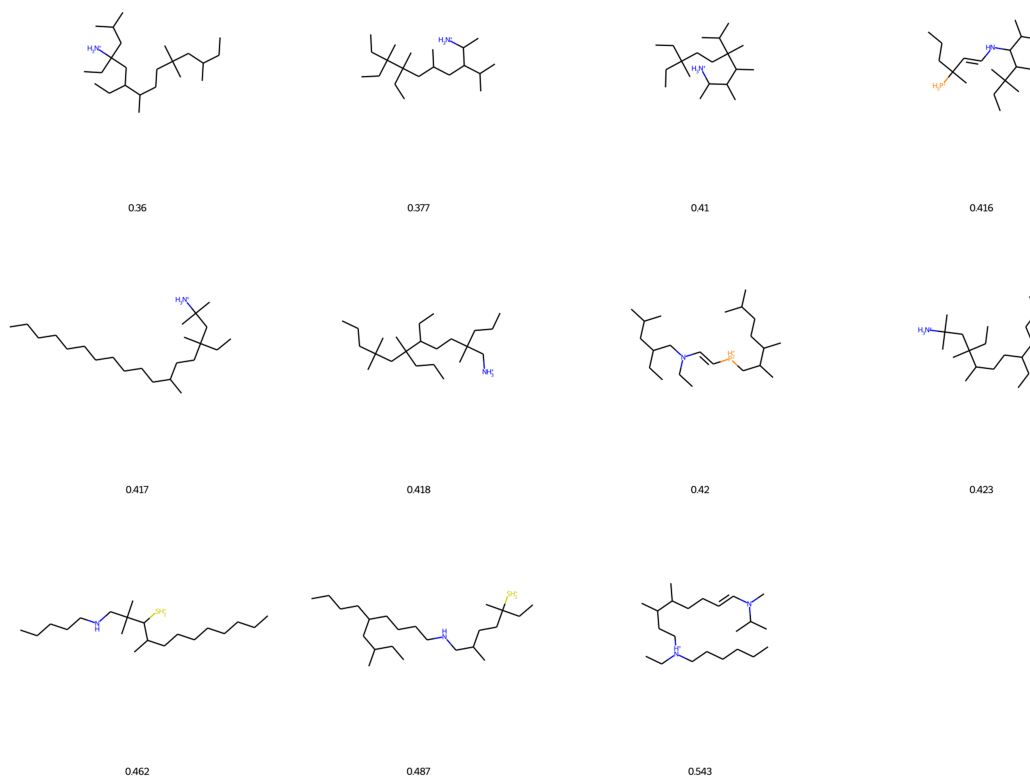


Figure D.7: VAE generated "Unknown" cation type as labeled by the classification model. The numbers below the cation denote the maximum of Tanimoto similarity index.

Aromatic/Cyclic cations

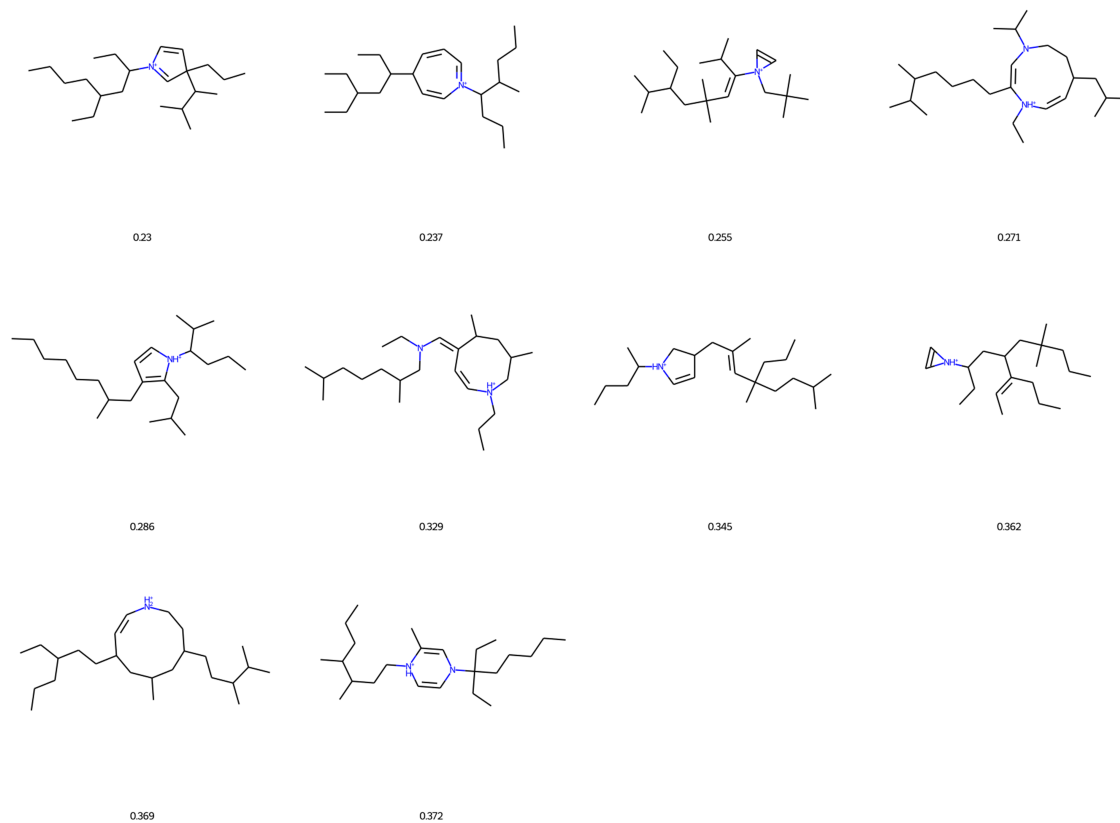


Figure D.8: Top 10 least similar aromatic/cyclic cations from the "Unknown" cation family based on Tanimoto similarity index. The numbers below the cations denote the maximum of Tanimoto similarity index.

Electrochemical Window Calculation

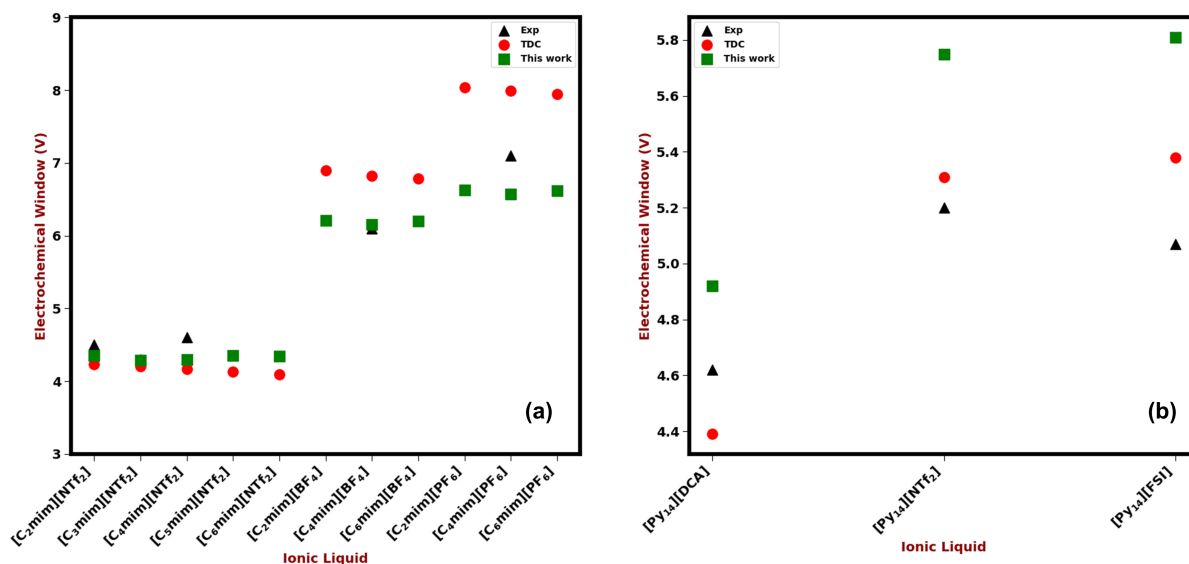


Figure D.9: Benchmarking electrochemical window calculations using the methodology adopted in this work against (a) The work of Kazemiabnavi et al. [42] as indicated by the red circle denoting thermodynamic cycle (TDC) approach calculated against Li^+/Li reference electrode. Experiment data are also added for some of the ionic liquids. [11]. (b) Asha et al. [104] work using TDC approach also calculated against Li^+/Li reference electrode for pyrrolidinium ionic liquids and experimental data. [22, 301]

MD simulation

The MD simulation involved five steps: Minimization, Annealing, *NVT* Equilibrium, *NPT* Equilibrium, and *NPT* Production Run. The *NPT* Production run lasted 60 ns, out of which the last 20 ns were used for analysis. Packmol was used to prepare the simulation box of 250 ion pairs [369]. The simulation was performed using the Gromacs package [370].

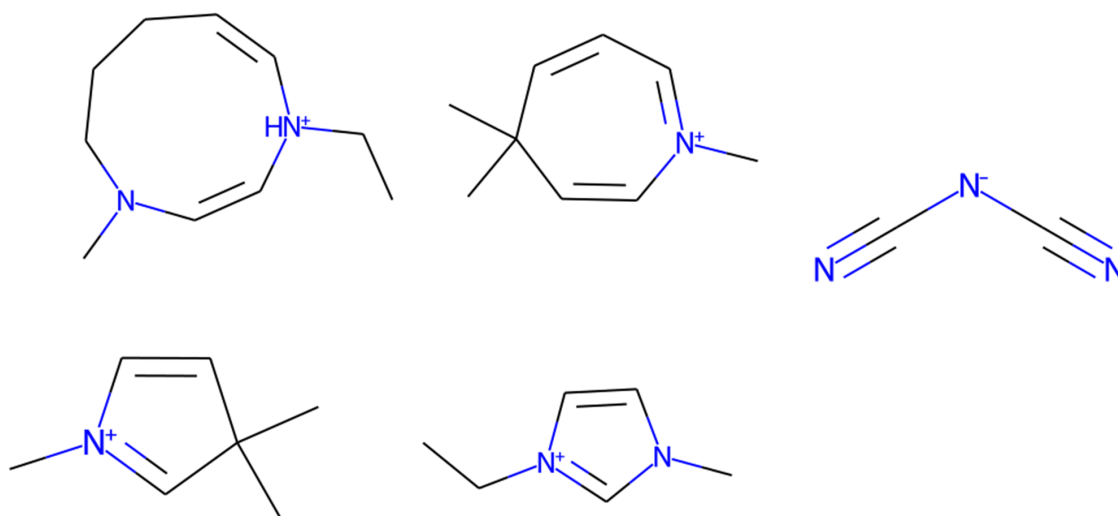


Figure D.10: Cations and anion involved in MD simulation.

Besides having high electrochemical stability, it is also crucial for ionic liquids to have faster dynamics as such specifications are essential for battery application. This section evaluates the dynamics of a few of the selected "Unknown" cyclic/aromatic cations that had the lowest Tanimoto similarity index. The dynamics, along with structural analysis, are examined using MD simulation on four of the cations shown in Figure D.10. Note that these four cations are the same as the ones shown in Figure D.8 but with alkyl chains trimmed on all sides to avoid sluggish dynamics during simulation.

The results of these cations are compared to 1-ethyl-3-methylimidazolium $[C_2mim]^+$ cation paired with dicyanamide $[DCA]^-$ anion as it has one of the fastest dynamics compared to other ionic liquids [11]. Since there are no forcefields available for the VAE generated cations, we obtained the forcefields from the work of Malde et al. [371] using Automatic Force Field Topology Builder (ATB). Although there are plenty of force field data for $[C_2mim]^+$ cation and $[DCA]^-$ anion available in the literature, to be consistent, we used ATB for this system as well. The electrostatic interaction was scaled to ± 0.8 to induce polarizability. The supporting information contains the entire MD simulation protocol and forcefield data. The

simulation temperature is set at 323.15 K to enable faster dynamics and to ensure the ionic liquids are in a liquid state. Out of the four VAE generated cations, one of the cations repeatedly failed to equilibrate; thus, it is discarded from further analysis.

Figure D.11 depicts the self-diffusion coefficient and center of mass (COM) radial distribution function (RDF) of the four ionic liquids. The self-diffusion constant is obtained by fitting 8-18 ns of the mean square displacement. Interestingly, the imidazolium-based ionic liquid [C₂mim][DCA] is still the fastest dynamics compared to the rest, followed by the modified imidazolium cation without the second nitrogen atom and the seven-member cation ring structure. The slowest cation is the bulkier protic cation with a large cyclic structure that significantly slows down the system's dynamics. A peculiar trait of the dicyanamide-based ionic liquid is that the anion has much faster dynamics than the cation, which is the opposite for most anions [236].

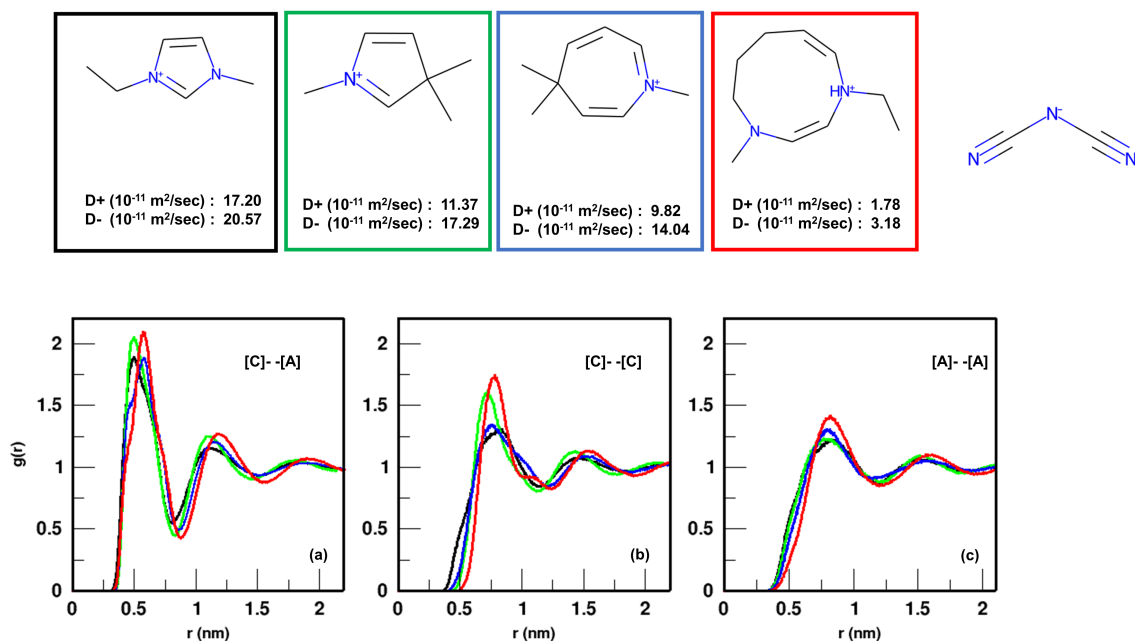


Figure D.11: Self diffusion constant and center of mass radial distribution function for four ionic liquids calculated using MD simulation at 323.15 K.

We also examined the com rdf for the cation-anion, cation-cation and anion-anion interaction as seen in Figure D.11 (a), (b) and (c). The cation-cation and cation-anion rdf demonstrate a sharp peak for the first solvation shell and a gradual decrease in peak height for the second solvation shell. However, the anion-anion interaction has a broader first solvation shell with a small peak height. For the cation-anion, the distance at which the first peak height occurs varies with the size of the cation. The bulkier cation (2) and (3) are located the furthest compared to cations (1) and (2) as these two cations have a relatively smaller size. As for the cation-cation interaction, cation (4) has the highest peak height indicating a possible aggregation of the cations because of the enormous ring structure.

CHAPTER E

Appendix: Mapping the Frontier Orbital Energies of Imidazolium-Based Cations

Cation Chemical Space

Table E.1: 1-alkyl-3-methylimidazolium ($[C_m C_n \text{mim}]^+$) branched isomer cation chemical space for alkyl chain length from 1 to 10.

	$[C_m C_n \text{mim}]^+$									
	10	9	8	7	6	5	4	3	2	1
10	128778									
9	106977	22366								
8	45123	18779	4005							
7	19773	8229	3471	780						
6	8619	3587	1513	663	153					
5	4056	1688	712	312	136	36				
4	2028	844	356	156	68	32	10			
3	1014	422	178	78	34	16	8	3		
2	507	211	89	39	17	8	4	2	1	
1	507	211	89	39	17	8	4	2	1	1

Energy Gap

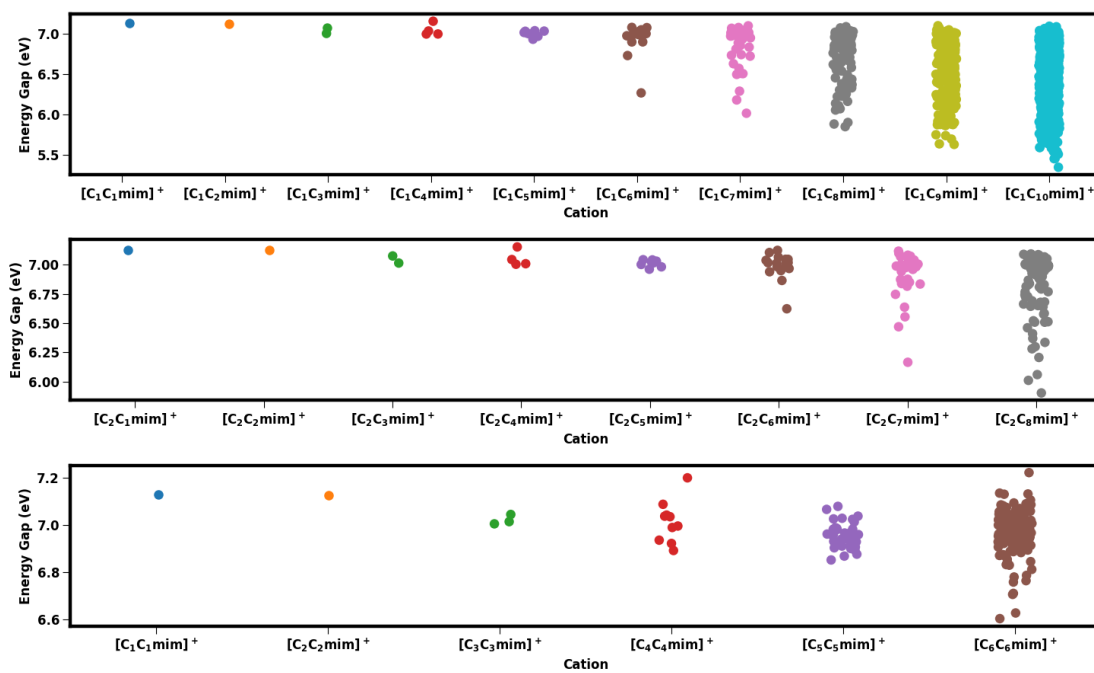


Figure E.1: Energy gap (eV) of $[C_n C_m \text{im}]^+$ cations calculated using DFT.

Hyperparameter Tuning

```
# Create the parameter grid based on the results of random search
params = { 'max_depth': [3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20],
          'learning_rate': [0.0005,0.0006,0.0007,0.0008,0.0009,0.001,0.002,0.003,0.004,0.005,0.006,0.007,
                           0.008,0.009,0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.09,0.1],
          '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': [50,100,150,200,250,300,350,400,450,500,550,600,650,750,800,850,900,950,1000,1050,1100,1200,
                           1300,1400,1500,1600,1700,1800,1900,2000,2100,2200]}
```

Figure E.2: XGBoost hyperparameter grid space for the HOMO model.

Final model parameter : n_estimators=1400, max_depth=3, learning_rate=0.06,
colsample_bytree=0.8, subsample=0.8, colsample_bylevel=0.9

```
# Create the parameter grid based on the results of random search
params = { 'max_depth': [3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20],
          'learning_rate': [0.0005,0.0006,0.0007,0.0008,0.0009,0.001,0.002,0.003,0.004,0.005,0.006,
                           0.007,0.008,0.009,0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.09,0.1],
          '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': [50,100,150,200,250,300,350,400,450,500,550,600,650,750,800,850,900,950,1000,1050,1100,1200,
                           1300,1400,1500,1600,1700,1800,1900,2000,2100,2200]}
```

Figure E.3: XGBoost hyperparameter grid space for the LUMO model.

Final model parameter : n_estimators=600, max_depth=5 ,learning_rate=0.07,
colsample_bytree=0.8, subsample=0.8, colsample_bylevel=0.1

Model Validation

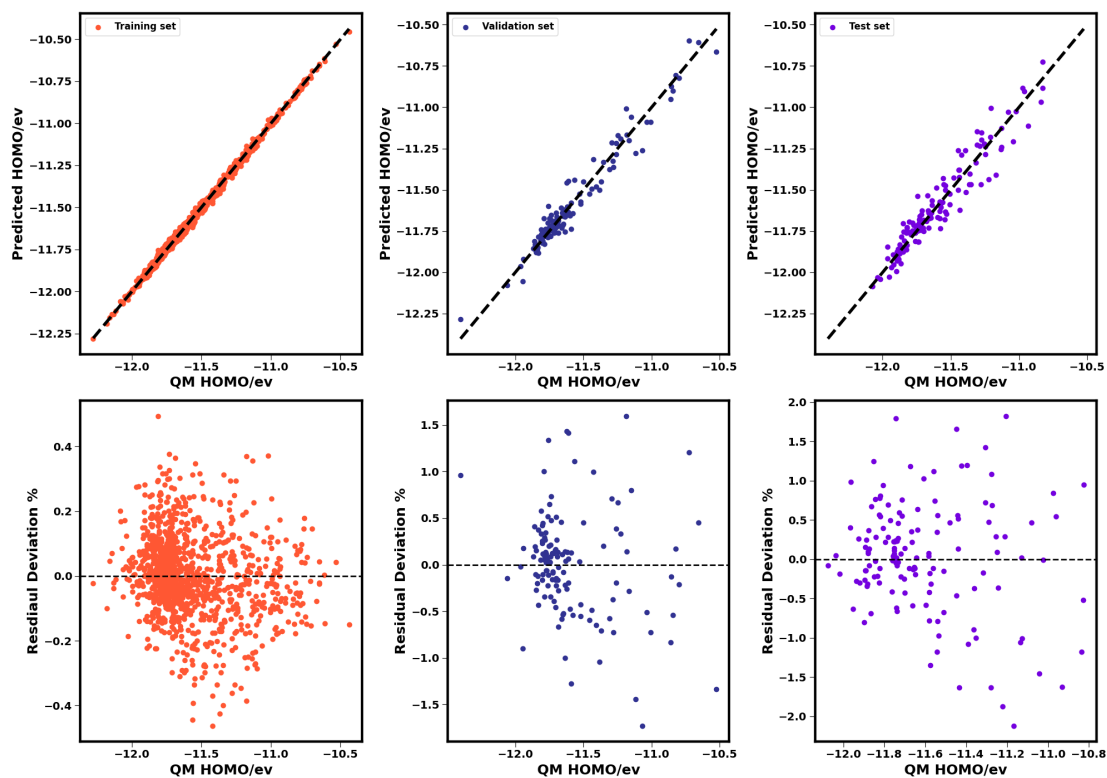


Figure E.4: Predicted vs DFT calculations of HOMO energy for training, validation and test set. Residual deviation % is the difference in DFT and model divided the DFT value.

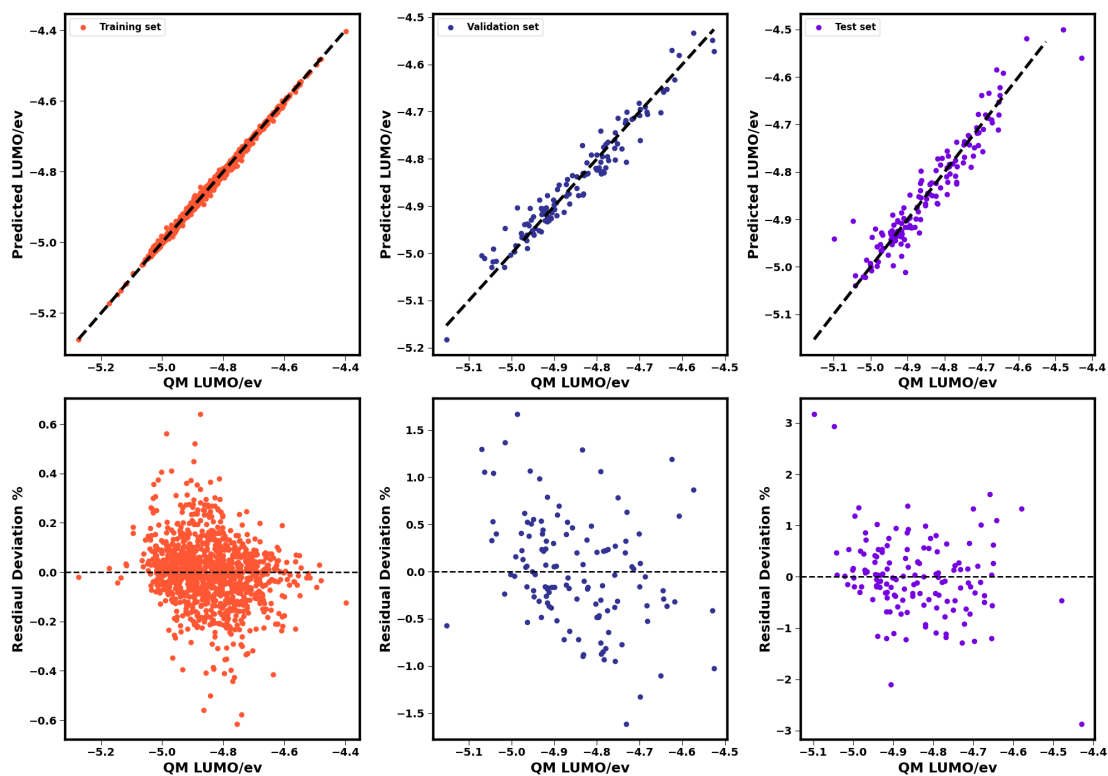


Figure E.5: Predicted vs DFT calculations of LUMO energy for training, validation and test set. Residual deviation % is the difference in DFT and model divided the DFT value.

Bibliography

- [1] Micah S Ziegler and Jessika E Trancik. Re-examining rates of lithium-ion battery technology improvement and cost decline. *Energy & Environmental Science*, 14(4):1635–1651, 2021.
- [2] Joshua Thomas Jameson Burd, Elizabeth A Moore, Hesham Ezzat, Randolph Kirchain, and Richard Roth. Improvements in electric vehicle battery technology influence vehicle lightweighting and material substitution decisions. *Applied Energy*, 283:116269, 2021.
- [3] Jianhui Wang, Yuki Yamada, Keitaro Sodeyama, Eriko Watanabe, Koji Takada, Yoshitaka Tateyama, and Atsuo Yamada. Fire-extinguishing organic electrolytes for safe batteries. *Nature Energy*, 3(1):22–29, 2018.
- [4] E Peter Roth and Christopher J Orendorff. How electrolytes influence battery safety. *The Electrochemical Society Interface*, 21(2):45, 2012.
- [5] P Isken, C Dippel, R Schmitz, RW Schmitz, M Kunze, S Passerini, M Winter, and A Lex-Balducci. High flash point electrolyte for use in lithium-ion batteries. *Electrochimica Acta*, 56(22):7530–7535, 2011.
- [6] Julian Kalhoff, Gebrekidan Gebresilassie Eshetu, Dominic Bresser, and Stefano Passerini. Safer electrolytes for lithium-ion batteries: state of the art and perspectives. *ChemSusChem*, 8(13):2154–2175, 2015.
- [7] Roger Atkinson and Janet Arey. Atmospheric degradation of volatile organic compounds. *Chemical reviews*, 103(12):4605–4638, 2003.

- [8] Yuanyuan Cao and Tiancheng Mu. Comprehensive investigation on the thermal stability of 66 ionic liquids by thermogravimetric analysis. *Industrial & engineering chemistry research*, 53(20):8651–8664, 2014.
- [9] Arunprakash T Karunanithi and Amirhossein Mehrkesh. Computer-aided design of tailor-made ionic liquids. *AIChE Journal*, 59(12):4627–4640, 2013.
- [10] Daniel M Lowe, Peter T Corbett, Peter Murray-Rust, and Robert C Glen. Chemical name to structure: Opsin, an open source solution, 2011.
- [11] Maciej Galiński, Andrzej Lewandowski, and Izabela Stępnia. Ionic liquids as electrolytes. *Electrochimica acta*, 51(26):5567–5580, 2006.
- [12] Thomas E Sutto. The electrochemical behavior of trialkylimidazolium imide based ionic liquids and their polymer gel electrolytes. *Journal of the Electrochemical Society*, 154(11):P130, 2007.
- [13] Ekaterina I Izgorodina, Radha Maganti, Vanessa Armel, Pamela M Dean, Jennifer M Pringle, Kenneth R Seddon, and Douglas R MacFarlane. Understanding the effect of the c2 proton in promoting low viscosities and high conductivities in imidazolium-based ionic liquids: part i. weakly coordinating anions. *The Journal of Physical Chemistry B*, 115(49):14688–14697, 2011.
- [14] Isabel Bandrés, Diego F Montaña, Ignacio Gascón, Pilar Cea, and Carlos Lafuente. Study of the conductivity behavior of pyridinium-based ionic liquids. *Electrochimica acta*, 55(7):2252–2257, 2010.
- [15] Qing-Shan Liu, Pei-Pei Li, Urs Welz-Biermann, Jian Chen, and Xiao-Xia Liu. Density, dynamic viscosity, and electrical conductivity of pyridinium-based hydrophobic ionic liquids. *The Journal of Chemical Thermodynamics*, 66:88–94, 2013.
- [16] Qing-Guo Zhang, Ying Wei, Si-Si Sun, Chang Wang, Miao Yang, Qing-Shan Liu, and Yan-An Gao. Study on thermodynamic properties of ionic liquid n-butyl-3-methylpyridinium bis (trifluoromethylsulfonyl) imide. *Journal of Chemical & Engineering Data*, 57(8):2185–2190, 2012.

- [17] Milan Vranes, Sanja Dozic, Vesna Djeric, and Slobodan Gadzuric. Physicochemical characterization of 1-butyl-3-methylimidazolium and 1-butyl-1-methylpyrrolidinium bis (trifluoromethylsulfonyl) imide. *Journal of Chemical & Engineering Data*, 57(4):1072–1077, 2012.
- [18] Sónia PM Ventura, Ana MM Gonçalves, Tânia Sintra, Joana L Pereira, Fernando Gonçalves, and Joao AP Coutinho. Designing ionic liquids: the chemical structure role in the toxicity. *Ecotoxicology*, 22(1):1–12, 2013.
- [19] IF Mena, E Diaz, J Palomar, JJ Rodriguez, and AF Mohedano. Cation and anion effect on the biodegradability and toxicity of imidazolium–and choline–based ionic liquids. *Chemosphere*, 240:124947, 2020.
- [20] Nareerat Plylahan, Manfred Kerner, Du-Hyun Lim, Aleksandar Matic, and Patrik Johansson. Ionic liquid and hybrid ionic liquid/organic electrolytes for high temperature lithium-ion battery application. *Electrochimica Acta*, 216:24–34, 2016.
- [21] Shaohua Fang, Zhengxi Zhang, Yide Jin, Li Yang, Shin-ichi Hirano, Kazuhiro Tachibana, and Shingo Katayama. New functionalized ionic liquids based on pyrrolidinium and piperidinium cations with two ether groups as electrolytes for lithium battery. *Journal of Power Sources*, 196(13):5637–5644, 2011.
- [22] Nedher Sanchez-Ramirez, Birhanu Desalegn Assresahegn, Daniel Belanger, and Roberto M Torresi. A comparison among viscosity, density, conductivity, and electrochemical windows of n-n-butyl-n-methylpyrrolidinium and triethyl-n-pentylphosphonium bis (fluorosulfonyl imide) ionic liquids and their analogues containing bis (trifluoromethylsulfonyl) imide anion. *Journal of Chemical & Engineering Data*, 62(10):3437–3444, 2017.
- [23] Muna Hassan Ibrahim, Maan Hayyan, Mohd Ali Hashim, Adeeb Hayyan, and Mohamed K Hadj-Kali. Physicochemical properties of piperidinium, ammonium, pyrrolidinium and morpholinium cations based ionic liquids paired with bis (trifluoromethylsulfonyl) imide anion. *Fluid Phase Equilibria*, 427:18–26, 2016.

- [24] Ramzi Zarrougui, Noureddine Raouafi, and Daniel Lemordant. New series of green cyclic ammonium-based room temperature ionic liquids with alkylphosphite-containing anion: Synthesis and physicochemical characterization. *Journal of Chemical & Engineering Data*, 59(4):1193–1201, 2014.
- [25] Jianyang Weng, Congmin Wang, Haoran Li, and Yong Wang. Novel quaternary ammonium ionic liquids and their use as dual solvent-catalysts in the hydrolytic reaction. *Green Chemistry*, 8(1):96–99, 2006.
- [26] Kevin J Fraser and Douglas R MacFarlane. Phosphonium-based ionic liquids: an overview. *Australian journal of chemistry*, 62(4):309–321, 2009.
- [27] Anthony JR Rennie, Vitor L Martins, Roberto M Torresi, and Peter J Hall. Ionic liquids containing sulfonium cations as electrolytes for electrochemical double layer capacitors. *The Journal of Physical Chemistry C*, 119(42):23865–23874, 2015.
- [28] Maral PS Mousavi, Benjamin E Wilson, Sadra Kashefolgheta, Evan L Anderson, Siyao He, Philippe Bühlmann, and Andreas Stein. Ionic liquids as electrolytes for electrochemical double-layer capacitors: structures that optimize specific energy. *ACS applied materials & interfaces*, 8(5):3396–3406, 2016.
- [29] Pedro J Carvalho, Sónia PM Ventura, Marta LS Batista, Bernd Schröder, Fernando Gonçalves, José Esperança, Fabrice Mutelet, and João AP Coutinho. Understanding the impact of the central atom on the ionic liquid behavior: phosphonium vs ammonium cations. *The Journal of chemical physics*, 140(6):064505, 2014.
- [30] Jose A Vega, Junfeng Zhou, and Paul A Kohl. Electrochemical comparison and deposition of lithium and potassium from phosphonium- and ammonium-tfsi ionic liquids. *Journal of The Electrochemical Society*, 156(4):A253, 2009.
- [31] Laura Katharina Scarbath-Evers, Patricia A Hunt, Barbara Kirchner, Douglas R MacFarlane, and Stefan Zahn. Molecular features contributing to the

- lower viscosity of phosphonium ionic liquids compared to their ammonium analogues. *Physical Chemistry Chemical Physics*, 17(31):20205–20216, 2015.
- [32] Chuan-Pei Lee, Jia-De Peng, D Velayutham, Jeffrey Chang, Ping-Wei Chen, V Suryanarayanan, and Kuo-Chuan Ho. Trialkylsulfonium and tetraalkylammonium cations-based ionic liquid electrolytes for quasi-solid-state dye-sensitized solar cells. *Electrochimica Acta*, 114:303–308, 2013.
- [33] Arijit Bhattacharjee, Andreia Luís, João H Santos, José A Lopes-da Silva, Mara G Freire, Pedro J Carvalho, and João AP Coutinho. Thermophysical properties of sulfonium-and ammonium-based ionic liquids. *Fluid phase equilibria*, 381:36–45, 2014.
- [34] Abner Massari Sampaio, Guilherme Ferreira Lemos Pereira, Mathieu Salanne, and Leonardo José Amaral Siqueira. Comparing the performance of sulfonium and phosphonium ionic liquids as electrolytes for supercapacitors by molecular dynamics simulations. *Electrochimica Acta*, 364:137181, 2020.
- [35] Shaokun Tang, Gary A Baker, and Hua Zhao. Ether-and alcohol-functionalized task-specific ionic liquids: attractive properties and applications. *Chemical Society Reviews*, 41(10):4030–4066, 2012.
- [36] Jianhao Zhang, Shaohua Fang, Long Qu, Yide Jin, Li Yang, and Shin-ichi Hirano. Synthesis, characterization, and properties of ether-functionalized 1, 3-dialkylimidazolium ionic liquids. *Industrial & Engineering Chemistry Research*, 53(43):16633–16643, 2014.
- [37] Alex R Neale, Sinead Murphy, Peter Goodrich, Christopher Hardacre, and Johan Jacquemin. Thermophysical and electrochemical properties of ethereal functionalised cyclic alkylammonium-based ionic liquids as potential electrolytes for electrochemical applications. *ChemPhysChem*, 18(15):2040–2057, 2017.
- [38] Kazuki Yoshii, Takuya Uto, Naoki Tachikawa, and Yasushi Katayama. The effects of the position of the ether oxygen atom in pyrrolidinium-based

- room temperature ionic liquids on their physicochemical properties. *Physical Chemistry Chemical Physics*, 22(35):19480–19491, 2020.
- [39] Tzi-Yi Wu, Shyh-Gang Su, Keng-Fu Lin, Yuan-Chung Lin, H Paul Wang, Ming-Wei Lin, Shr-Tusen Gung, and I-Wen Sun. Voltammetric and physicochemical characterization of hydroxyl- and ether-functionalized onium bis (trifluoromethanesulfonyl) imide ionic liquids. *Electrochimica acta*, 56(21):7278–7287, 2011.
- [40] Yukihiro Yoshida, Osamu Baba, and Gunzi Saito. Ionic liquids based on dicyanamide anion: influence of structural variations in cationic structures on ionic conductivity. *The Journal of Physical Chemistry B*, 111(18):4742–4749, 2007.
- [41] Vitor L Martins, Anthony JR Rennie, Nedher Sanchez-Ramirez, Roberto M Torresi, and Peter J Hall. Improved performance of ionic liquid supercapacitors by using tetracyanoborate anions. *ChemElectroChem*, 5(4):598, 2018.
- [42] Saeed Kazemiabnavi, Zhengcheng Zhang, Katsuyo Thornton, and Soumik Banerjee. Electrochemical stability window of imidazolium-based ionic liquids as electrolytes for lithium batteries. *The Journal of Physical Chemistry B*, 120(25):5691–5702, 2016.
- [43] Yong-Hui Tian, George S Goff, Wolfgang H Runde, and Enrique R Batista. Exploring electrochemical windows of room-temperature ionic liquids: a computational study. *The Journal of Physical Chemistry B*, 116(39):11943–11952, 2012.
- [44] Raquel FM Frade and Carlos AM Afonso. Impact of ionic liquids in environment and humans: an overview. *Human & experimental toxicology*, 29(12):1038–1054, 2010.
- [45] Jennifer Neumann, Chul-Woong Cho, Stephanie Steudte, Jan Köser, Marc Uerdingen, Jorg Thöming, and Stefan Stolte. Biodegradability of fluoroor-

- ganic and cyano-based ionic liquid anions under aerobic and anaerobic conditions. *Green chemistry*, 14(2):410–418, 2012.
- [46] Preston Griffin, Selene Ramer, Matthew Winfough, and Jakub Kostal. Practical guide to designing safer ionic liquids for cellulose dissolution using a tiered computational framework. *Green Chemistry*, 22(11):3626–3637, 2020.
- [47] Yuya Hiraga, Aya Kato, Yoshiyuki Sato, and Richard L Smith Jr. Densities at pressures up to 200 mpa and atmospheric pressure viscosities of ionic liquids 1-ethyl-3-methylimidazolium methylphosphate, 1-ethyl-3-methylimidazolium diethylphosphate, 1-butyl-3-methylimidazolium acetate, and 1-butyl-3-methylimidazolium bis (trifluoromethylsulfonyl) imide. *Journal of Chemical & Engineering Data*, 60(3):876–885, 2015.
- [48] Kenta Fukumoto, Masahiro Yoshizawa, and Hiroyuki Ohno. Room temperature ionic liquids from 20 natural amino acids. *Journal of the American Chemical Society*, 127(8):2398–2399, 2005.
- [49] Miriam Kunze, Sangsik Jeong, Giovanni B Appetecchi, Monika Schönhoff, Martin Winter, and Stefano Passerini. Mixtures of ionic liquids for low temperature electrolytes. *Electrochimica Acta*, 82:69–74, 2012.
- [50] H Every, AG Bishop, Maria Forsyth, and Douglas R Macfarlane. Ion diffusion in molten salt mixtures. *Electrochimica Acta*, 45(8-9):1279–1284, 2000.
- [51] Heiko Niedermeyer, Jason P Hallett, Ignacio J Villar-Garcia, Patricia A Hunt, and Tom Welton. Mixtures of ionic liquids. *Chemical Society Reviews*, 41(23):7780–7802, 2012.
- [52] Utkarsh Kapoor and Jindal K Shah. Macroscopic differentiators for microscopic structural nonideality in binary ionic liquid mixtures. *The Journal of Physical Chemistry B*, 124(36):7849–7856, 2020.
- [53] Nicholas J Brooks, Franca Castiglione, Cara M Doherty, Andrew Dolan, Anita J Hill, Patricia A Hunt, Richard P Matthews, Michele Mauri, Andrea

- Mele, Roberto Simonutti, et al. Linking the structures, free volumes, and properties of ionic liquid mixtures. *Chemical science*, 8(9):6359–6374, 2017.
- [54] Utkarsh Kapoor and Jindal K Shah. Preferential ionic interactions and microscopic structural changes drive nonideality in binary ionic liquid mixtures as revealed from molecular simulations. *Industrial & Engineering Chemistry Research*, 55(51):13132–13146, 2016.
- [55] Ying Huang, Xiangping Zhang, Xin Zhang, Haifeng Dong, and Suojian Zhang. Thermodynamic modeling and assessment of ionic liquid-based CO₂ capture processes. *Industrial & Engineering Chemistry Research*, 53(29):11805–11817, 2014.
- [56] Joan F Brennecke and Burcu E Gurkan. Ionic liquids for CO₂ capture and emission reduction. *The Journal of Physical Chemistry Letters*, 1(24):3459–3464, 2010.
- [57] H elene Olivier-Bourbigou, L Magna, and D Morvan. Ionic liquids and catalysis: Recent progress from knowledge to applications. *Applied Catalysis A: General*, 373(1-2):1–56, 2010.
- [58] Silvia Garc ıa, Marcos Larriba, Juli an Garc ıa, Jos e S Torrecilla, and Francisco Rodr ıguez. Liquid–liquid extraction of toluene from n-heptane using binary mixtures of n-butylpyridinium tetrafluoroborate and n-butylpyridinium bis(trifluoromethylsulfonyl) imide ionic liquids. *Chemical engineering journal*, 180:210–215, 2012.
- [59] Juli an Garc ıa, Silvia Garc ıa, Jos e S Torrecilla, Mercedes Oliet, and Francisco Rodr ıguez. Liquid–liquid equilibria for the ternary systems {heptane+ toluene+ N-butylpyridinium tetrafluoroborate or N-hexylpyridinium tetrafluoroborate} at t= 313.2 k. *Journal of Chemical & Engineering Data*, 55(8):2862–2865, 2010.
- [60] Juli an Garc ıa, Silvia Garc ıa, Jos e S Torrecilla, and Francisco Rodr ıguez. N-butylpyridinium bis-(trifluoromethylsulfonyl) imide ionic liquids as solvents

for the liquid–liquid extraction of aromatics from their mixtures with alkanes: Isomeric effect of the cation. *Fluid phase equilibria*, 301(1):62–66, 2011.

- [61] Mehmet Isik, Haritz Sardon, and David Mecerreyes. Ionic liquids and cellulose: dissolution, chemical modification and preparation of new cellulosic materials. *International journal of molecular sciences*, 15(7):11922–11940, 2014.
- [62] Masayoshi Watanabe, Morgan L Thomas, Shiguo Zhang, Kazuhide Ueno, Tomohiro Yasuda, and Kaoru Dokko. Application of ionic liquids to energy storage and conversion materials and devices. *Chemical reviews*, 117(10):7190–7239, 2017.
- [63] Gebrekidan Gebresilassie Eshetu, Michel Armand, Bruno Scrosati, and Stefano Passerini. Energy storage materials synthesized from ionic liquids. *Angewandte Chemie International Edition*, 53(49):13342–13359, 2014.
- [64] Bruce Dunn, Haresh Kamath, and Jean-Marie Tarascon. Electrical energy storage for the grid: a battery of choices. *Science*, 334(6058):928–935, 2011.
- [65] Xiaozhou Huang, Rui He, Matthew Li, Mason Oliver Lam Chee, Pei Dong, and Jun Lu. Functionalized separator for next-generation batteries. *Materials Today*, 41:143–155, 2020.
- [66] Naoki Nitta, Feixiang Wu, Jung Tae Lee, and Gleb Yushin. Li-ion battery materials: present and future. *Materials Today*, 18(5):252–264, 2015.
- [67] Xiaoen Wang, Robert Kerr, Fangfang Chen, Nicolas Goujon, Jennifer M Pringle, David Mecerreyes, Maria Forsyth, and Patrick C Howlett. Toward high-energy-density lithium metal batteries: opportunities and challenges for solid organic electrolytes. *Advanced Materials*, 32(18):1905219, 2020.
- [68] Susanne Wilken, Marcel Treskow, Johan Scheers, Patrik Johansson, and Per Jacobsson. Initial stages of thermal decomposition of lipf₆-based lithium ion battery electrolytes by detailed raman and nmr spectroscopy. *Rsc Advances*, 3(37):16359–16364, 2013.

- [69] Xuning Feng, Mingguo Ouyang, Xiang Liu, Languang Lu, Yong Xia, and Xi-angming He. Thermal runaway mechanism of lithium ion battery for electric vehicles: A review. *Energy Storage Materials*, 10:246–267, 2018.
- [70] Neha Chawla, Neelam Bharti, and Shailendra Singh. Recent advances in non-flammable electrolytes for safer lithium-ion batteries. *Batteries*, 5(1):19, 2019.
- [71] Akiko Tsurumaki, Marco Agostini, Ruggero Poiana, Lucia Lombardo, Ernestino Lufano, Cataldo Simari, Aleksandar Matic, Isabella Nicotera, Stefania Panero, and Maria Assunta Navarra. Enhanced safety and galvanostatic performance of high voltage lithium batteries by using ionic liquids. *Electrochimica Acta*, 316:1–7, 2019.
- [72] Huizhe Niu, Le Wang, Ping Guan, Nan Zhang, Chaoren Yan, Minling Ding, Xulong Guo, Tongtong Huang, and Xiaoling Hu. Recent advances in application of ionic liquids in electrolyte of lithium ion batteries. *Journal of Energy Storage*, 40:102659, 2021.
- [73] Qian Dong, Chris D Muzny, Andrei Kazakov, Vladimir Diky, Joseph W Magee, Jason A Widegren, Robert D Chirico, Kenneth N Marsh, and Michael Frenkel. Ilthermo: a free-access web database for thermodynamic properties of ionic liquids. *Journal of Chemical & Engineering Data*, 52(4):1151–1159, 2007.
- [74] Qian Dong, Andrei F Kazakov, Chris D Muzny, Robert D Chirico, Jason A Widegren, Vladimir Diky, Joe W Magee, Kenneth N Marsh, and Michael D Frenkel. Ionic liquids database (ilthermo). Technical report, 2006.
- [75] Tzi-Yi Wu, Shyh-Gang Su, H Paul Wang, Yuan-Chung Lin, Shr-Tusen Gung, Ming-Wei Lin, and I-Wen Sun. Electrochemical studies and self diffusion coefficients in cyclic ammonium based ionic liquids with allyl substituents. *Electrochimica Acta*, 56(9):3209–3218, 2011.

- [76] Byeongjin Baek, Sangjun Lee, and Cheolsoo Jung. Pyrrolidinium cation-based ionic liquids with different functional groups: butyl, butyronitrile, pentenyl, and methyl butyrate. *Int. J. Electrochem. Sci*, 6:6220–6234, 2011.
- [77] Fannie Alloin, Pierre Strobel, Jean-Claude Leprêtre, Laure Cointeaux, Carlos Pérez del Valle, et al. Electrolyte based on fluorinated cyclic quaternary ammonium ionic liquids. *Ionics*, 18(9):817–827, 2012.
- [78] My Loan Phung Le, Fannie Alloin, Pierre Strobel, Jean-Claude Leprêtre, Carlos Peérez del Valle, and Patrick Judeinstein. Structure- properties relationships of lithium electrolytes based on ionic liquid. *The Journal of Physical Chemistry B*, 114(2):894–903, 2010.
- [79] Ramzi Zarrougui, Rahma Hachicha, Refka Rjab, Sabri Messaoudi, and Ouassim Ghodbane. Physicochemical characterizations of novel dicyanamide-based ionic liquids applied as electrolytes for supercapacitors. *RSC advances*, 8(54):31213–31223, 2018.
- [80] Olga Russina, Ruggero Caminiti, Alessandro Triolo, Sunita Rajamani, Bernardo Melai, Alessandra Bertoli, and Cinzia Chiappe. Physico-chemical properties and nanoscale morphology in n-alkyl-n-methylmorpholinium dicyanamide room temperature ionic liquids. *Journal of molecular liquids*, 187:252–259, 2013.
- [81] Shiro Seki, Takeshi Kobayashi, Nobuyuki Serizawa, Yo Kobayashi, Katsuhito Takei, Hajime Miyashiro, Kikuko Hayamizu, Seiji Tsuzuki, Takushi Mitsugi, Yasuhiro Umebayashi, et al. Electrolyte properties of 1-alkyl-2, 3, 5-trimethylpyrazolium cation-based room-temperature ionic liquids for lithium secondary batteries. *Journal of Power Sources*, 195(18):6207–6211, 2010.
- [82] Ming Chai, Yide Jin, Shaohua Fang, Li Yang, Shin-ichi Hirano, and Kazuhiro Tachibana. Low-viscosity ether-functionalized pyrazolium ionic liquids as new electrolytes for lithium battery. *Journal of Power Sources*, 216:323–329, 2012.

- [83] Cinzia Chiappe, Angelo Sanzone, Daniele Mendola, Franca Castiglione, Antonino Famulari, Guido Raos, and Andrea Mele. Pyrazolium-versus imidazolium-based ionic liquids: structure, dynamics and physicochemical properties. *The Journal of Physical Chemistry B*, 117(2):668–676, 2013.
- [84] Lei Guo, Xu Pan, Changneng Zhang, Meng Wang, Molang Cai, Xiaqin Fang, and Songyuan Dai. Novel hydrophobic cyclic sulfonium-based ionic liquids as potential electrolyte. *Journal of Molecular Liquids*, 158(2):75–79, 2011.
- [85] Li Yang, ZhengXi Zhang, XuHui Gao, HuanQi Zhang, and Kiyotaka Mashita. Asymmetric sulfonium-based molten salts with tfsi- or pf₆- anion as novel electrolytes. *Journal of power sources*, 162(1):614–619, 2006.
- [86] Sinead Murphy, Flavien Ivol, Alex R Neale, Peter Goodrich, Fouad Ghamouss, Christopher Hardacre, and Johan Jacquemin. Acyclic and cyclic alkyl and ether-functionalised sulfonium ionic liquids based on the [tfsi]- and [fsi]- anions as potential electrolytes for electrochemical applications. *ChemPhysChem*, 19(23):3226–3236, 2018.
- [87] Johannes Landmann, Jan AP Sprenger, Philipp T Hennig, Rüdiger Bertermann, Matthias Grüne, Frank Würthner, Nikolai V Ignat’ev, and Maik Finze. Perfluoroalkyltricyanoborate and perfluoroalkylcyanofluoroborate anions: Building blocks for low-viscosity ionic liquids. *Chemistry–A European Journal*, 24(3):608–623, 2018.
- [88] Lisa A Bischoff, Michael Drisch, Christoph Kerpen, Philipp T Hennig, Johannes Landmann, Jan AP Sprenger, Rüdiger Bertermann, Matthias Grüne, Qinqin Yuan, Jonas Warneke, et al. Cyanohydridoborate anions: Synthesis, salts, and low-viscosity ionic liquids. *Chemistry–A European Journal*, 25(14):3560–3574, 2019.
- [89] Ramzi Zarrougui, Rahma Hachicha, Refka Rjab, and Ouassim Ghodbane. 1-allyl-3-methylimidazolium-based ionic liquids employed as suitable electrolytes for high energy density supercapacitors based on graphene nanosheets electrodes. *Journal of Molecular Liquids*, 249:795–804, 2018.

- [90] Gwan-Hong Min, Tae-eun Yim, Hyun-Yeong Lee, Dal-Ho Huh, Eun-joo Lee, Jun-young Mun, Seung M Oh, and Young-Gyu Kim. Synthesis and properties of ionic liquids: imidazolium tetrafluoroborates with unsaturated side chains. *Bulletin of the Korean Chemical Society*, 27(6):847–852, 2006.
- [91] Matthias Hilder, GMA Girard, K Whitbread, S Zavorine, M Moser, D Nucciarone, Maria Forsyth, DR MacFarlane, and PC Howlett. Physicochemical characterization of a new family of small alkyl phosphonium imide ionic liquids. *Electrochimica Acta*, 202:100–109, 2016.
- [92] Aleksandar Tot, Črtomir Podlipnik, Marija Bešter-Rogač, Slobodan Gadžurić, and Milan Vraneš. Influence of oxygen functionalization on physico-chemical properties of imidazolium based ionic liquids—experimental and computational study. *Arabian Journal of Chemistry*, 13(1):1598–1611, 2020.
- [93] Katsuhiko Tsunashima, Eiko Niwa, Shun Kodama, Masashi Sugiya, and Yasushi Ono. Thermal and transport properties of ionic liquids based on benzyl-substituted phosphonium cations. *The Journal of Physical Chemistry B*, 113(48):15870–15874, 2009.
- [94] Zhi-Bin Zhou, Hajime Matsumoto, and Kuniaki Tatsumi. Cyclic quaternary ammonium ionic liquids with perfluoroalkyltrifluoroborates: synthesis, characterization, and properties. *Chemistry—A European Journal*, 12(8):2196–2212, 2006.
- [95] Zachary P Rosol, Natalie J German, and Stephen M Gross. Solubility, ionic conductivity and viscosity of lithium salts in room temperature ionic liquids. *Green Chemistry*, 11(9):1453–1457, 2009.
- [96] Jakob Asenbauer, Nour Ben Hassen, Bryan D McCloskey, and John M Prausnitz. Solubilities and ionic conductivities of ionic liquids containing lithium salts. *Electrochimica Acta*, 247:1038–1043, 2017.

- [97] Jennifer M Pringle, Jake Golding, Craig M Forsyth, Glen B Deacon, Maria Forsyth, and Douglas R MacFarlane. Physical trends and structural features in organic salts of the thiocyanate anion. *Journal of Materials Chemistry*, 12(12):3475–3480, 2002.
- [98] J Brad Shotwell and Robert A Flowers. Electrochemical investigation of the solvolytic properties of ethylammonium nitrate (ean) and propylammonium nitrate (pan). *Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis*, 12(3):223–226, 2000.
- [99] Wen-Li Yuan, Xiao Yang, Ling He, Ying Xue, Song Qin, and Guo-Hong Tao. Viscosity, conductivity, and electrochemical property of dicyanamide ionic liquids. *Frontiers in chemistry*, 6:59, 2018.
- [100] Douglas R MacFarlane, Jake Golding, Stewart Forsyth, Maria Forsyth, and Glen B Deacon. Low viscosity ionic liquids based on organic salts of the dicyanamide anion. *Chemical communications*, (16):1430–1431, 2001.
- [101] Meriem Anouti, Magaly Caillon-Caravanier, Yosra Dridi, Herve Galiano, and Daniel Lemordant. Synthesis and characterization of new pyrrolidinium based protic ionic liquids. good and superionic liquids. *The Journal of Physical Chemistry B*, 112(42):13335–13343, 2008.
- [102] Mérièm Anouti and Laure Timperman. A pyrrolidinium nitrate protic ionic liquid-based electrolyte for very low-temperature electrical double-layer capacitors. *Physical Chemistry Chemical Physics*, 15(17):6539–6548, 2013.
- [103] Bernadette M Quinn, Zhifeng Ding, Roger Moulton, and Allen J Bard. Novel electrochemical studies of ionic liquids. *Langmuir*, 18(5):1734–1742, 2002.
- [104] Suseeladevi Asha, Kunduchi P Vijayalakshmi, and Benny K George. Pyrrolidinium-based ionic liquids as electrolytes for lithium batteries: A computational study. *International Journal of Quantum Chemistry*, 119(22):e26014, 2019.

- [105] Spyridon Koutsoukos, Frederik Philippi, Francisco Malaret, and Tom Welton. A review on machine learning algorithms for the ionic liquid chemical space. *Chemical science*, 12(20):6820–6843, 2021.
- [106] Bakhtyar Sepehri. A review on created qspr models for predicting ionic liquids properties and their reliability from chemometric point of view. *Journal of Molecular Liquids*, 297:112013, 2020.
- [107] David M Eike, Joan F Brennecke, and Edward J Maginn. Predicting infinite-dilution activity coefficients of organic solutes in ionic liquids. *Industrial & engineering chemistry research*, 43(4):1039–1048, 2004.
- [108] Mara G Freire, Catarina MSS Neves, Sónia PM Ventura, Maria J Pratas, Isabel M Marrucho, João Oliveira, João AP Coutinho, and Ana M Fernandes. Solubility of non-aromatic ionic liquids in water and correlation using a qspr approach. *Fluid phase equilibria*, 294(1-2):234–240, 2010.
- [109] Katsumi Tochigi and Hiroshi Yamamoto. Estimation of ionic conductivity and viscosity of ionic liquids using a qspr model. *The Journal of Physical Chemistry C*, 111(43):15989–15994, 2007.
- [110] Patrik Johansson, Leif Erik Fast, Aleksandar Matic, Giovanni B Appetecchi, and Stefano Passerini. The conductivity of pyrrolidinium and sulfonylimide-based ionic liquids: a combined experimental and computational study. *Journal of Power Sources*, 195(7):2074–2076, 2010.
- [111] Philipp Eiden, Safak Bulut, Tobias Koíchner, Christian Friedrich, Thomas Schubert, and Ingo Krossing. In silico predictions of the temperature-dependent viscosities and electrical conductivities of functionalized and nonfunctionalized ionic liquids. *The Journal of Physical Chemistry B*, 115(2):300–309, 2011.
- [112] Ramesh L Gardas and Joao AP Coutinho. A group contribution method for viscosity estimation of ionic liquids. *Fluid Phase Equilibria*, 266(1-2):195–201, 2008.

- [113] Yuqiu Chen, Yingjun Cai, Kaj Thomsen, Georgios M Kontogeorgis, and John M Woodley. A group contribution-based prediction method for the electrical conductivity of ionic liquids. *Fluid Phase Equilibria*, page 112462, 2020.
- [114] Farhad Gharagheizi, Poorandokht Ilani-Kashkouli, Mehdi Sattari, Amir H Mohammadi, Deresh Ramjugernath, and Dominique Richon. Development of a lssvm-gc model for estimating the electrical conductivity of ionic liquids. *Chemical Engineering Research and Design*, 92(1):66–79, 2014.
- [115] Ali Zeinolabedini Hezave, Mostafa Lashkarbolooki, and Sona Raeissi. Using artificial neural network to predict the ternary electrical conductivity of ionic liquid systems. *Fluid Phase Equilibria*, 314:128–133, 2012.
- [116] Oscar Nordness, Pratik Kelkar, Yuanyuan Lyu, Michael Baldea, Mark A Stadtherr, and Joan F Brennecke. Predicting thermophysical properties of dialkylimidazolium ionic liquids from sigma profiles. *Journal of Molecular Liquids*, 334:116019, 2021.
- [117] Zi Kang Koi, Wan Zaireen Nisa Yahya, and Kiki Adi Kurnia. Prediction of ionic conductivity of imidazolium-based ionic liquids at different temperatures using multiple linear regression and support vector machine algorithms. *New Journal of Chemistry*, 45(39):18584–18597, 2021.
- [118] Wesley Beckner and Jim Pfaendtner. Fantastic liquids and where to find them: Optimizations of discrete chemical space. *Journal of chemical information and modeling*, 59(6):2617–2625, 2019.
- [119] Wesley Beckner, Chowdhury Ashraf, James Lee, David AC Beck, and Jim Pfaendtner. Continuous molecular representations of ionic liquids. *The Journal of Physical Chemistry B*, 124(38):8347–8357, 2020.
- [120] Sang Uck Lee, Jaehoon Jung, and Young-Kyu Han. Molecular dynamics study of the ionic conductivity of 1-n-butyl-3-methylimidazolium salts as ionic liquids. *Chemical physics letters*, 406(4-6):332–340, 2005.

- [121] Th Dhileep N Reddy and Bhabani S Mallik. Ionic dynamics of hydroxylammonium ionic liquids: A classical molecular dynamics simulation study. *The Journal of Physical Chemistry B*, 124(24):4960–4974, 2020.
- [122] Brian Doherty, Xiang Zhong, Symon Gathiaka, Bin Li, and Orlando Acevedo. Revisiting opls force field parameters for ionic liquid simulations. *Journal of chemical theory and computation*, 13(12):6131–6145, 2017.
- [123] Amir Taghavi Nasrabadi and Lev D Gelb. Structural and transport properties of tertiary ammonium triflate ionic liquids: A molecular dynamics study. *The Journal of Physical Chemistry B*, 121(8):1908–1921, 2017.
- [124] Veronika Zeindlhofer, Leopold Zehetner, Werner Paschinger, Alexander Bismarck, and Christian Schroeder. Computational analysis of conductivity contributions in an ionic liquid mixture of 1-ethyl-3-methylimidazolium dicyanamide and tetrafluoroborate. *Journal of Molecular Liquids*, 288:110993, 2019.
- [125] Zhimin Xue, Li Qin, Jingyun Jiang, Tiancheng Mu, and Guohua Gao. Thermal, electrochemical and radiolytic stabilities of ionic liquids. *Physical Chemistry Chemical Physics*, 20(13):8382–8402, 2018.
- [126] Cheng Lian, Honglai Liu, Chunzhong Li, and Jianzhong Wu. Hunting ionic liquids with large electrochemical potential windows. *AIChE Journal*, 65(2):804–810, 2019.
- [127] Niranjana V Ilawe, Jia Fu, Shriram Ramanathan, Bryan M Wong, and Jianzhong Wu. Chemical and radiation stability of ionic liquids: A computational screening study. *The Journal of Physical Chemistry C*, 120(49):27757–27767, 2016.
- [128] Shanthi Pandian, SG Raju, Krishnan S Hariharan, Subramanya M Kolake, Da-Hye Park, and Myung-Jin Lee. Functionalized ionic liquids as electrolytes for lithium-ion batteries. *Journal of Power Sources*, 286:204–209, 2015.
- [129] Andrew Ng. Lecture notes for machine learning, 2010.

- [130] Andrew Ng. Introduction to machine learning, 2011.
- [131] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.
- [132] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [133] Lecture 2: The svm classifier. <https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf>. Accessed: 2022-03-11.
- [134] Support vector machines (svms) lecture 2. <https://people.csail.mit.edu/dsontag/courses/ml14/slides/lecture2.pdf>. Accessed: 2022-03-11.
- [135] Support vector machines. <https://www.math.hkust.edu.hk/~makchen/MSBD5013/Ch9slide.pdf>. Accessed: 2022-03-11.
- [136] Understanding the mathematics behind support vector machines. <https://shuzhanfan.github.io/2018/05/understanding-mathematics-behind-support-vector-machines/>. Accessed: 2022-03-11.
- [137] Alex Smola and SVN Vishwanathan. Introduction to machine learning. *Cambridge University, UK*, 32(34):2008, 2008.
- [138] Andrew Ng. Cs 229 lecture notes: Support vector machines. *University Lecture notes*, 2012.
- [139] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier, 1992.
- [140] Stephen Odaibo. Tutorial: Deriving the standard variational autoencoder (vae) loss function. *arXiv preprint arXiv:1907.08956*, 2019.
- [141] John R Hershey and Peder A Olsen. Approximating the kullback leibler divergence between gaussian mixture models. In *2007 IEEE International Con-*

ference on Acoustics, Speech and Signal Processing-ICASSP'07, volume 4, pages IV–317. IEEE, 2007.

- [142] <https://pypi.org/project/pyilt2/>. Accessed: 2020-09-03.
- [143] <http://www.rdkit.org/>. Accessed: 2020-03-28.
- [144] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [145] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.
- [146] Michael P Allen et al. Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins*, 23(1):1–28, 2004.
- [147] Edward J Maginn. Molecular theory and modeling chemical engineering, 698d. *Notas de aula. Disponível em:* < http://www.nd.edu/ed/Teaching/molecular_theory_notes.pdf >. Acesso em, 13, 2007.
- [148] Martin Brehm and Barbara Kirchner. Travis-a free analyzer and visualizer for monte carlo and molecular dynamics trajectories, 2011.
- [149] Annalisa Paolone and Sergio Brutti. Comparison of the performances of different computational methods to calculate the electrochemical stability of selected ionic liquids. *Materials*, 14(12):3221, 2021.
- [150] Marco Carboni, Riccardo Spezia, and Sergio Brutti. Perfluoroalkyl-fluorophosphate anions for high voltage electrolytes in lithium cells: Dft study. *The Journal of Physical Chemistry C*, 118(42):24221–24230, 2014.
- [151] Patrik Johansson. Intrinsic anion oxidation potentials. *The Journal of Physical Chemistry A*, 110(44):12077–12080, 2006.

- [152] Chang-Guo Zhan, Jeffrey A Nichols, and David A Dixon. Ionization potential, electron affinity, electronegativity, hardness, and electron excitation energy: molecular properties from density functional theory orbital energies. *The Journal of Physical Chemistry A*, 107(20):4184–4195, 2003.
- [153] MJE Frisch, GW Trucks, H Bernhard Schlegel, Gustavo E Scuseria, Michael A Robb, James R Cheeseman, Giovanni Scalmani, Vincenzo Barone, Benedetta Mennucci, GAe Petersson, et al. gaussian 09, revision d. 01, gaussian. *Inc., Wallingford CT*, 201, 2009.
- [154] Tobias Binniger, Aris Marcolongo, Matthieu Mottet, Valery Weber, and Teodoro Laino. Comparison of computational methods for the electrochemical stability window of solid-state electrolyte materials. *Journal of Materials Chemistry A*, 8(3):1347–1359, 2020.
- [155] Peter Wasserscheid and Thomas Welton. *Ionic liquids in synthesis*. John Wiley & Sons, 2008.
- [156] AB Pereiro, JMM Araújo, JMSS Esperança, IM Marrucho, and LPN Rebelo. Ionic liquids in separations of azeotropic systems—a review. *The Journal of Chemical Thermodynamics*, 46:2–28, 2012.
- [157] Andrew S Paluch and Pratik Dhakal. Thermodynamic assessment of the suitability of the limiting selectivity to screen ionic liquid entrainers for homogeneous extractive distillation processes. *ChemEngineering*, 2(4):54, 2018.
- [158] Pratik Dhakal, Jonathan A Ouimet, Sydnee N Roese, and Andrew S Paluch. Mosced parameters for 1-n-alkyl-3-methylimidazolium-based ionic liquids: Application to limiting activity coefficients and intuitive entrainer selection for extractive distillation processes. *Journal of Molecular Liquids*, 293:111552, 2019.
- [159] Pratik Dhakal, Anthony R Weise, Martin C Fritsch, Cassandra M O’Dell, and Andrew S Paluch. Expanding the solubility parameter method

- mosced to pyridinium-, quinolinium-, pyrrolidinium-, piperidinium-, bicyclic-, morpholinium-, ammonium-, phosphonium-, and sulfonium-based ionic liquids. *ACS omega*, 5(8):3863–3877, 2020.
- [160] Julián García, Silvia García, José S Torrecilla, and Francisco Rodríguez. N-butylpyridinium bis-(trifluoromethylsulfonyl) imide ionic liquids as solvents for the liquid–liquid extraction of aromatics from their mixtures with alkanes: Isomeric effect of the cation. *Fluid phase equilibria*, 301(1):62–66, 2011.
- [161] Antje R Hansmeier, Mark Jongmans, G Wytze Meindersma, and Andre B de Haan. VLE data for the ionic liquid 3-methyl-n-butyl pyridinium dicyanamide with several aromatic and aliphatic hydrocarbons. *The Journal of Chemical Thermodynamics*, 42(4):484–490, 2010.
- [162] Irene Domínguez, Emilio J Gonzalez, and Ángeles Domínguez. Liquid extraction of aromatic/cyclic aliphatic hydrocarbon mixtures using ionic liquids as solvent: literature review and new experimental VLE data. *Fuel processing technology*, 125:207–216, 2014.
- [163] Andreas Heintz, Jochen K Lehmann, Christiane Wertz, and Johan Jacquemin. Thermodynamic properties of mixtures containing ionic liquids. 4. VLE of binary mixtures of [c₂mim][ntf₂] with propan-1-ol, butan-1-ol, and pentan-1-ol and [c₄mim][ntf₂] with cyclohexanol and 1, 2-hexanediol including studies of the influence of small amounts of water. *Journal of Chemical & Engineering Data*, 50(3):956–960, 2005.
- [164] Douglas R MacFarlane, Maria Forsyth, Patrick C Howlett, Jennifer M Pringle, Jiazeng Sun, Gary Annat, Wayne Neil, and Ekaterina I Izgorodina. Ionic liquids in electrochemical devices and processes: managing interfacial electrochemistry. *Accounts of chemical research*, 40(11):1165–1173, 2007.
- [165] Andrzej Lewandowski and Agnieszka Świdarska-Mocek. Ionic liquids as electrolytes for li-ion batteries—an overview of electrochemical studies. *Journal of Power sources*, 194(2):601–609, 2009.

- [166] Alpha A Lee, Dominic Vella, Susan Perkin, and Alain Goriely. Are room-temperature ionic liquids dilute electrolytes? *The journal of physical chemistry letters*, 6(1):159–163, 2015.
- [167] Xiangping Zhang, Xiaochun Zhang, Haifeng Dong, Zhijun Zhao, Suojiang Zhang, and Ying Huang. Carbon capture with ionic liquids: overview and progress. *Energy & Environmental Science*, 5(5):6668–6681, 2012.
- [168] Alexia Finotello, Jason E Bara, Dean Camper, and Richard D Noble. Room-temperature ionic liquids: temperature dependence of gas solubility selectivity. *Industrial & Engineering Chemistry Research*, 47(10):3453–3459, 2008.
- [169] Cesar Cadena, Jennifer L Anthony, Jindal K Shah, Timothy I Morrow, Joan F Brennecke, and Edward J Maginn. Why is CO_2 so soluble in imidazolium-based ionic liquids? *Journal of the American Chemical Society*, 126(16):5300–5308, 2004.
- [170] Johan Jacquemin, Pascale Husson, Vladimir Majer, and Margarida F Costa Gomes. Influence of the cation on the solubility of CO_2 and H_2 in ionic liquids based on the bis (trifluoromethylsulfonyl) imide anion. *Journal of solution chemistry*, 36(8):967–979, 2007.
- [171] Hikari Sakaebe and Hajime Matsumoto. N-methyl-n-propylpiperidinium bis (trifluoromethanesulfonyl) imide ($\text{pp}_{13}\text{-tfsi}$)—novel electrolyte base for Li battery. *Electrochemistry Communications*, 5(7):594–598, 2003.
- [172] Silvia García, Marcos Larriba, Julián García, José S Torrecilla, and Francisco Rodríguez. Liquid–liquid extraction of toluene from n-heptane using binary mixtures of n-butylpyridinium tetrafluoroborate and n-butylpyridinium bis (trifluoromethylsulfonyl) imide ionic liquids. *Chemical engineering journal*, 180:210–215, 2012.
- [173] Pratik Dhakal and Jindal K Shah. Recent advances in molecular simulations of ionic liquid–ionic liquid mixtures. *Current Opinion in Green and Sustainable Chemistry*, 18:90–97, 2019.

- [174] Natalia V Plechkova and Kenneth R Seddon. Applications of ionic liquids in the chemical industry. *Chemical Society Reviews*, 37(1):123–150, 2008.
- [175] Kamil Paduszyński. Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 1. Density. *Industrial & Engineering Chemistry Research*, 58(13):5322–5338, March 2019.
- [176] Wesley Beckner, Coco M Mao, and Jim Pfaendtner. Statistical models are able to predict ionic liquid viscosity across a wide range of chemical functionalities and experimental conditions. *Molecular Systems Design & Engineering*, 3(1):253–263, 2018.
- [177] Kamil Paduszyński. Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity. *Industrial & Engineering Chemistry Research*, 58(36):17049–17066, August 2019.
- [178] Vishwesh Venkatraman, Sigvart Evjen, Hanna K Knuutila, Anne Fiksdahl, and Bjørn Kåre Alsberg. Predicting ionic liquid melting points using machine learning. *Journal of Molecular Liquids*, 264:318–326, 2018.
- [179] Lingdi Cao, Peng Zhu, Yongsheng Zhao, and Jihong Zhao. Using machine learning and quantum chemistry descriptors to predict the toxicity of ionic liquids. *Journal of hazardous materials*, 352:17–26, 2018.
- [180] Vishwesh Venkatraman and Bjørn Kåre Alsberg. Predicting CO_2 capture of ionic liquids using machine learning. *Journal of CO_2 Utilization*, 21:162–168, 2017.
- [181] Mohammad Mesbah, Shohreh Shahsavari, Ebrahim Soroush, Neda Rahaei, and Mashallah Rezakazemi. Accurate prediction of miscibility of CO_2 and supercritical CO_2 in ionic liquids using machine learning. *Journal of CO_2 Utilization*, 25:99–107, 2018.
- [182] Hadi Mokarizadeh, Saeid Atashrouz, Hamed Mirshekar, Abdolhossein Hemmati-Sarapardeh, and Ahmad Mohaddes Pour. Comparison of LSSVM

- model results with artificial neural network model for determination of the solubility of SO₂ in ionic liquids. *Journal of Molecular Liquids*, 304(C):112771, April 2020.
- [183] Kamil Padiuszyński and 2021. Extensive Databases and Group Contribution QSPRs of Ionic Liquid Properties. 3: Surface Tension. *Ind. Eng. Chem. Res.*
- [184] Dilek Yalcin, Tu C Le, Calum J Drummond, and Tamar L Greaves. Machine Learning Approaches for Further Developing the Understanding of the Property Trends Observed in Protic Ionic Liquid Containing Solvents. *The Journal of Physical Chemistry B*, 123(18):4085–4097, May 2019.
- [185] Mohammad Hashemkhani, Reza Soleimani, Hossein Fazeli, Moonyong Lee, Alireza Bahadori, and Mahsa Tavalaeian. Prediction of the binary surface tension of mixtures containing ionic liquids using Support Vector Machine algorithms. *Journal of Molecular Liquids*, 211(C):534–552, November 2015.
- [186] Mohsen Hosseinzadeh and Abdolhossein Hemmati-Sarapardeh. Toward a predictive model for estimating viscosity of ternary mixtures containing ionic liquids. *Journal of Molecular Liquids*, 200(PB):340–348, December 2014.
- [187] José S Torrecilla, Maggel Deetlefs, Kenneth R Seddon, and Francisco Rodríguez. Estimation of ternary liquid–liquid equilibria for arene/alkane/ionic liquid mixtures using neural networks. *Physical Chemistry Chemical Physics*, 10(33):5114–7, 2008.
- [188] Tamar L Greaves, Karin S Schaffarczyk McHale, Raphael F Burkart-Radke, Jason B Harper, and Tu C Le. Machine learning approaches to understand and predict rate constants for organic processes in mixtures containing ionic liquids. *Physical Chemistry Chemical Physics*, 23(4):2742–2752, February 2021.
- [189] Falola Yusuf, Teslim Olayiwola, and Clement Afagwu. Application of Artificial Intelligence-based predictive methods in Ionic liquid studies: A review. *Fluid Phase Equilib.*, 531:112898, March 2021.

- [190] Wesley Beckner and Jim Pfaendtner. Fantastic Liquids and Where To Find Them: Optimizations of Discrete Chemical Space. *Journal of Chemical Information and Modeling*, 59(6):2617–2625, April 2019.
- [191] Witali Beichel, Yang Yu, Günter Dlubek, Reinhard Krause-Rehberg, Jürgen Pionteck, Dirk Pfefferkorn, Safak Bulut, Dana Bejan, Christian Friedrich, and Ingo Krossing. Free volume in ionic liquids: a connection of experimentally accessible observables from pals and pvt experiments with the molecular structure from xrd data. *Physical Chemistry Chemical Physics*, 15(22):8821–8830, 2013.
- [192] Witali Beichel, Ulrich P Preiss, Sergey P Verevkin, Thorsten Koslowski, and Ingo Krossing. Empirical description and prediction of ionic liquids’ properties with augmented volume-based thermodynamics. *Journal of Molecular Liquids*, 192:3–8, 2014.
- [193] Ramesh L Gardas and Joao AP Coutinho. Group contribution methods for the prediction of thermophysical and transport properties of ionic liquids. *AIChE Journal*, 55(5):1274–1290, 2009.
- [194] Mitsuhiro Kanakubo, Kenneth R Harris, Noriaki Tsuchihashi, Kazuyasu Ibuki, and Masakatsu Ueno. Temperature and pressure dependence of the electrical conductivity of 1-butyl-3-methylimidazolium bis (trifluoromethanesulfonyl) amide. *Journal of Chemical & Engineering Data*, 60(5):1495–1503, 2015.
- [195] J Vila, B Fernández-Castro, E Rilo, J Carrete, M Domínguez-Pérez, JR Rodríguez, M García, LM Varela, and O Cabeza. Liquid–solid–liquid phase transition hysteresis loops in the ionic conductivity of ten imidazolium-based ionic liquids. *Fluid Phase Equilibria*, 320:1–10, 2012.
- [196] Sanja Dožić, Nebojša Zec, Aleksandar Tot, Snežana Papović, Ksenija Pavlović, Slobodan Gadžurić, and Milan Vraneš. Does the variation of the alkyl chain length on n1 and n3 of imidazole ring affect physicochemical

- features of ionic liquids in the same way? *The Journal of Chemical Thermodynamics*, 93:52–59, 2016.
- [197] Kenneth R Harris, Mitsuhiro Kanakubo, Noriaki Tsuchihashi, Kazuyasu Ibuki, and Masakatsu Ueno. Effect of pressure on the transport properties of ionic liquids: 1-alkyl-3-methylimidazolium salts. *The Journal of Physical Chemistry B*, 112(32):9830–9840, 2008.
- [198] BE Mbondo Tsamba, S Sarraute, Mounir Traïkia, and P Husson. Transport properties and ionic association in pure imidazolium-based ionic liquids as a function of temperature. *Journal of Chemical & Engineering Data*, 59(6):1747–1754, 2014.
- [199] Takashi Makino, Mitsuhiro Kanakubo, Yoshio Masuda, and Hiroshi Mukaiyama. Physical and CO₂-absorption properties of imidazolium ionic liquids with tetracyanoborate and bis (trifluoromethanesulfonyl) amide anions. *Journal of Solution Chemistry*, 43(9-10):1601–1613, 2014.
- [200] Ya-Hung Yu, Allan N Soriano, and Meng-Hui Li. Heat capacities and electrical conductivities of 1-n-butyl-3-methylimidazolium-based ionic liquids. *Thermochimica acta*, 482(1-2):42–48, 2009.
- [201] Ji-Guang Li, Yu-Feng Hu, Chuan-Wei Jin, Hong-Da Chu, Xiao-Ming Peng, and Ying-Guo Liang. Study on the conductivities of pure and aqueous bromide-based ionic liquids at different temperatures. *Journal of solution chemistry*, 39(12):1877–1887, 2010.
- [202] <https://opsin.ch.cam.ac.uk/>. Accessed: 2020-03-28.
- [203] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.
- [204] Mohammad Hashemkhani, Reza Soleimani, Hossein Fazeli, Moonyong Lee, Alireza Bahadori, and Mahsa Tavalaeian. Prediction of the binary surface tension of mixtures containing ionic liquids using support vector machine algorithms. *Journal of Molecular Liquids*, 211:534–552, 2015.

- [205] CY Zhao, HX Zhang, XY Zhang, MC Liu, ZD Hu, and BT Fan. Application of support vector machine (svm) for prediction toxic activity of different data sets. *Toxicology*, 217(2-3):105–119, 2006.
- [206] Kazuhiko Matsumoto, Erisa Nishiwaki, Takafumi Hosokawa, Shinya Tawa, Toshiyuki Nohira, and Rika Hagiwara. Thermal, physical, and electrochemical properties of $[\text{n}(\text{so}_2\text{f})_2]\text{-}[1\text{-ethyl-3-methylimidazolium}][\text{n}(\text{so}-2\text{f})_2]$ ionic liquid electrolytes for li secondary batteries operated at room and intermediate temperatures. *The Journal of Physical Chemistry C*, 121(17):9209–9219, 2017.
- [207] Gunzi Saito and Yukihiro Yoshida. Development of conductive organic molecular assemblies: organic metals, superconductors, and exotic functional materials. *Bulletin of the Chemical Society of Japan*, 80(1):1–137, 2007.
- [208] W Martino, J Fernandez De La Mora, Y Yoshida, G Saito, and J Wilkes. Surface tension measurements of highly conducting ionic liquids. *Green Chemistry*, 8(4):390–397, 2006.
- [209] Yuriy O Andriyko, Wolfgang Reischl, and Gerhard E Nauer. Trialkyl-substituted imidazolium-based ionic liquids for electrochemical applications: basic physicochemical properties. *Journal of Chemical & Engineering Data*, 54(3):855–860, 2009.
- [210] Alexander Stoppa, Richard Buchner, and Glenn Hefter. How ideal are binary mixtures of room-temperature ionic liquids? *Journal of Molecular Liquids*, 153(1):46–51, 2010.
- [211] Pedro DA Bastos, Filipe S Oliveira, Luís PN Rebelo, Ana B Pereiro, and Isabel M Marrucho. Separation of azeotropic mixtures using high ionicity ionic liquids based on 1-ethyl-3-methylimidazolium thiocyanate. *Fluid Phase Equilibria*, 389:48–54, 2015.
- [212] Mitsuhiro Kanakubo, Takashi Makino, Takayuki Taniguchi, Toshiki Nokami, and Toshiyuki Itoh. CO_2 solubility in ether functionalized ionic liquids on

- mole fraction and molarity scales. *ACS Sustainable Chemistry & Engineering*, 4(2):525–535, 2016.
- [213] Hui Ning, MinQiang Hou, QingQing Mei, YuanHui Liu, DeZhong Yang, and BuXing Han. The physicochemical properties of some imidazolium-based ionic liquids and their binary mixtures. *Science China Chemistry*, 55(8):1509–1518, 2012.
- [214] Matthew T Clough, Colin R Crick, John Gräsvik, Patricia A Hunt, Heiko Niedermeyer, Tom Welton, and Oliver P Whitaker. A physicochemical investigation of ionic liquid mixtures. *Chemical science*, 6(2):1101–1114, 2015.
- [215] Marc Bruessel, Martin Brehm, Alfonso S Pensado, Friedrich Malberg, Muhammad Ramzan, Annegret Stark, and Barbara Kirchner. On the ideality of binary mixtures of ionic liquids. *Physical Chemistry Chemical Physics*, 14(38):13204–13215, 2012.
- [216] Dong Xiao, Justin Rajesh Rajian, Larry G Hines Jr, Shengfu Li, Richard A Bartsch, and Edward L Quitevis. Nanostructural organization and anion effects in the optical kerr effect spectra of binary ionic liquid mixtures. *The Journal of Physical Chemistry B*, 112(42):13316–13325, 2008.
- [217] Iuliia V Voroshylova, Elisabete SC Ferreira, Michal Malčák, Renata Costa, Carlos M Pereira, and M Natália DS Cordeiro. Influence of the anion on the properties of ionic liquid mixtures: a molecular dynamics study. *Physical Chemistry Chemical Physics*, 20(21):14899–14918, 2018.
- [218] Xueping Wang, Yiping Huang, Li Li, Liangliang Huang, Xiangshu Chen, and Zhen Yang. Molecular-level insights into composition-dependent structure, dynamics, and hydrogen bonds of binary ionic liquid mixture from molecular dynamics simulations. *Chemical Physics*, 542:111051, 2021.
- [219] Sascha Gehrke, Michael von Domaros, Ryan Clark, Oldamur Hollóczki, Martin Brehm, Tom Welton, Alenka Luzar, and Barbara Kirchner. Structure and

- lifetimes in ionic liquids and their mixtures. *Faraday Discussions*, 206:219–245, 2018.
- [220] Paloma Navia, Jacobo Troncoso, and Luis Romani. Excess magnitudes for ionic liquid binary mixtures with a common ion. *Journal of Chemical & Engineering Data*, 52(4):1369–1374, 2007.
- [221] Paloma Navia, Jacobo Troncoso, and Luis Romani. Viscosities for ionic liquid binary mixtures with a common ion. *Journal of solution chemistry*, 37(5):677–688, 2008.
- [222] Eric T Fox, Joshua EF Weaver, and Wesley A Henderson. Tuning binary ionic liquid mixtures: linking alkyl chain length to phase behavior and ionic conductivity. *The Journal of Physical Chemistry C*, 116(8):5270–5274, 2012.
- [223] Karina Shimizu, M Tariq, Luís PN Rebelo, and José N Canongia Lopes. Binary mixtures of ionic liquids with a common ion revisited: a molecular dynamics simulation study. *Journal of Molecular Liquids*, 153(1):52–56, 2010.
- [224] Cameron C Weber, Nicholas J Brooks, Franca Castiglione, Michele Mauri, Roberto Simonutti, Andrea Mele, and Tom Welton. On the structural origin of free volume in 1-alkyl-3-methylimidazolium ionic liquid mixtures: A saxs and ^{129}Xe nmr study. *Physical Chemistry Chemical Physics*, 21(11):5999–6010, 2019.
- [225] Utkarsh Kapoor and Jindal K Shah. Globular, sponge-like to layer-like morphological transition in 1-n-alkyl-3-methylimidazolium octylsulfate ionic liquid homologous series. *The Journal of Physical Chemistry B*, 122(1):213–228, 2018.
- [226] Duncan W Bruce, Christopher P Cabry, José N Canongia Lopes, Matthew L Costen, Luciea D’Andrea, Isabelle Grillo, Brooks C Marshall, Kenneth G McKendrick, Timothy K Minton, Simon M Purcell, et al. Nanosegregation

- and structuring in the bulk and at the surface of ionic-liquid mixtures. *The Journal of Physical Chemistry B*, 121(24):6002–6020, 2017.
- [227] Tyler Cosby, Utkarsh Kapoor, Jindal K Shah, and Joshua Sangoro. Mesoscale organization and dynamics in binary ionic liquid mixtures. *The journal of physical chemistry letters*, 10(20):6274–6280, 2019.
- [228] Xueping Wang, Fangjia Fu, Kuilin Peng, Zhen Yang, Na Hu, Xiangshu Chen, and Wei Li. Understanding of structures, dynamics, and hydrogen bonds of imidazolium-based ionic liquid mixture from molecular dynamics simulation. *Chemical Physics*, 525:110391, 2019.
- [229] Meysam Mirarabrazi, Olga Stolarska, Marcin Smiglak, and Christian Robelin. Solid–liquid equilibria for a pyrrolidinium-based common-cation ternary ionic liquid system, and for a pyridinium-based ternary reciprocal ionic liquid system: an experimental study and a thermodynamic model. *Physical Chemistry Chemical Physics*, 20(1):637–657, 2018.
- [230] Maria Taige, Diana Hilbert, and Thomas JS Schubert. Mixtures of ionic liquids as possible electrolytes for lithium ion batteries. *Zeitschrift für physikalische Chemie*, 226(2):129–139, 2012.
- [231] Pankaj Bharmoria, Krishnaiah Damarla, Tushar J Trivedi, Naved I Malek, and Arvind Kumar. A reciprocal binary mixture of protic/aprotic ionic liquids as a deep eutectic solvent: physicochemical behaviour and application towards agarose processing. *RSC advances*, 5(120):99245–99252, 2015.
- [232] Praveenkumar Sappidi, Jason E Bara, and C Heath Turner. Molecular-level behavior of imidazolium-based ionic liquid mixtures. *Chemical Engineering Science*, 229:116073, 2021.
- [233] José N Canongia Lopes, Johnny Deschamps, and Agílio AH Pádua. Modeling ionic liquids using a systematic all-atom force field. *The journal of physical chemistry B*, 108(6):2038–2047, 2004.

- [234] José N Canongia Lopes and Agílio AH Pádua. Molecular force field for ionic liquids iii: Imidazolium, pyridinium, and phosphonium cations; chloride, bromide, and dicyanamide anions. *The Journal of Physical Chemistry B*, 110(39):19586–19592, 2006.
- [235] José N Canongia Lopes and Agílio AH Pádua. Molecular force field for ionic liquids composed of triflate or bistriflylimide anions. *The Journal of Physical Chemistry B*, 108(43):16893–16898, 2004.
- [236] Kateryna Goloviznina, José N Canongia Lopes, Margarida Costa Gomes, and Agílio AH Pádua. Transferable, polarizable force field for ionic liquids. *Journal of chemical theory and computation*, 15(11):5858–5871, 2019.
- [237] Sascha Gehrke, Michael von Domaros, Ryan Clark, Oldamur Hollóczy, Martin Brehm, Tom Welton, Alenka Luzar, and Barbara Kirchner. Structure and lifetimes in ionic liquids and their mixtures. *Faraday discussions*, 206:219–245, 2017.
- [238] H Bekker, HJC Berendsen, EJ Dijkstra, S Achterop, R Van Drunen, D Van der Spoel, A Sijbers, H Keegstra, B Reitsma, and MKR Renardus. Gromacs: A parallel computer for molecular dynamics simulations. 92:252–256, 1993.
- [239] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and computation*, 4(3):435–447, 2008.
- [240] Leandro Martínez, Ricardo Andrade, Ernesto G Birgin, and José Mario Martínez. Packmol: a package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry*, 30(13):2157–2164, 2009.
- [241] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52(2):255–268, 1984.

- [242] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.
- [243] Richard P Matthews, Ignacio J Villar-Garcia, Cameron C Weber, Jeraine Griffith, Fiona Cameron, Jason P Hallett, Patricia A Hunt, and Tom Welton. A structural investigation of ionic liquid mixtures. *Physical Chemistry Chemical Physics*, 18(12):8608 – 8624, 2016.
- [244] M Brehm, M Thomas, S Gehrke, and B Kirchner. Travis—a free analyzer for trajectories from molecular simulation. *The Journal of chemical physics*, 152(16):164105, 2020.
- [245] Sascha Gehrke and Barbara Kirchner. Robustness of the hydrogen bond and ion pair dynamics in ionic liquids to different parameters from the reactive flux method. *Journal of Chemical & Engineering Data*, 65(3):1146–1158, 2019.
- [246] Th Dhileep N Reddy and Bhabani S Mallik. Protic ammonium carboxylate ionic liquids: insight into structure, dynamics and thermophysical properties by alkyl group functionalization. *Physical Chemistry Chemical Physics*, 19(16):10358–10370, 2017.
- [247] Mara G Freire, Ana Rita R Teles, Marisa AA Rocha, Bernd Schroöder, Catarina MSS Neves, Pedro J Carvalho, Dmitry V Evtuguin, Luís MNBF Santos, and Joao AP Coutinho. Thermophysical characterization of ionic liquids able to dissolve biomass. *Journal of Chemical & Engineering Data*, 56(12):4813–4822, 2011.
- [248] Raquel G Seoane, Sandra Corderí, Elena Goómez, Noelia Calvar, Emilio J González, Eugenia A Macedo, and Ángeles Domínguez. Temperature dependence and structural influence on the thermophysical properties of eleven commercial ionic liquids. *Industrial & engineering chemistry research*, 51(5):2492–2504, 2012.

- [249] Azita Aghosseini and Aaron M Scurto. Viscosity of imidazolium-based ionic liquids at elevated pressures: cation and anion effects. *International Journal of Thermophysics*, 29(4):1222–1243, 2008.
- [250] Jason A Widegren and Joseph W Magee. Density, viscosity, speed of sound, and electrolytic conductivity for the ionic liquid 1-hexyl-3-methylimidazolium bis (trifluoromethylsulfonyl) imide and its mixtures with water. *Journal of Chemical & Engineering Data*, 52(6):2331–2338, 2007.
- [251] Y Zhang and E J Maginn. Direct Correlation between Ionic Liquid Transport Properties and Ion Pair Lifetimes: A Molecular Dynamics Study. *J. Phys. Chem. Lett.*, 6:700 – 705, 2015.
- [252] Hiroyuki Tokuda, Kikuko Hayamizu, Kunikazu Ishii, Md Abu Bin Hasan Susan, and Masayoshi Watanabe. Physicochemical properties and structures of room temperature ionic liquids. 2. variation of alkyl chain length in imidazolium cation. *The Journal of Physical Chemistry B*, 109(13):6103–6110, 2005.
- [253] Alain Berthod, MJ Ruiz-Angel, and Samuel Carda-Broch. Ionic liquids in separation techniques. *Journal of Chromatography A*, 1184(1-2):6–18, 2008.
- [254] Jiajian Peng and Youquan Deng. Cycloaddition of carbon dioxide to propylene oxide catalyzed by ionic liquids. *New Journal of Chemistry*, 25(4):639–641, 2001.
- [255] Dongbin Zhao, Min Wu, Yuan Kou, and Enze Min. Ionic liquids: applications in catalysis. *Catalysis today*, 74(1-2):157–189, 2002.
- [256] Eleanor D Bates, Rebecca D Mayton, Ioanna Ntai, and James H Davis. CO₂ capture by a task-specific ionic liquid. *Journal of the American Chemical Society*, 124(6):926–927, 2002.
- [257] José D Figueroa, Timothy Fout, Sean Plasynski, Howard McIlvried, and Rameshwar D Srivastava. Advances in CO₂ capture technology—the us de-

- partment of energy's carbon sequestration program. *International journal of greenhouse gas control*, 2(1):9–20, 2008.
- [258] Chenqian Xu and Zhenmin Cheng. Thermal stability of ionic liquids: Current status and prospects for future development. *Processes*, 9(2):337, 2021.
- [259] Deepak Chand, Magdalena Wilk-Kozubek, Volodymyr Smetana, and Anja-Verena Mudring. Alternative to the popular imidazolium ionic liquids: 1, 2, 4-triazolium ionic liquids with enhanced thermal and chemical stability. *ACS sustainable chemistry & engineering*, 7(19):15995–16006, 2019.
- [260] Liangbing Hu and Kang Xu. Nonflammable electrolyte enhances battery safety. *Proceedings of the National Academy of Sciences*, 111(9):3205–3206, 2014.
- [261] Taeun Yim, Mi-Sook Kwon, Junyoung Mun, and Kyu Tae Lee. Room temperature ionic liquid-based electrolytes as an alternative to carbonate-based electrolytes. *Israel Journal of Chemistry*, 55(5):586–598, 2015.
- [262] Catia Arbizzani, Giulio Gabrielli, and Marina Mastragostino. Thermal stability and flammability of electrolytes for lithium-ion batteries. *Journal of Power Sources*, 196(10):4801–4805, 2011.
- [263] AM Andersson and Kristina Edström. Chemical composition and morphology of the elevated temperature sei on graphite. *Journal of the Electrochemical Society*, 148(10):A1100, 2001.
- [264] Steven E Sloop, James K Pugh, S Wang, JB Kerr, and K Kinoshita. Chemical reactivity of pf_5 and lipf_6 in ethylene carbonate/dimethyl carbonate solutions. *Electrochemical and Solid State Letters*, 4(4):A42, 2001.
- [265] Zhen Song, Huaiwei Shi, Xiang Zhang, and Teng Zhou. Prediction of co_2 solubility in ionic liquids using machine learning methods. *Chemical Engineering Science*, 223:115752, 2020.

- [266] Ali Shafiei, Mohammad Ali Ahmadi, Seyed Hayan Zaheri, Alireza Baghban, Ali Amirfakhrian, and Reza Soleimani. Estimating hydrogen sulfide solubility in ionic liquids using a machine learning approach. *The Journal of Supercritical Fluids*, 95:525–534, 2014.
- [267] Pratik Dhakal and Jindal K Shah. Developing machine learning models for ionic conductivity of imidazolium-based ionic liquids. *Fluid Phase Equilibria*, 549:113208, 2021.
- [268] Swantje Lerch and Thomas Strassner. Expanding the electrochemical window: New tunable aryl alkyl ionic liquids (taails) with dicyanamide anions. *Chemistry—A European Journal*, 25(71):16251–16256, 2019.
- [269] Maria Assunta Navarra, Kanae Fujimura, Mirko Sgambetterra, Akiko Tsurumaki, Stefania Panero, Nobuhumi Nakamura, Hiroyuki Ohno, and Bruno Scrosati. New ether-functionalized morpholinium-and piperidinium-based ionic liquids as electrolyte components in lithium and lithium-ion batteries. *ChemSusChem*, 10(11):2496–2504, 2017.
- [270] Shumin Shen, Shaohua Fang, Long Qu, Dong Luo, Li Yang, and Shin-ichi Hirano. Low-viscosity ether-functionalized pyrazolium ionic liquids based on dicyanamide anions: properties and application as electrolytes for lithium metal batteries. *RSC advances*, 5(114):93888–93899, 2015.
- [271] Ming Chai, Yide Jin, Shaohua Fang, Li Yang, Shin-ichi Hirano, and Kazuhiro Tachibana. Ether-functionalized pyrazolium ionic liquids as new electrolytes for lithium battery. *Electrochimica acta*, 66:67–74, 2012.
- [272] Colin SM Kang, Ruhamah Yunis, Haijin Zhu, Cara M Doherty, Oliver E Hutt, and Jennifer M Pringle. Ionic liquids and plastic crystals utilising the oxazolidinium cation: the effect of ether functionality in the ring. *Materials Chemistry Frontiers*, 5(16):6014–6026, 2021.

- [273] Rebecca R Hawker, Ronald S Haines, and Jason B Harper. Variation of the cation of ionic liquids: the effects on their physicochemical properties and reaction outcome. *Targets Heterocycl. Syst. Prop.*, 18:141–213, 2015.
- [274] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.
- [275] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
- [276] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4, 2015.
- [277] Takeshi Kakibe, Nobuko Yoshimoto, Minato Egashira, and Masayuki Morita. Optimization of cation structure of imidazolium-based ionic liquids as ionic solvents for rechargeable magnesium batteries. *Electrochemistry communications*, 12(11):1630–1633, 2010.
- [278] Hyung-Tae Kim, Oh Min Kwon, Junyoung Mun, Seung M Oh, Taeun Yim, and Young Gyu Kim. Novel pyrrolinium-based ionic liquids for lithium ion batteries: effect of the cation on physicochemical and electrochemical properties. *Electrochimica Acta*, 240:267–276, 2017.
- [279] S Pohlmann, T Olyschläger, P Goodrich, J Alvarez Vicente, J Jacquemin, and A Balducci. Azepanium-based ionic liquids as green electrolytes for high voltage supercapacitors. *Journal of Power Sources*, 273:931–936, 2015.
- [280] Mohammad H Fatemi and Parisa Izadiyan. Cytotoxicity estimation of ionic liquids based on their effective structural features. *Chemosphere*, 84(5):553–563, 2011.
- [281] Dilek Yalcin, Tu C Le, Calum J Drummond, and Tamar L Greaves. Machine learning approaches for further developing the understanding of the property

- trends observed in protic ionic liquid containing solvents. *The Journal of Physical Chemistry B*, 123(18):4085–4097, 2019.
- [282] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.
- [283] Ya Sun, MinChun Chen, Yongmei Zhao, Zhenfeng Zhu, Han Xing, Peng Zhang, Xiaojian Zhang, and Yi Ding. Machine learning assisted qspr model for prediction of ionic liquid’s refractive index and viscosity: The effect of representations of ionic liquid and ensemble model development. *Journal of Molecular Liquids*, 333:115970, 2021.
- [284] Yi Ding, Minchun Chen, Chao Guo, Peng Zhang, and Jingwen Wang. Molecular fingerprint-based machine learning assisted qsar model development for prediction of ionic liquid properties. *Journal of Molecular Liquids*, 326:115212, 2021.
- [285] D Bonchev and N Trinajstić. Information theory, distance matrix, and molecular branching. *The Journal of Chemical Physics*, 67(10):4517–4533, 1977.
- [286] Lawien F Zubeir, Marisa AA Rocha, Niki Vergadou, Wilko MA Weggemans, Loukas D Peristeras, Peter S Schulz, Ioannis G Economou, and Maaïke C Kroon. Thermophysical properties of imidazolium tricyanomethanide ionic liquids: experiments and molecular simulation. *Physical Chemistry Chemical Physics*, 18(33):23121–23138, 2016.
- [287] David Rooney, Johan Jacquemin, and Ramesh Gardas. Thermophysical properties of ionic liquids. *Ionic liquids*, pages 185–212, 2009.
- [288] Lowell H Hall and Lemont B Kier. The molecular connectivity chi indexes and kappa shape indexes in structure-property modeling. *Reviews in computational chemistry*, pages 367–422, 1991.
- [289] Lemont B Kier and Lowell H Hall. An electrotopological-state index for atoms in molecules. *Pharmaceutical research*, 7(8):801–807, 1990.

- [290] Johann Gasteiger and Mario Marsili. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*, 36(22):3219–3228, 1980.
- [291] JM Tarascon and D Guyomard. New electrolyte compositions stable over the 0 to 5 v voltage range and compatible with the $\text{Li}_{1+x}\text{Mn}_2\text{O}_4$ /carbon li-ion cells. *Solid State Ionics*, 69(3-4):293–305, 1994.
- [292] Marcelo J Monteiro, Fernanda FC Bazito, Leonardo JA Siqueira, Mauro CC Ribeiro, and Roberto M Torresi. Transport coefficients, raman spectroscopy, and computer simulation of lithium salt solutions in an ionic liquid. *The Journal of Physical Chemistry B*, 112(7):2102–2109, 2008.
- [293] Robin D Rogers and Kenneth R Seddon. Ionic liquids—solvents of the future? *Science*, 302(5646):792–793, 2003.
- [294] Werner Kunz and Katharina Häckl. The hype with ionic liquids as solvents. *Chemical Physics Letters*, 661:6–12, 2016.
- [295] Mahinder Ramdin, Theo W de Loos, and Thijs JH Vlugt. State-of-the-art of CO_2 capture with ionic liquids. *Industrial & Engineering Chemistry Research*, 51(24):8149–8177, 2012.
- [296] Hans-Peter Steinrueck and Peter Wasserscheid. Ionic liquids in catalysis. *Catalysis Letters*, 145(1):380–397, 2015.
- [297] Zhen Yang and Wubin Pan. Ionic liquids: Green solvents for nonaqueous biocatalysis. *Enzyme and Microbial Technology*, 37(1):19–28, 2005.
- [298] A Balducci, SS Jeong, GT Kim, S Passerini, M Winter, M Schmuck, GB Appetecchi, R Marcilla, D Mecerreyes, V Barsukov, et al. Development of safe, green and high performance ionic liquids-based batteries (illibatt project). *Journal of Power Sources*, 196(22):9719–9730, 2011.

- [299] Xiaochuan Jiang, Yi Nie, Chunxi Li, and Zihao Wang. Imidazolium-based alkylphosphate ionic liquids—a potential solvent for extractive desulfurization of fuel. *Fuel*, 87(1):79–84, 2008.
- [300] Frank Wendler, Loredana-Nicoleta Todi, and Frank Meister. Thermostability of imidazolium ionic liquids as direct solvents for cellulose. *Thermochimica acta*, 528:76–84, 2012.
- [301] Maan Hayyan, Farouq S Mjalli, Mohd Ali Hashim, Inas M AlNashef, and Tan Xue Mei. Investigating the electrochemical windows of ionic liquids. *Journal of Industrial and Engineering Chemistry*, 19(1):106–112, 2013.
- [302] A Romero, A Santos, J Tojo, and AJJOHM Rodríguez. Toxicity and biodegradability of imidazolium ionic liquids. *Journal of hazardous materials*, 151(1):268–273, 2008.
- [303] Bernd Jastorff, Reinhold Störmann, Johannes Ranke, Kerstin Mölter, Frauke Stock, Boris Oberheitmann, Wolfgang Hoffmann, Jens Hoffmann, Matthias Nüchter, Bernd Ondruschka, et al. How hazardous are ionic liquids? structure–activity relationships and biological testing as important elements for sustainability evaluation. *Green Chemistry*, 5(2):136–142, 2003.
- [304] Ricardo Condemarin and Paul Scovazzo. Gas permeabilities, solubilities, diffusivities, and diffusivity correlations for ammonium-based room temperature ionic liquids with comparison to imidazolium and phosphonium rtil data. *Chemical Engineering Journal*, 147(1):51–57, 2009.
- [305] Giovanni B Appetecchi, Maria Montanino, Maria Carewska, Margherita Moreno, Fabrizio Alessandrini, and Stefano Passerini. Chemical–physical properties of bis (perfluoroalkylsulfonyl) imide-based ionic liquids. *Electrochimica Acta*, 56(3):1300–1307, 2011.
- [306] Jiazeng Sun, Douglas Robert Macfarlane, and Maria Forsyth. A new family of ionic liquids based on the 1-alkyl-2-methyl pyrrolinium cation. *Electrochimica acta*, 48(12):1707–1711, 2003.

- [307] Urszula Domańska, Marta Krolikowska, and Kamil Paduszyński. Physico-chemical properties and phase behaviour of piperidinium-based ionic liquids. *Fluid phase equilibria*, 303(1):1–9, 2011.
- [308] Jacob M Crosthwaite, Mark J Muldoon, JaNeille K Dixon, Jessica L Anderson, and Joan F Brennecke. Phase transition and decomposition temperatures, heat capacities and viscosities of pyridinium ionic liquids. *The Journal of Chemical Thermodynamics*, 37(6):559–568, 2005.
- [309] Pedro Verdía, Marta Hernaiz, Emilio J González, Eugénia A Macedo, Josefa Salgado, and Emilia Tojo. Effect of the number, position and length of alkyl chains on the physical properties of polysubstituted pyridinium ionic liquids. *The Journal of Chemical Thermodynamics*, 69:19–26, 2014.
- [310] JN Barisci, Gordon G Wallace, Douglas Robert Macfarlane, and RH Baughman. Investigation of ionic liquids as electrolytes for carbon nanotube electrodes. *Electrochemistry communications*, 6(1):22–27, 2004.
- [311] Pierre Bonhote, Ana-Paula Dias, Nicholas Papageorgiou, Kuppuswamy Kalyanasundaram, and Michael Grätzel. Hydrophobic, highly conductive ambient-temperature molten salts. *Inorganic chemistry*, 35(5):1168–1178, 1996.
- [312] Takashi Makino, Mitsuhiro Kanakubo, Yoshio Masuda, Tatsuya Umecky, and Akira Suzuki. CO₂ absorption properties, densities, viscosities, and electrical conductivities of ethylimidazolium and 1-ethyl-3-methylimidazolium ionic liquids. *Fluid Phase Equilibria*, 362:300–306, 2014.
- [313] Andreas Nazet, Sophia Sokolov, Thomas Sonnleitner, Sergej Friesen, and Richard Buchner. Densities, refractive indices, viscosities, and conductivities of non-imidazolium ionic liquids [et3s][tfsi],[et2mes][tfsi],[bupy][tfsi],[n8881][tfa], and [p14][dca]. *Journal of Chemical & Engineering Data*, 62(9):2549–2561, 2017.

- [314] My-Loan-Phung Le, Fannie Alloin, Pierre Strobel, Jean-Claude Leprêtre, Laure Cointeaux, and Carlos Perez del Valle. Electrolyte based on fluorinated cyclic quaternary ammonium ionic liquids. *Ionics*, 18(9):817–827, 2012.
- [315] Long Chen, Mahdi Sharifzadeh, Niall Mac Dowell, Tom Welton, Nilay Shah, and Jason P Hallett. Inexpensive ionic liquids:[hso 4]-based solvent production at bulk scale. *Green Chemistry*, 16(6):3098–3106, 2014.
- [316] Yabo Dan, Yong Zhao, Xiang Li, Shaobo Li, Ming Hu, and Jianjun Hu. Generative adversarial networks (gan) based efficient sampling of chemical composition space for inverse design of inorganic materials. *npj Computational Materials*, 6(1):1–7, 2020.
- [317] Jaechang Lim, Seongok Ryu, Jin Woo Kim, and Woo Youn Kim. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *Journal of cheminformatics*, 10(1):1–9, 2018.
- [318] Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.
- [319] Mykola Galushka, Chris Swain, Fiona Browne, Maurice D Mulvenna, Raymond Bond, and Darren Gray. Prediction of chemical compounds properties using a deep learning model. *Neural Computing and Applications*, pages 1–22, 2021.
- [320] Lee A. Hollaar. Direct implementation of asynchronous control units. *IEEE Transactions on Computers*, 31(12):1133–1141, 1982.
- [321] Carlos E Ferreira, Nieves MC Talavera-Prieto, Isabel MA Fonseca, António TG Portugal, and Abel GM Ferreira. Measurements of pvt, viscosity, and surface tension of trihexyltetradecylphosphonium tris (pentafluoroethyl)

- trifluorophosphate ionic liquid and modelling with equations of state. *The Journal of Chemical Thermodynamics*, 47:183–196, 2012.
- [322] Michael Moret, Lukas Friedrich, Francesca Grisoni, Daniel Merk, and Gisbert Schneider. Generative molecular design in low data regimes. *Nature Machine Intelligence*, 2(3):171–180, 2020.
- [323] Greg Landrum. Rdkit: Open-source cheminformatics.
- [324] Noel M O’Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1):1–14, 2011.
- [325] Shyue Ping Ong, Oliviero Andreussi, Yabi Wu, Nicola Marzari, and Gerbrand Ceder. Electrochemical windows of room-temperature ionic liquids from molecular dynamics and density functional theory calculations. *Chemistry of Materials*, 23(11):2979–2986, 2011.
- [326] Yong Zhang, Chaojun Shi, Joan F Brennecke, and Edward J Maginn. Refined method for predicting electrochemical windows of ionic liquids and experimental validation studies. *The Journal of Physical Chemistry B*, 118(23):6250–6255, 2014.
- [327] Mario Krenn, Florian Häse, AkshatKumar Nigam, Pascal Friederich, and Alan Aspuru-Guzik. Self-referencing embedded strings (selfies): A 100% robust molecular string representation. *Machine Learning: Science and Technology*, 1(4):045024, 2020.
- [328] Yanqiang Zhang, Suojiang Zhang, Xingmei Lu, Qing Zhou, Wei Fan, and XiangPing Zhang. Dual amino-functionalised phosphonium ionic liquids for CO₂ capture. *Chemistry—A European Journal*, 15(12):3003–3011, 2009.
- [329] Qi Li, Juner Chen, Lei Fan, Xueqian Kong, and Yingying Lu. Progress in electrolytes for rechargeable li-based batteries and beyond. *Green Energy & Environment*, 1(1):18–42, 2016.

- [330] Kang Xu. Nonaqueous liquid electrolytes for lithium-based rechargeable batteries. *Chemical reviews*, 104(10):4303–4418, 2004.
- [331] Ziqi Zeng, Bingbin Wu, Lifan Xiao, Xiaoyu Jiang, Yao Chen, Xinping Ai, Hanxi Yang, and Yuliang Cao. Safer lithium ion batteries based on non-flammable electrolyte. *Journal of Power Sources*, 279:6–12, 2015.
- [332] James H. Davis, Jr. Task-specific ionic liquids. *Chemistry letters*, 33(9):1072–1077, 2004.
- [333] Ralf Giernoth. Task-specific ionic liquids. *Angewandte Chemie International Edition*, 49(16):2834–2839, 2010.
- [334] M Moreno, E Simonetti, GB Appetecchi, M Carewska, M Montanino, G-T Kim, N Loeffler, and S Passerini. Ionic liquid electrolytes for safer lithium batteries. *Journal of The Electrochemical Society*, 164(1):A6026, 2016.
- [335] Kenneth N Marsh, John A Boxall, and Ruediger Lichtenthaler. Room temperature ionic liquids and their mixtures—a review. *Fluid phase equilibria*, 219(1):93–98, 2004.
- [336] Pratik Dhakal and Jindal Shah. A generalized machine learning model for predicting ionic conductivity for ionic liquids. 2022.
- [337] Wim Buijs, Geert-Jan Witkamp, and Maaïke C Kroon. Correlation between quantumchemically calculated lumo energies and the electrochemical window of ionic liquids with reduction-resistant anions. *International Journal of Electrochemistry*, 2012, 2012.
- [338] Yinghao Yu, Xingmei Lu, Qing Zhou, Kun Dong, Hongwei Yao, and Suojian Zhang. Biodegradable naphthenic acid ionic liquids: synthesis, characterization, and quantitative structure–biodegradation relationship. *Chemistry–A European Journal*, 14(35):11174–11182, 2008.
- [339] Elif Can, Ahsan Jalal, I Gulcin Zirhlioglu, Alper Uzun, and Ramazan Yildirim. Predicting water solubility in ionic liquids using machine learning

- towards design of hydro-philic/phobic ionic liquids. *Journal of Molecular Liquids*, page 115848, 2021.
- [340] Mahsa Aghaie and Sohrab Zendehboudi. Estimation of CO_2 solubility in ionic liquids using connectionist tools based on thermodynamic and structural characteristics. *Fuel*, 279:117984, 2020.
- [341] Ali Yousefi, Soheila Javadian, Nima Dalir, Jamal Kakemam, and Jafar Akbari. Imidazolium-based ionic liquids as modulators of corrosion inhibition of SDS on mild steel in hydrochloric acid solutions: experimental and theoretical studies. *RSC advances*, 5(16):11697–11713, 2015.
- [342] Guocai Tian and Kaitao Yuan. Performance and mechanism of alkylimidazolium ionic liquids as corrosion inhibitors for copper in sulfuric acid solution. *Molecules*, 26(16):4910, 2021.
- [343] Eapen Thomas, KP Vijayalakshmi, and Benny K George. Kinetic stability of imidazolium cations and ionic liquids: a frontier molecular orbital approach. *Journal of Molecular Liquids*, 276:721–727, 2019.
- [344] Huilong Dong, Fenglou Gu, Min Li, Bencai Lin, Zhihong Si, Tingjun Hou, Feng Yan, Shuit-Tong Lee, and Youyong Li. Improving the alkaline stability of imidazolium cations by substitution. *ChemPhysChem*, 15(14):3006–3014, 2014.
- [345] Tina Erdmenger, Jürgen Vitz, Frank Wiesbrock, and Ulrich S Schubert. Influence of different branched alkyl side chains on the properties of imidazolium-based ionic liquids. *Journal of Materials Chemistry*, 18(43):5267–5273, 2008.
- [346] Marta C Corvo, João Sardinha, Teresa Casimiro, Graciane Marin, Marcus Seferin, Sandra Einloft, Sonia C Menezes, Jairton Dupont, and Eurico J Cabrita. A rational approach to CO_2 capture by imidazolium ionic liquids: Tuning CO_2 solubility by cation alkyl branching. *ChemSusChem*, 8(11):1935–1946, 2015.

- [347] Kiki A Kurnia, Tânia E Sintra, Catarina MSS Neves, Karina Shimizu, José N Canongia Lopes, Fernando Gonçalves, Sónia PM Ventura, Mara G Freire, Luís MNBF Santos, and João AP Coutinho. The effect of the cation alkyl chain branching on mutual solubilities with water and toxicities. *Physical Chemistry Chemical Physics*, 16(37):19952–19963, 2014.
- [348] Sonia PM Ventura, Carolina S Marques, Andreia A Rosatella, Carlos AM Afonso, Fernando Goncalves, and João AP Coutinho. Toxicity assessment of various ionic liquid families towards vibrio fischeri marine bacteria. *Ecotoxicology and environmental safety*, 76:162–168, 2012.
- [349] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Gogings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian~16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.
- [350] Suseeladevi Asha, Kunduchi Periya Vijayalakshmi, and Benny K George. Electronic structural studies of pyrrolidinium-based ionic liquids for electrochemical application. *International Journal of Quantum Chemistry*, 119(17):e25972, 2019.

- [351] Guillaume Hoffmann, Muhammet Balcilar, Vincent Tognetti, Pierre Héroux, Benoît Gaüzère, Sébastien Adam, and Laurent Joubert. Predicting experimental electrophilicities from quantum and topological descriptors: A machine learning approach. *Journal of Computational Chemistry*, 41(24):2124–2136, 2020.
- [352] Alexandru T Balaban. Highly discriminating distance-based topological index. *Chemical physics letters*, 89(5):399–404, 1982.
- [353] Steven H Bertz. The first general index of molecular complexity. *Journal of the American Chemical Society*, 103(12):3599–3601, 1981.
- [354] Nathan D Phillip, Andrew S Westover, Claus Daniel, and Gabriel M Veith. Structural degradation of high voltage lithium nickel manganese cobalt oxide (nmc) cathodes in solid-state batteries and implications for next generation energy storage. *ACS Applied Energy Materials*, 3(2):1768–1774, 2020.
- [355] Bernardine LD Rinkel, David S Hall, Israel Temprano, and Clare P Grey. Electrolyte oxidation pathways in lithium-ion batteries. *Journal of the American Chemical Society*, 142(35):15058–15074, 2020.
- [356] <https://ilthermo.boulder.nist.gov/>. Accessed: 2020-06-22.
- [357] Elsa A Olivetti, Jacqueline M Cole, Edward Kim, Olga Kononova, Gerbrand Ceder, Thomas Yong-Jin Han, and Anna M Hiszpanski. Data-driven materials research enabled by natural language processing and information extraction. *Applied Physics Reviews*, 7(4):041317, 2020.
- [358] Daniel C Elton, Dhruv Turakhia, Nischal Reddy, Zois Boukouvalas, Mark D Fuge, Ruth M Doherty, and Peter W Chung. Using natural language processing techniques to extract information on the properties and functionalities of energetic materials from large text corpora. *arXiv preprint arXiv:1903.00415*, 2019.

- [359] Zhi-Bin Zhou, Hajime Matsumoto, and Kuniaki Tatsumi. Structure and properties of new ionic liquids based on alkyl- and alkenyltrifluoroborates. *ChemPhysChem*, 6(7):1324–1332, 2005.
- [360] S García-Garabal, J Vila, E Rilo, M Domínguez-Pérez, L Segade, E Tojo, P Verdía, LM Varela, and O Cabeza. Transport properties for 1-ethyl-3-methylimidazolium n-alkyl sulfates: Possible evidence of grotthuss mechanism. *Electrochimica Acta*, 231:94–102, 2017.
- [361] Veronika Strehmel, André Laschewsky, Hendrik Wetzel, and Eckhard Görnitz. Free radical polymerization of n-butyl methacrylate in ionic liquids. *Macromolecules*, 39(3):923–930, 2006.
- [362] Karina Shimizu, Carlos E S Bernardes, and José N Canongia Lopes. Structure and Aggregation in the 1-Alkyl-3-Methylimidazolium Bis(trifluoromethylsulfonyl)imide Ionic Liquid Homologous Series. *The Journal of Physical Chemistry B*, 118(2):567–576, January 2014.
- [363] Ramesh L Gardas, Mara G Freire, Pedro J Carvalho, Isabel M Marrucho, Isabel MA Fonseca, Abel GM Ferreira, and Joao AP Coutinho. P ρ t measurements of imidazolium-based ionic liquids. *Journal of Chemical & Engineering Data*, 52(5):1881–1888, 2007.
- [364] Catarina MSS Neves, Kiki Adi Kurnia, Joaáo AP Coutinho, Isabel M Marrucho, José N Canongia Lopes, Mara G Freire, and Luís Paulo N Rebelo. Systematic study of the thermophysical properties of imidazolium-based ionic liquids with cyano-functionalized anions. *The Journal of Physical Chemistry B*, 117(35):10271–10283, 2013.
- [365] José MSS Esperança, Henrique JR Guedes, José N Canongia Lopes, and Luís Paulo N Rebelo. Pressure- density- temperature (p- ρ - t) surface of [c₆mim][ntf₂]. *Journal of Chemical & Engineering Data*, 53(3):867–870, 2008.
- [366] Francois Chollet. Building autoencoders in keras. *The Keras Blog*, 14, 2016.
- [367] François Chollet et al. Keras. <https://keras.io>, 2015.

- [368] Mark Davies, Michał Nowotka, George Papadatos, Nathan Dedman, Anna Gaulton, Francis Atkinson, Louisa Bellis, and John P Overington. Chembl web services: streamlining access to drug discovery data and utilities. *Nucleic acids research*, 43(W1):W612–W620, 2015.
- [369] Leandro Martínez, Ricardo Andrade, Ernesto G Birgin, and José Mario Martínez. Packmol: a package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry*, 30(13):2157–2164, 2009.
- [370] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.
- [371] Alpeshkumar K Malde, Le Zuo, Matthew Breeze, Martin Stroet, David Poger, Pramod C Nair, Chris Oostenbrink, and Alan E Mark. An automated force field topology builder (atb) and repository: version 1.0. *Journal of chemical theory and computation*, 7(12):4026–4037, 2011.

Vita

Pratik Dhakal

Candidate for the Degree of

Doctor of Philosophy

Dissertation: LEVERAGING ATOMISTIC SIMULATIONS AND MACHINE LEARNING FOR THE DESIGN OF IONIC LIQUIDS AS ELECTROLYTES FOR BATTERY APPLICATION

Major Field: CHEMICAL ENGINEERING

Biographical

Education:

- Doctor of Philosophy in Chemical Engineering, Oklahoma State University, Stillwater, Oklahoma, USA, May 2022
- Master of Science in Chemical Engineering, Miami University, Oxford, Ohio, USA, August 2018
- Bachelor of Science in Chemical Engineering, Shandong University of Science and Technology, Qingdao, Shandong, China, August 2016