



Article Machine-Learning Model Prediction of Ionic Liquids Melting Points

Zafer Acar¹, Phu Nguyen² and Kah Chun Lau^{1,*}

- ¹ Department of Physics and Astronomy, California State University, Northridge, Los Angeles, CA 91330, USA; zafer.acar.334@my.csun.edu
- ² Department of Computer Science, California State University, Northridge, Los Angeles, CA 91330, USA; phu.nguyen.713@my.csun.edu
- * Correspondence: kahchun.lau@csun.edu; Tel.: +1-81-8677-7433

Abstract: Ionic liquids (ILs) have great potential for application in energy storage and conversion devices. They have been identified as promising electrolytes candidates in various battery systems. However, the practical application of many ionic liquids remains limited due to the unfavorable melting points (T_m) which constrain the operating temperatures of the batteries and exhibit unfavorable transport property. To fine tune the T_m of ILs, a systematic study and accurate prediction of T_m of ILs is highly desirable. However, the T_m of an IL can change considerably depending on the molecular structures of the anion and cation and their combination. Thus, a fine control in T_m of ILs can be challenging. In this study, we employed a deep-learning model to predict the T_m of various ILs that consist of different cation and anion classes. Based on this model, a prediction of the melting point of ILs can be made with a reasonably high accuracy, achieving an R^2 score of 0.90 with RMSE of ~32 K, and the T_m of ILs are mostly dictated by some important molecular descriptors, which can be used as a set of useful design rules to fine tune the T_m of ILs.

Keywords: ionic liquids; deep-learning; chemoinformatics; melting points



Citation: Acar, Z.; Nguyen, P.; Lau, K.C. Machine-Learning Model Prediction of Ionic Liquids Melting Points. *Appl. Sci.* **2022**, *12*, 2408. https://doi.org/10.3390/ app12052408

Academic Editors: Dongkyu Lee and Dahyun Oh

Received: 30 January 2022 Accepted: 23 February 2022 Published: 25 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

With various valuable physicochemical properties (e.g., hydrophobic/hydrophilic, ionic conductivity, thermal stability, etc.), ionic liquids (ILs) have attracted a lot of attention within the scientific community in various applications, such as energy storage, CO₂ capture, catalysis, lubricant additives, pharmaceuticals, and foods and bioproducts [1–8]. The main advantage of these novel compounds is that ILs are molten salts which consists of cations and anions with a wide variation of possible combination. Thus, the novel properties of the ILs are therefore determined by the unique molecular structures and interactions between ions. Based on the purpose of applications, the composition of ILs and their properties can be selected and fine-tuned from a large diversity of candidates of inorganic or organic cations and anions [8–11]. ILs are usually known as "magic solvents" due to their high flexibilities in designed synthesis and wide tunable molecular structures and properties.

ILs present great potential for energy storage applications, e.g., lithium-ion batteries, lithium-oxygen (Li-O₂), lithium-sulfur (Li-S) batteries and redox flow batteries (RFBs) [1,12–15] as promising supporting electrolyte materials. The solvent viscosity and melting point are critical factors to determine the device's efficiency. However, the practical application of many ionic liquids remains restricted which is attributed to the unfavorable transport property due to the high melting points [1,15–17]. To tune the melting point of ILs, a fundamental understanding at the molecular level and a systematic in-depth study of structural properties is needed. For any given chemical compound, the melting point is a determinant character in solid-to-liquid transition, dictated by several important factors, e.g., molecular structures, the configurations of atoms, ions and molecules in a crystalline structures

(e.g., symmetry, crystal packing, molecular conformation flexibility), and the important interplay of various interactions (e.g., intra-molecular bonds, electrostatics, van der Waals, hydrogen bonding) among the molecular constituents. Thus, an accurate prediction of the melting point of ILs is highly anticipated and is challenging as well.

According to the study of Katritzky et al. [18], there exist approximately 10¹⁸ combinations of ions that could lead to useful ionic liquids. However, the wide diversity of IL compounds with various complex structures and physicochemical properties makes a systematic in-depth study on IL compounds extremely challenging. Therefore, there is a need to pursue predictive computational tools to aid the experimental design and synthesis of ILs with desirable properties, such as practically low melting points (T_m) . To explore and predict melting points of a wide range of ILs, a comprehensive study adopting a rigorous thermodynamic approach that provides an accurate quantitative prediction with high computational cost [19–22] in all systems using atomistic or molecular simulation, might not be feasible in practices. To overcome this challenge, a systematic high-throughput screening and in-depth systematic analysis on a vast amount of reported ILs dataset with a specific focus on their physicochemical properties is an important baseline study. An affordable solution, in terms of the present state-of-the-art methodology, is to develop predictive modeling to estimate and predict some important physicochemical properties, such as melting point, viscosity, ionic conductivity of the ILs and their mixtures by utilizing advanced statistical learning methods over sufficiently large datasets available in studies using quantitative structure–property relationships, QSPR; or quantitative structure– activity relationships, QSAR [23,24]. Together with the continued exponential growth in available datasets from the published literature, the utilization of statistical machine learning models for the prediction of various physicochemical properties of ILs is definitely a timely approach.

To accommodate for the ever-increasing dataset from literatures, the deep-learning model [25–28] is generally known to outperform traditional machine-learning models because of its capacity to process a vast number of features to construct an effective datadriven model. Thus, the development of robust modeling tools for the high-throughput screening of a large amount of ILs data from the literature using the advanced data mining and machine-learning techniques can be very helpful to identify and solve the important material design problems related to the ILs in many applications. This novel approach will help to significantly speed up the exploration of materials, molecular design, discovery and the development process of ILs, complementary to more in-depth fundamental focus studies based on other methods, such as industrial-process modeling, thermodynamic process and atomistic simulation. For this reason, as a baseline study, we propose the adoption of a chemoinformatic approach and deep-learning model [25–28] to model and predict the melting points (T_m) of a wide range of ILs, based on the descriptors of the molecular constituents, with the aim of providing new insights to complement the available theory-driven models in the field [19–22,29,30].

2. Methods

The data used in this study were collected from the Ionic Liquids Database—ILThermo (v2.0) (https://ilthermo.boulder.nist.gov/) (accessed on 25 October 2021) [31,32]. According to the latest updates of this database, it contains 2175 types of ILs that comprises about 4200 compounds compiled from nearly 3500 published references. For pure ionic liquids alone, the database contains nearly 1800 IL systems. For the datapoints related to pure ILs, there are nearly 120,000 datapoints available. These datapoints cover various aspects in thermodynamic, thermochemical and transport properties. The melting points data we collected, which contain 1253 reported ILs, covers a large variety of ILs families. For such a large collection, predicting melting points accurately without using costly computational resources is highly desirable.

Figure 1 provides a schematic overview of the workflow we adopted in this work to predict the melting points (T_m) of ILs based on the published dataset in ILThermo database

using a machine-learning (ML) model. As an initial step, we extracted the physical and chemical properties in which we were interested, i.e., melting temperature (T_m), from the ILThermo database by utilizing pyilt2 library [33] with our in-house code written in python. For the conversion of IUPAC names of ILs to SMILES representations, the functionality in the OPSIN library [34] was used. Based on the Dragon7 software [35], there were 5272 molecular descriptors based on a quantitative structure–activity relationship (QSPR) calculated for each ionic liquid molecule in the dataset.



Figure 1. The workflow of basic methodology that adopted in this work to predict melting point (T_m) of various ionic liquids based on machine-learning (ML) model (i.e., deep-learning model).

These molecular descriptors were then subsequently analyzed using statistical and machine-learning (ML) models based on the python libraries available in scikit-learn (scikit-learn 1.0.2), TensorFlow (TensorFlow 2) and Keras (Keras 2.7) [36–38]. Prior to ML modeling, all the low variance molecular descriptors columns and those containing missing values or empty columns were excluded. To prevent overfitting, we further reduced the dimensionality of the original descriptor matrix, and the Pearson correlation matrix was used to identify a set of significant molecular descriptors that show a statistical significance with high correlation in determining the melting points of ILs. By excluding the molecular descriptors with low correlation (<0.20) and high correlation (>0.90), a set of important molecular features which consists of 137 molecular descriptors were identified and were used to fit the ML model. After the correlation features selection and normalization of molecular descriptors, the dataset was randomly divided into training (80%) and testing or validation (20%) data for the validation of the ML model's prediction.

In this study, our primary structure-property predictive machine-learning model is based on a deep learning (DL) model composed of recurrent and recursive neural networks, RNNs [25,26,39], which are a family of neural networks specialized to process sequential data. A deep learning model is a subset of machine-learning models [25–28,39] which mimics how the human brain processes information and learns based on a set of algorithms which 'learns in layers'. It involves learning through layers which allows a computer to develop a hierarchy of learning process by developing several layers of information processing states in hierarchical structures to learn and infer. A typical DL model contains multiple hidden layers including input and output layers represented by neural networks. In terms of implementing and training the DL model, it relies on parallelized matrix and tensor operations, as well as computing gradients and optimization [25–28,39]. Thus, to construct this DL model, the libraries and utilities including pre-trained models available in Keras and TensorFlow were used. The DL model based on the sequential model is represented by multiple linear stack of input layers with each layer consisting of certain number of neurons that provides training and inference features. Figure 2 shows a general structure of our DL model, which consists of one input layer, five hidden layers and one output layer, and each layer has a different number of neurons (i.e., input layer: 137 neurons, 1st hidden layer: 512 neurons, 2nd hidden layer: 512 neurons, 3rd hidden layer: 512, 4th hidden layer: 256 neurons, 5th hidden layer: 64 neurons, output layer: 1 neuron). To update all the network weights or parameters iteratively in the model training, we used an adaptive moment estimation (Adam) optimizer which is a stochastic gradient descent method used to speed up the optimization process [40]. To fit the RNN on to training sets, we set the number of iterations, i.e., epochs to be 15,000 and batch size as equal to 32 in each epoch. All calculations were carried out based on a Linux workstation with Intel i7-8700 6-core 3.70 GHz CPU and 32 GB RAM. In this work, training a DL model takes longer, i.e., ~18 h, based on the number of parameters we used (~744,000) in the DL algorithm. However, after being trained once, the model can be used repetitively. Furthermore, testing is extremely fast and takes only seconds to make the prediction.



Figure 2. The schematic plot of the deep learning (DL) model adopted in this work, which consists of one input layer, five hidden layers and one output layer.

3. Results

3.1. Clustering and Melting Points Distribution

As mentioned in the previous section, 1253 different ILs have been considered in this work. To help improve the accuracy of the subsequent ML prediction on the melting points, the similarity among ILs molecules based on the filtered 137 molecular descriptors were analyzed and grouped based on the clustering method. Thus, the entire dataset was separated into several different clusters or groups based on their similarities using k-means algorithm implemented in scikit-learn libraries [36]. With the Elbow method [41], we found that the optimal number of clusters into which the dataset may be grouped

or clustered is 5, and a condensed view of the distribution of these clusters/groups in multidimensional descriptor space can be visualized through dimensionality reduction using principal component analysis (PCA) (Figure 3a). As shown in Figure 3, the score plot for the two principal components shows significant groupings that correspond to five distinct clusters. Therefore, throughout this work, we separated our entire dataset into five clusters or groups for the training and testing in the DL-model. Among these five clusters or groups, the size of the dataset for each cluster varies and consist of 605, 186, 134, 297, and 31 different ILs separately, which consists of with various combination of cations (e.g., ammonium, imidazolium, phosphonium, pyridinium, etc.) and anions (e.g., sulfonate, phosphate, hexafluorophosphate, borate, acetate, dicyanamide, triazolide, etc.) families (Figure 3b).



Figure 3. (a) The score plot for the first two principal components (PCA1, PCA2) with respect to 5 clusters that representing 1253 ILs dataset. (b) The randomly selected representative ILs molecules from cluster 1 to 5 dataset.

From Figure 4, a wide distribution of melting points (T_m) can be found for this large diverse dataset of 1253 ILs, whose melting points vary from 30 K to 550 K. As shown, a "bi-modal"-like distribution in melting points of the entire dataset is found, which could be assigned to two apparently distinct ranges, i.e., low melting points (i.e., $T_m < 273$ K) and high melting points, i.e., $T_m > 273$ K (Figure 4a). Meanwhile, the distribution of the melting points among these five clusters (i.e., cluster 1–5) is different, and only cluster 1 has a similar T_m distribution to the entire dataset which consists of both low and high T_m ILs candidates (Figure 4a,b). In cluster 1, a wide variation of T_m for ILs can be found (i.e., from 30 K to 550 K) with a median $T_m \sim 228$ K. For cluster 2–5, the distribution of the melting points mostly consists of high T_m ILs candidates, which generally yield a higher median T_m , i.e., 317 K to 333 K (Figure 4).



Figure 4. The distribution of melting points based on the (**a**) total 1253 ILs dataset, (**b**–**f**) cluster/group 1–5 ILs dataset that found from k-means clustering analysis. The orange dotted line highlights a boundary with T = 273 K that distinguish the low melting ($T_m < 273$ K) and high melting point ($T_m > 273$ K) ILs system. The median T_m from distribution of (**a**–**f**) is 279 K, 228 K, 322 K, 333 K, 317 K, 328 K.

3.2. Deep-Learning (DL) Model Performance

To quantify the accuracy of the prediction from the deep-learning (DL) model we described in Sect. 2, the model performance was assessed based on two metrics such as the square coefficient of correlation (R^2), and the root mean square error (RMSE), which estimates statistical accuracies in the predictions. A summary of the model's performance is presented in Table 1. The performance metrics (e.g., R^2 , RMSE) for the test sets among different clusters/groups or the total dataset were found to be sensitive to the size of the individual dataset (Table S1). As shown in Table S1, the DL-model is not suitable for small datasets since the R^2 score is only ~0.60. This suggests that a sample size that is too small in terms of training data will result in poor performance. Compared to the individual small clusters, the R^2 score of the entire dataset is significantly better (i.e., $R^2 \sim 0.90$) and might be attributed to the larger size in the training dataset (i.e., with sample size, N = 1253), which contains a wide varieties of ILs system. Interestingly, the similarity in the distribution of T_m among cluster 1 and the entire dataset (Figure 4) guarantees a similar good performance in the high R^2 score among these datasets (i.e., $R^2 \sim 0.90 - 0.94$), despite a smaller sample size, i.e., N = 605 (Table S1). To highlight the good predictive capability of our DL model to the entire dataset (N = 1253) in terms of high R^2 (i.e., $R^2 \sim 0.90$ in Table 1), Figure S1 depicts the small deviation between the predicted values of melting points and experimentally measured values obtained from the literature.

Compared to recent works in the literatures [42–46], the test R^2 score and RMSE reported in this work are outstanding. The previous works in the literature have reported comparable (~25–33 K) or higher (~39–45 K) RMSE values for the melting point prediction (Table 1). Despite being based on different methodologies and datasets, the R^2 score reported in the literature [42–46] (~0.54–0.82) are relatively low or less accurate compared to this work (i.e., R^2 ~0.90), as shown in Table 1. One of the best performing models in the literatures is based on the kernel ridge regression (KRR) model [46] which uses a significantly lower number of features yet achieves comparable accuracy (i.e., R^2 ~0.76, RMSE~39 K) to the rest of the studies. As reported by Low et al. [46], the KRR method only

depends on four molecular features or descriptors (e.g., Coulomb matrix, molecular orbital energies) which can, unfortunately, using the ab initio or first-principles calculation data, be computationally costly if a much larger IL dataset and larger IL molecules are considered. For the QSPR model [44], the low R^2 (i.e., R^2 ~0.72) might possibly be attributed to the commonly known sign change problem of descriptors in QSPR when the contributions of a set of selected descriptors or features are analyzed using a multivariate regression model [47].

Table 1. The performance metrics for predictions of the ILs melting temperature obtained in this work compared with that of the literature. N refers to the size of the entire dataset, which is the total number of ILs. GC refers to group contribution; ANN is artificial neural network; QSPR is quantitative structure-property relationship; RF is random forest; KRR is kernel ridge regression. For the RMSE, the unit is in Kelvin (K).

Reference	Ν	Features	Model	RMSE	R^2
[42]	799	80	GC	24.86	0.82
[43]	799	40	ANN	33.33	0.54
[44]	808	12	QSPR	26.85	0.72
[45]	2212	226	RF	45.00	0.66
[46]	2212	5	KRR	38.54	0.76
This work	1253	137	DL	32.88	0.90

For the model based on the group contribution (GC) method reported by Gharagheizi et al. [42], the RMSE is the smallest, i.e., \sim 25 K, with a high R^2 score \sim 0.82 compared to the rest (Table 1). For this method, the contributions of cation or anion functional groups are used to predict the targeted physicochemical property (e.g., T_m). In the literature [29,42,48–50], the implementation of the GC model varies, and the reported accuracies differ (i.e., $R^2 \sim 0.5$ –0.8) which substantially depends on certain groups of ILs in the datasets, and the work reported by Gharagheizi et al. [42] (Table 1) is probably the most accurate among the reported GC methods. However, in order to make predictions on new ILs dataset that have new chemical substructures or functional groups outside of the original dataset, the group contribution (GC) scheme would need to be re-devised and most probably could be time-consuming. Thus, by comparison to RMSE, the R^2 score and sample size of the ILs dataset, our current DL-model falls comfortably in between these two best predictive models, i.e., KRR and GC (Table 1). In particular, the advantage of the DL model in terms of sample size is discernable. As shown in Table 1, the DL model out-performed other ML techniques (i.e., KRR and RF in Table 1) with significantly higher accuracy (i.e., $R^2 \sim 0.90$, N = 1253) relative to RF (i.e., $R^2 \sim 0.66$, N = 2212) [45] and KRR (i.e., $R^2 \sim 0.76$, N = 2212) [46] despite the smaller size of the dataset, N (Table 1). Specifically, the DL model is generally considered to be very suitable to process large datasets [25-28,39], thus, we expect the accuracy of the DL model will improve significantly if a larger dataset is employed.

3.3. The Important Molecular Descriptors

To further examine which variables or molecular features influence the model's performance, the top ranking molecular features for the DL- model were computed based on the permutation importance as implemented in the ELI5 (ELI5 0.11.0) library [51]. Here, it is important to note that although the influence and correlation of the different molecular descriptors or features on the melting points of ILs is not directly obvious, their contributive effects are nonetheless comprehensible by conducting a feature importance analysis. Based on the DL-model (Table 1) and features filtering and ranking scores (Figure S2), important molecular descriptors that have a significant impact on the target property (i.e., T_m) can be obtained. Table 2 shows the top 10 most important molecular features or descriptors that have large impacts on the melting points of ILs obtained from three complementary correlation models, i.e., Pearson, Spearman and Kendall. The top 20 most influential molecular descriptors based on these correlation models can be found in Figure S2 and Table S2. As shown in Table 2 and Table S2, a consistent finding regarding the topmost influential molecular descriptors can be found using three different correlation methods. To highlight the complementary molecular features, the molecular descriptors listed are colored based on descriptor types or models defined in the literature [23,24,35].

Table 2. The top 10 most important molecular descriptors obtained from DL-model based on 3 different correlation methods, i.e., Pearson, Spearman and Kendall. Color code: red (P-VSA); blue (constitutional); green (topological indices); purple (molecular properties); black (atom-type E-state indices).

Pearson	Spearman	Kendall	
P_VSA_logP_1	DELS	P_VSA_ppp_hal	
P_VSA_ppp_N	MW	P_VSA_m_4	
DELS	P_VSA_logP_1	P_VSA_ppp_N	
P_VSA_m_4	P_VSA_ppp_N	P_VSA_e_4	
P_VSA_ppp_hal	MAXDP	DELS	
P_VSA_e_3	P_VSA_m_4	P_VSA_e_3	
P_VSA_m3	TPSA(Tot)	MW	
P_VSA_i_1	P_VSA_ppp_hal	TIE	
P_VSA_m_5	P_VSA_m_3	S2K	
S1K	SsF	P_VSA_m_5	

From these top 10 descriptors (Table 2), the important contributions from both cations and anions in ILs are captured and found to be closely related to geometrical structures, shapes, the size of the IL molecules, partial charges of IL cations or anions and the hydrogen bonds that influences intermolecular interactions. As shown in Table 2, the key molecular features found in the wide varieties of IL families in the datasets are the molecular descriptors related to the geometrical structures, branching and shapes or topological characters described by topological indices (e.g., DELS, MAXDP, S1K, S2K, etc.) and constitutional molecular weights and size (e.g., MW). For instance, the DELS is the topological indices is a representation of the measure of molecular electrotopological variation [52], which is related to the sum of overall atoms of the intrinsic state differences and could be interpreted as a measure of total charge transfer in ILs molecules. The S1K and S2K are the shape indices [53] which are a measure of the relative cyclicity of IL molecules, and are particularly relevant to cyclic compounds commonly found in IL cationic scaffolds (e.g., imidazolium, pyridinium, etc.). As another leading molecular descriptors, the P-VSA based molecular descriptors in Table 2 and Table S2 is based on atomic contributions to the van der Waals surface area, octanol-water partition coefficient (logP), molar refractivity and atomic partial charge-based [54] descriptors which encode the factors that could possibly influence the melting points of ILs, with a strong correlation to intermolecular hydrogen bonds and the hydrophobic and hydrophilic, polarizability, and electrostatic interaction of the ILs.

In our opinion, the high accuracy (high R^2 score) of the current DL-model, as shown in Table 1, might be attributed to the better utilization of all important molecular features in the prediction of melting points of ILs. To further understand the underlying interaction or correlation between these important molecular features/descriptors which are implicitly incorporated in the DL-model, an analysis of the distribution of the melting points of ILs using the best combination of two molecular descriptors from the topmost important descriptors (Tables 2 and S2) may lead us to some insights to control the physicochemical properties, such as melting point, T_m of ILs. In Figures 5 and S3, the important interplay dictated by the structural properties of the cations and anions of ILs that determine the distribution of T_m can be seen. As shown in Figure S3, the high T_m of ILs tends to favor the regime of small values in S2K and S3K which are the shape indices [53] that measure the relative cyclicity of IL molecules relevant to cyclic compounds commonly found in ILs cationic counterparts (e.g., imidazolium, pyridinium, phosphonium, etc.). To maximize the electrostatic interaction in high T_m ILs, the cationic scaffolds tend to favor smaller molecular weights/mass (i.e., smaller values in P_VSA_m_4) and more potential pharmacophore points in their positive charge distribution (i.e., larger values in P_VSA_ppp_P), as shown in Figure S3.



Figure 5. The distribution of the ILs melting points (1253 datapoints) which distinguish the low-($T_m < 273$ K) and high-melting ($T_m > 273$ K) points ILs using the best combination of two molecular descriptors: (a) P_VSA_LogP_1 vs. Chi_Dz(e); (b) TPSA(Tot) vs. P_VSA_ppp_A. The color of data points indicates the low (blue color) and high (brown color) melting point of the corresponding ILs dataset. The yellow region highlights the high melting point regime.

By considering the effects on the anions from the entire dataset (1253 ILs system), it was found that the high melting point ILs generally accumulate at the regime which has a high Chi_Dz(e) value (Figure 5a), with a high autocorrelation of Sanderson electronegativity (e.g., presence F, O, N, Cl, etc.) which might be attributed to the presence of anionic counterparts, e.g., halides (e.g., chlorides, bromide, or iodide anions), borates, and perfluorinated anionic species (e.g., $[BF_4]^-$, $[PF_6]^-$, $[CF_3SO_3]^-$, $[N-(SO_2CF_3)_2]^-$). Complementary to Figure 5a, most of the high melting point IL candidates are also found at the regime of high TPSA(Tot) and P_VSA_ppp_A values (Figure 5b). From the descriptors' model [23,35], it can be seen that the high TPSA(Tot) is related to a topological polar surface area based on polar constituents (e.g., N, O, S, and P) contributions, and more polarized molecules or constituents typically yield larger TPSA(Tot) values. The P_VSA_ppp_A is attributed to the P-VSA based molecular descriptors which relate to potential pharmacophore points of hydrogen-bond acceptors in ILs that account for hydrogen bonding ability. Thus, based on the trend observed in Figure 5b, one can find that by incorporating appropriate anionic species with a more localized negative charge, and a stronger hydrogen bonding in intermolecular interaction, the melting point for ILs will increase.

4. Discussion

A stated in the reported literature [8–11], it is known that the ILs with low melting points normally tend to have low viscosities which might be beneficial to the transport properties in battery application. It is also known that the melting points of ILs are primarily dictated by the structural properties of the scaffolds of cations, anions and the mutual interaction between them, and these interesting features can be seen from our findings based on the topmost important molecular descriptors analysis (Section 3.3). With these analyses, some useful design principles to fine tune the T_m of ILs can be obtained. To reduce the melting point of ILs, one can reduce the Coulombic interaction between the anions and cations (e.g., increasing the interionic separation, through the charge delocalization or shielding), lower the symmetry among the configuration of ions, increase the volume of the ions and reduce the efficient packing among the anions and cations by utilizing the ions

with high conformational flexibility [55,56]. Thus, to fine tune the melting temperatures of ILs, one can incorporate a variety of counterions with different molecular shapes (e.g., linear, spherical, etc.), structures (e.g., single chain, multiple chains, linear or branched), charge coordination through the functional groups design and engineering during the synthesis process of ILs.

In this perspective, finding a set of useful design-principles that can guide the functionalgroup design with an optimal selection of anion-cation combinations is critical to attain new useful ILs with desirable physicochemical properties (e.g., melting points, ionic conductivity, thermal stability, etc.). Thus, we believe the predictive models based on a set of dominant molecular descriptors [42–46] included in this work can be helpful and will provide some useful insights in IL design. However, due to a vast selection and combination of cations and anions, the construction of a large and sustainable ILs database (e.g., ILThermo v2.0) [31,32] is deemed important. To advance the design and development of ILs, finding an optimal predictive model that can simultaneously predict several target physicochemical properties (e.g., melting point, viscosity, solubility, electrical and thermal conductivity) of various ILs accurately, using a minimal set of a few important descriptors remains a challenge, and will be an active research focus in the future. Specifically, to further improve the accuracy of the predictive models [29,30,42–46], the current state-of-the-art deep learning (DL) model [25–28] which is robust in processing vast amounts of features and datasets and supports highly parallelized and distributed algorithms that utilize graphic processing unit (GPU) machines, could be a promising method to achieve these goals.

5. Conclusions

Ionic liquids (ILs) have great potential for application in energy storage and conversion devices, and they have been identified as promising electrolyte candidates in various batteries systems. However, the practical application of many ionic liquids remains limited due to the unfavorable melting points that constrain the operating temperatures and exhibit unfavorable transport property in batteries. With this as our motivation, we carried out a baseline study to investigate the trend of melting points (T_m) of a wide variety of ILs, with the aim to search for insights that will lead us to fine tune the T_m of ILs using highthroughput screenings of large a ILs dataset and machine-learning model. Based on the dataset (1253 ILs) obtained from an established ILs database, i.e., ILThermo (v2.0) [31,32], we managed to construct a predictive model to predict the melting points (T_m) of ILs with a reasonably high accuracy, achieving an R^2 score of 0.90 with an RMSE of ~32 K by utilizing a set of important quantitative structure-property relationship (QSPR) molecular features or descriptors based on the deep-learning (DL) model (Section 3.2). Despite a wide variation in the distribution of melting points and a wide variety of anion-cation combination in the ILs dataset (Section 3.1), we found the melting points T_m of various ILs can be determined based on a limited set of molecular descriptors. These molecular descriptors consist of 137 descriptors that highlight several important molecular features that have significant influence in determining the melting points of ILs, e.g., the presence of electronegative constituents, geometrical structures, branching and shapes, hydrogen-bonding ability, polarizabilities, etc. (Section 3.3).

Based on the DL model, the important interplay dictated by the structural properties of the cations and anions of ILs that determine the distribution of T_m can be seen (Section 3.3). For example, the high T_m of ILs tends to favor small values of shape indices which measure the relative cyclicity of ILs molecules that could relate to cyclic compounds commonly found in ILs cationic counterparts (e.g., imidazolium, pyridinium, phosphonium, etc.). We elucidated the effects of anionic counterparts by incorporating appropriate anionic species with a more localized negative charge, and stronger hydrogen bonding in intermolecular interaction, which can lead to an increasing melting point for ILs (Section 3.3). Thus, with a fine selection of anion–cation combination in ILs, we believe that the design and engineering of functional groups is the key to fine tune the melting points, and further studies in the development of predictive models that are able to accurately predict other

physicochemical properties (e.g., viscosity, hydrophilicity/hydrophobicity, conductivity) relevant to battery application will be conducted.

Supplementary Materials: The following supporting information can be downloaded at: https://www.action.com/actionals //www.mdpi.com/article/10.3390/app12052408/s1, Table S1: The performance metrics of DL-model that applied to different testing sets which based on 137 molecular features/descriptors. N is the sample size of the dataset. For RMSE, the unit is in Kelvin (K); Figure S1: DL model predicted melting point (i.e., T_m (*Prediction*)/K) values (in blue color) of the entire ILs dataset versus the experimental measured melting point (i.e., T_m (Exp)/K) (in red color) obtained from the literatures (i.e., ILThermo database); Figure S2: A correlation matrix comparing the melting point (T_m) with top 20 important molecular descriptors based on the Pearson correlation method. The numbers found in the correlation matrix is the feature correlation coefficients or ranking scores computed based on Pearson correlation method; Table S2: The top 20 most important molecular descriptors obtained from DL-model based on 3 different correlation methods, i.e., Pearson, Spearman and Kendall; Figure S3: The distribution of the ILs melting points plot using the best combination of two molecular descriptors that mostly focus on the cations contribution: (a) S2K vs. P_VSA_ppp_hal; (b) S3K vs. P_VSA_m; (c) S2K vs. P_VSA_m_4; (d) P_VSA_ppp_P vs. P_VSA_m_4. The color of data points indicates the low melting point (i.e., $T_m < 273$ K) (in blue color) and high melting point (i.e., $T_m > 273$ K) (in brown color) of the corresponding ILs. The yellow region highlights the high melting point regime.

Author Contributions: Conceptualization, K.C.L.; methodology, Z.A., P.N. and K.C.L.; validation, Z.A., K.C.L.; formal analysis, Z.A., P.N. and K.C.L.; investigation, Z.A., K.C.L.; data curation, P.N.; writing—original draft preparation, Z.A., and K.C.L.; writing—review and editing, K.C.L.; visualization, Z.A., P.N. and K.C.L.; supervision, K.C.L.; project administration, K.C.L.; funding acquisition, K.C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Research Corporation for Science Advancement (RCSA) through a Cottrell Scholar Award (Award# 26829) and faculty start-up fund through California State University Northridge (CSUN).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: K.C.L. acknowledges the support of the U.S. Research Corporation for Science Advancement and California State University Northridge. The contribution during the initial development of our in-house code for this work by Michael Munje is acknowledged. Besides, the valuable comments given by Qing Guo is acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Watanabe, M.; Thomas, M.L.; Zhang, S.; Ueno, K.; Yasuda, T.; Dokko, K. Application of Ionic Liquids to Energy Storage and Conversion Materials and Devices. *Chem. Rev.* **2017**, *117*, 7190–7239. [CrossRef] [PubMed]
- 2. Lei, Z.; Dai, C.; Chen, B. Gas Solubility in Ionic Liquids. *Chem. Rev.* 2014, 114, 1289–1326. [CrossRef]
- 3. Zhang, Q.; Zhang, S.; Deng, Y. Recent advances in ionic liquids catalysis. Green Chem. 2011, 13, 2619–2637. [CrossRef]
- 4. Qu, J.; Zhou, Y. Ionic Liquids as Lubricant Additives: A Review. ACS Appl. Mater. Interfaces 2017, 9, 3209–3222.
- Hough, W.L.; Smiglak, M.; Rodríguez, H.; Swatloski, R.P.; Spear, S.K.; Daly, D.T.; Pernak, J.; Grisel, J.E.; Carliss, R.D.; Soutullo, M.D.; et al. The third evolution of ionic liquids: Active pharmaceutical ingredients. *New. J. Chem.* 2007, *31*, 1429–1436. [CrossRef]
- Sahbaz, Y.; Williams, H.D.; Nguyen, T.H.; Saunders, J.; Ford, L.; Charman, S.A.; Scammells, P.J.; Porter, C.J.H. Transformation of Poorly Water-Soluble Drugs into Lipophilic Ionic Liquids Enhances Oral Drug Exposure from Lipid Based Formulations. *Mol. Pharm.* 2015, *12*, 1980–1991. [CrossRef]
- Gupta, K.M.; Jiang, J. Cellulose dissolution and regeneration in ionic liquids: A computational perspective. *Chem. Eng. Sci.* 2015, 121, 180–189. [CrossRef]
- Venkatraman, V.; Evjen, S.; Lethesh, K.C. The Ionic Liquid Property Explorer: An Extensive Library of Task-Specific Solvents. Data 2019, 4, 88. [CrossRef]
- Hallett, J.P.; Welton, T. Room-Temperature Ionic Liquids: Solvents for Synthesis and Catalysis. 2. Chem. Rev. 2011, 111, 3508–3576.
 [CrossRef]
- 10. Seddon, K.R. Ionic Liquids for Clean Technology. J. Chem. Technol. Biotechnol. 1997, 68, 351–356. [CrossRef]

- Greaves, T.L.; Drummond, C.J. Protic Ionic Liquids: Evolving Structure–Property Relationships and Expanding Applications. *Chem. Rev.* 2015, 115, 11379–11448. [CrossRef] [PubMed]
- 12. Balducci, A. Ionic Liquids in Lithium-Ion Batteries. Top. Curr. Chem. 2017, 375, 20. [CrossRef]
- Zhang, J.; Sun, B.; Zhao, Y.; Tkacheva, A.; Liu, Z.; Yan, K.; Guo, X.; McDonagh, A.M.; Shanmukarai, D.; Wang, C.; et al. A versatile functionalized ionic liquid to boost the solution-mediated performances of lithium-oxygen batteries. *Nat. Commun.* 2019, 10, 602. [CrossRef]
- 14. Josef, E.; Yan, Y.; Stan, M.C.; Wellmann, J.; Visintin, A.; Winter, M.; Johansson, P.; Dominko, R.; Guterman, R. Ionic Liquids and their Polymers in Lithium-Sulfur Batteries. *Israel J. Chem.* **2019**, *59*, 832–842. [CrossRef]
- 15. Ortiz-Martínez, V.M.; 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. [CrossRef]
- Martin, S.; Pratt, H.D., III; Anderson, T.M. Screening for High Conductivity/Low Viscosity Ionic Liquids Using Product Descriptors. *Mol. Inf.* 2017, 36, 1600125. [CrossRef] [PubMed]
- Tiago, G.A.O.; Matias, I.A.; Ribeiro, A.P.C.; Martins, L.M.D.R.S. Application of Ionic Liquids in Electrochemistry—Recent Advances. *Molecules* 2020, 25, 5812. [CrossRef]
- Katrizky, A.R.; Jain, R.; Lomaka, A.; Petrukhin, R.; Karelson, M.; Visser, A.E.; Rogers, R.D. 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, 225–231. [CrossRef]
- Rabideau, B.D.; Soltani, M.; Parker, R.A.; Siu, B.; Salter, E.A.; Wierzbicki, A.; West, K.N.; Davis, J.H., Jr. Tuning the melting point of selected ionic liquids through adjustment of the cation's dipole moment. *Phys. Chem. Chem. Phys.* 2020, 22, 12301–12311. [CrossRef]
- Zhang, Y.; Maginn, E.J. Molecular dynamics study of the effect of alkyl chain length on melting points of [CnMIM][PF6] ionic liquids. *Phys. Chem. Chem. Phys.* 2014, 16, 13489–13499. [CrossRef]
- Karu, K.; Elhi, F.; Pohako-Esko, K.; Ivaništšev, V. Predicting Melting Points of Biofriendly Choline-Based Ionic Liquids with Molecular Dynamics. *Appl. Sci.* 2019, 9, 5367. [CrossRef]
- 22. Valderrama, J.O.; Cardona, L.F. Predicting the melting temperature and the heat of melting of ionic liquids. *J. Ion. Liq.* 2021, *1*, 100002. [CrossRef]
- 23. Roy, K.; Kar, S.; Das, R.N. A Primer on QSAR/QSPR Modeling, 1st ed.; Springer: New York, NY, USA, 2015; pp. 1–36.
- 24. Karelson, M. Molecular Descriptors in QSAR/QSPR, 1st ed.; Wiley: New York, NY, USA, 2000; pp. 1–448.
- 25. Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning, 1st ed.; MIT Press: Cambridge, MA, USA, 2016; pp. 363-405.
- 26. Schmidhuber, J. Deep learning in neural networks: An overview. Neural Netw. 2015, 61, 85–117. [PubMed]
- 27. Sarker, I.H. Deep Learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput. Sci.* **2021**, *2*, 420. [CrossRef]
- Sejnowski, T.J. The unreasonable effectiveness of deep learning in artificial intelligence. *Proc. Natl. Acad. Sci. USA* 2020, 117, 30033–30038. [CrossRef]
- Lazzús, J.A. A group contribution method to predict the melting point of ionic liquids. *Fluid Phase Equilibria* 2012, 313, 1–6. [CrossRef]
- 30. Varnek, A.; Kireeva, N.; Tetko, I.V.; Baskin, I.I.; Solovev, V.P. Exhaustive QSPR Studies of a Large Diverse Set of Ionic Liquids: How Accurately Can We Predict Melting Points? *J. Chem. Inf. Model.* **2007**, *47*, 1111–1122. [CrossRef]
- Dong, Q.; Kazakov, A.; Muzny, C.; Chirico, R.; Widegren, J.; Diky, V.; Magee, J.; Marsh, K.; Frenkel, M. Ionic Liquids Database (ILThermo), Ionic Liquids Database (ILThermo). 2006. Available online: https://ilthermo.boulder.nist.gov/ILThermo/mainmenu. uix (accessed on 28 January 2022).
- Kazakov, A.; Magee, J.; Chirico, R.; Diky, V.; Kroenlein, K.; Muzny, C.; Frenkel, M. Ionic Liquids Database—ILThermo (v2.0). Ionic Liquids Database—ILThermo (v2.0). 2013. Available online: https://trcsrv1.boulder.nist.gov/ilthermo/ilthermo.html (accessed on 28 January 2022).
- 33. Roemer, F. pyILT2. Available online: http://wgserve.de/pyilt2/ (accessed on 28 January 2022).
- Lowe, D.M.; Corbett, P.T.; Murray-Rust, P.; Glen, R.C. Chemical Name to Structure: OPSIN, an Open Source Solution. J. Chem. Inf. Model. 2011, 51, 739–753. [CrossRef]
- Talete srl Dragon, Version 7.0 Software for Molecular Descriptor Calculation. Available online: https://chm.kode-solutions.net/ pf/dragon-7-0/ (accessed on 28 January 2022).
- 36. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: https://tensorflow.org (accessed on 28 January 2022).
- 38. Keras. Available online: https://keras.io (accessed on 28 January 2022).
- 39. Dive into Deep Learning. Available online: https://d2l.ai/index.html (accessed on 28 January 2022).
- 40. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. arXiv 2017, arXiv:1412.6980.
- 41. Yuan, C.; Yang, H. Research on K-value Selection Method of K-Means Clustering Algorithm. J 2019, 2, 226–235. [CrossRef]

- 42. Gharagheizi, F.; Ilani-Kashkouli, P.; Mohammadi, A.H. Computation of normal melting temperature of ionic liquids using a group contribution method. *Fluid Phase Equilibria* **2012**, *329*, 1–7. [CrossRef]
- 43. Valderrama, J.O. Myths and Realities about Existing Methods for Calculating the Melting Temperatures of Ionic Liquids. *Ind. Eng. Chem. Res.* 2014, 53, 1004–1014. [CrossRef]
- 44. Farahani, N.; Gharagheizi, F.; Mirkhani, S.A.; Tumba, K. Ionic liquids: Prediction of melting point by molecular-based model. *Thermochim. Acta* **2012**, *549*, 17–34. [CrossRef]
- 45. Venkatraman, V.; Evjen, S.; Lethesh, K.C.; Raj, J.J.; Knuutila, H.; Fiksdahl, A. Rapid, comprehensive screening of ionic liquids towards sustainable applications. *Sustain. Energy Fuels* **2019**, *3*, 2798–2808. [CrossRef]
- Low, K.; Kobayashi, R.; Izgorodina, E.I. The effect of descriptor choice in machine learning models for ionic liquid melting point prediction. J. Chem. Phys. 2020, 153, 104101. [CrossRef] [PubMed]
- 47. Kiralj, R.; Ferreira, M.M.C. Is your QSAR/QPSR descriptor real or trash? J. Chemom. 2010, 24, 681–693. [CrossRef]
- 48. Yamamoto, H. Structure Properties Relationship of Ionic Liquid. J. Comput. Aided Chem. 2006, 7, 18–30. [CrossRef]
- 49. Huo, Y.; Xia, S.; Zhang, Y.; Ma, P. Group Contribution Method for Predicting Melting Points of Imidazolium and Benzimidazolium Ionic Liquids. *Ind. Eng. Chem. Res.* 2009, *48*, 2212–2217. [CrossRef]
- Preiss, U.P.; Beichel, W.; Erle, A.M.T.; Paulechka, Y.U.; Krossing, I. Simple Melting Point Prediction Possible? *ChemPhysChem* 2011, 12, 2959–2972. [CrossRef]
- 51. ELI5. Available online: https://eli5.readthedocs.io/en/latest/index.html (accessed on 28 January 2022).
- Gramatica, P.; Corradi, M.; Consonni, V. Modelling and prediction of soil sorption coefficients of non-ionic organic pesticides by molecular descriptors. *Chemosphere* 2000, 41, 763–777. [CrossRef]
- 53. Kier, L.B. Distinguishing Atom Differences in a Molecular Graph Shape Index. Quant. Struct. Act. Relat. 1986, 5, 7–12. [CrossRef]
- 54. Labute, P. A widely applicable set of descriptors. J. Mol. Graph. Model. 2000, 18, 464–477. [CrossRef]
- 55. Krossing, I.; Slattery, J.M.; Daguenet, C.; Dyson, P.J.; Oleinikova, A.; Weingartner, H. Why Are Ionic Liquids Liquid? A Simple Explanation Based on Lattice and Solvation Energies. *J. Am. Chem. Soc.* **2006**, *128*, 13427–13434. [CrossRef]
- Holbrey, J.D.; Reichert, W.M.; Nieuwenhuyzen, M.; Johnson, S.; Seddon, K.R.; Rogers, R.D. Crystal polymorphism in 1-butyl-3methylimidazolium halides: Supporting ionic liquid formation by inhibition of crystallization. *Chem. Commun.* 2003, *9*, 1636–1637. [CrossRef]