Post-print version of the paper by Venkatraman et al. in J Mol. Liq, 264 (2018) 318-326: http://dx.doi.org/10.1016/j.molliq.2018.03.090

Predicting Ionic Liquid Melting Points using Machine Learning

Vishwesh Venkatraman[a,*], Sigvart Evjen[a], Hanna K. Knuutila[b], Anne Fiksdahl[a], Bjørn Kåre Alsberg[a]

[a] *Department of Chemistry, Norwegian University of Science and Technology, 7491, Trondheim, Norway*
[b] *Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway*

Abstract

The melting point (T_m) of an ionic liquid (IL) is of crucial importance in many applications. The T_m can vary considerably depending on the choice of the anion and cation. This study explores the use of various machine learning (ML) methods to predict the melting points (-96°C - 359°C range) of structurally diverse 2212 ILs based on a combination of 1369 cations and 141 anions. Among the ML models applied to independent training and test sets, tree-based ensemble methods (Cubist, random forest and gradient boosted regression) were found to demonstrate slightly better performance over support vector machines and k-nearest neighbour approaches. In comparison, quantum chemistry based COSMOtherm predictions were generally found to have significant deviations with respect to the experimental values. However, classification models were more efficient in discriminating between ILs with $T_m > 100$°C and those below 100°C.

Keywords: QSPR, ionic liquids, melting point, machine learning, experimental, quantum chemistry

1. Introduction

In recent years, there has been a dramatic increase in the number of ionic liquids (ILs) synthesized and tested. This explosion of interest has been due to the various advantageous properties that ILs possess such as a negligible vapour pressure, a liquid range of up to more than 400 K and non-flammability under ambient conditions. This has lead to a wide range of applications such as energy storage[1], solvents for CO_2 capture[2], catalysis [3], lubricant additives[4], pharmaceuticals[5, 6], cellulose dissolution[7] and, foods and bioproducts[8].

*Corresponding author
Email address: vishwesh.venkatraman@ntnu.no (Vishwesh Venkatraman)

Ionic liquids are molten salts composed of cations (mostly organic) and anions(inorganic or organic). The unique properties of the ILs are largely determined by the structure and interaction between the ions.

Of particular interest in a number of applications such as gas separation, catalysis and electrochemistry, there is a demand for room-temperature ionic liquids that have melting points below 100°C[9]. Factors such as the charge distribution on the ions, hydrogen bonding ability etc. are seen to strongly influence melting points. Given the large variety of cations and anions to choose from, the potential combinations[10] (assuming all to be feasible) of the two are indeed quite staggering ($\sim 10^{18}$). For such a large collection, the ability to predict melting points accurately in a short time without expensive and laborious experiments would therefore be of great value. In this regard, quantitative structure-property relationship-based (QSPR) methods have had reasonable success[11]. These methods rely on correlating physicochemical features (ranging from fragment based counts to quantum chemistry-based descriptors) with the melting point values that are summarised in Table 1. Compared with the regression based approaches listed in Table 1, Kireeva et al.[12] propose a classification model based on splitting the ionic liquids into individual classes according to the values of their melting points. The models reported accuracies ranging between $78 - 85\%$ depending on the data set. In summary, while the models for closely related groups of ionic liquids have yielded reasonable predictions, the performances for cases where there was significant structural diversity were found to be less robust. Other efforts have attempted to create thermodynamic melting point prediction models [13, 14] based on COSMO-RS [15] computations. For a set of 520 organic salts, the thermodynamic model yielded a mean absolute error of 34°C. In recent years, approaches based on molecular dynamics simulations[16–18] and density functional theory[19] (DFT) have also been used to determine melting points. Given the high computational expense associated with such calculations, the studies so far have been limited to a very small sets of ionic liquids.

Table 1: Table summarises QSPR approaches for IL melting point prediction. Here MAD is the mean absolute deviation, MLR - multiple linear regression, PLSR- partial least-squares regression, and SVR-support vector regression.

Dataset	**Descriptors**	**Regression**	**Results**	**Ref.**
33 ILs	molecular orbital and electrostatic descriptors	MLR	13 Triazolium bromides: $R^2 = 0.89$, 13 nitrates: $R^2 = 0.84$, 7 nitrocyanamides: $R^2 = 0.96$	[20]
126 pyridinium bromides	constitutional, 2D and 3D	decision trees	$R^2 = 0.82$	[21]
126 pyridinium bromides	positional trees	recurrent neural network	$R^2_{train} = 0.96$, $R^2_{test} = 0.87$	[22]
717 bromides	fragment based,E-state indices	neural network	$R^2 = 0.71$	[23]
394 ILs	topological indices	MLR	$R^2 = 0.78$	[24]
400 ILs	group contribution	genetic algorithm	$R^2 = 0.81$	[25]
808 ILs	constitutional, 2D and 3D	genetic function approximation	705 ILs:$R^2_{train} = 0.66$, 143 ILs:$R^2_{test} = 0.72$	[26]
97 imidazolium ILs	constitutional, 2D and 3D descriptors	multilayer perceptron	$R^2 = 0.99$	[27]
667 ILs	group contribution	back-propagation neural networks	$MAD_{train} = 3.7\%$,$MAD_{test} = 14.6\%$	[28]
61 ILs	quantum chemistry-based descriptors	MLR	$R^2 = 0.61$	[29]
136 ILs	group contribution	optimization algorithm	$MAD = 22.6K$	[30]
45 ILs	electrostatic, quantum mechanical and topological descriptors	MLR	16 imidazolium tetrafluoroborates $R^2 = 0.90$, 25 imidazolium hexafluorophosphates $R^2 = 0.92$	[31]
799 ILs	group contribution	optimization algorithm	598 ILs: $R^2_{train} = 0.81$, 201 ILs: $R^2_{test} = 0.82$	[32]
37 ILs	quantum chemistry-based descriptors	genetic programming	$R^2 = 0.91$	[33]
62 imidazolium ILs	constitutional, topological, and geometric descriptors	PLSR	$R^2 = 0.869$	[34]
126 pyridinium bromides	quantum chemistry-based descriptors	MLR	$R^2 = 0.79$	[35]
190 ILs	group contribution	optimization algorithm	$R^2 = 0.90$	[36]
288 ILs	quantum chemistry-based descriptors	projection pursuit	$R^2 = 0.81$	[37]
75 alkyl-ammonium bromides	quantum chemistry-based descriptors	genetic function approximation	$R^2 = 0.79$	[38]

Post-print version of the paper by Venkatraman et al. in J Mol. Liq, 264 (2018) 318-326: http://dx.doi.org/10.1016/j.molliq.2018.03.090

In this article, we evaluate the performance of various machine learning models for melting point predictions of a large diverse set of 2212 ILs for which the melting points vary in the range of -96°C - 359°C. Given the relative success of quantum chemistry-based descriptors (see Table 1) we focus primarily on such variables for establishing structure-property relationship models based on both regression and classification schemes. Among the modelling schemes analysed, non-linear methods, in particular regression tree-based approaches were found to have moderate performance. In comparison, both linear and non-linear classification models were able to better distinguish between ILs with $T_m > 100$°C and those below 100°C.

2. Methods and Materials

2.1. Ionic Liquid Melting Points

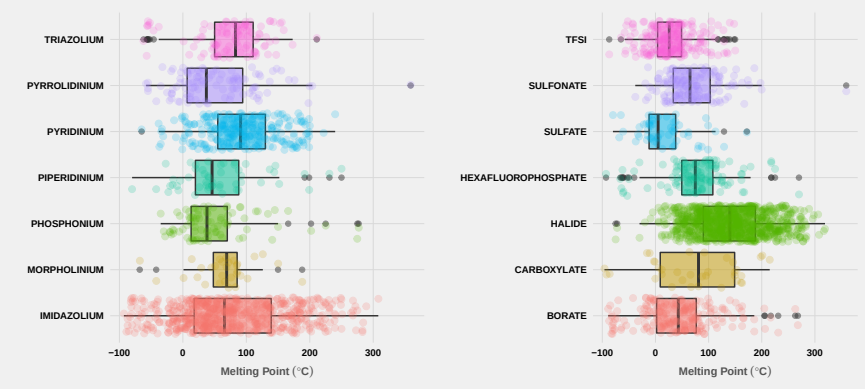

Figure 1: Boxplot shows the distribution of the melting points based on the different cation (left) and anion (right) classes. The distribution of the temperatures is summarised by the minimum, median, maximum and the first and third quartiles.

Experimental melting point temperatures (T_m) for a set of 2212 ILs comprising 1369 cations and 141 anions were extracted from over ∼300 references in the literature with a primary source being Zhang et al[39] and Varnek et al[23]. Additional experimental data was obtained by performing a literature search on the ISI Web of Science using the keywords "ionic liquids and melting points". Most measurements are carried out using differential scanning calorimetry or differential thermal analysis. As ionic liquid melting points are influenced factors such as presence of impurities and measurement protocols[40], discrepancies in the T_m data are seen with associated experimental uncertainties seldom reported[41]. In order to ensure a chemically diverse data set with relatively low noise levels, we have adopted the following protocol: (i) ILs for which a melting point range was specified, the mid-point of this range was chosen, (ii) For ILs with multiple T_m values, the most frequent value was retained (iii) ILs for which only glass transitions or alternatively those for which the T_m was reported as below or

above a certain threshold, were omitted. In addition, ILs composed of multiple cations or anions were excluded. The structures of the cations and anions, the experimental T_m values and associated references are listed in Tables S1 and S2 in the supplementary material. The cations span several compound classes such as imidazoliums, azaniums, triazoliums, pyridiniums and piperidiniums while the anions consist largely of halides, hexfluorophosphates and borates. Boxplots in Figure 1 summarize the melting point distributions for the ILs. The temperatures range between -90°C for trialkylammonium perfluoroalkyl β-diketonate salts[42] to over 300°C for pyrrolidinium perfluorobutanesulfonates [43] (see Figure F1 in the supplementary material).

2.2. Molecular Representation

Quantum mechanical descriptors have met with reasonable success in modelling a number of IL properties[44]. While previous studies [14, 20, 27, 33] have made use of *ab initio* methods as a starting point for descriptor generation, here, we employ molecular descriptors[45] obtained from computationally low-cost PM6[46] calculations. The initial structures of the ions were drawn using the MarvinSketch[47] software and subsequently converted to three-dimensions using OpenBabel[48] (based on the Universal Force Field[49]). Each ion was then optimized using the PM6 Hamiltonian in MOPAC[50] with the keywords: "PM6 XYZ PRECISE STATIC POLAR MMOK SUPER ENPART LARGE". Using the data supplied by MOPAC output files, the software KRAKENX(downloadable from `www.krakenminer.com`) was used to calculate various quantum chemical and molecular orbital based descriptors such as the HOMO/LUMO energies, polarizabilities, superdelocalizabilities, charge partial surface areas (CPSA) and geometrical indices. For computational ease, the cations and anions are treated independently and ion-ion interaction effects are ignored in the current work. A total of 113 descriptors were calculated for each anion and cation.

2.3. Statistical Modelling

Variables Pretreatment. Prior to modelling, low variance columns and those containing missing values were excluded. With a view to reducing the dimensionality of the original descriptor matrix, only one among the highly correlated pair of variables ($R^2 > 0.95$) was retained[51]. The remaining variables were then autoscaled to zero mean and unit variance. The data set was further split randomly (67:33) into independent calibration and test sets containing 1486 and 726 ILs respectively.

Methods. Both linear and non-linear approaches were used to create the structure-property models. The statistical software *R*[52] was used to build the models using available packages such as *pls*[53] for partial least squares regression (PLSR), *kernlab*[54] for support vector regression (SVR), *randomForest*[55] for random forests regression (RF), *gbm*[56] for generalized boosted regression (GBM), *Cubist* [57] for tree-based regression for and *KernelKnn*[58] for k-nearest neighbours (k-nn) regression. In addition, classification models were also attempted

by discretizing the melting points into two levels. Here, we focused on partial least squares discriminant analysis (PLSDA), random forests and gradient boosting. Variable importance was further used as a means to interpret models and where possible, reduce model complexity by way of feature selection.

PLSR Given the descriptor matrix (X) partial least squares regression[59] calculates latent variables that are oriented along the directions of maximum covariance between the independent variables in X and the response Y. For categorical responses, partial least squares discriminant analysis[60] was applied. The optimal number of latent variables was determined using cross-validation.

RF In the random forests (RF) approach, multiple decision trees are trained on random sub-samples of the data set and predictions from each tree are aggregated to determine the final outcome[61]. For RF, the variable selection step was carried out using the package *VSURF*[62] which makes use of permutation-based score of importance combined with a stepwise forward strategy for variable introduction.

k-nn k-nearest neighbours regression is a non-parametric method that calculates an output based on a weighted mean of the number of neighbours ranked by the Euclidean distance. In this study, the weighting scheme combines a tricube weight function (which gives the greatest weight to the closest observations) with a Gaussian kernel to optimize the output predictions[63]. The optimal value of k (varied between 2-20 neighbours) was determined using cross-validation. A simulated annealing based feature selection[64] was further used to identity simpler models with better predictive performance.

GBM In generalized boosted regression[65], regression trees are fit sequentially with respect to a differentiable loss function (residual error plus function complexity) that is being minimized. During each iteration, a subsample of the training data is drawn at random and is used to calculate the model update. The prediction model is thus an ensemble of weak prediction models.

Cubist Cubist regression is an extension of the M5 model tree approach by Quinlan[66, 67]. It is a rule-based model where each rule is associated with a multivariate linear model. In this study, both the number of committees (a boosting scheme where iterative model trees are created in sequence) and the number of nearest-neighbors were optimized to adjust the predictions from the model. The committees are made up of several rule-based models and used to fine-tune the models. Each member of the committee predicts the target value for a given case and the predictions from the members are averaged to give a final prediction.

Model performances were assessed using multiple metrics such as the squared coefficient of correlation (R^2), root mean square error ($RMSE$) and the mean

absolute error which estimates the bias in the predictions. The robustness of the models were assessed using 5-fold cross-validation which were repeated three times to account for the randomness of the splits. Further to guard against chance correlations, $y-$randomization tests[68] (repeated 500 times) were also performed. For both GBM and SVR, recursive feature elimination protocol (which recursively considers smaller sets of features) as implemented in the *caret*[69] package was used. Source code implementing the above metrics and methods was written in-house. All calculations were carried out on a desktop PC with Intel i5-2400 Quad-Core 3.10GHz CPU and 8GB RAM. The final set of descriptors and the model predictions are provided in the Supplementary material.

2.4. COSMO-RS

For comparison, the software COSMOtherm[70] which implements the melting point prediction model proposed by Preiss et al.[13] was also used. Each cation and anion was subjected to DFT calculations using the ORCA quantum chemistry program[71]. Here, the BP functional B88-p86 with a triple-ζ valence polarized basis set[72] (TZVP) and the resolution of identity standard (RI) approximation was employed. The output from ORCA (COSMO files) were then used as an input to the software COSMOtherm[70] along with the parameterization set BP_TZVP_C30_01601.

2.5. Distribution of the ILs

As an initial step, the distribution of the analysed ILs in the multidimensional descriptor space was studied using principal component analysis (PCA). As can be seen from the plots in Figure 2, the first 5 principal components (PCs) explain nearly 73% of the variance in the data. The score plots (Figure 2B-D) with respect to the first two principal components show groupings that correspond to the different cation and anion classes. For example, the halide-based ILs occupy a significant cluster (shown in orange) while imidazoliums are a little scattered but are largely concentrated in the largest cluster located in the center. Variable contributions summarising the information shared by each variable (Figure 2E-F) and a particular component were additionally analysed[73]. The first PC is dominated by contributions from both cations and anions while the second is more influenced by the cations with descriptors focusing on the geometrical and quantum-chemical characteristics. For instance, the energy related terms such as the HOMO-LUMO gap reflect the stability of the IL while the electronic structural features (most negative or most positive atomic charge) and partial surface area descriptors encode factors that influence intermolecular interactions. The radius of gyration, here calculated for the anion, is a size descriptor based on the distribution of atomic masses[74]. It has been observed that the increase in the size of the anion can lead to a decrease in the melting point[75].

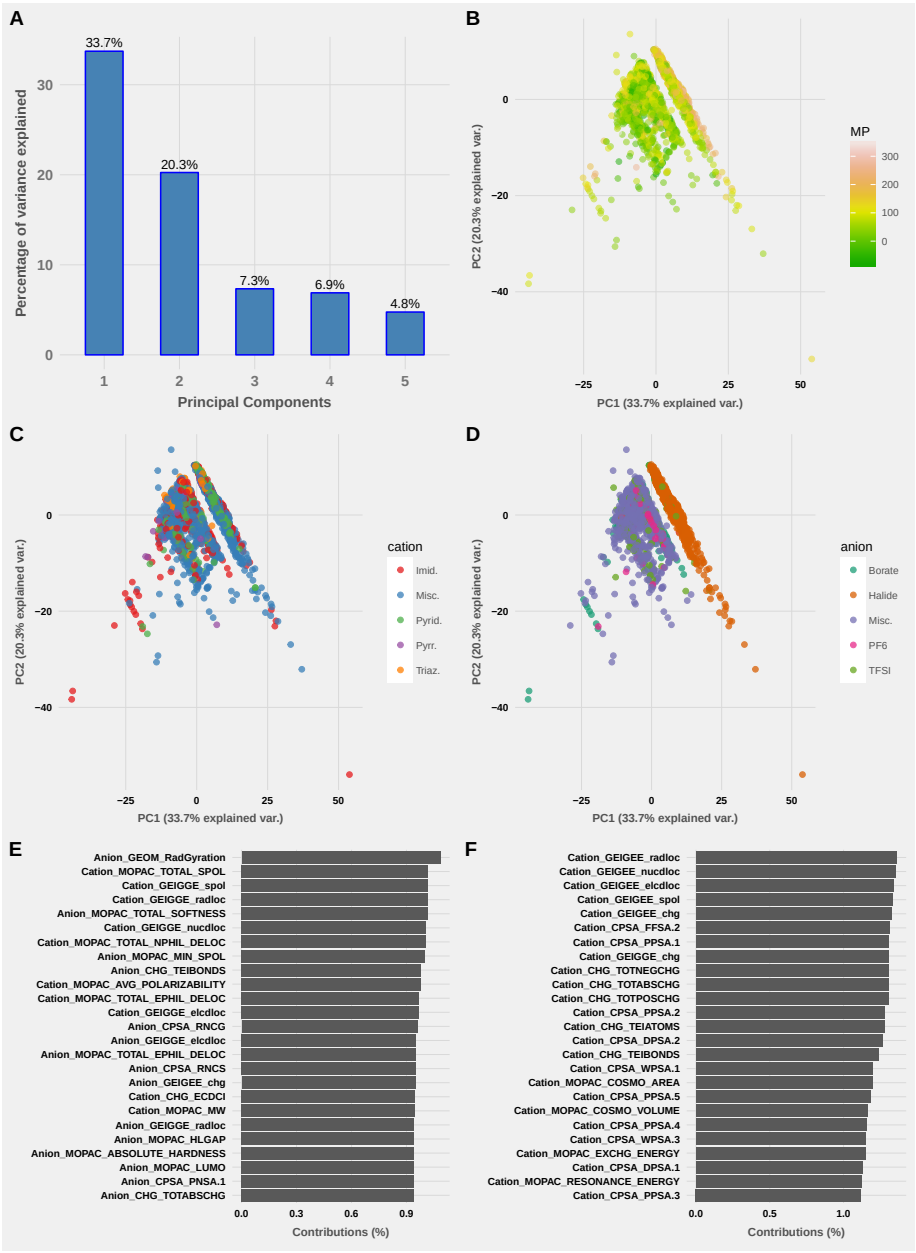

Figure 2: Graphical summary of the principal component analysis. (B) shows the score plot for the first two PCs. (C) and (D) show the score plot with respect to the prominent cationic (imidazolium, pyridinium, pyrrolidinium, triazolium) and anionic (borates, halides, hexafluorophosphates, TFSI) groups. For the variable contributions, only the top 25 variables are shown.

Table 2: Table summarises the machine learning performances for different regression methods applied to independent calibration and test sets. R^2_{cv} is the 5-fold cross-validated squared correlation coefficient averaged over 3 independent runs. Numbers in brackets in the RMSE column are the corresponding mean absolute errors.

Method	**Training**		**Testing**	
	$RMSE$ (MAE)	R^2	$RMSE$ (MAE)	R^2
PLSR	55 (41)	0.50	58 (41)	0.53
SVR	47 (16)	0.64	46 (34)	0.65
k-nn	47 (35)	0.64	49 (36)	0.61
RF	44 (14)	0.67	46 (34)	0.66
GBM	45 (13)	0.66	45 (33)	0.67
CUBIST	45 (12)	0.67	47 (33)	0.64

2.6. Regression modelling

Summaries of the model performances are presented in Table 2. A complete list of the model parameters and results is presented in Table S4 in the supplementary material. Among the methods used, the ensemble tree-based Cubist, GBM and RF models were found to have the best performance averaging $R^2 > 0.65$ for both calibration and test sets. Mean absolute errors for the Cubist model were seen to be the lowest across both calibration and test sets. While both k-nn and SVR models showed similar trends, the linear PLSR model is the least accurate. Interestingly, both k-nn and RF models use less than 35 variables after feature selection, while for the other approaches the variable reduction was not found to improve calibration results. Randomization tests applied to the response (MP) yielded $p-$values $<$ 0.001 (based on 500 iterations) suggesting that the chances of overfitting are small. The performance of models, obtained using descriptors calculated for the more recent PM7[76] method was also investigated. However, no discernible improvement improvement in the results was observed (see Table S7 in the Supplementary information).

In order to examine which variables are influencing the performances, an importance measure was computed based on the corresponding reduction in error when the predictor of interest is permuted. Figure 3 shows the top ranking features for the Cubist, GBM and RF models. Although, the influence of the different descriptors on the melting point behaviour is not directly obvious, they are nonetheless suggestive of a contributory effect. For both RF and GBM models, a mixture of both cation and anion-based descriptors are prominent and mirror the trends seen with respect to the first principal component. On the other hand, the top ranking features for the Cubist method are largely based on the cation. The chemical reactivity parameters such as the HOMO/LUMO energies are closely related to electrophilic/nucleophilic attack and the chemical hardness that is associated with the stability and reactivity of a chemical system feature prominently in all models. The atom specific delocalizabilities, which are dynamic reactivity indices are reflective of the energy stabilization due to electron redistribution during the formation of the complex

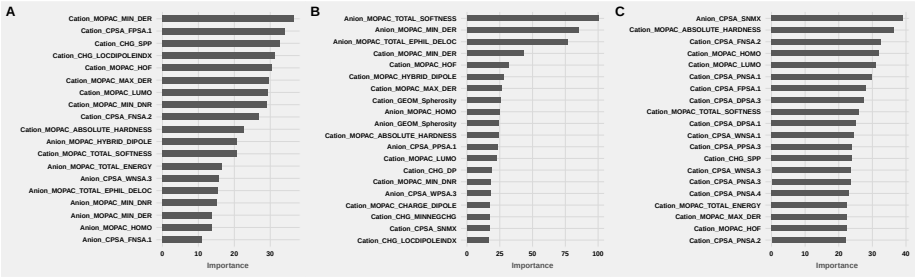

Figure 3: Prominent variables (top 20) influencing the model predictions (A) RF (B) GBM (C) Cubist. For Cubist models, the importance is taken as the percentage of times each variable was used in a condition and/or a linear model. For RF, the importance is calculated by permuting each predictor variable and computing the mean square error on the out-of-bag data for each tree.

with a second molecule[77]. These descriptors can be understood as quantifying the interactions between the cation and anion and are likely to influence the molecular packing and extent of hydrogen bonding[20, 75]. For instance, the alkylation pattern on the cations/anions can lead to a poor packing efficiency and contributes to low melting points[78]. Other geometric descriptors such as the spherosity which varies between 0 (flat molecules) and 1 (totally spherical) can be interpreted as the relative surface on which the charge is accessible to the counterion wherein the charge distribution can impact the melting points[34, 79].

Although a comparison with other published models was attempted, the lack of sufficient detail and unavailability of required parameters made it difficult to do so. Melting point estimations were also carried out using COSMOtherm which uses the mathematical model published by Preiss et al[13]. Of the 2212 ILs analysed, the COSMOtherm model was unable to provide estimates for 951 cases. For the remaining ILs, a mean absolute deviation of 89°C was obtained. Corresponding values for the RF, Cubist and GBM approaches are 21, 17 and 20 °C respectively. A popular approach to developing QSPRs has been the use of congeneric series of structures. Given that the imidazoliums form (645 ILs) the biggest structural group in the analysed data, ML models for only this class were examined. The predictive abilities of the models (RF and Cubist) see only a marginal improvement. However, for other structural groups such as the pyridiniums and bistriflimides, the results were relatively inferior (see Tables S4-S6 in the supplementary material). These results mirror the trends observed in an earlier article by Varnek et al[23].

The variability in the predictions can be attributed to a number of reasons. A better choice of descriptors such as those focusing on cation-anion interactions may be beneficial[80]. Attempts to include this by way of creating a product matrix containing product of every pair of cation-anion descriptor values, was however not found to improve the quality of predictions. A closer inspection of the literature shows that for a number of ILs, multiple melting point values were reported. 1-butyl-3-ethylimidazolium bis((trifluoromethyl)sulfonyl)imide for in-

Table 3: Summary of the classification model performances in terms of the κ statistic and accuracies (ACC).

Method	**Training**		**Testing**	
	κ	ACC	κ	ACC
PLSDA	0.64	0.83	0.57	0.79
RF	0.69	0.84	0.61	0.81
GBM	0.69	0.85	0.63	0.82

stance has reported melting points of -40 and -9°C[81, 82]. The experimental deviations are much larger in the case of 1-dodecyl-3-methylimidazolium chloride which has reported values of 97 and -3 °C[83, 84]. These variations have been attributed to a number of factors such as the purity of the ionic liquid sample being tested, its water content, formation of liquid-crystalline phases etc[14, 85]. In addition, there have been instances where glass transition temperatures (T_g) have been confused with T_m[40]. From a modelling perspective, these factors can cause the models to have poor generalizability[41].

2.7. Classification models

Given the moderate accuracy of regression models, a classification-based approach was used understand the structure-property relationships. The ionic liquids were divided into two classes according to the values of their melting points wherein those with melting points above 100°C were assigned to class A (876 ILs) and those below to class B (1336 ILs). Performance of classification models was gauged using both the *kappa* statistic[86] and accuracy. While both tree-based approaches yield similar metrics, the linear partial least squares discriminant analysis model performs comparably with accuracies for both calibration and test sets $> 80\%$. Comparing the performance with the regression models, one notes that the correlation coefficient is indeed much smaller and suggests a less than perfect relationship between the observed and predicted. However, the Spearman ranked correlations (calculating Pearson's correlation on the ranked values of the data) between the experimental and ML-predicted values are considerably better with values of 0.93, 0.93 and 0.94 for RF, GBM and Cubist regression models (across the entire data) and indicates a strong positive association between the trends seen for the two sets of values.

2.8. Experimental validation

Given the relative differences in performances of the classification and regression models, an independent verification of the predictive abilities was carried out. Five new ILs were synthesized and their melting points recorded (see Supplementary information for experimental details). A further 14 ILs were taken from recent literature[87–89]. The new ILs (cations and anions shown in Figure 4) span a temperature range of $-84 - 195$°C with nine ILs having very low melting points (close to 0°C or less), the rest had $MP > 50$. Table 4 shows the prediction results for the new ILs. Mean absolute errors of 56, 50 and 45°C

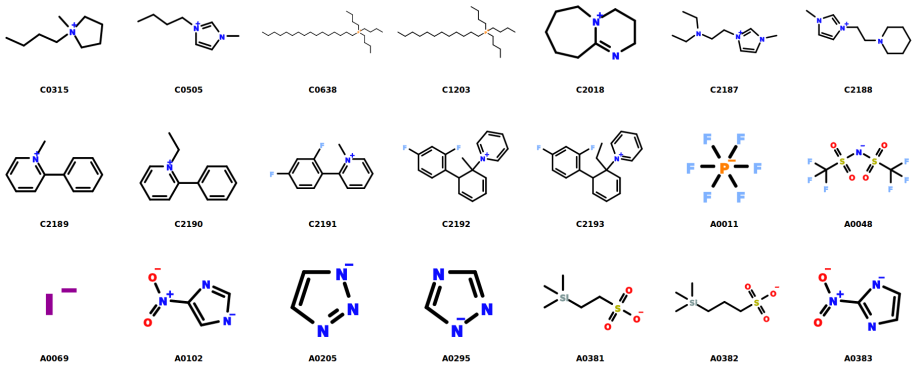

Figure 4: List of cations and anions used as part of the external validation.

are seen for the RF, Cubist and GBM models respectively. Classification models based on the GBM and PLSDA yield better performances with the former showing an accuracy of 84%.

2.9. Applicability Domain Analysis

Owing to a constrained chemical space coverage of the calibration data, the developed machine learning models need to be applied with caution[90, 91]. In order to establish the boundaries within which the model predictions can be trusted, a probability-oriented distance-based[92] approach was examined. Here, the structural similarity of a given compound (calculated as the average Euclidean distance in the $n-$dimensional descriptor space) to the training set ILs is used as a means to gauge whether the model prediction is reliable. In addition, the prediction uncertainty for each observation was calculated using bootstrapping[93, 94] as: $\widetilde{y_i} = \sqrt{\frac{\sum_{j=1}^{N}(\widehat{y_j}-\bar{y})}{N-1}}$ where, $\widehat{y_j}$ is the uncertainty in prediction associated with object i, N is the number of bootstrap ML models (set to 100 in this study), $\widehat{y_j}$ is the prediction for the j^{th} model and $\bar{y}$, the mean of predictions obtained by using the complete training set. Predictions with small standard deviations are typically considered to be reliable.

Figure 5 shows the applicability domain regions for the GBM model, wherein the reliable prediction space is obtained by computing the 95% and 99% confidence interval (based on Student's t-distribution) boundaries with respect to the calibration data. For the larger test set, 64 predictions are flagged as being outside the 95% confidence interval boundary while the corresponding number for the second validation set is 4. Thus, the GBM model accounts for nearly 90% and 79% reliable predictions across the two test sets. Examination of the structures for which the GBM model predictions were flagged as unreliable, shows that a number of cation/anion chemistries are potentially underrepresented in the calibration. For instance, the ILs based on the organosilanesulfonate anions (A0381-A0382) have large deviations and are also flagged as outside the respective applicability domains for the Cubist, SVR and RF models (see Figures

Table 4: Experimental and predicted melting points for 19 new ionic liquids. Predictions for both regression and classification models are reported. Here $MP > 100°C$ are assigned to class A (above) and B (below) otherwise.

Cation	Anion	Regression					Classification			Reference
		Expt.	RF	Cubist	GBM	COSMO	PLSDA	GBM	RF	
C0505	A0295	-68	-10	-20	-21	-105	B	B	B	This study
C0505	A0205	-70	-5	1	2	-105	B	B	B	This study
C0315	A0295	-68	41	-27	-14	-102	B	B	B	This study
C2187	A0048	-84	17	16	2	-13	B	B	B	This study
C2188	A0048	-67	20	15	15	-5	B	B	B	This study
C2189	A0069	150	120	159	147	NA	A	A	A	[88]
C2189	A0011	170	112	119	130	115	B	A	A	[88]
C2190	A0069	70	126	170	152	NA	A	A	A	[88]
C2190	A0011	130	112	112	142	89	B	A	A	[88]
C2191	A0069	195	113	104	161	NA	A	A	A	[88]
C2192	A0011	225	89	155	135	143	A	A	B	[88]
C2193	A0069	130	115	157	167	NA	A	B	B	[88]
C2193	A0011	145	95	158	149	98	A	A	B	[88]
C1203	A0381	-6	29	32	17	4	B	B	B	[87]
C0638	A0381	5	29	41	28	2	B	B	B	[87]
C1203	A0382	2	32	46	33	-11	B	B	B	[87]
C0638	A0382	-5	31	53	41	-11	B	B	B	[87]
C2018	A0102	50	94	95	40	-118	B	B	B	[89]
C2018	A0383	100	73	94	30	-98	B	B	B	[89]

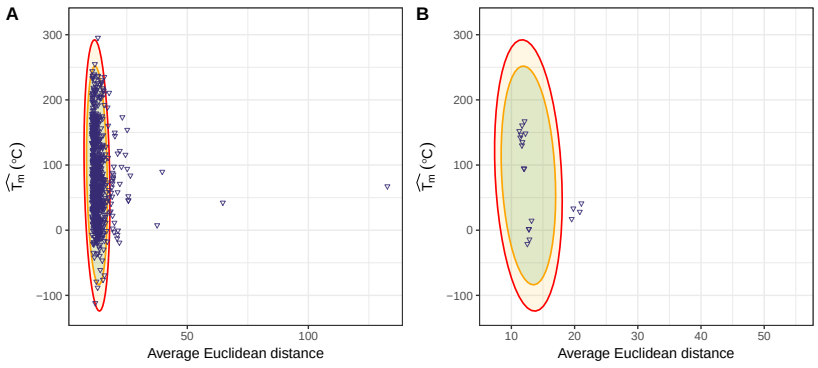

Figure 5: Applicability domain boundaries for the (A) test set and (B) external validation set based on the GBM model. Compounds located within the 95% (green zone) and 99% (orange zone) confidence intervals are said to be inside the domain (reliable) of the model.

F2-F4 in the Supplementary material). Examination of the prediction uncertainties with respect to Cubist, GBM, RF and SVR models (see Figure F5 in the Supplementary material) show that approximately 39%,33%,62% and 40% of the compounds exhibit uncertainties and absolute deviations less than 20°. This illustrates the trend towards compounds with the smallest uncertainties having relatively small prediction errors. For cases with large deviations (say greater than 100°), the uncertainty estimates are less precise and therefore less effective as a technique to identify unreliable predictions.

3. Conclusions

This paper presents a machine learning framework for the prediction of the melting temperatures of a large and diverse set of ionic liquids. To the best of our knowledge, the ILs used in this study form one of the largest data sets analysed. Despite the moderate accuracy of prediction, the developed models can predict the melting point trends reasonably well and should facilitate high-throughput screening of compounds in search for new potential ILs. In this article, we have mainly focused on ILs with a single cation and anion. Recently, there has been much interest in mixtures of ionic liquids[95] that not only increases the synthetic flexibility but also provides new avenues for their application. Although experimental data in this case is somewhat limited, it would nonetheless be interesting to see whether machine learning models can be extended successfully to the modelling of properties of IL mixtures.

Dedication

This manuscript is dedicated to the memory of Prof. Bjørn Kåre Alsberg who passed away shortly after submission. He is greatly missed by those of us who worked closely with him.

Acknowledgements

The Norwegian Research Council (NFR) is acknowledged for financial support from the CLIMIT (Grant No. 233776). We also thank ChemAxon (http://www.chemaxon.com) for free academic use of the Marvin package.

References

[1] M. Watanabe, M. L. Thomas, S. Zhang, K. Ueno, T. Yasuda, K. Dokko, Application of ionic liquids to energy storage and conversion materials and devices, Chem. Rev. 117 (10) (2017) 7190–7239. doi:10.1021/acs.chemrev.6b00504.

[2] Z. Lei, C. Dai, B. Chen, Gas solubility in ionic liquids, Chem. Rev. 114 (2) (2014) 1289–1326.

[3] Q. Zhang, S. Zhang, Y. Deng, Recent advances in ionic liquid catalysis, Green Chem. 13 (10) (2011) 2619. doi:10.1039/c1gc15334j.

[4] Y. Zhou, J. Qu, Ionic liquids as lubricant additives: A review, ACS Appl. Mater. Interfaces 9 (4) (2017) 3209–3222. doi:10.1021/acsami.6b12489.

[5] W. L. Hough, M. Smiglak, H. Rodríguez, R. P. Swatloski, S. K. Spear, D. T. Daly, J. Pernak, J. E. Grisel, R. D. Carliss, M. D. Soutullo, J. James H. Davis, R. D. Rogers, The third evolution of ionic liquids: active pharmaceutical ingredients, New J. Chem. 31 (8) (2007) 1429. doi:10.1039/b706677p.

[6] Y. Sahbaz, H. D. Williams, T.-H. Nguyen, J. Saunders, L. Ford, S. A. Charman, P. J. Scammells, C. J. H. Porter, Transformation of poorly water-soluble drugs into lipophilic ionic liquids enhances oral drug exposure from lipid based formulations, Mol. Pharm. 12 (6) (2015) 1980–1991. doi:10.1021/mp500790t.

[7] K. M. Gupta, J. Jiang, Cellulose dissolution and regeneration in ionic liquids: A computational perspective, Chem. Eng. Sci. 121 (2015) 180–189. doi:10.1016/j.ces.2014.07.025.

[8] A. A. C. T. Hijo, G. J. Maximo, M. C. Costa, E. A. C. Batista, A. J. A. Meirelles, Applications of ionic liquids in the food and bioproducts industries, ACS Sustainable Chem. Eng. 4 (10) (2016) 5347–5369. doi:10.1021/acssuschemeng.6b00560.

[9] J. P. Hallett, T. Welton, Room-temperature ionic liquids: Solvents for synthesis and catalysis. 2, Chem. Rev. 111 (5) (2011) 3508–3576. doi:10.1021/cr1003248.

[10] K. R. Seddon, Ionic liquids for clean technology, J. Chem. Technol. Biotechnol. 68 (4) (1997) 351–356. doi:10.1002/(SICI)1097-4660(199704)68:4<351::AID-JCTB613>3.0.CO;2-4.

[11] T. Le, V. C. Epa, F. R. Burden, D. A. Winkler, Quantitative structure–property relationship modeling of diverse materials properties, Chem. Rev. 112 (5) (2012) 2889–2919. doi:10.1021/cr200066h.

[12] N. Kireeva, S. L. Kuznetsov, A. Y. Tsivadze, Toward navigating chemical space of ionic liquids: Prediction of melting points using generative topographic maps, Ind. Eng. Chem. Res. 51 (44) (2012) 14337–14343. doi:10.1021/ie3021895.

[13] U. Preiss, S. Bulut, I. Krossing, In silico prediction of the melting points of ionic liquids from thermodynamic considerations: A case study on 67 salts with a melting point range of 337°c, J Phys. Chem. B 114 (34) (2010) 11133–11140. doi:10.1021/jp104679m.

[14] U. P. Preiss, W. Beichel, A. M. T. Erle, Y. U. Paulechka, I. Krossing, Is universal, simple melting point prediction possible?, Chem. Phys. Chem. 12 (16) (2011) 2959–2972. doi:10.1002/cphc.201100522.

[15] F. Eckert, A. Klamt, Fast solvent screening via quantum chemistry: COSMO-RS approach, AIChE Journal 48 (2) (2002) 369–385. doi:10.1002/aic.690480220.

[16] S. Alavi, D. L. Thompson, Molecular dynamics studies of melting and some liquid-state properties of 1-ethyl-3-methylimidazolium hexafluorophosphate [emim][PF6], J Chem. Phys. 122 (15) (2005) 154704. doi:10.1063/1.1880932.

[17] Y. Zhang, E. J. Maginn, The effect of c2 substitution on melting point and liquid phase dynamics of imidazolium based-ionic liquids: insights from molecular dynamics simulations, Phys. Chem. Chem. Phys. 14 (35) (2012) 12157. doi:10.1039/c2cp41964e.

[18] E. J. Maginn, Molecular simulation of ionic liquids: current status and future opportunities, J. Phys. Condens. Matter 21 (37) (2009) 373101. doi:10.1088/0953-8984/21/37/373101.

[19] L. Chen, V. S. Bryantsev, A density functional theory based approach for predicting melting points of ionic liquids, Phys. Chem. Chem. Phys. 19 (5) (2017) 4114–4124. doi:10.1039/c6cp08403f.

[20] S. Trohalaki, R. Pachter, Prediction of melting points for ionic liquids, Mol. Inf. 24 (4) (2005) 485–490. doi:10.1002/qsar.200430927.

[21] G. Carrera, J. ao Aires-de Sousa, Estimation of melting points of pyridinium bromide ionic liquids with decision trees and neural networks, Green Chem. 7 (1) (2005) 20. doi:10.1039/b408967g.

[22] R. Bini, C. Chiappe, C. Duce, A. Micheli, R. Solaro, A. Starita, M. R. Tiné, Ionic liquids: prediction of their melting points by a recursive neural network model, Green Chem. 10 (3) (2008) 306. doi:10.1039/b708123e.

[23] A. Varnek, N. Kireeva, I. V. Tetko, I. I. Baskin, V. P. Solov'ev, Exhaustive QSPR studies of a large diverse set of ionic liquids: how accurately can we predict melting points?, J Chem. Inf. Model. 47 (3) (2007) 1111–1122. doi:10.1021/ci600493x.

[24] F. Yan, S. Xia, Q. Wang, Z. Yang, P. Ma, Predicting the melting points of ionic liquids by the quantitative structure property relationship method using a topological index, J Chem. Therm. 62 (2013) 196–200. doi:10.1016/j.jct.2013.03.016.

[25] J. A. Lazzús, A group contribution method to predict the melting point of ionic liquids, Fluid Phase Equilib. 313 (2012) 1–6. doi:10.1016/j.fluid.2011.09.018.

[26] N. Farahani, F. Gharagheizi, S. A. Mirkhani, K. Tumba, Ionic liquids: Prediction of melting point by molecular-based model, Thermochimica Acta 549 (2012) 17–34. doi:10.1016/j.tca.2012.09.011.

[27] J. S. Torrecilla, F. Rodríguez, J. L. Bravo, G. Rothenberg, K. R. Seddon, I. López-Martin, Optimising an artificial neural network for predicting the melting point of ionic liquids, Phys. Chem. Chem. Phys. 10 (38) (2008) 5826. doi:10.1039/b806367b.

[28] J. O. Valderrama, C. A. Faúndez, V. J. Vicencio, Artificial neural networks and the melting temperature of ionic liquids, Ind. Eng. Chem. Res. 53 (25) (2014) 10504–10511. doi:10.1021/ie5010459.

[29] H. Yamamoto, Structure properties relationship of ionic liquid, J Comput. Aided Chem. 7 (2006) 18–30. doi:10.2751/jcac.7.18.

[30] C. L. Aguirre, L. A. Cisternas, J. O. Valderrama, Melting-point estimation of ionic liquids by a group contribution method, Int. J. Thermophys 33 (1) (2011) 34–46. doi:10.1007/s10765-011-1133-5.

[31] N. Sun, X. He, K. Dong, X. Zhang, X. Lu, H. He, S. Zhang, Prediction of the melting points for two kinds of room temperature ionic liquids, Fluid Phase Equilib. 246 (1-2) (2006) 137–142. doi:10.1016/j.fluid.2006.05.013.

[32] F. Gharagheizi, P. Ilani-Kashkouli, A. H. Mohammadi, Computation of normal melting temperature of ionic liquids using a group contribution method, Fluid Phase Equilib. 329 (2012) 1–7. doi:10.1016/j.fluid.2012.05.017.

[33] A. Mehrkesh, A. T. Karunanithi, New quantum chemistry-based descriptors for better prediction of melting point and viscosity of ionic liquids, Fluid Phase Equilib. 427 (2016) 498–503. doi:10.1016/j.fluid.2016.07.006.

[34] I. López-Martin, E. Burello, P. N. Davey, K. R. Seddon, G. Rothenberg, Anion and cation effects on imidazolium salt melting points: A descriptor modelling study, Chem. Phys. Chem. 8 (5) (2007) 690–695. doi:10.1002/cphc.200600637.

[35] A. R. Katritzky, A. Lomaka, R. Petrukhin, R. Jain, M. Karelson, A. E. Visser, R. D. Rogers, QSPR correlation of the melting point for pyridinium bromides, potential ionic liquids, J Chem. Inf. Model. 42 (1) (2002) 71–74. doi:10.1021/ci0100503.

[36] Y. Huo, S. Xia, Y. Zhang, P. Ma, Group contribution method for predicting melting points of imidazolium and benzimidazolium ionic liquids, Ind. Eng. Chem. Res. 48 (4) (2009) 2212–2217. doi:10.1021/ie8011215.

[37] Y. Ren, J. Qin, H. Liu, X. Yao, M. Liu, QSPR study on the melting points of a diverse set of potential ionic liquids by projection pursuit regression, Mol. Inf. 28 (11-12) (2009) 1237–1244. doi:10.1002/qsar.200710073.

[38] D. M. Eike, J. F. Brennecke, E. J. Maginn, Predicting melting points of quaternary ammonium ionic liquids, Green Chem. 5 (3) (2003) 323.

[39] S. Zhang, X. Lu, Q. Zhou, X. Li, X. Zhang, S. Li, Ionic Liquids Physicochemical Properties, Elsevier, Amsterdam, 2009.

[40] P. Wasserscheid, T. Welton, Ionic Liquids in Synthesis, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2008.

[41] J. O. Valderrama, Myths and realities about existing methods for calculating the melting temperatures of ionic liquids, Ind. Eng. Chem. Res. 53 (2) (2014) 1004–1014. doi:10.1021/ie403293z.

[42] O. D. Gupta, B. Twamley, J. M. Shreeve, Low melting and slightly viscous ionic liquids via protonation of trialkylamines by perfluoroalkyl *beta*-diketones, Tetrahedron Lett. 45 (8) (2004) 1733–1736. doi:10.1016/j.tetlet.2003.12.090.

[43] A. B. Pereiro, J. M. M. Araújo, S. Martinho, F. Alves, S. Nunes, A. Matias, C. M. M. Duarte, L. P. N. Rebelo, I. M. Marrucho, Fluorinated ionic liquids: Properties and applications, ACS Sustainable Chem. Eng. 1 (4) (2013) 427–439. doi:10.1021/sc300163n.

[44] E. I. Izgorodina, Z. L. Seeger, D. L. A. Scarborough, S. Y. S. Tan, Quantum chemical methods for the prediction of energetic, physical, and spectroscopic properties of ionic liquids, Chem. Rev. 117 (10) (2017) 6696–6754. doi:10.1021/acs.chemrev.6b00528.

[45] V. Venkatraman, B. K. Alsberg, Krakenx: software for the generation of alignment-independent 3d descriptors, J. Mol. Model. 22 (4) (2016) 1–8.

[46] J. J. P. Stewart, Optimization of parameters for semiempirical methods v: Modification of NDDO approximations and application to 70 elements, J. Mol. Model. 13 (12) (2007) 1173–1213. doi:10.1007/s00894-007-0233-4.

[47] Marvin 5.9.3, chemAxon (http://www.chemaxon.com) (2012).

[48] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open babel: An open chemical toolbox, J. Cheminf. 3 (1) (2011) 33.

[49] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. G. III, W. M. Skiff, Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations, J. Am. Chem. Soc. 114 (25) (1992) 10024–10035.

[50] J. J. P. Stewart, Mopac2016, stewart Computational Chemistry, Colorado Springs, CO, USA, (http://OpenMOPAC.net) (2016).

[51] M. Shen, Y. Xiao, A. Golbraikh, V. K. Gombar, A. Tropsha, Development and validation ofk-nearest-neighbor QSPR models of metabolic stability of drug candidates, J. Med. Chem. 46 (14) (2003) 3013–3020.

[52] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2017).
URL https://www.R-project.org/

[53] B.-H. Mevik, R. Wehrens, The pls package: Principal component and partial least squares regression in r, J. Stat. Soft. 18 (2) (2007) 1–24.
URL http://www.jstatsoft.org/v18/i02

[54] A. Karatzoglou, A. Smola, K. Hornik, A. Zeileis, kernlab – an S4 package for kernel methods in R, J. Stat. Soft. 11 (9) (2004) 1–20.

[55] A. Liaw, M. Wiener, Classification and regression by randomforest, R News 2 (3) (2002) 18–22.

[56] G. R. with contributions from others, gbm: Generalized Boosted Regression Models, r package version 2.1.1 (2015).
URL https://CRAN.R-project.org/package=gbm

[57] M. Kuhn, S. Weston, C. Keefer, N. C. C. code for Cubist by Ross Quinlan, Cubist: Rule- And Instance-Based Regression Modeling, r package version 0.0.19 (2016).
URL https://CRAN.R-project.org/package=Cubist

[58] L. Mouselimis, KernelKnn: Kernel k Nearest Neighbors, r package version 1.0.5 (2017).
URL https://CRAN.R-project.org/package=KernelKnn

[59] J. M. Andrade-Garda, A. Carlosena-Zubieta, R. Boqué-Martí, J. Ferré-Baldrich, CHAPTER 5. partial least&#8208:squares regression, in: RSC Analytical Spectroscopy Series, Royal Society of Chemistry, 2013, pp. 280–347. doi:10.1039/9781849739344-00280.

[60] R. G. Brereton, G. R. Lloyd, Partial least squares discriminant analysis: taking the magic away, J. Chemometrics 28 (4) (2014) 213–225. doi:10.1002/cem.2609.

[61] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[62] R. Genuer, J.-M. Poggi, C. Tuleau-Malot, VSURF: Variable Selection Using Random Forests, r package version 1.0.3 (2016).

[63] K. Yu, L. Ji, X. Zhang, Kernel nearest-neighbor algorithm, Neural Process. Lett. 15 (2) (2002) 147–156. doi:10.1023/A:1015244902967.

[64] M. Shen, A. LeTiran, Y. Xiao, A. Golbraikh, H. Kohn, A. Tropsha, Quantitative structure-activity relationship analysis of functionalized amino acid anticonvulsant agents UsingkNearest neighbor and simulated annealing PLS methods, J. Med. Chem. 45 (13) (2002) 2811–2823. doi:10.1021/jm010488u.

[65] J. H. Friedman, Stochastic gradient boosting, Comput. Stat. Data Anal. 38 (4) (2002) 367–378. doi:10.1016/s0167-9473(01)00065-2.

[66] R. J. Quinlan, Learning with continuous classes, in: 5th Australian Joint Conference on Artificial Intelligence, World Scientific, Singapore, 1992, pp. 343–348.

[67] G. Holmes, M. Hall, E. Frank, Generating rule sets from model trees, in: Twelfth Australian Joint Conference on Artificial Intelligence, Springer, 1999, pp. 1–12.

[68] V. Venkatraman, B. K. Alsberg, Quantitative structure-property relationship modelling of thermal decomposition temperatures of ionic liquids, J. Mol. Liquids 223 (2016) 60–67.

[69] M. K. C. from Jed Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, T. Hunt., caret: Classification and Regression Training, r package version 6.0-73 (2016). URL https://CRAN.R-project.org/package=caret

[70] F. Eckert, A. Klamt, Cosmotherm version c3.0, release 16.01, cOSMOlogic GmbH & Co. KG, Leverkusen, Germany (2015).

[71] F. Neese, F. Wennmohs, Orca version 3.0.3, max Planck Institute for Chemical Energy Conversion, Germany (2015).

[72] F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy, Phys. Chem. Chem. Phys. 7 (18) (2005) 3297.

[73] H. Abdi, L. J. Williams, Principal component analysis, Wiley Interdiscip. Rev. Comput. Stat. 2 (4) (2010) 433–459. doi:10.1002/wics.101.

[74] R. Todeschini, V. Consonni, Descriptors from molecular geometry, in: Handbook of Chemoinformatics, Wiley-VCH Verlag GmbH, 2008, pp. 1004–1033. doi:10.1002/9783527618279.ch37.

[75] P. Wasserscheid, W. Keim, Ionic liquids-new solutions for transition metal catalysis, Angew Chem. Int. Ed. 39 (21) (2000) 3772–3789. doi:10.1002/1521-3773(20001103)39:21<3772::AID-ANIE3772>3.0.CO;2-5.

[76] J. J. P. Stewart, Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters, J. Mol. Model. 19 (1) (2012) 1–32. doi:10.1007/s00894-012-1667-x.

[77] M. Karelson, V. S. Lobanov, A. R. Katritzky, Quantum-chemical descriptors in QSAR/QSPR studies, Chem. Rev. 96 (3) (1996) 1027–1044. doi:10.1021/cr950202r.

[78] A. S. Larsen, J. D. Holbrey, F. S. Tham, C. A. Reed, Designing ionic liquids: imidazolium melts with inert carborane anions, J. Am. Chem. Soc. 122 (30) (2000) 7264–7272. doi:10.1021/ja0007511.

[79] H. Stegemann, A. Rohde, A. Reiche, A. Schnittke, H. Fllbier, Room temperature molten polyiodides, Electrochimica Acta 37 (3) (1992) 379–383. doi:10.1016/0013-4686(92)87025-u.

[80] S. Martin, H. D. Pratt, T. M. Anderson, Screening for high conductivity/low viscosity ionic liquids using product descriptors, Mol. Inf. 36 (7) (2017) 1600125. doi:10.1002/minf.201600125.

[81] U. Domańska, E. Bogel-Łukasik, R. Bogel-Łukasik, 1-octanol/water partition coefficients of 1alkyl-3-methylimidazolium chloride, Chem. Euro. J. 9 (13) (2003) 3033–3041. doi:10.1002/chem.200204516.

[82] A. E. Bradley, C. Hardacre, J. D. Holbrey, S. Johnston, S. E. J. McMath, M. Nieuwenhuyzen, Small-angle x-ray scattering studies of liquid crystalline 1-alkyl-3-methylimidazolium salts, Chem. Mater. 14 (2) (2002) 629–635. doi:10.1021/cm010542v.

[83] P. Bonhôte, A.-P. Dias, N. Papageorgiou, K. Kalyanasundaram, M. Grätzel, Hydrophobic, highly conductive ambient-temperature molten salts, Inorg. Chem. 35 (5) (1996) 1168–1178. doi:10.1021/ic951325x.

[84] A. Berthod, M. Ruiz-Ángel, S. Carda-Broch, Ionic liquids in separation techniques, J. Chromatogr. A 1184 (1-2) (2008) 6–18. doi:10.1016/j.chroma.2007.11.109.

[85] S. Handy (Ed.), Ionic Liquids - Classes and Properties, InTech, 2011. doi:10.5772/853.

[86] M. L. McHugh, Interrater reliability: the kappa statistic, Biochemia Medica (2012) 276–282doi:10.11613/bm.2012.031.

[87] N. Saurín, I. Minami, J. Sanes, M. Bermúdez, Study of the effect of tribomaterials and surface finish on the lubricant performance of new halogen-free room temperature ionic liquids, Appl. Surf. Sci. 366 (2016) 464–474. doi:10.1016/j.apsusc.2016.01.127.

[88] P. Dreyse, A. Alarcón, A. Galdámez, I. González, D. Cortés-Arriagada, F. Castillo, A. Mella, Influence of the anion nature and alkyl substituents in the behavior of ionic liquids derived from phenylpyridines, J. Mol. Struct. 0 (0) (2017) 0–0. doi:10.1016/j.molstruc.2017.10.062.

[89] X. Zhu, M. Song, Y. Xu, DBU-based protic ionic liquids for CO2 capture, ACS Sustainable Chem. Eng. 5 (9) (2017) 8192–8198. doi:10.1021/acssuschemeng.7b01839.

[90] M. Toplak, R. Močnik, M. Polajnar, Z. Bosnić, L. Carlsson, C. Hasselgren, J. Demšar, S. Boyer, B. Zupan, J. Stålring, Assessment of machine learning reliability methods for quantifying the applicability domain of QSAR regression models, J. Chem. Inf. Model. 54 (2) (2014) 431–441. doi:10.1021/ci4006595.

[91] U. Sahlin, M. Filipsson, T. Öberg, A risk assessment perspective of current practice in characterizing uncertainties in QSAR regression predictions, Mol. Inf. 30 (6-7) (2011) 551–564. doi:10.1002/minf.201000177.

[92] A. Gajewicz, How to judge whether QSAR/read-across predictions can be trusted: a novel approach for establishing a model's applicability domain, Env. Sci. Nanodoi:10.1039/c7en00774d.

[93] V. Venkatraman, M. Gupta, M. Foscato, H. F. Svendsen, V. R. Jensen, B. K. Alsberg, Computer-aided molecular design of imidazole-based absorbents for CO_2 capture, International Journal of Greenhouse Gas Control 49 (2016) 55–63. doi:10.1016/j.ijggc.2016.02.023.

[94] S. Wager, T. Hastie, B. Efron, Confidence intervals for random forests: The jackknife and the infinitesimal jackknife, J. Mach. Learn. Res. 15 (1) (2014) 1625–1651.

[95] H. Niedermeyer, J. P. Hallett, I. J. Villar-Garcia, P. A. Hunt, T. Welton, Mixtures of ionic liquids, Chem. Soc. Rev. 41 (23) (2012) 7780. doi:10.1039/c2cs35177c.