

Article

Open Access



# Room temperature ionic liquids viscosity prediction from deep-learning models

Zafer Acar<sup>1,3</sup> , Phu Nguyen<sup>2,3</sup>, Xiaoqi Cui<sup>3</sup>, Kah Chun Lau<sup>3,\*</sup> 

<sup>1</sup>Department of Physics, Michigan Technological University, Houghton, MI 49931, USA.

<sup>2</sup>Department of Computer Science, San José State University, San José, CA 95192-0249, USA.

<sup>3</sup>Department of Physics and Astronomy, California State University Northridge, Northridge, CA 91330, USA.

\*Correspondence to: Prof./Dr. Kah Chun Lau, Department of Physics and Astronomy, California State University Northridge, Northridge, CA 91330, USA. E-mail: kahchun.lau@csun.edu

**How to cite this article:** Acar Z, Nguyen P, Cui X, Lau KC. Room temperature ionic liquids viscosity prediction from deep-learning models. *Energy Mater* 2023;3:300039. <https://dx.doi.org/10.20517/energymater.2023.38>

**Received:** 16 May 2023 **First Decision:** 25 Jun 2023 **Revised:** 18 Jul 2023 **Accepted:** 10 Aug 2023 **Published:** 4 Sep 2023

**Academic Editors:** Giovanni Battista Appetecchi, Hao Liu, Sen Xin, Yuping Wu **Copy Editor:** Fangling Lan **Production Editor:** Fangling Lan

## Abstract

Ionic liquids (ILs) are a new group of novel solvents with great potential in design-synthesis. They are promising electrolyte candidates in energy storage applications, especially in rechargeable batteries. However, in practice, their usage remains limited due to the unfavorable high-viscosity ( $\eta$ ) property at ambient conditions. To optimize the design synthesis of ILs, a systematic fundamental study of their structure-property relationship is deemed necessary. In this study, we employed a deep-learning (DL) model to predict the room-temperature viscosity of a wide range of ILs that consist of various cationic and anionic families. Based on this DL model, accurate prediction of IL viscosity can be realized, reaching an  $R^2$  score of 0.99 with a root mean square error of  $\sim 45$  mPa·s. To further help identify low- and high- $\eta$  ILs, a low/high- $\eta$  binary classification model with an overall accuracy of 93% for test prediction is obtained based on the DL model. From the important structure-property relationship analysis governed by the top-rank molecular descriptors of this model, a list of very low- $\eta$  ILs (i.e.,  $\eta < 30$  mPa·s) that could be potentially useful in battery electrolytes is identified. Based on the finding of the DL model, it suggests that in order to achieve low- $\eta$ , grafting IL cations into smaller sizes (e.g., smaller head rings) and short alkyl chains and reducing ionization potentials/energies will help. Meanwhile, for the same cations, further reducing anions in sizes, chain lengths, and hydrogen bonds might be useful to further reduce the viscosity. Thus, with a fine selection and molecular grafting of anionic and cationic species in ILs, we believe fine-tuning IL viscosities can be achieved through the proper design synthesis of functional groups in ILs.

**Keywords:** Machine learning, deep learning, ionic liquids, batteries, viscosity



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



## INTRODUCTION

In current modern society, increasing energy demand is inevitable and causes a lot of unwanted environmental problems due to the strong dependence on fossil fuels used to generate electricity in various applications. In particular, the consequences of climate change can be dramatic. To be better tailored to this challenge, utilizing and diversifying renewable energy resources is the key to sustainable growth of current society. To achieve this goal, wide applications of electrochemical energy storage technologies (e.g., batteries) are one of the promising strategies because of their several attractive features, including excellent round-trip efficiency, high power flexibility in various grid applications, durable cycle life, *etc.*<sup>[1,2]</sup>. Meanwhile, to provide intermittent renewable energy resources (e.g., solar, wind) into the grid and various applications, rechargeable batteries are known to be a practical solution to energy storage technology. They are energy-saving and environmentally friendly and can be utilized on a large scale, supplying the world with clean and sustainable electricity.

Due to their high volumetric energy density<sup>[3,4]</sup>, lithium (Li)-based batteries have been widely used in many portable electronic devices since the 1990s and are now powering many battery electric vehicles. To improve their performance and capacity, there is tremendous progress in the exploration and design-development of new electrode materials in Li-based batteries, especially the cathodes, including Li-O<sub>2</sub> and Li-S batteries<sup>[4-6]</sup>. However, one of the limiting factors remains to be the electrolytes, which inherently govern the current or power density, cycling performance, electrochemical stability, and safety<sup>[6-13]</sup>. For a working electrolyte, it has to be both a good ionic conductor and an electronic insulator; therefore, ionic liquids (ILs) are one of the promising candidates for electrolytes<sup>[7-13]</sup>.

For rechargeable batteries, the crucial parameters that determine their IL-based electrolytes include the viscosity ( $\eta$ ), ionic conductivity, electrochemical stability, and safety<sup>[7-15]</sup>. Among these key parameters, ionic conductivity is one of the important performance metrics for these novel liquids in energy storage applications, which is usually determined by the viscosity. In general, low  $\eta$  is strongly correlated with high ionic conductivity<sup>[14,15]</sup> and is an important feature critically needed in battery applications. The viscosity determines the resistance to flow in ILs and is generally governed by several factors: (1) size; (2) shape; and (3) ionic interactions among the constituent anions and cations. These interactions are determined by electrostatic forces, van der Waals forces, and hydrogen bonds, which depend on the molecular structures of ILs<sup>[16-20]</sup>. In addition, it was also found that the fluidity of ILs is complicated. Besides the motion of free cations or anions, the influences of ion pairs, clusters, or aggregates are significant in ILs<sup>[21,22]</sup>. Although these factors are known, finding optimum and practical design principles to perfect the related physicochemical properties (e.g., viscosity) of ILs is not trivial.

It is known that ILs are a specific class of molecular electrolytes characterized by the absence of co-solvent in solution due to their unique interplay between electrostatic and van der Waals interactions<sup>[23]</sup>. However, understanding their unique viscosity features can be challenging. Particularly, an accurate measurement of viscosity is never a trivial task<sup>[24]</sup>. Although using viscometers has proven to be an effective means of determining the IL viscosity over a wide range of temperatures and pressures, measuring an extensive selection of ILs at various thermodynamic conditions can be extremely challenging. Meanwhile, to better understand their fundamental properties, the advanced molecular dynamics (MD) simulations based on polarizable force fields can provide thermodynamics, structural correlations, ions dynamics, and collective dynamics description, including viscosity prediction<sup>[17,25-29]</sup> with great accuracy. However, the huge computational cost generally limits their use and is only suitable for specific problem-based focus studies. According to Katritzky *et al.*, about 10<sup>18</sup> combinations of cations or anions could be used to form ILs<sup>[30]</sup>. The large diversity of IL species and their physicochemical properties make a systematic detailed investigation

on these IL compounds exceptionally difficult, especially for both experimental and theoretical studies.

To address this challenge, it is important to develop robust computational tools to benefit the experimental design and synthesis of new ILs with desirable structural properties, such as ILs with practically low  $\eta$  and high ionic conductivity within a room-temperature range. To explore and predict the viscosities of large varieties of ILs, a comprehensive study based on advanced atomistic or molecular simulation can be a challenge and might not be feasible in practice<sup>[25-29]</sup>. To overcome this challenge, a systematic high-throughput screening supported by a detailed study on a large amount of reported IL datasets, with a specific focus on viscosity analysis, has been proposed in recent years<sup>[31-38]</sup>. As an affordable solution and a predictive model to estimate or predict the viscosity of various ILs, advanced statistical models following Arrhenius, Litovitz, Andrade, Vogel-Fucher-Tammann (VFT) equations using quantitative structure-property relationships (QSPR) are one of the practical methods<sup>[33-35]</sup>. In particular, advanced traditional machine-learning (ML) methods, such as support vector machine (SVM) and least-squares SVM (LSSVM) approaches, are found to be very useful<sup>[34,38]</sup> when combined with structural data of ILs through the group contribution (GC) theory. In terms of the accuracy and prediction capacity, this approach is found to be comparable to classical QSPR methods<sup>[33-35]</sup>.

However, due to the ever-increasing data on ILs in reported literature, a well-timed strategy is to utilize advanced ML methods for a systematic study of various types of IL properties. To accommodate the huge datasets from reported literature, the contemporary deep-learning (DL) models<sup>[39-42]</sup> are known to outperform traditional statistical methods or ML models because of their capacity to process a huge number of feature properties from big data and intelligent big data analysis in materials science for design and discovery<sup>[43,44]</sup>. Thus, the development of robust simulation methods that integrate high-throughput screening of huge feature properties of ILs using state-of-the-art data-mining approaches and DL models can be extremely valuable in their fundamental studies. This unique strategy will help us to significantly speed up the exploration and discovery of new ILs from currently known data, complementary to more specific case studies using advanced atomistic simulations. With this as a motivation, we propose to adopt a combined data mining method, chemoinformatic approach, and DL models to high-throughput screen and predict the viscosities for a large variety of ILs, with the hope to benefit the design and development of electrolytes in energy storage applications.

## METHODS

### Dataset extraction and preparation

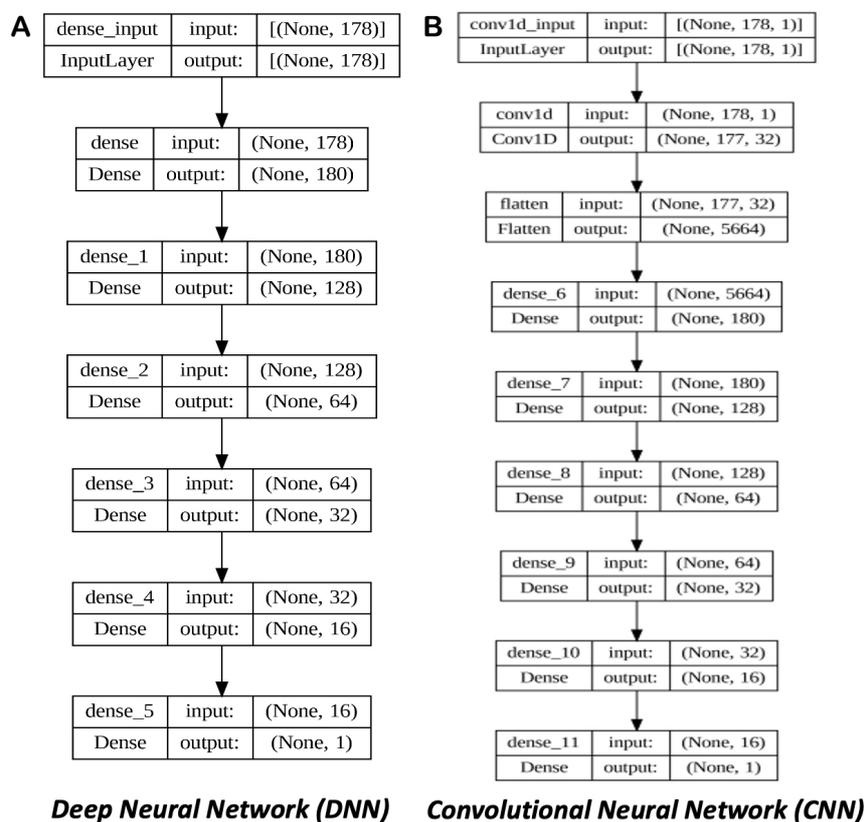
In this work, all the viscosity datasets we used were gathered from the Ionic Liquids Database - ILThermo (v2.0.)<sup>[45,46]</sup>, which is a comprehensive database of thermophysical and thermodynamic properties of ILs in the field. According to the latest update (by 28th Dec 2022), the ILThermo contains 2,732 types of ILs and includes 5,177 compounds, with a total of 870,304 datapoints collected from 4,230 published works of literature. For pure ILs, the database has 2,332 IL systems with 145,602 datapoints related to thermodynamic, thermochemical, and transport properties. For this study, the room-temperature viscosity collected dataset contains 922 types of ILs and includes a wide range of IL families. For this wide range of IL candidates, predicting room-temperature viscosities accurately and understanding their useful structure-property relationship without depending on costly computational resources are important in practice. The details of data extraction, conversion, and post-processing of the ILThermo dataset can be found in our recent work<sup>[47]</sup>. To account for an extensive description of molecular structure features of each individual IL in the dataset, all the 5,272 molecular descriptors based on chemoinformatic QSPR approach<sup>[48,49]</sup>, e.g., constitutional indices, topological indices, connectivity indices, walk and path counts, *etc.*, were generated based on Dragon7 software<sup>[48]</sup>.

Prior to the ML study, it is important to remove the unnecessary data. During this process, all the molecular descriptor columns with low variance and those containing missing values or empty columns were removed. All the molecular descriptors and viscosity values were normalized using the Standard Scaler function from scikit-learn (scikit-learn 1.2.0)<sup>[50]</sup>. To overcome the problem of overfitting, the dimension of the original molecule descriptors matrix was further reduced using the Pearson correlation. This helps us to further identify important molecular descriptors that exhibit a statistical significance with a high correlation to the viscosity of ILs. To achieve this goal, the molecular descriptors with low correlations ( $< 0.20$ ) and high correlations ( $> 0.90$ ) were further excluded from the dataset. Throughout this process, a set of important molecular structure features consisting of 179 molecular descriptors was identified. After the correlation feature selection and normalization of these molecular descriptors, the dataset was randomly split into training and testing datasets with a ratio of 80/20 (or 80% for training and 20% for testing) for the evaluation of our ML models. To improve the accuracy (i.e.,  $R^2$ ) of our ML models in regression analysis, we have removed some of the outliers (i.e., 13 in total). These outliers are not restricted to particular cationic or anionic species. They are mostly ILs with high  $\eta$  ( $\eta \sim 150\text{-}2,030$  mPa·s) and might not be very useful in batterie applications. Generally, all these outliers are large molecules and have large molecular weight ( $190 \leq MW \leq 583$ ), large Bertz branching index (i.e.,  $11 \leq BBI \leq 52$ ), and large Zagreb indices<sup>[51,52]</sup> (e.g.,  $44 \leq ZM1 \leq 176$ ) due to their high complexity in molecular substructures (e.g., N,N,N-tributyloctan-1-aminium trifluoromethanesulfonate, decyldimethyl(4-vinylbenzyl)ammonium bis(trifluoromethanesulfonyl)amide), etc.). Therefore, for the  $\eta$ -prediction based on regression model of ML in the subsequent study (Section "Viscosity prediction from deep-learning (DL) models"), it will base on 909 types instead of 922 types of ILs found in this work.

#### Machine-learning model: deep-learning

In this work, we have considered two types of DL models, i.e., deep neural networks (DNN) and convolutional neural networks (CNN), which are based on the algorithms implemented in TensorFlow (i.e., TensorFlow 2)<sup>[53]</sup>, a popular ML framework that provides a high-level python API to construct and train DL models. DL is a subfield of ML and is an artificial neural network (ANN) that is essentially represented by multiple layers of neural networks. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accurate prediction<sup>[39-41]</sup>. These neural networks attempt to mimic how the human brain processes information progressively with higher-level features from large data and are able to develop a hierarchy of learning processes based on a set of algorithms defined within each layer.

For a DL model, a multi-layer feedforward neural network is constructed with multiple hidden layers, of which each layer contains predefined numbers of neurons to capture the non-linear relationship between the input features (e.g., molecular descriptors of ILs) and the output viscosities. The network is trained using stochastic gradient descent with backpropagation to minimize the mean squared error between the predicted and actual viscosity values. Figure 1 shows a general structure of our DNN and CNN models, which consists of multiple layers, with each layer containing a different number of neurons. As shown in Figure 1A, our DNN model consists of one input layer (179 neurons), 1st hidden layer (128 neurons), 2nd hidden layer (64 neurons), 3rd hidden layer (32 neurons), 4th hidden layer (16 neurons), and output layer (1 neuron). In contrast, our CNN model [Figure 1B] is a neural network with basic building blocks/layers consisting of tensors except for the output layer. Specifically, Figure 1B features one input layer represented by 1D convolutional layer with tensor (179, 1), 1st hidden layer based on 1D convolutional layer with tensor (177, 32), 2nd hidden layer as flattened layer (5,664 neurons), 3rd hidden layer (180 neurons), 4th hidden layer (128 neurons), 5th hidden layer (64 neurons), 6th hidden layer (32 neurons), 7th hidden layer (16 neurons) and output layer (1 neuron). To update all the neural network weights or hyperparameters iteratively during the training of the DL model, an adaptive moment estimation (Adam) optimizer, a



**Figure 1.** The schematic plot of the deep-learning (DL) models adopted in this work: (A) deep neural network (DNN) and (B) convolutional neural network (CNN).

stochastic gradient descent method, is used to speed up the optimization process<sup>[54]</sup>. To avoid an over-fitting problem during the training process of DL models (i.e., DNN and CNN), we have carried out a systematic study on the accuracy of data validation or error rates for a range of epochs through an iterative process. As shown in [Supplementary Figure 1](#), an optimal number of epochs (i.e., epoch = 1,500) are identified when convergence is achieved (i.e., when the accuracy in the DL model stops improving). Thus, in this work, to fit the DNN and CNN with a training dataset, the number of iterations, i.e., epochs, is set to 1,500, and the batch size equal to 32 in each epoch is used throughout the process.

## RESULTS AND DISCUSSION

### Data consolidation and evaluation

In this work, we are only interested in the viscosity of pure ILs that are close to the room temperature (i.e.,  $T = 298 \pm 5$  K) and at ambient pressure (i.e.,  $P = 101 \pm 2$  kPa), a condition relevant to the operating condition of conventional batteries. By neglecting the binary and ternary mixtures, there are 2,119 entries of the viscosity dataset of pure ILs, which are within this room-temperature range collected from the ILThermo database<sup>[45,46]</sup>. Among these experimentally measured  $\eta$  data reported by different authors at ILThermo<sup>[45,46]</sup>, we found that some entries are duplicated or belong to the same ILs. For these duplicated entries, it is noted that the collected  $\eta$  data reported by different authors do exhibit some small discrepancies or variations, which might be possibly related to impurities (e.g., water), differences in purification procedures, or experimental techniques employed in viscosity measurements. To reconcile with the discrepancies reported by different authors that belong to the same ILs, the reported  $\eta$  values, which are duplicated by different studies or duplicated entries, were averaged and retained only if the minimum and

maximum reported values were within the average error (i.e., reported standard deviation) of the average value ( $\bar{\eta}$ ), i.e.,  $\bar{\eta} + \bar{\eta}_{std} < \eta_{max}$  and  $\bar{\eta} - \bar{\eta}_{std} > \eta_{min}$ , where  $\bar{\eta}$  is the average duplicated reported values,  $\bar{\eta}_{std}$  is the average measurement error, and  $\eta_{min}$  and  $\eta_{max}$  are the minimum and maximum reported value for the duplicated values reported by different studies, respectively.

In this dataset, a wide diversity of structures and functional groups in ILs with various combinations of cations (e.g., imidazolium, ammonium, pyridinium, phosphonium, pyrrolidinium, *etc.*) and anions (e.g., bistriflimide (NTF2) derivatives, sulfonate, phosphate, hexafluorophosphate, borate, sulfate, acetate, dicyanamide, triazolide, *etc.*) families can be found and is highlighted in [Figure 2](#). From this large dataset (922 types of ILs), a wide distribution of room temperature measured  $\eta$  can be found, whose viscosities vary from 2 to 97,000 (in mPa·s) based on various combinations of anionic and cationic families of ILs. Although the imidazolium-based cation is the dominant cationic family in this dataset [[Figure 2](#)], it covers a wide variety of ILs. Among the 320 types of ILs, those consisting of imidazolium-related cations exhibit a wide range of viscosity values, i.e., spanning from 1-ethyl-1H-imidazolium acetate to 1-(2-cyanoethyl)-3-(phenylmethyl)-1H-imidazolium chloride with  $\eta \sim 4$ -69,000 mPa·s. Meanwhile, for the imidazolium-based ILs that with viscosity  $< 50$  mPa·s, there are only 77 candidates are found from 320, as shown in [Figure 2](#). Similarly, a wide range of  $\eta \sim 12$ -20,100 mPa·s is also found among the dominant NTF2 anionic species [[Figure 2](#)] related ILs. Therefore, in our opinion, the issue of a strongly skewed IL dataset toward a narrow range of viscosities restricted to the dominant imidazolium cationic species or NTF2 anionic species should not be a major concern.

### Viscosity prediction from deep-learning (DL) models

To measure the accuracy of the prediction from the DL models (i.e., DNN and CNN) described in Section "Machine-learning model: deep-learning", three metrics, such as the square coefficient of correlation ( $R^2$ ), the root mean square error (RMSE), and mean absolute error (MAE), were used to assess the performance of the DL models for the test dataset. Based on the selected 179 molecular descriptors for each IL (Section "Dataset extraction and preparation"), a summary of the performance of DL models on test dataset prediction can be found in [Figure 3](#) and [Table 1](#). As shown in [Figure 3](#), a linear regression with high  $R^2$  value [[Table 1](#)] can be found. Supported by the high  $R^2$  value, the predicted viscosity values from both DNN and CNN models are found to be randomly distributed along the diagonal line with only a small deviation, compared with experimentally reported viscosity values at room temperature.

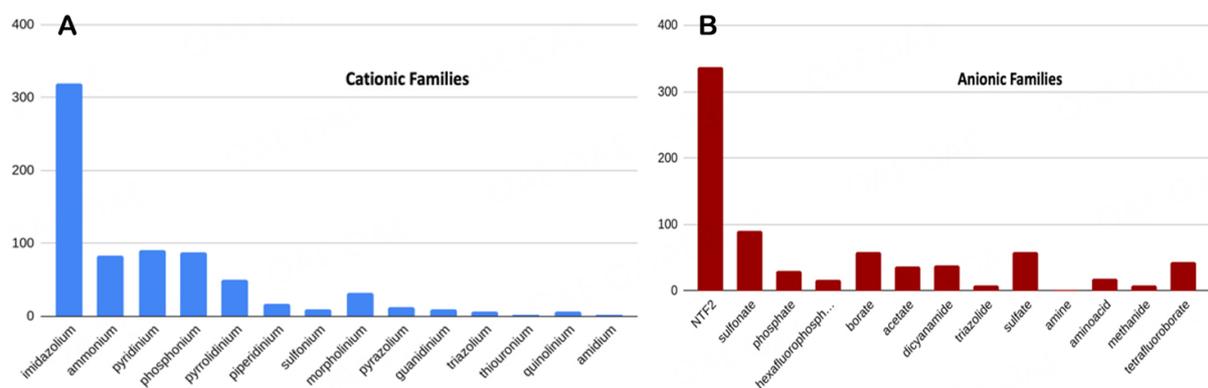
As shown in [Table 1](#), we found that the accuracy of viscosity prediction is, in general, dictated by the sample size of distinct ILs ( $n$ ), ML models, and the representation of structure features or molecular descriptors of ILs. Compared to the reported literature<sup>[34,36-38]</sup>, the test  $R^2$  score and RMSE for the prediction of viscosity around room temperature ( $T \sim 298$  K) reported in this work are found to be outstanding. Among the DL models we considered, the performance metrics obtained from CNN models are found to be among the best in reported literature, especially for the prediction of room-temperature viscosities (i.e.,  $R^2 \sim 0.9980$ ,  $RMSE \sim 45.27$ ,  $MAE \sim 30.42$ ) [[Table 1](#)]. For the DNN model, the prediction accuracy is also found to be excellent, i.e.,  $R^2 \sim 0.9869$ ,  $RMSE \sim 63.78$ ,  $MAE \sim 41.68$  [[Table 1](#)], compared to the reported literature, especially for the room-temperature viscosity prediction.

To further examine the influences of molecular descriptors on the performance of DL models, the top important molecular descriptors for these models were computed, and the top 20 most important molecular descriptors can be found in [Supplementary Table 1](#). Based on Pearson correlation, these 20 descriptors are found to have important impacts on the room-temperature viscosity of ILs obtained from DNN and CNN models. For DNN models, the performance of metrics is found to be nearly similar (i.e.,  $R^2 \sim 0.988$ ,

**Table 1.** The comparison of the performance metrics for test predictions of IL viscosities obtained in this work with reported literature. ANN is a feedforward artificial neural network<sup>[36]</sup>, and GC/FFANN-LSSVM is a combined two-layer feedforward artificial neural network and least-squares support vector machine based on a group contribution scheme for ILs<sup>[34]</sup>. MTL-TransCNN is a multi-task learning model based on transformer convolutional neural networks using QSPR models<sup>[37]</sup>. GC/SVM is a support vector machine model using a newly improved scheme of group contributions of ILs<sup>[38]</sup>

Reference	ML model	Number of distinct ILs (n)	Temperature (K)	R <sup>2</sup>	RMSE (mPa·s)	MAE (mPa·s)
This work	DNN	909	298 ± 5	0.9869	63.78	41.68
This work	CNN	909	298 ± 5	0.9980	45.27	30.42
Beckner <i>et al.</i> , 2018 <sup>[36]</sup>	ANN	723	273.15-373.15	0.9290 <sup>a</sup>	0.6856	N/A
Paduszyński <i>et al.</i> , 2019 <sup>[34]</sup>	GC/FFANN-LSSVM	1,596	290-410	0.9120 <sup>b</sup>	203.43	N/A
Baskin <i>et al.</i> , 2022 <sup>[37]</sup>	MTL-TransCNN	988	298	0.690 <sup>c</sup>	0.40	0.28
Baskin <i>et al.</i> , 2022 <sup>[37]</sup>	MTL-TransCNN	988	288-343	0.674 <sup>d</sup>	0.375	0.265
Boualem <i>et al.</i> , 2022 <sup>[38]</sup>	GC/SVM	1,654	253-571	0.9859 <sup>e</sup>	57.92	N/A

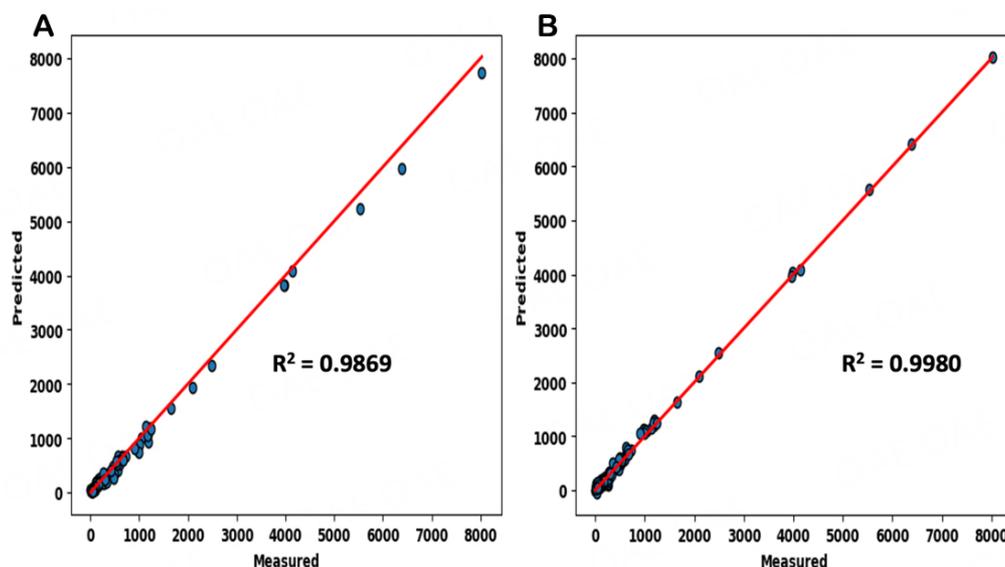
This part is footer. <sup>a</sup>Consider pressure range of 60-160 kPa. <sup>b</sup>Calculated R<sup>2</sup> based on the viscosity,  $\eta_0$  referenced at  $T = 298.15$  K is 0.8637, with RMSE = 0.51. <sup>c,d</sup>Calculated using log(MPa·s). <sup>e</sup>Calculated using log(mPa·s).



**Figure 2.** The schematic plot of the distribution of cationic (A) and anionic (B) families from various ILs in the dataset.

RMSE ~ 108.92 mPa·s, MAE ~ 65.96 mPa·s) compared to 179 molecular descriptors [Table 1] if only the selected top 20 molecular descriptors for each IL are used for test prediction [Supplementary Table 1]. Whereas for the CNN model, the performance of metrics based on the top 20 molecular descriptors [Supplementary Table 1] is found to be decreased slightly (i.e.,  $R^2 \sim 0.963$ , RMSE ~ 195.2 mPa·s, MAE ~ 115.41 mPa·s) compared to the prediction based on 179 molecular descriptors [Table 1]. Therefore, this suggests that the performance metrics of our DNN and CNN models are very similar and robust if an optimum number of important molecular descriptors are used.

Prior to our study, a systematic study of IL viscosity prediction across a wide range of IL families based on the ILThermo dataset<sup>[45,46]</sup> has been reported by Beckner *et al.* Based on the dataset, consisting of 723 distinct ILs for temperature range 273.15-373.15 K, pressure range of 60-160 kPa, and  $\eta$  range of 3.5-993 mPa·s, a reasonably good model of viscosity prediction can be obtained based on traditional ML models<sup>[36]</sup>. According to Beckner *et al.*, high accuracy of viscosity prediction ( $R^2 \sim 0.93$ , RMSE ~ 0.69 mPa·s) [Table 1] can be achieved based on feedforward ANN (FFANN)<sup>[36]</sup>. For that work, the result is based on 11 molecular descriptors selected by the least absolute shrinkage and selection operator (LASSO) from 633 molecular structure descriptors of ILs using the QSPR model.



**Figure 3.** The DL model predicted viscosity values vs. experimentally measured viscosity values (both in mPa·s unit) in the test dataset, with (A) representing the DNN model and (B) representing the CNN model separately. The dataset used consists of 909 types of ILs with 179 molecular descriptors for each IL.

The ability to predict the viscosities of a wide range of ILs accurately using the QSPR model has inspired further studies of physicochemical properties of other ILs in combination with diverse ML methods and molecular representations derived from QSPR models. To achieve this goal, a large-scale benchmark study of QSPR models combining several ML methods (e.g., random forest regression, RFR; extreme gradient boosting, XGBoost; TransCNN, *etc.*) with different types of molecular representations to predict several key physical properties of ILs (i.e., electrical conductance, density, refractive index, melting point, viscosity, and surface tension) was reported by Baskin *et al.* recently<sup>[37]</sup>. As shown by Baskin *et al.*, it is possible to predict  $N$  different properties of ILs at the same condition simultaneously or the same property under  $N$  different conditions (e.g., temperature) simultaneously based on multi-task learning (MTL) models<sup>[37]</sup>. From this MTL model, it was found that the accuracy, unfortunately, is low for viscosity prediction ( $R^2 = 0.69$ ,  $RMSE = 0.40$ ,  $MAE = 0.28$ ) [Table 1] when making a prediction for five physical properties (i.e., electrical conductance, density, surface tension, viscosity, and refractive index) at  $T = 298$  K during the test prediction. Whereas for MTL models in making a simultaneous prediction for viscosities at different temperatures within the range of  $T = 288$ – $343$  K, the performance metrics were found to be nearly similar (i.e.,  $R^2 = 0.674$ ,  $RMSE = 0.375$ ,  $MAE = 0.265$ ) [Table 1] during the cross-validation test prediction. Thus, this suggests that while more robust than a single-task QSPR model that only manages to predict one physical property at room temperature, how to systematically improve the accuracy of an MTL model capable of providing simultaneous predictions of several room-temperature properties or predicting a property at several different temperatures remains a challenge.

In addition to QSPR models adopted by Beckner *et al.*, one of the best approaches reported in the literature was based on the GC method developed by Paduszyński *et al.*<sup>[33,34,36]</sup>. Based on the GC method, the complex functional groups used to describe a diverse family of ILs can be represented as their cations and anions fragmented into a set of predefined molecular groups or fragments. And each group has a direct influence on the property (e.g.,  $\eta(T) = \eta_0 f(T)$  where  $\eta_0$  denotes the viscosity at a reference temperature,  $T_0 = 298.15$  K) value as a function of temperature ( $f(T)$ ), and can be combined with ML algorithms (e.g., FFANN; stepwise multiple linear regression, LSSVM) for a systematic study. Based on an extensive IL dataset, the FFANN is

found to be the best ML model<sup>[34]</sup> to predict the viscosity at room temperature ( $\eta_0$ ) with test accuracy of  $R^2 \sim 0.8637$  and  $RMSE \sim 0.51$  mPa·s. Based on the LSSVM best fitted  $f(T)$  with  $R^2 \sim 0.97$ , the overall prediction of  $\eta(T)$  with accuracy of  $R^2 \sim 0.91$ ,  $RMSE \sim 203$  mPa·s [Table 1] can be obtained modeled with FFANN and LSSVM. As noted by Padaszyński *et al.*, the large  $RMSE$  is attributed to a wide range of IL datasets across wide temperature ranges ( $T \sim 290$ -410 K)<sup>[34]</sup>.

To improve upon the GC method inspired by Padaszyński *et al.*<sup>[33,55-57]</sup>, a newly improved GC fragmentation scheme has been proposed by Boualem *et al.* recently<sup>[38]</sup>. This new GC scheme is capable of defining a large variety of ILs and is able to differentiate among the isomers. According to this scheme, the molecular structure of each IL can be considered as a collection of three separate groups, i.e., cationic, anionic, and substituent groups. The cationic and anionic groups are the constituent groups, representing charged components/fragments of ions, whereas the substituent groups are those representing neutral components/fragments of various types of side chains<sup>[33,38,58]</sup>. With this new approach, it is possible to overcome the limitation related to structural representation that is not sufficiently described in a conventional GC scheme<sup>[33,34]</sup>. With this new fragmentation scheme combined with SVM, the overall accuracy of the test prediction of  $\eta(T)$  reported by Boualem *et al.* was impressive, i.e.,  $R^2 \sim 0.9859$  with  $RMSE \sim 57.92$  mPa·s, despite being the largest dataset considered to date [Table 1]<sup>[38]</sup>. Thus, this suggests that besides the ML model itself, a systematic classification and robust identification of molecular structural features of ILs are also critical factors in improving the accuracy of the model.

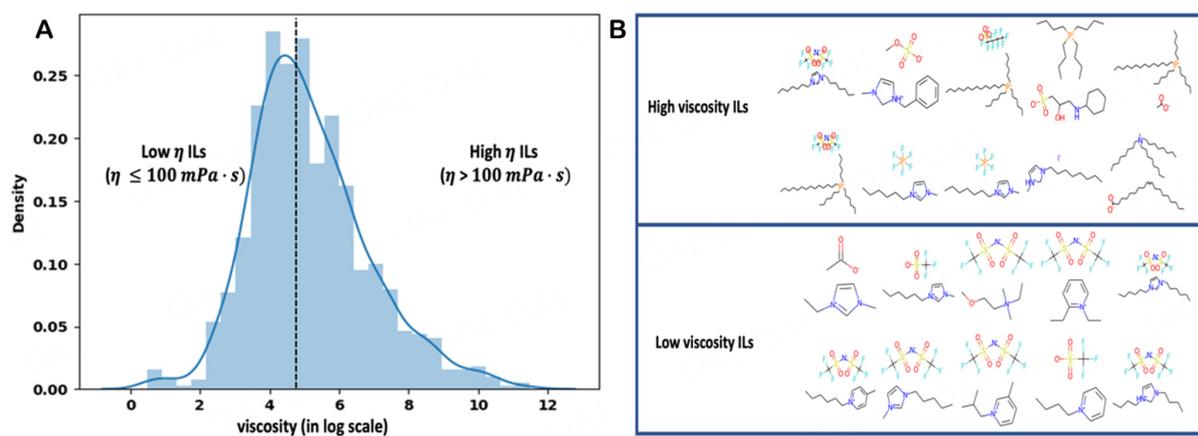
#### Low- and high-viscosity binary classification prediction from deep-learning model

In addition to the room-temperature viscosity prediction (Section "Viscosity prediction from deep-learning (DL) models"), it is also necessary to investigate the important structure-property relationship that determines the low- or high- $\eta$  of various ILs at room temperature based on a robust binary classification of ML models. It is noteworthy to point out that the extremely low- $\eta$  ( $< 2$ -3 mPa·s) ILs at room temperature comparable to traditional solvents (e.g., water, ethanol) are very rare in the ILThermo (v2.0.) database<sup>[45,46]</sup>, i.e., only six candidates (i.e., from 2-methylpyridinium acetate to N-methyl-2-oxopyrrolidinium butanoate) out of 922 types of ILs are found in this work. Whereas for the ILs below the range of  $\eta \sim 10$  mPa·s at room temperature, only about 19 types of ILs can be found (e.g., 2-methylpyridinium acetate, trihexylammonium hexanoate, 1-ethyl-3-methyl-1H-imidazolium tricyanomethanide, *etc.*). The very limited dataset makes ML model prediction, especially for the extremely low- $\eta$  ILs ( $< 2$ -3 mPa·s or  $\leq 10$  mPa·s), becomes very difficult. According to a recent comprehensive review of room-temperature viscosities of ILs by Jiang *et al.*, typical room-temperature low-to-medium viscosities of ILs were commonly found below 100 mPa·s, whereas for high- $\eta$ , the  $\eta$  is generally larger than 100 mPa·s<sup>[18]</sup>. In addition, it is also found that the most commonly used low- $\eta$  ILs in metal-ion batteries at room temperature are within the range of  $\eta \sim 19$ -156 mPa·s<sup>[59]</sup>. With this as a reference<sup>[18,59]</sup>, all the ILs studied in this work can be classified into two groups, i.e., low- $\eta$  ILs ( $\eta \leq 100$  mPa·s) and high- $\eta$  ILs ( $\eta > 100$  mPa·s), accordingly across a wide range of viscosities in our dataset [Figure 4A], which spans from two to 97,000 mPa·s based on various combinations of anionic and cationic families of ILs [Figure 4B]. Based on this low/high- $\eta$  binary classification, there are 540 high- $\eta$  and 382 low- $\eta$  ILs can be found in the dataset.

To further identify the hidden correlation of the structure-property relationship of these low- $\eta$  and high- $\eta$  ILs, a DL model based on a DNN model is constructed [Supplementary Figure 2], which is similar to the DNN model discussed in Section "Machine-learning model: deep-learning". Based on the 179 selected molecular descriptors ( $M = 179$  in Table 2), similar to our aforementioned viscosity prediction model (Section "Viscosity prediction from deep-learning (DL) models"), the overall accuracy of the low/high- $\eta$  binary classification reaches 93% for test prediction, even with a skewed dataset (i.e.,  $\sim 58.6\%$  high- $\eta$  ILs relative to  $\sim 41.4\%$  low- $\eta$  ILs). To better evaluate the performance of this DNN model in low/high- $\eta$  binary

**Table 2. The DNN model performance metrics for test predictions of low/high- $\eta$  binary classification for 922 types of ILs obtained in this work with different numbers ( $M$ ) of molecular descriptors for each IL. In this table, the negative case is low- $\eta$ , while the positive case is high- $\eta$ . The dataset of low/high- $\eta$  binary classification for 922 types of ILs with the top 20 molecular descriptors can be found in File2.csv in the Supplementary Material**

$M = 179$	precision	recall	$M = 20$	precision	recall
low- $\eta$	0.91	0.94	low- $\eta$	0.92	0.86
high- $\eta$	0.95	0.92	high- $\eta$	0.89	0.94



**Figure 4.** (A) The schematic plot of the room-temperature viscosity distribution (in log(mPa·s) scale) that classifies the low- $\eta$  ( $\eta \leq 100$  mPa·s) and high- $\eta$  ( $\eta > 100$  mPa·s) ILs from various types in the dataset. (B) A few randomly selected ILs that represent low- $\eta$  and high- $\eta$  ILs in the dataset.

classification, the *positive precision* (positive predictive value, *PPV*) (i.e., the accuracy of positive prediction/high- $\eta$  prediction), *negative precision* (negative predictive value, *NPV*) (i.e., the accuracy of negative prediction/low- $\eta$  prediction), *recall for sensitivity* of positive class (i.e., true positive rate, *TPR*), and *recall for specificity* of negative class (i.e., true negative rate, *TNR*) are used. These metrics are defined in the following equations<sup>[60]</sup>:

*positive precision* =  $\frac{TP}{TP+FP}$ , where *TP* is the number of true positives (i.e., true high- $\eta$ ), and *FP* is the number of false positives (i.e., false high- $\eta$ )

*negative precision* =  $\frac{TN}{TN+FN}$ , where *TN* is the number of true negatives (i.e., true low- $\eta$ ), and *FN* is the number of false negatives (i.e., false low- $\eta$ )

*recall (sensitivity)* =  $\frac{TP}{TP+FN}$ , where *FN* is the number of false negatives (i.e., false low- $\eta$ ).

*recall (specificity)* =  $\frac{TN}{TN+FP}$ , where *FP* is the number of false positives (i.e., false high- $\eta$ ).

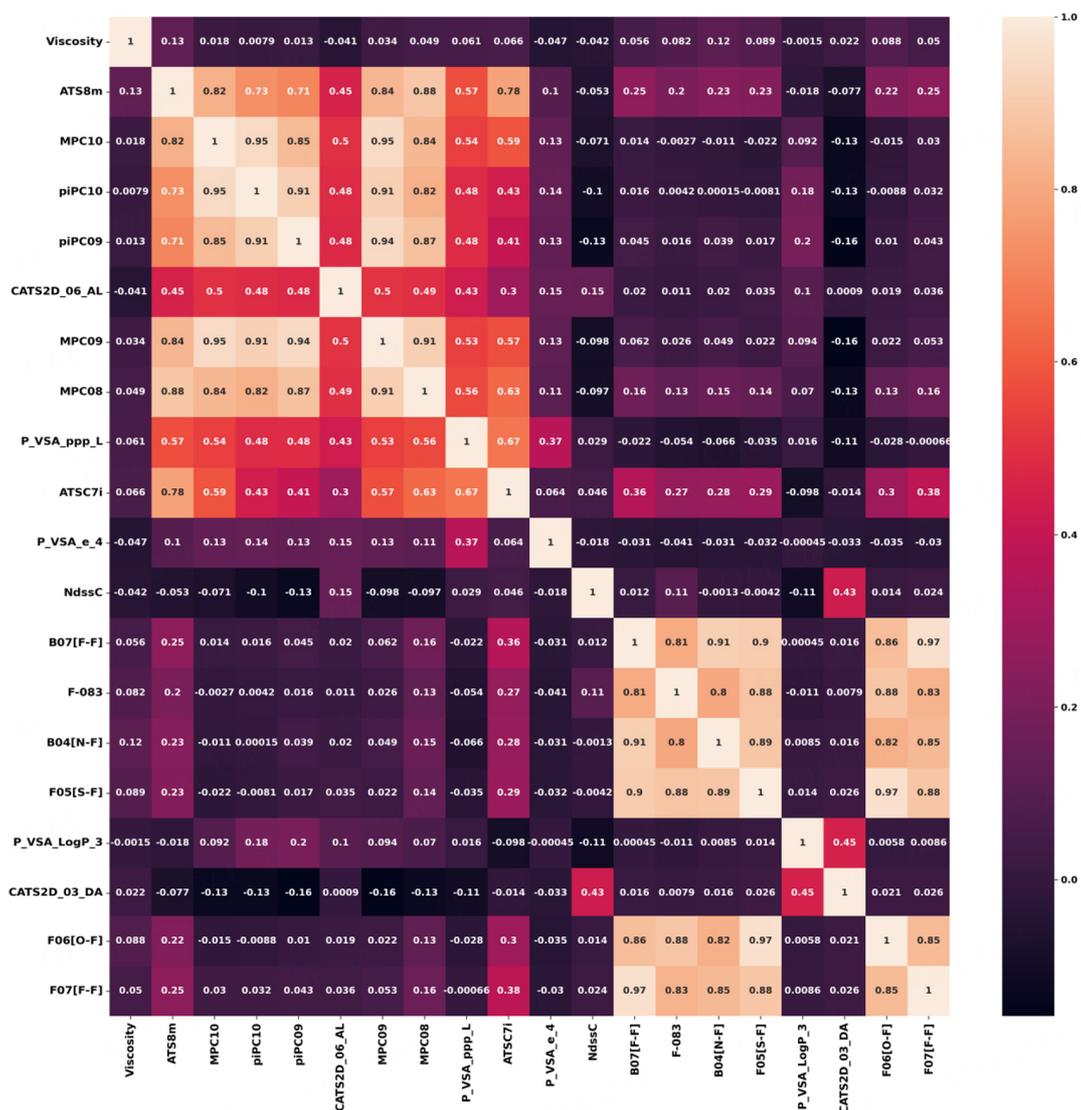
Based on the computed confusion matrix [Supplementary Figure 3], all the values of *TP*, *TN*, *FP*, and *FN* can be obtained. Based on the test sample (i.e., 20% of dataset), the overall performance metrics of DNN models [Supplementary Figure 2] in low/high- $\eta$  binary classification are shown in Table 2. Despite the skewed dataset, a balanced accuracy in both negative (*NPV* ~ 91%) and positive precision (*PPV* ~ 95%) was

found [Table 2] based on our DNN model. For positive (i.e., high- $\eta$ ) prediction, the *TPR* of 92% is slightly lower than *TNR* (i.e., 94%), even though a larger sample size in positive cases (high- $\eta$ ). To further examine the influence of the number of molecular descriptors on the DNN model performance in low/high- $\eta$  binary classification, the top-ranking important molecular descriptors were computed. From the feature filtering and ranking scores [Figure 5] of DNN models, the important correlation among the top 20 molecular descriptors that are important in low/high- $\eta$  binary classification of ILs was obtained using Pearson correlation. Based on the DNN model with only the top 20 molecular descriptors [Figure 5], the overall accuracy of the low/high- $\eta$  binary classification is 90 % for test prediction and is slightly lower than the 179 selected descriptor-based test accuracy, i.e., 93%. As shown in Table 2, the overall performance metrics (i.e., *PPV* = 89%, *NPV* = 92%, *TPR* = 94%, *TNR* = 86%) of the DNN model with  $M = 20$  are also generally lower than the accuracy obtained with  $M = 179$ . This suggests that the top 20 important molecular descriptors might not be sufficient to achieve the high accuracy anticipated in low/high- $\eta$  binary classification prediction according to our current DNN model.

### Important molecular descriptors that determine viscosities of ILs

According to the DNN model, the top 20 important molecular descriptors that determine the low/high- $\eta$  of ILs are highlighted in both Figure 5 and Supplementary Table 2. From Supplementary Tables 1 and 2, some common important molecular descriptors of ILs can be found in the findings of viscosity prediction [Supplementary Table 1] and low/high- $\eta$  binary classification [Supplementary Table 2] using the DNN model, e.g., MPC10, piPC09, NdssC, etc., defined by cheminformatics molecular descriptors<sup>[48,49]</sup> of ILs. From these top ranked molecular descriptors [Supplementary Tables 1 and 2], the significant influences of both cations and anions of ILs are found to be closely related to the geometrical structures, shapes, or branching characters (e.g., piPC10, piPC09, MPC10). Besides, the molecular weights/sizes (e.g., ATs8m) of ILs, partial charges (e.g. ATsC7i), and functional groups and bonds (e.g., NdssC, B02[C-S], B04[N-S], B07[F-F], F05[S-F]) features of ILs are also found to be important factors that could possibly affect their viscosity due to the strong correlation of local bonding with molecular dipole/polarizability and electrostatic interaction (e.g., cation-cation, anion-anion, cation-anion interaction). In addition, the important structure features governing the intermolecular, hydrogen bonding, and van der Waals interactions of ILs can be captured based on the P-VSA and CATS2D-based molecular descriptors (e.g., P\_VSA\_ppp\_L, CATS2D\_03DA, CATS2D\_07\_DL) [Supplementary Tables 1 and 2].

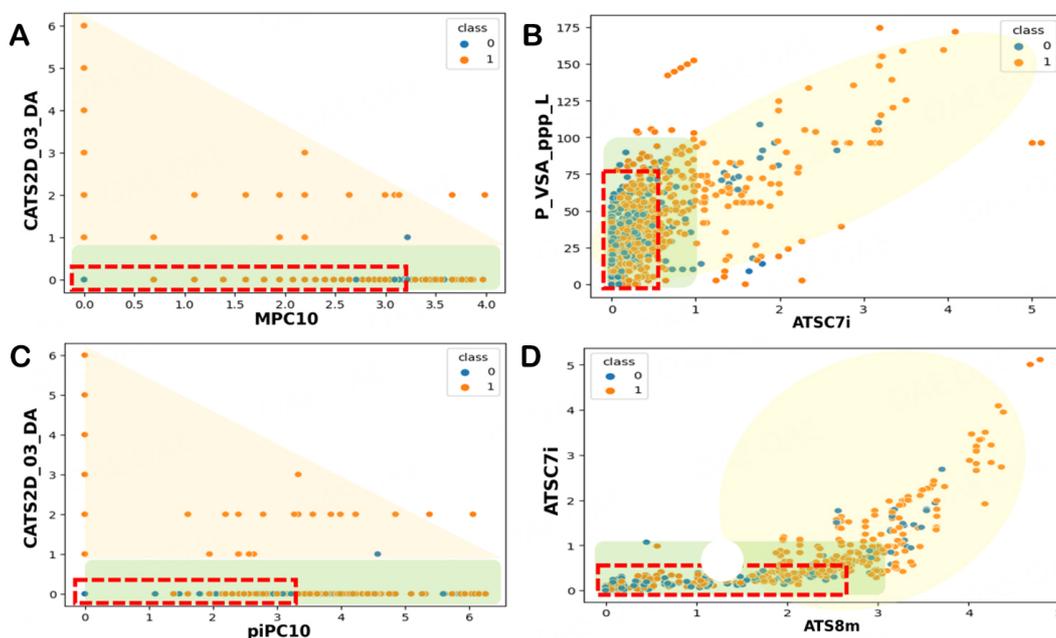
To improve our basic understanding on how these top ranked molecular descriptors [Supplementary Table 2] determine the low/high- $\eta$  of ILs, the distribution of low- and high- $\eta$  ILs was analyzed based on the combination of two important descriptors selected from Supplementary Table 2. From the analysis of the distribution of low/high- $\eta$  ILs using some of the best combinations of two top ranked molecular descriptors from Figure 5, we believe useful insights regarding design rules of IL viscosities can be found. In Figure 6, the unique interplay governed by the inherent structural features of the cations and anions of ILs that determine the distribution of low- and high- $\eta$  ILs is obvious. From Figure 6, an opposite trend between the low- and high- $\eta$  ILs represented by these molecular descriptors is evident. As shown in Figure 6A and C, the majority of low- $\eta$  ILs favor the regime of smaller values in CATS2D\_03\_DA (i.e., CATS2D\_03\_DA = 0), MPC10 (i.e., MPC10  $\leq$  3.25) or piPC10 (i.e., piPC10  $\leq$  3). From the cheminformatics model<sup>[48,49,61]</sup>, the CATS2D\_03\_DA is belonged to a CATS2D-based descriptor. It is a topological descriptor quantifying the number of hydrogen bond donor-acceptors (i.e., an integer) at a given topological distance (i.e., lag three) of an IL within the hydrogen-depleted molecular graph<sup>[48,49,61]</sup>. As highlighted in Figure 6A and C, nearly all the low- $\eta$  ILs ( $\eta \leq 100$  mPa·s) are found to have zero value (“0”) in CATS2D\_03\_DA values, whereas there is a substantial amount of high- $\eta$  ( $\eta > 100$  mPa·s) ILs are found to have nonzero (i.e., 1-6) values, despite some with zero values. Thus, this suggests that the majority of low- $\eta$  ILs tend to have a minimal number of hydrogen bond donor-acceptors, whereas the high- $\eta$  ILs tend to have



**Figure 5.** A Pearson correlation matrix comparing the viscosity with the top 20 important molecular descriptors for low/high- $\eta$  binary classification of ILs based on DNN models. The numbers shown in the Pearson correlation matrix are the feature correlation coefficients among different molecular descriptors. A brief description of these molecular descriptors can be found in [Supplementary Table 2](#).

a larger number of hydrogen bond donor-acceptors (e.g., 2-hydroxy-N-(2-hydroxyethyl)-N-methylethanaminium phosphate, 1-(2-hydroxyethyl)-3-methylimidazolium L-serinate, etc.).

Whereas for MPC10, it is a molecular path count (MPC) of order ten topological descriptors counting the total number of molecule paths of length  $m$  (in this case is ten). The length  $m$  of the path is the number of edges along the molecular path and is related to path order in an IL<sup>[48,49,62]</sup>. This is complementary to piPC10, which quantifies multiple path counts. piPC10 is a descriptor that can capture bond order features (e.g., aromatic bonds) and belongs to the path count descriptor group<sup>[48,49,63]</sup>. It is a count of molecular graph weighted paths of a given length (in this case is ten) in the molecular path, where each path is weighted by the product of the conventional bond order of the involved edges and, therefore, can account for multiple bonds in an IL<sup>[48,49,63]</sup>. Large values in MPC10 imply large values in molecular branching or presence of long chain branches (e.g., 17-hydroxy-N-(17-hydroxy-3,6,9,12,15-pentaoxaheptadec-1-yl)-N-methyl-N-



**Figure 6.** The distribution of the IL viscosity that distinguishes the low- $\eta$  (represented as 0) and high- $\eta$  (represented as 1) using the combination of two important molecular descriptors selected from Pearson correlation matrix: (A) CATS2D\_03\_DA vs. MPC10; (B) P\_VSA\_ppp\_L vs. ATSC7i; (C) CATS2D\_03\_DA vs. piPC10; and (D) ATSC7i vs. ATS8m. The yellow region highlights the dominance of high- $\eta$  ILs, whereas the green region highlights the prevalence of low- $\eta$  ILs in the distribution. For the ILs with very low  $\eta$  ( $\eta < 30$  mPa-s) that could be potentially useful in battery electrolytes, they are mostly found in the red dotted region. For the complete list of these potentially useful ILs ( $\eta < 30$  mPa-s), it can be found in [File1.csv](#) in the [Supplementary Material](#).

tetradecyl-3,6,9,12,15-pentaoxaheptadecan-1-aminium methyl sulfate) in ILs. Whereas larger values in piPC10 imply significant presence of multiple bonds (e.g., single, double, aromatic bonds) in molecular branches of ILs (e.g., L-phenylalanine benzyl ester bis(perfluoroethylsulfonyl)imide). For those ILs consisting of anionic and cationic species, this suggests that despite large molecular branching in IL cations, a reduced chain length or a minimal number of hydrogen bonds acceptor-donor in IL anions (e.g., low viscous trioctylammonium butanoate with  $\eta \sim 13$  mPa-s, MPC10 = 3.1, CATS2D\_03\_DA = 0 in [Figure 6A](#)) helps reduce the IL viscosities, analogous to dicyanamide-based ILs<sup>[18,64]</sup>.

The P-VSA-like descriptors are the molecular descriptors that define the amount of van der Waals surface area (VSA) having a property in a certain range<sup>[65]</sup>. According to P-VSA-based models<sup>[48,49,65]</sup>, P\_VSA\_ppp\_L is the descriptor measuring the potential pharmacophore points of lipophilic, an important factor that estimates the level of lipophilicity. High-lipophilicity molecules tend to be hydrophobic or less polar, whereas for ATSC7i, it is a 2D autocorrelation descriptor [[Supplementary Table 2](#)] based on autocorrelation of a topological structure (ATS) that describes how a property is distributed along the topological structure<sup>[48,49,66]</sup>. For ATSC7i, it is a descriptor measuring the centered Broto-Moreau autocorrelation of lag seven used to weigh all the contributions of the ionization potential of each different path length (lag) in the molecular graph. Smaller values of ATSC7i imply a strong tendency in the formation of cations. Therefore, from [Figure 6B](#), it is found that the small values in both P\_VSA\_ppp\_L and ATSC7i tend to yield low- $\eta$  with highly ionic ILs (e.g., N-methyl-2-oxopyrrolidinium acetate with  $\eta \sim 2.4$  mPa-s, P\_VSA\_ppp\_L = 13.9, ATSC7i = 0 in [Figure 6B](#)). In contrast, for high- $\eta$  ILs (e.g., trihexyl(tetradecyl)phosphonium chloride with  $\eta \sim 1,400$  mPa-s, ATSC7i = 3.2, P\_VSA\_ppp\_L = 148.6), large values in both P\_VSA\_ppp\_L and ATSC7i are generally found, as shown in [Figure 6B](#). In addition, it is also interesting to observe that low- $\eta$  ILs tend to favor the regime of small values in both ATSC7i and ATS8m, as shown in [Figure 6D](#) (e.g.,

trimethylammonium hydrogen sulfate with  $\eta \sim 4.7$  mPa·s, ATSC7i = 0, ATS8m = 0). Similar to ATSC7i, ATS8m belongs to 2D autocorrelation-based descriptors and represents a molecular feature that measures the centered Broto-Moreau autocorrelation of lag eight weighted by mass<sup>[48,49,66]</sup>. Thus, ILs with small molecular sizes generally have small values in ATS8m. From Figure 6D, the trend suggests that heavier and less polar ILs with large molecular sizes tend to have high- $\eta$  (e.g., tributyl(tetradecyl)phosphonium 2-dodecylbenzenesulfonate with  $\eta \sim 1,400$  mPa·s, ATSC7i = 3.5, ATS8m = 4.0), and this observation is consistent with the observation that the presence or increase of alkyl group or alkyl chain (e.g., imidazolium cations) will increase viscosities of ILs due to increasing intermolecular van der Waals interactions<sup>[18,35]</sup>.

Thus, from the general trend of molecular descriptors for low/high- $\eta$  ILs highlighted in Figure 6A-D, a set of useful design rules to tune the IL viscosity through molecular grafting can be obtained. Specifically for low- $\eta$  ILs, small values in all these important descriptors, i.e., CATS2D\_03\_DA, MPC10, piPC10, P\_VSA\_ppp\_L, ATSC7i, and ATS8m, are generally preferred (i.e., the red dotted region in Figure 6A-D and File1.csv in the Supplementary Material). From this observation, a list of 96 IL candidates with very low viscosities (i.e., below  $\sim 30$  mPa·s), which might be potentially useful electrolytes in battery applications, can be identified [File1.csv in the Supplementary Material]. Thus, the observation [Figure 6] suggests that in order to reduce the IL viscosity, grafting IL cations into smaller sizes (e.g., smaller head rings) and short alkyl chains and reducing ionization potentials/energies will help. Meanwhile, for the same cations, further reducing anions in sizes, chain length, and hydrogen bonds [e.g., TFSI<sup>-</sup>, N(CN)<sub>2</sub><sup>-</sup>] might be useful to further help to reduce the viscosities. Thus, to fine-tune the IL viscosity, a synergistic effect to achieve an optimum design of both cations and anions in ILs is deemed necessary.

## CONCLUSIONS

ILs are a new group of solvents with great potential in design synthesis. They are promising electrolytes in energy storage applications, especially in rechargeable batteries. It is known that the variation in viscosity of ILs can lead to subtle effects in their transport properties, e.g., ionic conductivity, charge transfer rate, *etc.*, which are extremely important factors in the development of novel electrolytes for energy storage applications<sup>[14,15,59]</sup>. However, in practice, the usages of ILs remain limited due to their unfavorable viscosity property at ambient conditions. To optimize their design synthesis, a systematic fundamental study of structure-property relationships in ILs is deemed necessary. With this as our motivation, we pursued a baseline study to investigate the trend of room-temperature viscosities for various types of ILs. To search for important insights that will lead us to fine-tune the viscosity of ILs, an integrated approach that combines high-throughput screenings of a large IL dataset, chemoinformatics, and DL models is adopted in this work.

Based on the dataset obtained from ILThermo (v2.0)<sup>[43,44]</sup>, we have constructed a robust DL model (Section "Viscosity prediction from deep-learning (DL) models") to predict the viscosity of ILs at ambient conditions with high accuracy. We achieved  $R^2$  scores of  $\sim 0.9869$  ( $RMSE \sim 63.78$  mPa·s) and  $0.9980$  ( $RMSE \sim 45.27$  mPa·s) using chemoinformatics descriptors based on DNN and CNN models separately. To help to identify the low- $\eta$  (i.e.,  $\eta \leq 100$  mPa·s) and high- $\eta$  (i.e.,  $\eta > 100$  mPa·s) among various ILs over a wide range of viscosities spanning from twp to 97,000 mPa·s (based on various combinations of cationic and anionic families of ILs), a low/high- $\eta$  binary classification DNN model (Section "Low- and high-viscosity binary classification prediction from deep-learning model") has been developed. Based on this low/high- $\eta$  binary classification model, 540 high- $\eta$  and 382 low- $\eta$  ILs can be found in the dataset (i.e., 922 types of ILs). Despite this skewed dataset (i.e., 59% high- $\eta$  vs. 41% low- $\eta$ ), a balanced accuracy in both low- $\eta$  and high- $\eta$  prediction with an overall accuracy of 93 % for test prediction is achieved.

Despite a large deviation in viscosity distributions and a huge variety of anion-cation combinations presented in the IL dataset, we found the viscosities of various ILs can be determined based on a limited number of important molecular descriptors. From the top ranked molecular features/descriptors obtained from DL models (Section "Low- and high-viscosity binary classification prediction from deep-learning model"), the molecular features, i.e., the geometrical structures, shapes or branching characters, constitutional molecular weights or sizes, partial charges, functional groups, local bonds, hydrogen bonding, and van der Waals interactions, related to both anions and cations of ILs are found equally important. Similarly, based on the DL model, an important structure-property relationship governed by a set of important molecular descriptors (e.g., ATSC7i, ATSC8m, MPC10, P\_VSA\_ppp\_L, etc. in [Figures 5 and 6](#)) of the cations and anions of ILs in defining the low/high- $\eta$  ILs can be found (Section "Important molecular descriptors that determine viscosities of ILs"). The analysis of DL model prediction suggests that in order to reduce the IL viscosity, grafting IL cations into smaller sizes (e.g., smaller head rings) and short alkyl chains and also reducing ionization potentials/energies will help. Meanwhile, for the same cations, further reducing anions in size, chain length, and hydrogen bonds (e.g., TFSI,  $N(CN)_2^-$ ) might be useful to further help reduce the viscosity of ILs. This is supported by a list of potentially useful ILs with very low  $\eta$  (i.e.,  $\eta < 30$  mPa·s) we found in this study [[File1.csv](#) in the [Supplementary Material](#)].

Thus, with a fine selection and molecular grafting of anionic and cationic species in ILs, we believe the design synthesis of appropriate molecular functional groups in ILs is vital to fine-tune the viscosity for potentially useful electrolytes. However, in order to obtain robust ILs for battery applications in practice, simply fine-tuning IL viscosity is insufficient. To identify robust IL-based electrolytes, considering a synergistic effect of achieving an optimum design of both cations and anions in ILs to simultaneously fulfill other important physicochemical properties (e.g., redox stability, salt concentration, ionic conductivity, melting/boiling points, thermal conductivity, heat capacity) relevant to specific energy storage applications, further development of more sophisticated multi-modal DL models will be necessary.

## DECLARATIONS

### Authors' contributions

Conceptualization; supervision; project administration; funding acquisition: Lau KC

Methodology: Acar Z, Nguyen P, Cui X, Lau KC

Formal analysis: Acar Z, Cui X, Lau KC

Data curation: Nguyen P

Writing-original draft preparation; writing-review and editing: Cui X, Lau KC

Visualization: Acar Z, Nguyen P, Lau KC

### Availability of data and materials

The data supporting our work can be found in the [Supplementary Material](#).

### Financial support and sponsorship

This work was supported by the U.S. Research Corporation for Science Advancement (RCSA) through the Cottrell Scholar Award (Award No. 26829).

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Copyright

© The Author(s) 2023.

## REFERENCES

1. Goodenough JB. Electrochemical energy storage in a sustainable modern society. *Energy Environ Sci* 2014;7:14-8. DOI
2. Trahey L, Brushett FR, Balsara NP, et al. Energy storage emerging: a perspective from the joint center for energy storage research. *Proc Natl Acad Sci USA* 2020;117:12550-7. DOI PubMed PMC
3. Aurbach D, Markevich E, Salitra G. High energy density rechargeable batteries based on Li metal anodes. the role of unique surface chemistry developed in solutions containing fluorinated organic co-solvents. *J Am Chem Soc* 2021;143:21161-76. DOI
4. Li Q, Chen J, Fan L, Kong X, Lu Y. Progress in electrolytes for rechargeable Li-based batteries and beyond. *Green Energy Environ* 2016;1:18-42. DOI
5. Lu J, Lee YJ, Luo X, et al. A lithium-oxygen battery based on lithium superoxide. *Nature* 2016;529:377-82. DOI
6. Ponnada S, Kiai MS, Gorle DB, Nowduri A. History and recent developments in divergent electrolytes towards high-efficiency lithium-sulfur batteries - a review. *Mater Adv* 2021;2:4115-39. DOI
7. Watanabe M, Thomas ML, Zhang S, Ueno K, Yasuda T, Dokko K. Application of ionic liquids to energy storage and conversion materials and devices. *Chem Rev* 2017;117:7190-239. DOI PubMed
8. Balducci A. Ionic liquids in lithium-ion batteries. *Top Curr Chem* 2017;375:20. DOI PubMed
9. Zhang J, Sun B, Zhao Y, et al. A versatile functionalized ionic liquid to boost the solution-mediated performances of lithium-oxygen batteries. *Nat Commun* 2019;10:602. DOI PubMed PMC
10. Josef E, Yan Y, Stan MC, et al. Ionic liquids and their polymers in lithium-sulfur batteries. *Isr J Chem* 2019;59:832-42. DOI
11. Ortiz-martínez V, Gómez-coma L, Pérez G, Ortiz A, Ortiz I. The roles of ionic liquids as new electrolytes in redox flow batteries. *Sep Purif Technol* 2020;252:117436. DOI
12. Giffin GA. The role of concentration in electrolyte solutions for non-aqueous lithium-based batteries. *Nat Commun* 2022;13:5250. DOI PubMed PMC
13. Cao X. Important factors for the reliable and reproducible preparation of non-aqueous electrolyte solutions for lithium batteries. *Commun Mater* 2023;4:10. DOI
14. Martin S, Pratt HD 3rd, Anderson TM. Screening for high conductivity/low viscosity ionic liquids using product descriptors. *Mol Inform* 2017;36:1600125. DOI PubMed
15. Tiago GAO, Matias IAS, Ribeiro APC, Martins LMDRS. Application of ionic liquids in electrochemistry-recent advances. *Molecules* 2020;25:5812. DOI PubMed PMC
16. Barthen P, Frank W, Ignatiev N. Development of low viscous ionic liquids: the dependence of the viscosity on the mass of the ions. *Ionics* 2015;21:149-59. DOI
17. Tsuzuki S, Shinoda W, Saito H, Mikami M, Tokuda H, Watanabe M. Molecular dynamics simulations of ionic liquids: cation and anion dependence of self-diffusion coefficients of ions. *J Phys Chem B* 2009;113:10641-9. DOI PubMed
18. Jiang S, Hu Y, Wang Y, Wang X. Viscosity of typical room-temperature ionic liquids: a critical review. *J Phys Chem Ref Data* 2019;48:033101. DOI
19. Philippi F, Rauber D, Eliassen KL, et al. Pressing matter: why are ionic liquids so viscous? *Chem Sci* 2022;13:2735-43. DOI PubMed PMC
20. Koutsoukos S, Philippi F, Malaret F, Welton T. A review on machine learning algorithms for the ionic liquid chemical space. *Chem Sci* 2021;12:6820-43. DOI PubMed PMC
21. Hayes R, Warr GG, Atkin R. Structure and nanostructure in ionic liquids. *Chem Rev* 2015;115:6357-426. DOI PubMed
22. Marullo S, D'Anna F, Rizzo C, Billeci F. Ionic liquids: "normal" solvents or nanostructured fluids? *Org Biomol Chem* 2021;19:2076-95. DOI PubMed
23. Lei Z, Chen B, Koo YM, MacFarlane DR. Introduction: ionic liquids. *Chem Rev* 2017;117:6633-5. DOI PubMed
24. Diogo JCF, Caetano FJP, Fareleira JMNA, Wakeham WA. Viscosity measurements on ionic liquids: a cautionary tale. *Int J Thermophys* 2014;35:1615-35. DOI
25. Bedrov D, Piquemal JP, Borodin O, MacKerell AD Jr, Roux B, Schröder C. Molecular dynamics simulations of ionic liquids and electrolytes using polarizable force fields. *Chem Rev* 2019;119:7940-95. DOI PubMed PMC
26. Zhang Y, Otani A, Maginn EJ. Reliable viscosity calculation from equilibrium molecular dynamics simulations: a time decomposition method. *J Chem Theory Comput* 2015;11:3537-46. DOI
27. Kirova EM, Norman GE. Viscosity calculations at molecular dynamics simulations. *J Phys Conf Ser* 2015;653:012106. DOI
28. Goloviznina K, Canongia Lopes JN, Costa Gomes M, Pádua AAH. Transferable, polarizable force field for ionic liquids. *J Chem Theory Comput* 2019;15:5858-71. DOI PubMed
29. Vázquez-Montelongo EA, Vázquez-Cervantes JE, Cisneros GA. Current status of AMOEBA-IL: a multipolar/polarizable force field

- for ionic liquids. *Int J Mol Sci* 2020;21:697. DOI PubMed PMC
30. Katritzky AR, Jain R, Lomaka A, et al. Correlation of the melting points of potential ionic liquids (imidazolium bromides and benzimidazolium bromides) using the CODESSA program. *J Chem Inf Comput Sci* 2002;42:225-31. DOI
  31. Berrod Q, Ferdeghini F, Zanotti JM, et al. Ionic liquids: evidence of the viscosity scale-dependence. *Sci Rep* 2017;7:2241. DOI PubMed PMC
  32. Chen Y, Peng B, Kontogeorgis GM, Liang X. Machine learning for the prediction of viscosity of ionic liquid-water mixtures. *J Mol Liq* 2022;350:118546. DOI
  33. Padaszyński K, Domańska U. Viscosity of ionic liquids: an extensive database and a new group contribution model based on a feed-forward artificial neural network. *J Chem Inf Model* 2014;54:1311-24. DOI
  34. Padaszyński K. Extensive databases and group contribution QSPRs of ionic liquids properties. 2. viscosity. *Ind Eng Chem Res* 2019;58:17049-66. DOI
  35. Chen B, Liang M, Wu T, Wang HP. A high correlate and simplified QSPR for viscosity of imidazolium-based ionic liquids. *Fluid Phase Equilibria* 2013;350:37-42. DOI
  36. Beckner W, Mao CM, Pfaendtner J. Statistical models are able to predict ionic liquid viscosity across a wide range of chemical functionalities and experimental conditions. *Mol Syst Des Eng* 2018;3:253-63. DOI
  37. Baskin I, Epshtein A, Ein-eli Y. Benchmarking machine learning methods for modeling physical properties of ionic liquids. *J Mol Liq* 2022;351:118616. DOI
  38. Boualem AD, Argoub K, Benkouider AM, Yahiaoui A, Toubal K. Viscosity prediction of ionic liquids using NLR and SVM approaches. *J Mol Liq* 2022;368:120610. DOI
  39. Goodfellow I, Bengio Y, Courville A. Deep learning, 1st ed. Cambridge, MA: MIT Press; 2016, pp. 363-405.
  40. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85-117. DOI PubMed
  41. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput Sci* 2021;2:420. DOI PubMed PMC
  42. Sejnowski TJ. The unreasonable effectiveness of deep learning in artificial intelligence. *Proc Natl Acad Sci USA* 2020;117:30033-8. DOI PubMed PMC
  43. Liu Y, Zhao T, Ju W, Shi S. Materials discovery and design using machine learning. *J Materiomics* 2017;3:159-77. DOI
  44. Liu Y, Guo B, Zou X, Li Y, Shi S. Machine learning assisted materials design and discovery for rechargeable batteries. *Energy Stor Mater* 2020;31:434-50. DOI
  45. Dong Q, Kazakov A, Muzny C, et al. Ionic liquids database - ILThermo. 2006. Available online: <https://ilthermo.boulder.nist.gov/ILThermo/mainmenu.uix> [Last accessed on 16 August 2023].
  46. Kazakov A, Magee J, Chirico R, et al. Ionic liquids database - ILThermo (v2.0). 2013. Available online: <https://trcsrv1.boulder.nist.gov/ilthermo/ilthermo.html> [Last accessed on 16 August 2023].
  47. Acar Z, Nguyen P, Lau KC. Machine-learning model prediction of ionic liquids melting points. *Appl Sci* 2022;12:2408. DOI
  48. Talet srl dragon. Version 7.0 software for molecular descriptor calculation. Available online: <https://chm.kode-solutions.net/pf/dragon-7-0/> [Last accessed on 16 August 2023].
  49. Todeschini R, Consonni V. Molecular descriptors for chemoinformatics, 1st ed. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co; 2009.
  50. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825-30. Available online: <https://github.com/scikit> [Last accessed on 16 August 2023]
  51. Gutman I, Milovanović E, Milovanović I. Beyond the zagreb indices. *AKCE Int J Graphs Comb* 2020;17:74-85. DOI
  52. Bertz SH. Branching in graphs and molecules. *Discret Appl Math* 1988;19:65-83. DOI
  53. Abadi M, Agarwal A, Barham P, et al. TensorFlow: large-scale machine learning on heterogeneous systems. 2015. Available online: <https://research.google/pubs/pub45166/> [Last accessed on 16 August 2023].
  54. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv* 2017:1412.6980. DOI
  55. Chen Y, Kontogeorgis GM, Woodley JM. Group contribution based estimation method for properties of ionic liquids. *Ind Eng Chem Res* 2019;58:4277-92. DOI
  56. Baghban A, Kardani MN, Habibzadeh S. Prediction viscosity of ionic liquids using a hybrid LSSVM and group contribution method. *J Mol Liq* 2017;236:452-64. DOI
  57. Lazzús JA, Pulgar-villaruel G. A group contribution method to estimate the viscosity of ionic liquids at different temperatures. *J Mol Liq* 2015;209:161-8. DOI
  58. Huang Y, Dong H, Zhang X, Li C, Zhang S. A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids. *AIChE J* 2013;59:1348-59. DOI
  59. Zhou W, Zhang M, Kong X, Huang W, Zhang Q. Recent advance in ionic-liquid-based electrolytes for rechargeable metal-ion batteries. *Adv Sci* 2021;8:2004490. DOI PubMed PMC
  60. Geron A. Hands-on machine-learning with scikit-learn, Keras and TensorFlow, 2nd ed. Canada: O'Reilly; 2019.
  61. Reutlinger M, Koch CP, Reker D, et al. Chemically advanced template search (CATS) for scaffold-hopping and prospective target prediction for 'orphan' molecules. *Mol Inform* 2013;32:133-8. DOI PubMed PMC
  62. Randić M, Jerman-blazić B, Grossman S, Rouvray D. A rational approach to the optimal design of drugs. *Math Model* 1987;8:571-82. DOI

63. Randić M, Jurs PC. On a fragment approach to structure-activity correlations. *Quant Struct Act Relat* 1989;8:39-48. DOI
64. Yuan WL, Yang X, He L, Xue Y, Qin S, Tao GH. Viscosity, conductivity, and electrochemical property of dicyanamide ionic liquids. *Front Chem* 2018;6:59. DOI PubMed PMC
65. Labute P. A widely applicable set of descriptors. *J Mol Graph Model* 2000;18:464-77. DOI PubMed
66. Broto P, Moreau G, Vandicke C. Molecular structures: perception, autocorrelation descriptor and sar studies. Autocorrelation descriptor. *Eur J Med Chem* 1984;19:66-70. Available online: <https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=9624511> [Last accessed on 18 August 2023].