Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity

Kamil Paduszyński*

Department of Physical Chemistry, Faculty of Chemistry, Warsaw University of Technology, Noakowskiego 3, 00-664 Warsaw, Poland

S *Supporting Information*

ABSTRACT: New quantitative structure−property relationships (QSPRs) for estimating dynamic viscosity (η) of pure ionic liquids (ILs) as a function of temperature and group contributions (GCs) are presented and evaluated. The correlations were established using three common machine learning algorithms (stepwise multiple linear regression, feed-forward artificial neural network, and least-squares support vector machine) on the basis of the largest database reported thus far, including the data for 2068 distinct ILs (3236 data sets and 22 268 data points). The GC scheme as well as two-stage modeling protocol (representing the property using separate reference term and temperature correction models) were applied consistently with the previous contribution [*Ind. Eng. Chem. Res.* **2019**, *58*, 5322−5338]. Standard internal and external validation techniques (such as, *K*-fold cross-validation, *y*-scrambling, "hold-out" testing, and the Williams plot) were adopted to select the best set of GCs, hence statistically the most significant model. The impact of the chemical structure of both cations and anions (as well as their combination) on the accuracy of prediction and classification (with respect to the order of magnitude of η) is analyzed in detail. The obtained models are compared with other methods reported in the literature. In particular, a broad comparison of the finally recommended model with the QSPR, employing descriptors derived from molecular geometry and charge distribution [*J. Phys. Chem. B* **2011**, *115*, 300−309] is given.
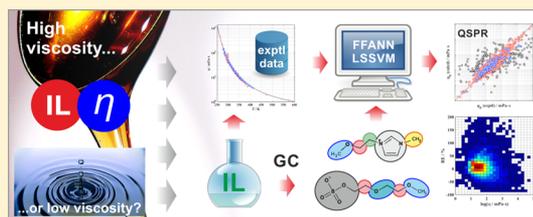
INTRODUCTION

Ionic liquids (henceforth abbreviated as ILs) are organic salts exhibiting a number of unique and peculiar properties compared to conventional solvents known in organic chemistry.[1−3] The key property which makes ILs special and distinguishes them from inorganic salts is the relatively low temperature at which they form a liquid phase (in many cases lower than 0 °C). As being ionic rather than molecular fluids, ILs are characterized by extremely low volatility and enhanced thermal stability. These features have made ILs being perceived as potential solvents for many applications in green chemistry and sustainable chemical engineering,[4−6] particularly in separations.[7−9] Furthermore, ILs are often named as "designer" solvents, because their properties can be tailored by a proper selection of chemical structure of constituting ions, for example, cation/anion cores, the length of alkyl chains, and task-specific functional groups. Thus far, more than two thousand distinct ILs have been characterized.[10] However, considering only the ions described in the literature, the number of all possible binary cation−anion combinations is at least two orders of magnitudes higher. Therefore, there is a substantial need to explore relevant physical and thermodynamic properties of the ILs still unknown. Of course, there is no doubt that experimental studies (syntheses, measurements, etc.) will always be the most reliable source on data on structure−property relationships in chemistry. Nevertheless, in the times of rapid growth of the impact of computers on research and development of diverse branches of science, the field of ILs as well has to be effectively and systematically investigated using the predictive models and other computational tools.[11−17]

Dynamic viscosity (η) is a key process parameter from the point of view of basically all unit operations that can be encountered in chemical engineering, in particular in those where effective fluid transport and/or mixing are crucial. In the case of ILs, a detailed knowledge of η and its temperature/pressure dependence is particularly important. In fact, high η is usually one of the obstacles in real-world applications of these solvents. Therefore, this is an urgent scientific task for the ILs community to develop the tools and quantitative structure−property relationships (QSPRs) allowing for performing computer-aided molecular design (CAMD) of low-viscous ILs as well as optimizing the operating conditions in such way to get favorable energy and material balance of the process, thus its economics. Since the beginning of this century, a significant number of contributions on modeling $p−\eta−T$ behavior of ILs was reported.[18−55] In particular, five years ago I co-authored the article presenting extensive machine learning-based correlation for predicting $p−\eta−T$ data for ILs.[36] The method has been recently successfully applied in CAMD of IL-based separations as an auxiliary tool for thermodynamic calculations.[56−58] Since

ref 36, many interesting articles on modeling IL viscosity have been published. In the following paragraphs, I present a concise (but critical) review of the recent, that is, since 2015, progress in this field.[37−55] Former contributions[18−35] are described and discussed in detail elsewhere.[36]

In 2015, Alcalde et al.[37] reported a simple QSPR for viscosity of 27 common ILs at a single temperature $T = 298.15$ K. The model was not validated using external data set, so it is quite difficult to assess its predictive capacity. In the same year, Lazzús and Pulgar-Villarroel presented a new group contribution (GC) method for predicting temperature-dependent viscosity data for ILs.[38] The model employs the common Andrade equation, namely, $\eta = A \exp(B/T)$, where A and B are the fluid-specific coefficients. Separate contributions to this equation due to cations and anions were defined, whereas the ion-specific coefficients were assumed to be the linear combinations of the GCs. In total, 20 cationic and 67 anionic functional groups were defined to represent the chemical structures of 326 distinct ILs. The final model was developed by regressing 1110 data points reported in the literature for 183 ILs (correlation set), whereas the remaining data (prediction set) were used for testing purposes. The values of overall average absolute relative error (AARE) between calculated and experimental η were 20.6 and 27.3% for correlation and prediction, respectively; it should be noted that the AARE values presented in ref 38 were much lower as they were computed using η expressed in ln-units. Although the results obtained by Lazzús and Pulgar-Villarroel[38] may be seen as very promising, one should note that their model was derived on the basis on relatively small number of data. Furthermore, the authors did not perform any cross-validation (CV) to check the model's stability toward the input data and they also did not provide any information of statistical significance of all GCs involved. Therefore, the model should be used with care, especially when applied beyond the range of the structures based on which it has been derived. Other QSPRs models for predicting η of ILs as a function of both temperature and pressure were proposed in the same year by Zhao et al.[39] Within the model, an IL was pictured in terms of molecular descriptors derived from screening charge density (σ) profiles, commonly used in conductor-like screening model for real solvents (COSMO-RS) thermodynamic model. Two machine learning algorithms, namely multiple linear regression (MLR) and support vector machine (SVM), were applied to find the relationship between η and the descriptors. The overall AAREs obtained (using log-units) were 10.7 and 6.6% in the case of MLR and SVM, respectively. When recomputed using absolute viscosity data, the corresponding values look less optimistic, namely, 59.5% for MLR and 30.2% in the case of SVM. Nevertheless, this level of accuracy should be perceived as satisfactory because it was achieved using only seven parameters. However, the number of ILs involved in the study was too small (89 ILs) to rely on the final model when CAMD of new ILs were used. Two years later, the same research group followed exactly the same methodology to develop new QSPRs, this time using other machine learning algorithm, namely, extreme learning machine (ELM).[42] A change in the computational method from SVM to ELM resulted in an improvement of prediction accuracy from 6.6 to 2.2%, when η is given in log-units. On the basis of the reported root-mean-squared error (RMSE), one can estimate that the latter value corresponds to AARE approximately equal to 20%, when η is expressed in absolute scale. In 2016, Esmaeili-Jaghdan et al.[40] utilized "committee machine intelligent system" (CMIS) to model viscosity of ILs. In this approach, viscosity is

computed by combining four different machine learning methods. Although the idea is very interesting, there are many drawbacks of ref 40 that should be highlighted in this review. First of all, the models were established by using a very small database (only 579 data points for 23 ILs), hence their applicability range is rather limited; this is in fact very surprising, taking into account the contributions published previously,[36] hence readily available data compilations. The next problem is that the models' input comprises a set of properties of ILs which are immeasurable, for example, critical temperature, boiling point temperature, and acentric factor; thus, their quality and uncertainty cannot be assessed or evaluated. What is even worse, the authors did not clearly state how these properties were estimated and did not provide any references to them. Furthermore, none of the details on the finally developed models (parameters, code) are presented, except the basic statistical measures of training and testing. Unfortunately, the authors did not provide any supplementary data (I have contacted them but received no response), so that this is unfeasible to reproduce results they presented. Finally, it looks like some calculations shown by Esmaeili-Jaghdan et al.[40] are simply of bad quality. In particular, for multilayer perceptron network, there are less than 3 data points per model parameter (0.8 × 579 ≈ 464 data points and 188 weights/biases for the networks with 17 nodes in the hidden layer), so that one may expect that the final model is over-fitted. Of course, this can be argued by very good results of the external validation (testing). However, such result can be due to an accidentally good testing set as well; in fact, the effect of data split on the modeling performance was not checked. Summing up, I strongly recommend not to rely on the models presented in ref 40 in making any predictions of viscosity of ILs. A novel method of temperature-dependent viscosity assessment for ILs was proposed by Barycki et al.[41] The approach consists of a typical QSPR for the value of η at reference temperature $T_0 = 298$ K and a separate equation accounting for an effect of T other than T_0. Temperature-correction term, however, requires the density of an IL at the temperature under study. In ref 41, it was estimated from molecular mechanics, which, in my opinion, is not a good choice for so complex fluids like ILs. The calculations were performed only for 23 ILs, so that its predictive capacity is doubtful. Furthermore, the authors did not compare their model's outcome with other available methods because they concluded:[41] "To our knowledge, this complex approach using QSPR and correction equation algorithm is the first one, allowing predicting the viscosity of ILs in various temperatures, based only on ILs' molecular structures." Obviously, this is not true, as till 2016, there were many better predictive QSPR viscosity models reported in the literature. Despite very limited data set used by Barycki et al.[41] in modeling the reference point viscosity, rather poor accuracy of AARE = 52% (absolute η) was achieved. When the effect of temperature is taken into account, AARE = 41%, which is lower but still very high for such a limited library of structures. Another QSPR, this time based on least squares SVM (LSSVM), was proposed in 2017 by Baghban et al.[43] Basically, their work is exactly the same as the article published previously by Gharagheizi et al.[30] except for the computational approach used. In fact, the authors used exactly the same database (1672 data points for 443 different ILs) and input variables. It has been confirmed that more sophisticated LSSVM approach gives much better results compared to MLR, that is "classical" QSPR tool, used by Gharagheizi et al.[30] Overall the AARE between computed and experimental η was about

1.5%, if η is given in log-units. Unfortunately, this is another article for which supplementary data were not provided. However, upon e-mail request, the table of calculated versus experimental viscosities was provided to me by the authors. On the basis of these data, one can easily get that AARE in the absolute η values is around 7.9%. In the same year, Fatehi et al.[44] proposed the model for estimation of both temperature- and pressure-dependent viscosity of ILs by using feed-forward artificial neural network (FFANN). For a relatively limited set of ILs (59 structures only), they obtained AARE in absolute η of 0.06 and 1.3% for training (612 data points) and testing (124 data points), respectively. These results seem to be extraordinary as they suggest that the optimal strategy for predicting viscosity of ILs is found. However, when the results published by Fatehi et al.[44] are analyzed more deeply, one may conclude that the ANN proposed by them is probably overfitted; this is the same case as in the aforementioned article by Esmaeili-Jaghdan et al.[40] In fact, the net comprising two hidden layers with 333 weights/biases was trained based on 551 data points, so there is less than 2 data points per parameter. Furthermore, the input data of the model by Fatehi et al.[44] are based mostly on molecular weight, whereas there is no information of an impact of specific functional groups on η. The way of numbering the chains required to define the input is also unclear and limited to a few common cation families only—for sure, this drawback will substantially hinder the model's applications. Recently, Yan and co-workers[45] presented the predictions of IL viscosity with the new QSPR based on descriptors derived from molecular chemical graphs and geometries of the ions. The authors analyzed the performance of their model by comparing its outcomes with 3228 experimental data points reported for 349 ILs. It is worth mentioning that the final equation incorporated only statistically significant coefficients selected using stepwise regression. Furthermore, the model was exhaustively validated by using standard techniques, such as CV, external validation, and y-scrambling. The AARE values of 4.7 and 5.7% for training and testing, respectively, for η are expressed in log-units. Overall AARE in absolute η was also quite higher, that is 19.0%. In general, the model proposed by Yan et al.[45] seems to be a very promising predictive tool, mostly due to its generalization abilities. However, it requires the optimized geometries of ions. Of course, different quantum chemical (QC) methods used to optimize the structures yield different geometries which may affect the value of η of a given IL resulting from Yan's method. Unfortunately, sensitivity of the predictions due to varying geometry was not studied by the authors. Furthermore, the structures used in the model development were not provided, so that reproducing the results presented in ref 45 is impossible. Very recently, Haghbakhsh and Raeissi[46] proposed a simple correlation for estimating temperature dependence of η for 1-alkyl-3-methylimidazolium-based ILs, similar to the correlations proposed previously by Dutt et al.[33] Compared to the previously discussed models, this method is capable of calculating η at any temperature, T given that viscosity at reference temperature, T_0 = 298 K and molecular weight of the IL is known. Therefore, this method cannot be used in CAMD of new ILs but can be utilized as a supporting tool for optimizing the process conditions. The model was examined by using a database containing 2073 data points extracted from the literature for 38 "plain" imidazolium ILs based on 10 distinct anions. The final equation with only four coefficients reproduces the experiment within 6.9% of AARE. Such accuracy is substantially better compared to a number of correlations of similar type published previously.

However, it accounts for temperature dependence only, rather than structural diversity of ions constituting ILs; the latter contribution is obviously much more difficult to be properly captured by both empirical correlations and theories.

Apart from the solely empirical models, mainly based on machine learning, some theoretical or semiempirical approaches for computing viscosity of ILs have been recently proposed as well. In particular, thermodynamic models such as cubic equations of state (EoS) or statistical associating fluid theory (SAFT) have been recently extensively applied as supporting tools for already known approaches for viscosity calculations, such as Eyring's absolute rate theory,[47,48] modified Yarranton–Satyro correlation (MYSC),[49,50] free-volume theory (FVT)[50–53] and frictional theory (FT).[50] Eyring's absolute rate theory combined with two common cubic EoS (Soave–Redlich–Kwong and Peng–Robinson) for calculating compressibility factor (Z) was applied by He et al.[47] to account for the pressure effect on viscosity of ILs. The proposed correlation was applied to regress the p–η–T data (up to T = 433 K and p = 300 MPa) for 25 ILs (1070 data points). The model allows to fit the experimental data with an excellent AARE of 1.5%, regardless of the method applied for estimating Z. Thus, it can serve as a very useful tool for correlating viscosity data for new ILs or other fluids. In order to increase its predictive capacity, the authors developed generalized correlations for the model coefficients. However, this correlations take into account alkyl chain length only, so that potential applications of the final model in CAMD of ILs are still very limited. A very similar work was presented shortly afterwards by Macias-Salinas.[48] More advanced, that is molecular-based, EoS-based models involving different versions of the SAFT approach were reported as well.[49–53] Polishuk[49] combined MYSC with "SAFT + Cubic" and showed that with this approach accurate viscosity predictions can be obtained for two homologous series of common (imidazolium +[BF$_4$], or [PF$_6$]) ILs in a wide range of pressure values, using a single set of universal parameters. MYSC was also applied by Abolala et al.,[50] who used using SAFT-VR-Mie EoS to represent the p–η–T data of an extended (by [NTf$_2$] anion) set of ILs. Overall AARE between calculated (fitted) and computed η was about 2.4%. This was an improvement compared to the work of Polishuk[49] whose approach yield AARE of 4.0%. However, Polishuk adopted the same model parameters for all fluids,[49] whereas Abolala et al.[50] fitted them to each IL separately. It is noteworthy that Abolala et al.[50] tested other viscosity models as well, including FVT and FT. According to their findings, MYSC seems to be the most promising approach.[50] FVT has been so far the most extensively used approach to compute η using SAFT-based EoS. For the very first time, it was adopted by the group of Vega[51,52] to model viscosity using soft-SAFT. Summing up, their contributions describe the calculations for 12 common ILs (imidazolium cation + [Cl], [BF$_4$], [CH$_3$SO$_3$] anions), accurate within 5% with regard to experimental data. Sun et al.[53] combined FVT with electrolyte ePC-SAFT EoS and provided a comprehensive study of this method comprising 89 ILs, diverse with respect to both cations and anions. The proposed parameterization scheme and modeling strategies allowed for predicting p–η–T with AARE around 5%, with the values varying from (1 to 10)% for individual ILs. To the best of my knowledge, the article by Sun et al.[53] presents the most advanced achievement in semitheoretical studies of viscosity of ILs. The predictive capacity of FVT-ePC-SAFT is limited only by the pure-ion parameters, which can be obtained from fundamental physical data (e.g., density) of respective pure ILs and their

Table 1. Summary of Cationic and Anionic Families of Ionic Liquids Contained in the $p-\eta-T$ Database Used in This Work

family name	code	number of distinct				
		molecules	counterions	ionic liquids	data sets	data points
		Cations				
imidazolium	[im]	331	149	781	1499	10 410
ammonium	[n]	198	83	416	540	2670
pyridinium	[py]	105	24	160	245	2393
phosphonium	[p]	69	106	233	292	2112
pyrrolidinium	[pyr]	54	45	133	247	1914
piperidinium	[pip]	35	25	67	102	725
morpholinium	[mo]	17	17	36	38	344
sulfonium	[s]	20	4	26	32	271
cyclopropanium	[cprop]	21	2	35	35	263
azepanium	[azp]	7	7	15	25	219
guanidinium	[guan]	19	12	28	30	211
triazolium	[trz]	17	8	24	26	175
bicyclic	[bic]	8	8	16	16	127
pyrazolium	[pz]	15	4	19	25	120
thiouronium	[thur]	14	2	15	17	98
cyclic sulfonium	[cs]	8	3	10	13	69
quinolinium	[quin]	7	4	9	9	58
thiazolium	[thz]	7	2	7	7	37
amidium	[amd]	5	2	9	9	16
cyclic amidium	[camd]	6	2	10	10	14
oxazolidinium	[ox]	4	6	12	12	12
piperazinium	[pipz]	2	2	2	2	5
cyclic phosphonium	[cp]	4	1	4	4	4
tetrazolium	[tetraz]	1	1	1	1	1
		Anions				
[NTf$_2$] derivatives	[NTf$_2$]	11	701	765	1321	8409
[BF$_4$] derivatives	[BF$_4$]	36	93	186	328	2195
[PF$_6$] derivatives	[PF$_6$]	5	45	56	145	2151
carboxylates	[RCO$_2$]	32	91	205	268	1794
sulfonates	[RSO$_3$]	27	95	153	205	1461
inorganics	[X]	10	123	156	223	1428
sulfates	[RSO$_4$]	14	74	100	167	1246
dicyanamides	[dca]	2	119	124	191	1170
heterocyclic amines	[hca]	17	29	57	65	537
phosphates	[RPO$_4$]	16	56	71	92	524
aminoacids	[AA]	20	20	66	79	471
metal complexes	[ML$_n$]	20	34	66	76	436
methanides	[CR$_3$]	6	8	13	26	225
alcoholates	[O]	26	10	33	33	125
organic borates	[B]	7	10	17	17	96

binary mixtures. Zhao et al.[54,55] demonstrated that other thermodynamic approach based on ion-specific parameters, namely, UNIFAC-VISCO, can be considered as a promising method for estimating viscosity of ILs. With the current version of the model, one is capable to estimate temperature-dependent viscosity data for 255 ILs composed from 17 distinct cations and 15 distinct anions. For 70 ILs whose data were used to adjust the model parameters, the AARE was 1.4% (based on 819 data points). Analysis of the distribution of the deviations showed that the accuracy of the predictions made for other compounds is expected to be below 5%, including pure ILs and their binary mixtures.

This work is a continuation of my research on comprehensive databases and QSPRs for ILs. Very recently, I published the extensive model for predicting IL density as a function of temperature and pressure, developed based on the largest data collection reported so far.[10] In this work, the study is extended to

viscosity—another property of ILs crucial for their applications in chemical engineering. In part, this article presents the revision and extension of the previously published model.[36] In particular, the new model presented herein is more universal in terms of functional groups, hence the chemical structures which can be represented as well. As in the case of my previous study on density,[10] two-stage modeling protocol is applied in order to investigate separately performance of the used tools when applied to predict the effect of chemical structure and temperature on η. Finally, much more attention (compared to ref 36) is paid to the revision and evaluation of the experimental data used in the modeling as well as to the proper validation of the final models with well-defined statistical procedures.

■ DATABASE

Overview. In the very first step of this study, I have carried out an extensive review of the open literature in order to prepare
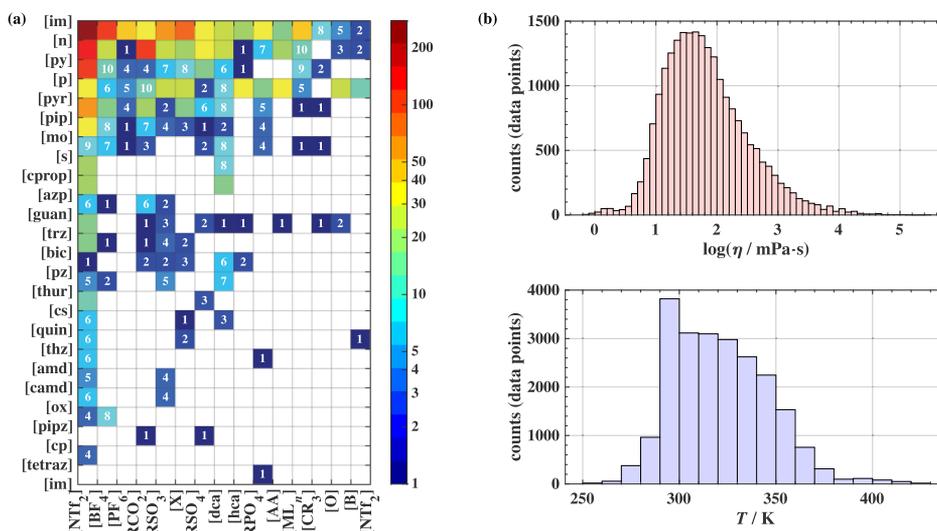
Figure 1. Overview of the viscosity (η) database used in this work. (a) Distribution of the number of ILs composed of the cations and anions belonging to different chemical families. Cationic and anionic families put on the ordinate and abscissa axes, respectively, abbreviated using the codes explained in Table 1. (b) Distribution of viscosity (upper panel) and temperature (lower panel) values in the database.

the IL viscosity database. The data for 2068 ILs were extracted from 680 papers published since 1984 to present; in total, 3236 data sets comprising 22 268 data points were collected. All details regarding the database can be found in the Microsoft Excel spreadsheet provided in the Supporting Information, including the lists of ions as well as their IUPAC names and SMILES codes. A brief summary of it is given in tabular form in Table 1 and concisely illustrated in Figure 1. From Table 1, one can easily notice that despite large diversity in both cationic and anionic families of ILs (974 and 249 moieties belonging to 24 and 15 distinct classes, respectively), there are actually only a few types of ions which has been investigated extensively so far. In particular, imidazolium-, ammonium-, pyridinium-, phosphonium-, and pyrrolidinium-based ILs cover more than 80% of all structures. Among anions, bistriflamide (commonly abbreviated by $[NTf_2]$) and its derivatives are the most "abundant" moieties in the database. In fact, almost one third of all ILs are the $[NTf_2]$-based ones. As seen from Figure 1a, great majority of the combinations of cation−anion families has not been studied at all yet—only in the case of 135 cation−anion pairs (out of 24 × 15 = 360), one can find any viscosity data in the literature. In many regions of the "chemical space" of ILs, the data are rather sparse. In fact, for 756 ILs, a single data point only can be found in the database, whereas only in the case of 372 ILs, at least two data sets reported by different authors are present. Figure 1b shows the distribution of viscosity data as well as temperature at which they were measured. As can be easily noticed, the η values cover around five orders of magnitude, namely, η = (0.7 to 257 000) mPa·s. In the case of temperature, almost all data ($\approx$90%) were measured and reported in the range from T = (290 to 370) K. Only for 1299 ILs, temperature-dependent η data were available.

The data on pressure-dependence of viscosity were available only for 29 ILs (in total, 47 data sets and 1593 data points). Because of such limited number of data, I decided to not account for an effect of pressure in the modeling presented in further text. In my opinion, applying the aforementioned methods such as FT, FVT, and MYSC combined with thermodynamic models is a much more proper and consistent way to represent $p−\eta−T$

data of fluids, not only ILs. This is one of the key differences of this work compared to my previous contribution.[36]

Data Preparation. Prior to the modeling, the database was revised and reduced using the same procedures as those described in my previous article on IL density.[10] In further text, I present some details specific to viscosity data only in brief.

Data Revision. The goal of the revision step was to exclude the data sets which: (i) seem to have very high or unknown uncertainty; (ii) do not fit either the data published by other authors or structure−property trends within a given family of ILs (e.g., homologous series). The case (i) regards the data sets with no clear statements on the sample's purity (e.g., water content) or experimental method used to measure η—such data sets have been rejected in the first place, unless there were no other data available. In order to address the criterion (ii), and in some part (i) as well, all data sets on dicationic and tricationic ILs (104 data sets and 315 data points for 94 ILs) were rejected. In fact, it was quite difficult to assess quality of the data, first of all, due to its sparsity. Furthermore, the data sets consisting of single data points were also excluded, if temperature-dependent data were available. For each IL, the remaining data were regressed, along with the Williams plot analysis following each regression used as a criterion for suspending some data sets from further consideration. The following Andrade-like or Vogel−Fülcher−Tammann (VFT) equations for representing temperature-dependent η data were applied

$$\ln(\eta/\mathrm{mPa·s}) = c_0 + c_1 (T/\mathrm{K})^{-1} + c_2 (T/\mathrm{K})^{-2} \qquad (1)$$

$$\ln(\eta/\mathrm{mPa·s}) = c_0 + c_1 (T/\mathrm{K} - c_2)^{-1} \qquad (2)$$

In the equations above, the symbols c_i (i = 0, 1, 2) stand for adjustable coefficients, which were iteratively included/excluded to provide the most significant fit. The type of equation, eqs 1 or 2 was established based on Akaike information criterion for model selection. Given the fitted versus experimental values of η, the data sets with the studentized residual higher than 3 for at least one data point were rejected. The process was iterated until none of the data sets met this criterion.

The initial database (i.e., only with dicationic/tricationic ILs and pressure-dependent data sets removed) covers 1974 ILs,

3085 data sets, and 20 045 data points. The collection of the data accepted based on the revision process covers all 1974 ILs, 2163 data sets, and 15 372 data points. In the Microsoft Excel spreadsheet provided in the Supporting Information, one can find all further details on the revised database as well as on the finally obtained fits for all ILs considered in this work, in particular the coefficients c_i, RMSEs, and AAREs between regressed and experimental values of viscosity

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\ln\frac{\hat{\eta}_i}{\eta_i}\right)^2} \tag{3}$$

$$\text{AARE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{\eta}_i}{\eta_i} - 1\right| \times 100\% \tag{4}$$

where "hat" is assigned to calculated property, whereas N stands for the number of data points. In general, the elaborated fits offer very accurate representation of the experimental (accepted) viscosity data. In particular, the 95th percentile of AARE is 3.6%. In the case of 63 ILs, higher AARE was obtained, mostly due to significant scatter of the measured data or significant amount of the data in the high-temperature region, where the values of η are much lower, hence relative errors (REs) are higher. Nevertheless, I still recommend the fits to be served as the IL-specific correlations for generating reference viscosity data, for example, for comparative analyses of different modeling tools.

An example of the performed data revision is shown in Figure 2 for 1-butyl-3-methylimidazolium bistriflamide, $[C_4C_1Im]$-$[NTf_2]$, that is, the IL with the highest number of data reported in the database and $[C_4C_1Im][PF_6]$, that is also extensively studied IL with extraordinarily significant data scatter. As seen from Figure 2a, a majority of the data for $[C_4C_1Im][NTf_2]$ (namely, 29 out of 44 data sets) was rejected and followed by establishing statistically significant and accurate fit (RMSE = 0.0225, AARE = 1.7%), with only a few points that could be suspended due to their high leverage. The data for $[C_4C_1Im]$-$[PF_6]$ (in total, 33 data sets with only 11 of them accepted) were regressed with noticeably worse accuracy (RMSE = 0.130, AARE = 10.4%), see Figure 2b. Such result was obtained following the above-described iterative procedure, and I did not find any further premises to improve it, for example, by rejecting other data sets. Additionally, I checked the consistency of viscosity data within the $[C_nC_1Im][PF6]$ series—the results are presented in the inset in Figure 2b. As seen, the datum obtained for the case $n = 4$ follows the observed trend (approximately linear increase of $\ln\eta$ with n), so that one may presume that the experimental data were revised and selected properly. However, the prediction confidence interval (95%) is quite wide because of the scatter of the accepted data. This can be explained by noticeable differences in sample purity and/or its purification procedure—this is a well-known fact that $[PF_6]$-based ILs may be chemically unstable in the presence of moisture.[59]

It is worth mentioning that the finally revised data are treated in further modeling as the data for pure ILs. However, one should keep in mind that the data correspond to real samples containing impurities, thus the data show uncertainty. Because not all ILs were characterized the same way (some data sets are published but do not contain any sample information at all), any variables accounting for impurity/uncertainty content were not taken into account. Nevertheless, the models proposed using such simplification can still be used as either benchmark or design supporting tools, as suggested elsewhere.[10]
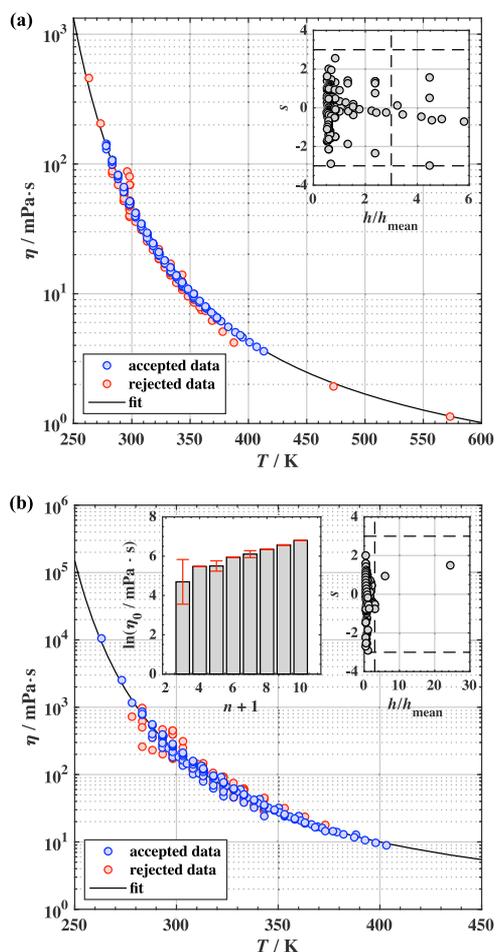
Figure 2. An example of viscosity (η) vs temperature (T) data revision for (a) 1-butyl-3-methylimidazolium bistriflamide and (b) 1-butyl-3-methylimidazolium hexafluorophosphate. Inset in panels (a,b): Williams plot of studentized residuals s resulting from the fit vs leverage h of each data point scaled by its mean value; dashed lines determined by the critical values of both h and s for data acceptance. Inset in panel (b): reference viscosity η_0 of $[C_nC_1Im][PF_6]$ series, where n stands for the number of carbon atoms in n-alkyl chain.

Data Reduction. Within the database "reduction", the raw and accepted viscosities were replaced by the values generated using the fits obtained in the revision step. Such procedure guaranties that all of the ILs will have approximately the same impact on the final models. For each IL with the fit available, five data points were computed at equidistant values of T between upper and lower limit of the experimental data pool. In the case of ILs with no temperature-dependent data available, raw data were directly employed. In total, the size of the reduced database used in the following modeling was 6826 data samples.

Each data point (T, η) was transformed into a pair (η_0, f), where η_0 denotes viscosity at reference temperature, $T_0 = 298.15$ K, whereas f is correction accounting for temperature change from T_0 to any T, defined as

$$\eta(T) = \eta_0 f(T) \tag{5}$$

The value of η_0 was obtained from the fits developed during the revision stage, whereas the corrections were simply calculated following the definition given in eq 5, that is $f = \eta/\eta_0$. In the case of ILs with the fit unavailable, raw experimental data were taken as η_0, given that the temperature at which they were measured

was $T_0 \pm 5$ K. The ILs not fulfilling this condition were excluded from further analyses. Those compounds possibly exist are solid at temperature as low as T_0, so that modeling them using the same approach as applied for liquids does not make any sense. Summary of the reduced viscosity database along with the corresponding values of η_0 and f is presented in the Microsoft Excel spreadsheet provided in the Supporting Information.

■ **MODELING**

GC Scheme. In this work, the QSPR for viscosity is expressed as a function of the numbers of predefined functional groups (so-called GCs) occurring in an IL's cations and anions, that is, the vector of integers, $\mathbf{n} = [n_1, ..., n_G]^T$ ($n_i = 0$ if the group i is not present in the chemical structure). A scheme of GC treatment of ILs is exactly the same as in my previous articles.[10,36] Therefore, only a brief summary of the methods and the numbers specific to the above-mentioned database is presented in the following paragraphs. The group assignments adopted to represent each ion from the database can be found in the Microsoft Excel spreadsheet provided in the Supporting Information.

The total number of functional groups defined to represent the ILs from the database was $G = 315$, including 76 cation cores, 128 anions/anion cores, and 111 substituted (side-chain) groups. Such a large set of groups allows to extrapolate the models based on it to basically an infinite number of ILs of any size and complexity of their ions. However, not all groups were used in this work in the final models, as for many of them only a few occurrences in the entire reduced database were found. In order to select the groups with the highest explanatory power, stepwise MLR (SWMLR) of η_0 (ln-units) versus $\mathbf{n}$ was applied; stepwisefit function implemented in MATLAB (MathWorks Inc., version 2019a) was utilized in the calculations. The protocol was basically the same as that applied previously,[10] that is SWMLR was run iteratively until sufficiently low studentized residuals ($|s| < 3$, with at most 5 outliers allowed) and leverages ($h/h_{mean} < 5$) for all ILs were obtained. This resulted in a significant reduction of both dimensionality of the problem (from $G = 315$ to $G = 72$ with only 24/28/20 cationic/anionic/substituted groups) and the number of ILs that could be considered in further modeling (from 1974 to 1596 ILs). Based on this result, one may conclude that viscosity of ILs is indeed a challenging property for modeling. In fact, many compounds seem to follow unclear patterns which are either inexpressible in terms of the simple scheme based on the proposed GCs or due to high uncertainty of the data. On the other hand, this can be also due to oversimplified relationship between η_0 and $\mathbf{n}$ assumed within the SWMLR method. That is why I decided to apply other non-linear methods described in the following subsection.

Methods. The key novelty of the proposed methodology is that each datum (T, η) is factorized into the reference term, η_0 and correction, f, according to the following definition

$$\hat{\eta}(T; \mathbf{n}) = \hat{\eta}_0(\mathbf{n})\hat{f}(T; \mathbf{n}) \tag{6}$$

Therefore, the model's capabilities of capturing effect of the chemical structure (reference, depending on GCs only) and working conditions (correction) can be investigated individually. Each term of eq 6 is modeled separately[10] as an estimated (hat) mapping $\hat{F}_y$: $(\mathbf{x}; \mathbf{q}) \rightarrow \hat{y}$, where $\mathbf{x}$ denotes the vector of explanatory variables, $\mathbf{q}$ stands for the vector of some parameters specific to the selected method, whereas y stands for the dependent variable. It is very important to note that in this work

$y \equiv \ln(\eta/\text{mPa·s})$ and $\mathbf{x} \equiv \mathbf{n}$ in the case of the reference term, whereas $y \equiv \ln f$ and $\mathbf{x} \equiv [T - T_0; \mathbf{n}]$ for the correction. The real (no hat) relationship between known set of values of y and $\mathbf{x}$ (here, the values from the reduced database) can be expressed as

$$y = F_y(\mathbf{x}) + \varepsilon \tag{7}$$

where ε is the residual that can be estimated as $\varepsilon = y - \hat{y}$. In general, parameters $\mathbf{q}$ are adjusted so that the sum of squares of residuals $\hat{\varepsilon}_i$ ($i = 1, ..., N$) is minimized.

In this work, the set of methods applied to express the functions $\hat{F}_y$ and determine their parameters is exactly the same as previously.[10] In the case of the reference term η_0, MLR using the GCs selected by means of the stepwise method (SWMLR) and two more advanced machine learning algorithms, namely, FFANN and LSSVM, were adopted. For the correction f, only FFANN and LSSVM were utilized. All computational details on these methods and their software implementations can be found in my previous contributions[10,36] and references therein.

Both internal and external validation of the models proposed in this work followed exactly the same protocol as described in detail in ref 10. In particular, 90% of the data pool was used as training set (i.e., 1437 and 4479 data points in the case of η_0 and f, respectively) for developing the models (i.e., adjusting their parameters $\mathbf{q}$) and checking its stability via K-fold (with $K = 9$) CV and y-scrambling. Testing set built of the remaining 10% of the data pool (i.e., 159 and 453 data points in the case of η_0 and f, respectively) was adopted in external validation of predictive (generalization) capacity of the resulting correlations. The data split (for both η_0 and f) was generated by using the DUPLEX algorithm.[60] Furthermore, statistics of the models based on it were compared to a large number of those based on randomized training/testing partitions. It is worth noting that the same data split was eventually used for all modeling methods applied. Performance of the finally obtained models was evaluated using common statistical measures such as R^2, $Q_{K\text{-}CV}^2$ (training determination coefficient), RMSE, AARE, Q^2 (prediction determination coefficient[61]), defined in detail elsewhere.[10]

All calculations presented in this paper were performed using MATLAB (MathWorks Inc., version 2019a, academic license) in-house subroutines using some built-in functionality. Only the LSSVM calculations were performed using external toolbox.[62]

■ **RESULTS AND DISCUSSION**

Reference Term (η_0). In the very first stage of this study, preliminary analysis of the data split and significance of the selected explanatory variables was carried out. The MLR-based model with the 72 GCs selected using the above-mentioned stepwise procedure was utilized for this purpose. The results are summarized in Figure 3.

An effect of training/testing split on η_0 modeling performance was checked. In Figure 3a, the change in the prediction determination coefficient, Q^2 (expressing predictivity; defined by Consonni et al.[61]) with respect to the DUPLEX-based split[60] is plotted as a function of the respective change in training determination coefficient, R^2 (expressing correlative ability) for models obtained from 10 000 randomly generated data partitions. As one might expect, a general trend was observed that an increase in R^2 yields in a decrease in Q^2. As can be seen, the DUPLEX algorithm resulted in a very representative data split, whose statistics are very close to those averaged over all randomized models. Following this finding, we conclude that the DUPLEX method indeed provides a very reasonable guess of
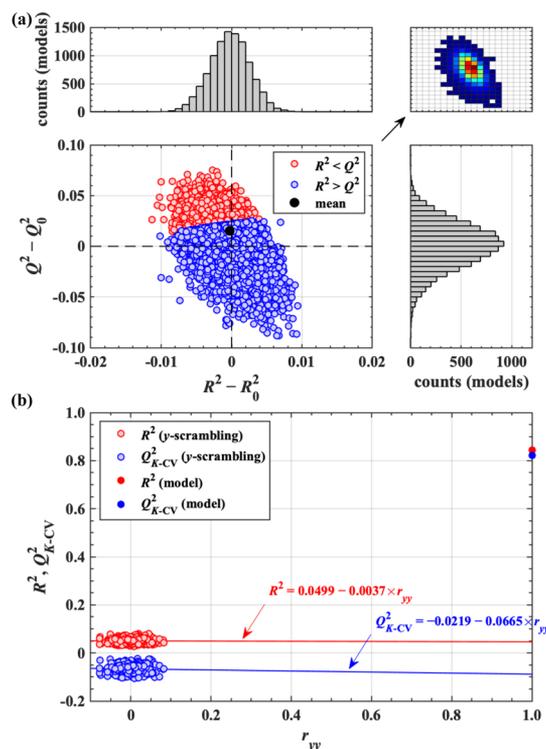
Figure 3. Summary of preliminary testing of the GC approach using SWMLR method: (a) dependence and distribution of testing determination coefficient (Q^2) vs training determination coefficient (R^2) for 10 000 randomized data splits for reference term (η_0) of eq 6, with respect to the statistics of the data split generated using DUPLEX algorithm[60] (R_0^2 and Q_0^2); (b) y-scrambling validation method results—training and K-fold validation ($K = 9$) determination coefficients (Q_{K-CV}^2) vs the Pearson's coefficient of correlation between real and permuted η_0 data (r_{yy}).

the data partition, also for further regressions employing FFANN and LSSVM. Furthermore, it is noteworthy that R^2 varies less significantly compared to Q^2. This means that correlative power of the set of selected GCs is approximately the same irrespective of the data used to adjust the model parameters. In the case of Q^2, the spread of the values is quite large (approximately from 0.74 to 0.88), so that the data split seems to have a noticeable impact on the conclusions on the predictive power of the model; in particular, for approximately one fifth of all data splits drawn, "paradoxical" models with $R^2 < Q^2$ were obtained.

An additional test adopting the so-called y-scrambling was performed to check whether one or more variables are correlated with the response by chance. The results are summarized in Figure 3b. In this method, random permutations of y data are generated and regressed with unchanged **x** data. Statistical measures of correlation and stability (i.e., R^2 and Q_{K-CV}^2) are recorded for each permutation and plotted versus Pearson's correlation coefficient between true and randomized response (r_{yy}). For completely randomized y, that is $r_{yy} = 0$, the values of $R^2 < 0.3$ and $Q_{K-CV}^2 < 0.05$[63] indicate that there are no chance correlations between y and **x**. In the case of the data plotted in Figure 3b, the respective values are 0.05 and −0.02, so that they are substantially lower compared to the recommended criteria.

Given the data split and the set of finally selected GCs, training of the models for reference term of eq 6 was performed using SWMLR, FFANN (with four nodes in the hidden layer, so that the number of data samples per model parameter does not exceed five[64]), and LSSVM methods. The results are shown in Figure 4 in the form of parity plots. All of the experimental (reduced) versus calculated values of η_0 as well as all equations/parameters incorporated to obtain them can be found in the Microsoft Excel spreadsheet provided in the Supporting Information. The most important result evidenced in Figure 4 is that the quality of prediction is similar to that of training, although all models tend to display some problems with predictions for very viscous ILs. Furthermore, one can see in Figure 4 that not rejecting a part of the data (as discussed above) would result in even worse correlations. The scatter of the computed data is quite high irrespective of the method. However, it is noticeably lower in the case of FFANN and LSSVM compared to SWMLR. This means that the problem of modeling viscosity of ILs is very complex, particularly when one wants to describe a great variety of structures using a single equation with a single set of parameters. Surprisingly, this is not only the case with "trivial" SWMLR approach, but also more sophisticated methods such as FFANN and LSSVM, as can be concluded from the summary of the statistics of the obtained models presented in Table 2. A slightly better accuracy of both correlation and prediction was displayed by the LSSVM-based model. However, overall AARE on the order of 30% is still high and on no account satisfactory. Nevertheless, one should bear in mind that: (i) the values listed in Table 2 regard only the η_0 values, so that they represent accuracy of prediction of viscosity based on the chemical structure of ILs only; (ii) the proposed model covers the greatest and the most diverse set of structures (1596 ILs) among other reported thus far. In fact, almost none
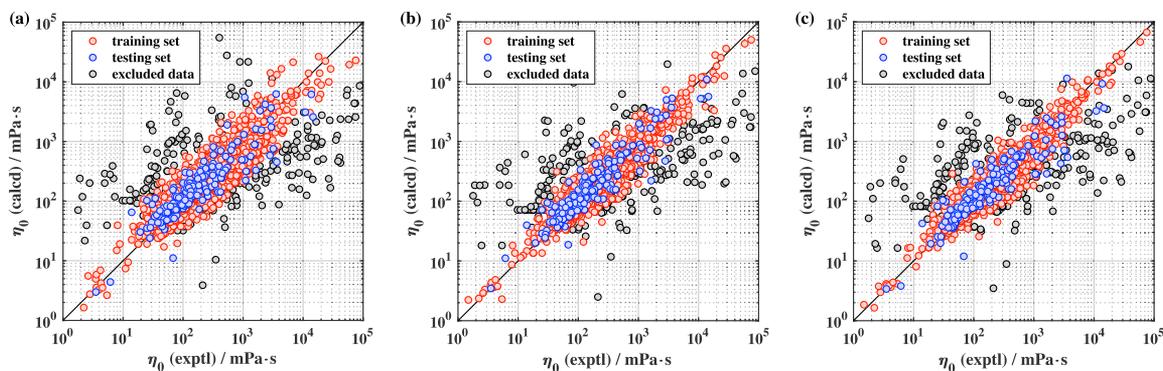
Figure 4. Calculated versus reduced experimental data on the viscosity reference ($T = 298.15$ K) term (η_0), see eq 6: (a) SWMLR; (b) FFANN; and (c) LSSVM.

Table 2. Statistics of the Proposed Models for Reference Term of Eq 6 Given in Terms of Training, K-fold CV and Testing Determination Coefficients (R^2, $Q_{K\text{-}CV}^2$, Q^2), RMSE, and AARE between Computed and Reduced Experimental Data

method	R^2	$Q_{K\text{-}CV}^2$	Q^2	RMSE/AARE (%) training	test
SWMLR	0.8435	0.8212	0.8067	0.55/46.3	0.61/47.2
FFANN	0.9092	0.8514	0.8637	0.42/32.1	0.51/39.5
LSSVM	0.9172	0.8606	0.8374	0.40/30.2	0.56/37.7

of the articles mentioned in the introduction (including my previous article[36] as well) account for point (i), as they report the statistics computed based on temperature-dependencies of η—this results in a bias toward lower values of AARE, for example, due to high number of data for accurately described ILs or low number of poorly described ones. It is also important to note that in spite of point (ii), such AARE values are comparable with some much less extensive GC models.[30,38,39]

In order to analyze an actual predictive capacity more specifically, distribution of the RE in η_0 rather than its average should be analyzed. Figure 5 presents a brief summary of such analysis for the LSSVM-based model, including the distribution of absolute RE observed for all ILs (Figure 5a) as well as the REs
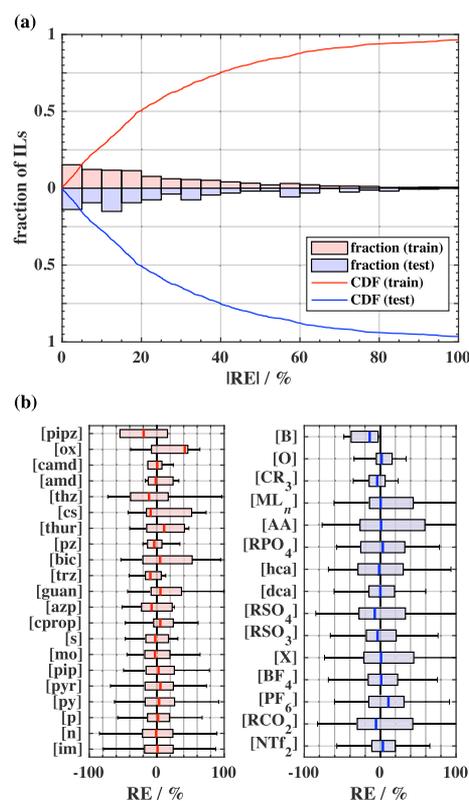
Figure 5. Distribution of RE between calculated (LSSVM-based GC model) and reduced experimental reference viscosity (η_0): (a) probability and cumulative distribution function of absolute RE, including all ILs studied; (b) distribution of RE within different chemical families of cations (left panel) and anions (right panel)—the central mark on each box is the median, the lower and upper edges of the box correspond to the first and the third quartiles, the whiskers extend to the most extreme data points not considered as outliers (1.5 of the box width). Abbreviations explained in Table 1.

observed for each cationic and anionic chemical family individually (Figure 5b). As seen from Figure 5a, the distribution of absolute RE values is quite broad. The value of AARE corresponds approximately to the third quartile. This regards both the training set and testing set, so that it is confirmed that the model developed is stable and poses good generalization capabilities. Furthermore, in the case of half of all ILs considered, the value of RE is lower than 20%. From Figure 5b, one can note that for the most extensively studied cationic families of ILs (imidazolium, ammonium, phosphonium, pyridinium, pyrrolidinium, and piperidinium), the distributions are quite similar, that is, they are symmetrical and relatively narrow. On the other hand, one should be careful when relying on the predictions for ILs based on less known cations, for example bicyclic ones. The performance of the model seems to be very similar regardless of the type of anion. For carboxylate-based ones (including amino acids) or typical inorganic anion-based ones (e.g., halides, metal complexes), the scatter of the errors is exceptionally large. However, this result should not be perceived as surprising. In fact, such ILs are usually strongly hydrophilic, so that the samples from different sources may differ in purity, hence the measured viscosity.

Based on the results given in Figures 4 and 5 and Table 2, one may argue that the applied GC approach is ineffective, when quantitative description is taken into account. However, an exact numerical value of η_0 is not always needed to be known. From a practical point of view, qualitative information on the viscosity is very often sufficient. In other words, η_0 can be expressed as a categorical variable, that is, a variable having a value from a finite set. In particular, one could say that ILs can have "low", "moderate", or "high" viscosity. This approach requires definition of the boundaries between the classes of ILs in such way, thus is ambiguous. An alternative approach relies on the following classification of η_0 with respect to the order of the magnitude $O(\eta)$ defined as

$$O{:}\eta \rightarrow O(\eta) \equiv \min\{\max\{O_{\min}, \lfloor\log(\eta/\mathrm{mPa\cdot s})\rfloor\}, O_{\max}\} \tag{8}$$

where $\lfloor x \rfloor$ denotes the "floor of x", that is, the highest integer lower than x, whereas the symbols $O_{\min}$ and $O_{\max}$ stand for lower and upper limiting values. Based on the analysis of the accepted experimental data, it was assumed that $O_{\min} = 0$ and $O_{\max} = 4$. I decided to apply eq 8 to both experimental and predicted values of η_0 and check the accuracy of such "quasi-classification" using the developed models. The values of $O(\eta_0)$ that emerged from all reduced experimental and computed η_0 data are listed in the Microsoft Excel spreadsheet provided in the Supporting Information. Herein, the results are depicted in Figure 6 in the form of the so-called "confusion matrix" presenting different metrics of classification performance. An overall classification accuracy is on the order of 80% in the case of SWMLR and 85% for both FFANN and LSSVM, with slightly higher value for the former approach. It is noteworthy, however, that the simplest model based on linear relationship between η_0 and GCs exhibits much lower precision in the case of extremely low viscous and extremely high viscous ILs. In those cases, FFANN-based model seems to be the most promising—nevertheless, one should bear in mind that the number of data points with $O(\eta_0) = 0$ and $O(\eta_0) = 4$ is relatively low. Within the range of $O(\eta_0)$ from 1 to 3, both FFANN and LSSVM display similar rates of true and false discoveries.

Classification accuracy computed for each family of cations and anions is listed in Table 3. In the case of cations, the
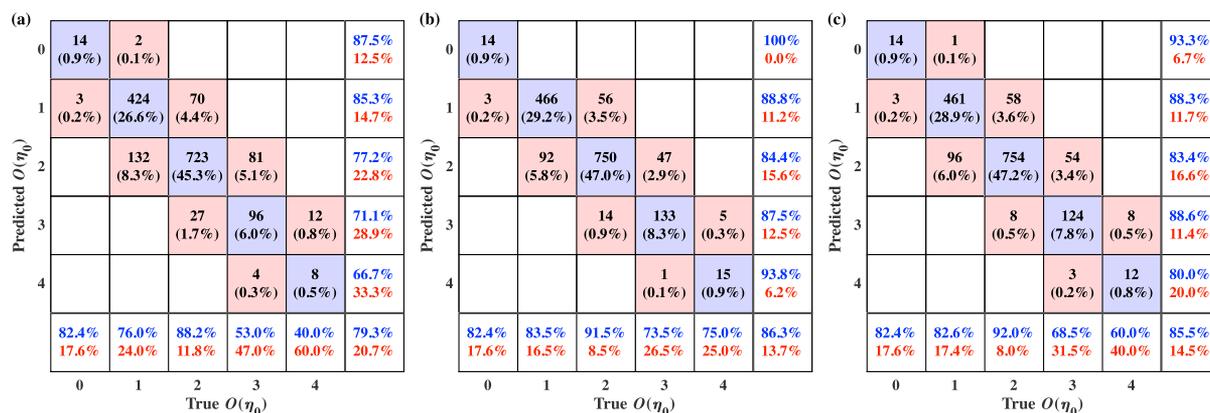
Figure 6. Confusion matrices of GC models for the reference viscosity order $O(\eta_0)$, see eq 8: (a) SWMLR; (b) FFANN; and (c) LSSVM. Precisions and false discovery rates are given in the far right column (upper and lower values, respectively). True positive and false positive rates are given in the bottom row (upper and lower values, respectively). The cell in the bottom right of the plot shows the overall classification accuracy.

Table 3. Percentages of Correctly Classified $O(\eta_0)$ Data, See Eq 8, with Regard to Different Families of Cations and Anions as Well as Three Computational Approaches Applied

cations[a]				anions[a]			
Family	SWMLR	FFANN	LSSVM	family	SWMLR	FFANN	LSSVM
[im]	75.2	82.9	82.0	[NTf$_2$]	84.3	87.6	87.1
[n]	75.9	84.6	85.0	[RCO$_2$]	70.3	80.4	80.4
[p]	77.0	91.0	91.0	[PF$_6$]	80.9	87.2	89.4
[py]	84.4	89.3	86.9	[BF$_4$]	83.3	89.6	89.6
[pyr]	81.1	86.8	85.8	[X]	67.5	84.2	75.4
[pip]	76.8	85.7	83.9	[RSO$_3$]	77.1	89.5	87.6
[mo]	93.5	83.9	87.1	[RSO$_4$]	77.5	83.8	86.3
[s]	100.0	100.0	100.0	[dca]	85.8	93.4	92.5
[cprop]	97.0	93.9	81.8	[hca]	80.9	80.9	85.1
[azp]	100.0	100.0	100.0	[RPO$_4$]	76.5	84.3	80.4
[guan]	90.9	90.9	95.5	[AA]	58.5	84.9	81.1
[trz]	88.2	88.2	88.2	[ML$_n$]	68.5	75.9	75.9
[bic]	88.9	100.0	100.0	[CR$_3$]	90.0	80.0	80.0
[pz]	100.0	100.0	100.0	[O]	95.8	91.7	100.0
[thur]	90.9	90.9	72.7	[B]	50.0	66.7	66.7
[cs]	90.0	90.0	90.0				
[thz]	80.0	80.0	80.0				
[amd]	100.0	100.0	100.0				
[camd]	100.0	100.0	100.0				
[ox]	91.7	100.0	91.7				
[pipz]	100.0	50.0	100.0				

[a]Abbreviations defined in Table 1.

FFANN-based model turned out to be more accurate compared to the LSSVM-based one for almost all families (except piperazinium-based ILs)—therefore FFANN is henceforth treated as the final reference model. In turn, the GC model developed using the SWMLR method yields in surprisingly good precision for a number of classes of ILs, such as morpholinium or cyclopropenium. An effect of anions is, however, captured more accurately by using either the FFANN- or LSSVM-based model. Finally, influence of the type of both cation and anion family on the expected classification accuracy is depicted in Figure 7. As seen from Figure 7a, for the SWMLR-based model, there are many cation−anion combinations for which the accuracy is below 70%, namely, 29 out of 117 cases. For FFANN and LSSVM GC models, there are only 12 and 11 such pairs, respectively. I strongly recommend to use the diagrams plotted in Figure 7 in estimating how accurate the predictions using the

proposed models can be for ILs with the type of cation core and the type of anion given.

Correction Term (f). The temperature correction term, $f \equiv \eta/\eta_0$, see eq 6, was modeled using FFANN and LSSVM approaches only because linear regression resulted in errors so unsatisfactory that I decided to not present them at all. In the case of FFANN, the number of nodes in the hidden layer (S) was determined by means of testing error optimization, as shown in Figure 8; details can be found elsewhere.[65] As can be noticed, the finally used FFANN is built of two layers, including the hidden layer with $S = 7$ nodes and the output layer composed on a single identity node. In fact, the models with more hidden nodes do not significantly improve testing mean squared error (MSE)

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\ln f_i - \ln \hat{f}_i)^2 \qquad (9)$$
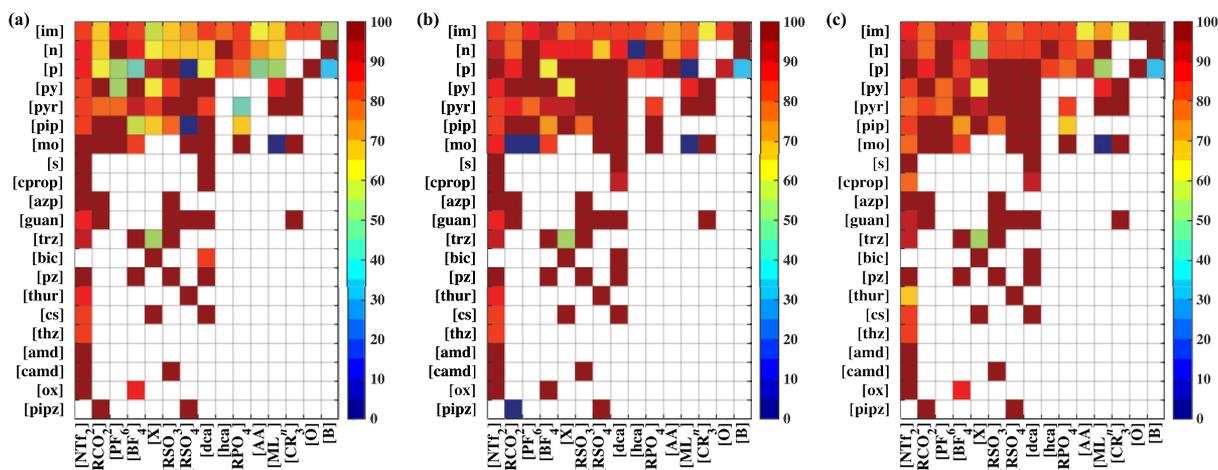
Figure 7. Overall accuracy of classification of the reference viscosity order $O(\eta_0)$, see eq 8, computed for different combinations of chemical families of cations and anions: (a) SWMLR; (b) FFANN; (c) LSSVM. Cationic and anionic families put on the ordinate and abscissa axis, respectively. Abbreviations explained in Table 1.
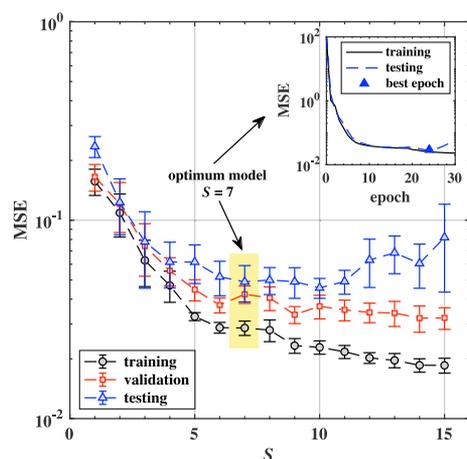
Figure 8. Determination of optimum number of hidden nodes (S) of FFANN-based model for the viscosity correction term f, see eq 6, based on the training, validation, and testing MSE between computed and experimental data. Performance plot of the best model (highlighted) presented in the inset. Error bars designated by the standard deviations of the MSE obtained from different training/validation splits adopted in K-fold cross validation ($K = 9$).

Furthermore, higher S results in a more noticeable scatter of MSE obtained from different training/validation data splits used for K-fold CV. For the finally proposed net, each model parameter corresponds to approximately 8 data points of the training set (on average). The "rule of thumb" I consequently referred to in my previous articles[10,36] is that the ratio should be higher than five.[64] The proposed network topology meets that criterion, so one can state that it is not overfitted.

The computed versus pseudoexperimental (i.e., obtained from the reduced η data) values of f are plotted in Figure 9 (the data for the ILs excluded in the GC selection process are not taken into account); detailed data allowing to reproduce plots shown can be found in the Microsoft Excel spreadsheet provided in the Supporting Information. As seen, in the case of both FFANN (Figure 9a) and LSSVM (Figure 9b), the calculated data are symmetrically and tightly scattered along the identity line within the wide range of 5 orders of magnitude of f. First of all, this statement applies to the training set employed in the model development process. Thus, the proposed GC scheme
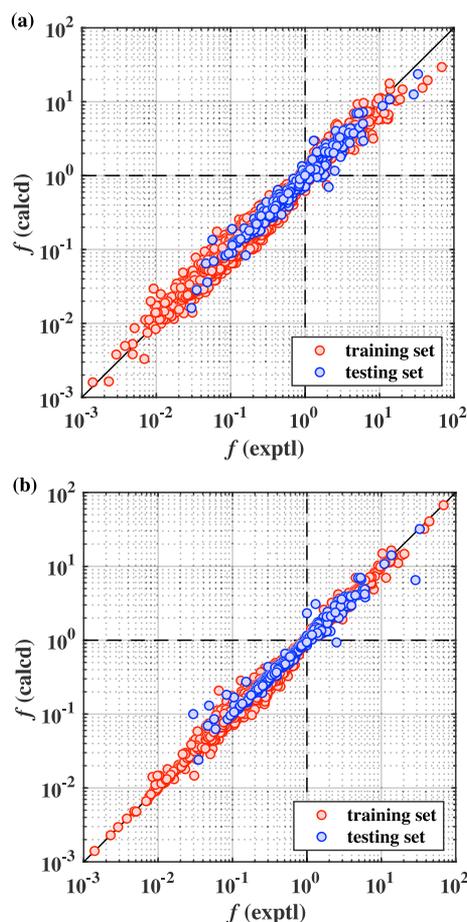
Figure 9. Calculated vs reduced experimental data on viscosity correction ($T \neq 298.15$ K) term (f), see eq 6: (a) FFANN; (b) LSSVM.

serves an excellent correlative capability with $R^2 = 0.9799$ and 0.9934 for the FFANN- and LSSVM-based models, respectively. For both approaches, the CV determination coefficients $Q_{K\text{-CV}}^2$ ($K = 9$) are also very high and close to R^2, namely, 0.9695 and 0.9816 for FFANN and LSSVM, respectively. Finally, a good predictive performance of the proposed models is evidenced by accurate representation of the target f data from the testing set;

in fact, prediction determination coefficients, Q^2 are also very high, namely, 0.9745 for FFANN and 0.9802 for LSSVM.

A very important finding that can be observed from Figure 9 is that the proposed models provide qualitative correct viscosity behavior compared to the experimental data. In fact, only in the case of approximately 2% of all data points, the model predicts that computed and experimental ln f differ in sign (more specifically, the corresponding percentages were 2.1% for FFANN and 1.6% in the case of LSSVM).

On the basis of all results mentioned above, it is quite difficult to recommend a single GC model for capturing an effect of temperature on viscosity, as both FFANN and LSSVM display very similar correlation/prediction performance. However, to the best of my knowledge and experience with machine learning methods, FFANN-based models tend to be unstable, particularly when applied in extrapolation. Furthermore, the GC scheme adopted in this work has recently turned out be very effective when combined with LSSVM.[66] Given that rationale, I conclude that the LSSVM-based GC model should be utilized as a tool for predicting the viscosity correction term of eq 6. Advantage of LSSVM over FFANN is additionally illustrated in Figure 10, where calculated versus experimental viscosity data
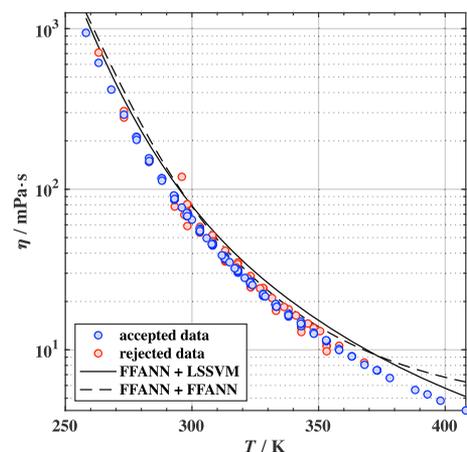
Figure 10. Calculated (by using FFANN-based reference term + FFANN/LSSVM-based correction term) vs experimental viscosity (η) of 1-hexyl-3-methylimidazolium bistriflamide as a function of temperature (T). References to experimental data as well as full list of accepted and reject data can be found in the Supporting Information.

for 1-hexyl-3-methylimidazolium bistriflamide, that is the IUPAC reference IL,[67] are presented; references to experimental data (in total, 27 data sets and 184 data points) as well as the full list of accepted and reject data can be found in the Microsoft Excel spreadsheet provided in the Supporting Information. As seen, the LSSVM-based correction provides much more accurate outcomes at temperatures much higher as compared to the reference conditions. In turn, at temperature not differing too much from T = 298.15 K, the effect of machine learning algorithm is not so evident.

Calculated Versus Raw Data. The finally proposed "FFANN + LSSVM" GC model, that is composed of FFANN-based reference term and LSSVM-based correction term of eq 6, was applied to recalculate the original ("raw") data points in the database considered in this work, including only the data revised for ILs not excluded in the dimensionality reduction process. All calculated data points are presented in the Microsoft Excel spreadsheet provided in the Supporting Information.

Dependence of REs between calculated and experimental viscosity on both η and temperature is given in Figure 11. As
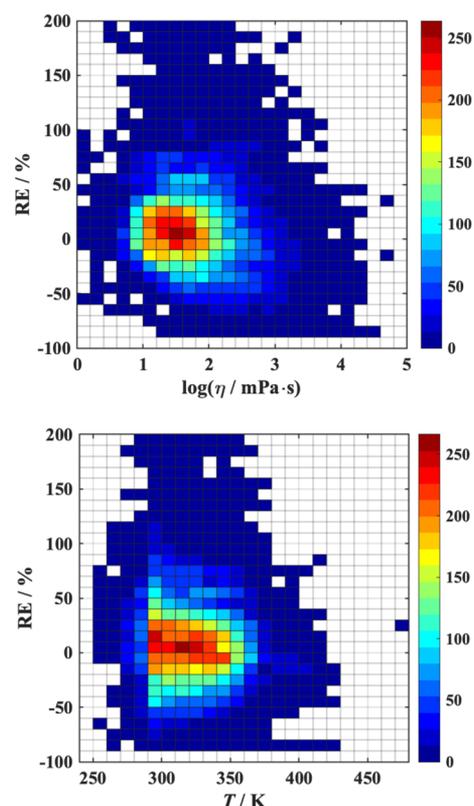
Figure 11. Distribution of RE between calculated and raw experimental viscosity (η) data as a function of η (upper panel) and temperature (T). The results shown for the reference and correction term modeled with FFANN and LSSVM, respectively, see eq 6.

seen, the distributions are rather symmetrical with respect to the RE equal to zero, so that it is not possible to unequivocally state whether the model tends to provide underestimated or overestimated outcome. Furthermore, a significant fraction of data points with relatively low RE ($\pm50\%$) is distributed within the range from η = 10 to 1000 mPa·s and from T = 290 to 360 K. Therefore, the model predictions outside these intervals should be treated with special care.

More detailed summary of the final model performance is presented in Table 4, where the values of AARE between computed and measured viscosity are listed for each of our 117 cation−anion combinations (with respect to chemical families of ions). As seen, overall AARE is about 32%, hence very similar to that obtained from the reduced reference data of η_0. However, accuracy of calculation strongly depends on the type of either cation or anion. In fact, for some classes of ILs, it is very low (e.g., AARE < 20% for 32 classes), whereas for some others, the computed data are very inaccurate (e.g., AARE > 100% only for five classes). It is also important to note that the model applied to the ILs based on "major" types of cations such as imidazolium, pyridinium, and pyrrolidinium discloses roughly similar AARE irrespective of the anion; if there are exceptions from this finding, they usually refer to the ILs composed of anions derived from carboxylic acids or amino acids, that is strongly hydrophylic compounds with highly uncertain viscosity data. The model is also quite stable when one considers an effect of the cation's family of viscosity. The most accurate calculations were

Table 4. AARE between Calculated (FFANN as Reference Term + LSSVM as Correction Term, See Eq 6) and Experimental Viscosity for Ionic Liquids Belonging to Different Cationic and Anionic Families

cation	[NTf$_2$]	[RCO$_2$]	[PF$_6$]	[BF$_4$]	[X]	[RSO$_3$]	[RSO$_4$]	[dca]	[hca]	[RPO$_4$]	[AA]	[ML$_n$]	[CR$_3$]	[O]	[B]	overall
[im]	22.3	52.6	17.2	42.9	54.0	23.2	41.6	22.5	36.7	27.3	46.9	38.9	15.0	33.4	17.5	29.7
[n]	32.6	55.9	27.3	19.9	6.3	25.8	39.9	38.6	40.7	46.3	158.1	19.4			136.4	45.1
[p]	25.4	157.6	60.7	84.4	48.7	65.0	17.1	18.5	27.3	45.2	34.1	52.6		26.1	30.6	46.7
[py]	22.2	82.0	32.4	20.5	42.7	33.8	35.0	25.0				25.3	9.7			25.0
[pyr]	22.9	36.5	24.7	10.3	51.5	10.2	17.2	30.6		73.9		4.5	36.7			29.1
[pip]	32.4	38.7	98.0	11.0	23.5	24.2	39.2	72.3		33.1						33.0
[mo]	21.5	65.0	83.6	30.3			50.4	23.4		24.4		72.4	29.2			32.8
[s]	15.8							25.6								18.1
[cprop]	14.7							19.5								16.7
[azp]	15.1	25.8				28.8										20.4
[guan]	15.8	121.1				140.9	67.9	1.8					19.8			20.6
[trz]	10.8			13.7	34.1	11.9										14.0
[bic]					73.6			30.3								44.7
[pz]	9.8			10.0		14.6		10.2								10.5
[thur]	45.0						24.0									41.1
[cs]	23.9				36.6			8.2								17.1
[quin]	25.7															25.7
[thz]	24.9															24.9
[amd]	9.1					5.6										7.7
[camd]	43.2			22.1												29.1
[ox]		57.6					18.4									49.7
overall	23.5	58.2	27.7	35.4	47.2	27.4	38.7	22.8	30.0	39.9	73.2	34.7	16.0	29.0	40.2	**32.3**

observed for methanides, dicyanoamides, and bistriflamides. This is a very good result taking into account the growing interest in ILs based on these anions in recent years. As one might expect, the worst results were recorded for ILs with the anions formed by carboxylates, amino acids, and inorganics, so special care should be taken when using the proposed model for such systems.

Summing up, I strongly encourage all potential users of the proposed model to check Table 4 prior to making any predictions. In particular, I do not recommend to apply the model to cation−anion combinations not present in Table 4—in my opinion, those "gaps" in our knowledge of ILs viscosity behavior should be studied experimentally first. On the other hand, one can try to carry out the extrapolations, only after checking how is the input chemical structure (i.e., GC assignment vector) related to the applicability domain (AD) of the model. The AD can be established by employing diverse methods extensively described in the literature,[68−70] along with all data used in the model development process, provided in the Supporting Information.

Finally, the proposed GC model was used in classification of ILs with respect to $O(\eta)$, see eq 8, computed based on all raw η values, that is also at temperatures other than the reference one. The results of the calculations are depicted in Figure 12. Figure 12a shows the confusion matrix of $O(\eta)$. As seen, the overall accuracy obtained from FFANN (reference) + LSSVM (correction) calculations is at the level of 87%. The lowest precision was observed for $O(\eta) = 0$ and $O(\eta) = 3$, so such model's outcomes should be treated with special care. In turn, quite low true positive rates were recorded for limiting values of $O(\eta)$, namely, 70% for $O(\eta) = 0$ and 57% for $O(\eta) = 4$. However, one should keep in mind that in the latter case, the number of data points is very low (<1% of the whole data pool). Besides, this is the case of extremely high viscosities, and thus also values with higher uncertainty. An effect of temperature on

accuracy of classification is demonstrated in Figure 12b. As can be noticed, the accuracy is roughly equal to the average value in the wide range of T, namely, from $T = 270$ to 410 K. Only at the lowest and the highest temperatures, the modeling performance is significantly deteriorated. In fact, this is consistent with the data shown in Figure 12 because low and high T correspond to high and low η, respectively.

Comparison with Literature Models. Table 5 presents a comparison of basic metrics of the IL viscosity models described in the open literature so far, namely, the number of ILs and data points involved in the model development process as well as the AARE between calculated and measured viscosity, with those characterizing the final GC model given in this work. All of the data listed therein can be used at least as the very first suggestion for selecting of the best methodology for prediction of viscosity of ILs.

The very first remark that should be noticed from Table 5 is that the model presented in this work has been derived from the extraordinarily large database. In fact, a great majority of the models (including even the most recent ones) has been established and tested on the basis of the data for less than a 100 ILs. In the case of theories and semiempirical models, it is quite understandable—such models usually require IL-specific parameters required to be fitted to the experimental temperature-dependent data, although such data are not accessible for all ILs. However, it is difficult to understand the reason for not using larger data sets in developing empirical QSPRs, particularly because the previous version of the database from this work has been published more than five years ago.[36] Apart from the large size, the database data covers the greatest number of chemical structures of both IL cations and anions among the models published so far. The models established on the basis of such large and diverse data collection can be used to represent significantly larger pool of ILs not synthesized yet, in particular when compared to the models involving cation/anion-specific
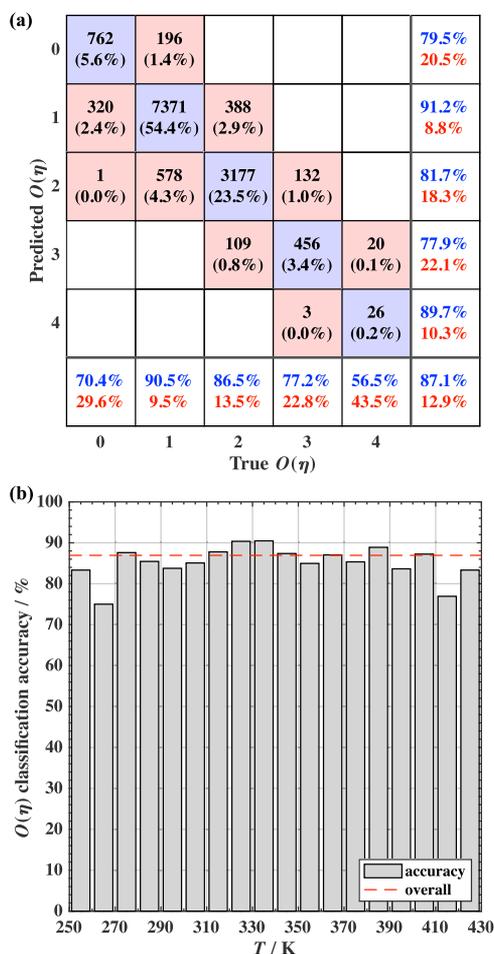
Figure 12. Classification of viscosity order data $O(\eta)$, see eq 8, computed from raw experimental viscosity values: (a) confusion matrix of the final GC model; (b) dependence of the classification accuracy on temperature (T). The results shown for the reference and correction term modeled with FFANN and LSSVM, respectively, see eq 6.

parameters. Therefore, in my opinion, the study presented (along with the Supporting Information) will be certainly beneficial to the IL community, mainly to the researchers employing modern CAMD protocols to obtain novel ILs having desired properties.

The overall AARE significantly depends on the model and varies from relatively low values <10% up to more than >50%. In my opinion, the value of ≈32% obtained in this work for such a large and diverse library of compounds should be perceived as a success; actually, this order of magnitude of AARE was obtained by other authors from much more limited viscosity data sets.[22,27,30,35,38,42]

Apart from the very general comparison given in Table 5, I decided to present a detailed analysis of the new model's accuracy when confronted with a purely in silico calculation approach proposed by Eiden et al.[27]—this was the only method that could be applied regardless of the chemical structure of ions; I have also tried to apply the model of Yan et al.[45] but due to complexity of its formulation as well as no examples provided by the authors, I was not able to reproduce the results shown by them. In the method of Eiden et al.,[27] viscosity is estimated on the basis of three molecular descriptors of ILs obtained solely from QC calculations: molecular radius (r_m; estimated from averaging the ions' volumes), symmetry number (σ), and the

sum of the solvation Gibbs energies of ions with a static dielectric constant at infinity (ΔG_{solv}^{∞}). The following simple QSPRs have been proposed in order to express viscosity in terms of those variables[27]

$$\ln\frac{\eta}{\eta_0} = -4.396 + 11.532\ln\frac{r_m}{r_m^0} + 1.338\ln\sigma$$
$$- 0.028\frac{\Delta G_{solv}^{\infty}}{G_0} \tag{10}$$

where $\eta_0 = 1$ mPa·s, $r_m^0 = 1$ nm, and $G_0 = 1$ kJ·mol^{-1}. The coefficients seen in eq 10 were adjusted to experimental viscosity data at $T = 298$ K for a relatively small set of 72 ILs, nevertheless quite diverse as containing cations and anions belonging, respectively, to 9 and 11 distinct chemical families. The values of r_m, σ, and ΔG_{solv}^{∞} were obtained at RI-BP86/def-TZVP/COSMO.[27] In this work, implementation of the original model given by eq 10 in COSMOtherm software (COSMOlogic GmbH & Co. KG, Leverkusen, Germany)[71] was used to calculate the reference viscosity for 1596 ILs considered in this work. All computed values are listed in the Microsoft Excel spreadsheet provided in the Supporting Information, whereas the optimized geometries of ions utilized in descriptor calculations are available upon e-mail request. The results of comparative analysis are depicted in Figure 13. From the parity plot shown Figure 13a, one can easily observe that the model reported in this work displays significantly lower scatter of the computed versus measured viscosity. In fact, Eiden's QSPR provides very good estimations (|RE| < 30%) for less than half as much ILs as the new GC model. Based on the distribution of the REs seen in Figure 13b, one can observe that the model given in eq 10 tends to return underestimated estimates of the viscosity. However, there are many ILs for which very high positive deviations were detected, but they are not plotted in Figure 13b for the sake of clarity. In terms of AARE, the model reported in ref 27 is far less accurate. Considering only the data points not suspected to be outliers (i.e., being in the 1.5 times interquartile range), the corresponding values are AARE = 26.4% for the reference term of eq 6 and AARE = 49.4% in the case of eq 10. The latter model also fails when applied as a classifier of the viscosity data with respect to the order of magnitude, as illustrated in Figure 13c. In fact, overall classification accuracy is at the level of 60%, with a noticeably high false discovery rates (thus low precision) observed for both low and high viscous ILs.

Despite unsatisfactory results obtained from the model of Eiden et al.,[27] one should bear in mind that they come from pure predictions obtained on the basis of molecular geometry only. The comparison summarized above is therefore another argument for the thesis that QSPRs involving molecular-based or QC-based descriptors can be perceived as robust predictive tools for CAMD, but only locally, that is when applied to the compounds not differing much from those employed to calibrate the working equation such as eq 10. To model extensive databases of a great variety of structures of ILs, the GC approach combined with machine learning seems to be a good or at least very promising choice.

■ CONCLUSIONS

A new QSPR for calculating viscosity of ILs as a function of temperature in terms of GCs has been proposed, validated, and recommended. The model involves FFANN-based reference term and LSSVM-based correction term of eq 6. The results shown and discussed in detail above suggest that the finally

Table 5. Comparison of Ionic Liquids Viscosity Models Reported in the Literature with the Methods Presented in This Work Given in Terms of AARE

authors	method[a]	year	number of		T/K	AARE/%[b]
			ILs	data points		
Empirical Models						
Matsuda et al.[19]	GC/MLR	2007	[c]	300	273−353	
Matsuda et al.[20]	GC/MLR	2007	77	147	283−362	18.7
Tochigi & Yamamoto[22]	QSPR + GC/non-linear	2007	162	376	273−353	21.7
	QSPR + GC/MLR	2007	162	376	273−353	28.4
Bini et al.[23]	QSPR/MLR	2008	32	32	293	
Gardas & Coutinho[24,25]	GC	2009	24	482	293−393	7.5
Billard et al.[26]	QSPR	2011	99	99	298	
Eiden et al.[27]	QSPR	2011	72	655	253−409	34.0
Han et al.[28]	QSPR	2011		1732	258−433	13.6[d,e]
Valderrama et al.[29]	QSPR/ANN	2011	58	358	298−333	4.8
Gharagheizi et al.[30]	GC/MLR	2012	443	1672	253−433	31.0
Mirkhani & Gharagheizi[31]	QSPR/MLR	2012	293	435	253−373	8.8[f]
Yu et al.[32]	QSPR/MLR	2012	696	5046		[e]
Dutt et al.[33]	correlation/ANN	2013	81	736	273−388	6.6[g]
Chen et al.[34]	QSPR/MLR	2013	26	304	258−433	9.5
Yu et al.[35]	QSPR/MLR	2013	146	325	283−343	[e]
Paduszyński & Domańska[36]	GC/ANN	2014	1484	12 405	253−573	11.4
Lazzús & Pulgar-Villarroel[38]	GC	2015	326	1445	253−395	22.3
Zhao et al.[39]	QSPR/MLR	2015	89	1502	253−395	59.4
	QSPR/LSSVM	2015	89	1502	253−395	30.2
Alcalde et al.[37]	QSPR/MLR	2015	27	27	298	
Barycki et al.[41]	QSPR/MLR	2016	23	138	298−343	41.3
Esmaeili-Jaghdan et al.[40]	QSPR/CMIS	2016	23	579	273−388	
Baghban et al.[43]	GC/LSSVM	2017	443	1672	253−433	7.9
Fatehi et al.[44]	QSPR/ANN	2017	59	736	273−393	1.3
Kang et al.[42]	QSPR/ELM	2017	89	1502	253−395	≈20
Yan et al.[45]	QSPR/MLR	2018	349	3228	253−573	8.2
Haghbakhsh & Raeissi[46]	correlation/ANN	2019	38	2073	258−408	6.9[g]
Theories and Semiempirical Models						
Abbott[18]	hole theory	2004	12	12	298	>100%
Slattery et al.[21]	VBT	2007	27	27	293	8.7
He et al.[47]	cubic EoS + Eyring	2018	25	1070	273−433	1.5
Macías-Salinas[48]	cubic EoS + Eyring	2018	16	703	273−353	1.6
Polishuk[49]	SAFT-cubic + MYS	2012	5			
Abolala[50]	SAFT-VR-Mie + MYS	2015	11	822	293−373	2.4
	SAFT-VR-Mie + FVT	2015	11	822	293−373	3.8
	SAFT-VR-Mie + FT	2015	11	822	293−373	4.9
Dowell et al.[51]	soft-SAFT + VFT	2014	6			≈6
Llovell and Vega[52]	soft-SAFT + VFT	2014	12			≈5
Sun et al.[53]	ePC-SAFT + VFT	2018	89			≈5
Zhao et al.[54,55]	UNIFAC-VISCO	2016	70	819	273−373	1.4

[a]Details in the text. [b]AARE defined in eq 4. [c]Not stated clearly in the paper. [d]Maximum deviation in the test set; cation/anion specific models. [e]Cation/anion/temperature-specific models. [f]Regarding η given in log-units. [g]Temperature-dependence described only.

developed tool can be used in predicting η itself as well as classification of the viscosity order, $O(\eta)$, see eq 8.

The study presented clearly proves that modeling viscosity of ILs in terms of GCs is not an easy problem, irrespective of the computational method applied. In particular, a simple additivity rule, that is the basis of the SWMLR, yielded the worst results, including prediction of both η and $O(\eta)$. More sophisticated machine learning methods (FFANN and LSSVM) resulted in a noticeably better modeling performance, but physical or mechanistic interpretation of parameters of the resulting models is basically unfeasible. Thus, these models are more valuable from the purely utilitarian point of view rather than for the fundamental knowledge or physical insight into the viscous behavior of ILs. Furthermore, it was shown that AARE between computed and experimental η (in absolute, i.e., neither ln- nor log-units) on the order of 30% seems to be the limit one can currently achieve in the modeling of IL viscosity-GC relationships, especially if universal correlations accounting for ILs belonging to various cationic families and based on diverse anions (see Table 5). In fact, the values of AARE obtained from modeling of a great diversity of ILs are very close to, in some cases even lower, those resulted from the models developed on the basis of much smaller databases and much more limited library of chemical structures—such finding should be also perceived as an extra argument evidencing improvement in estimating viscosity of ILs using empirical equations.
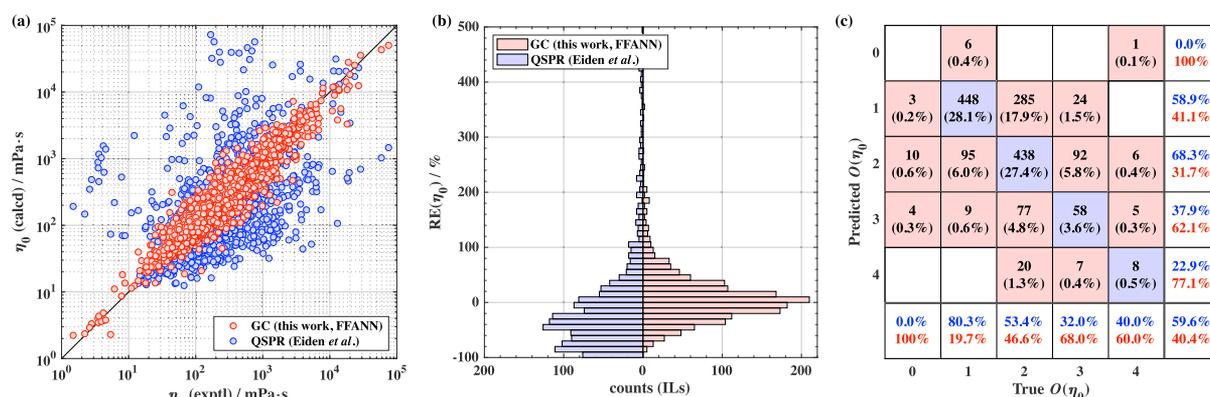
Figure 13. Comparison of the GC model developed in this work for the reference viscosity (η_0), see eq 6, with the QSPR method by Eiden et al.:[27] (a) parity plot of the calculated versus reduced experimental data; (b) distribution of RE between computed and experimental η_0; and (c) confusion matrix of the model by Eiden et al.[27] for the reference viscosity order $O(\eta_0)$, see eq 8. The results shown for the reference term modeled with FFANN.

One should keep in mind that despite enormous number of viscosity data sets measured so far, there are still many cation−anion pairs not checked yet. Indeed, one can imagine about 250 thousand ILs that could be synthesized combining 974 cations and 249 anions present in the database attached with this article (see the Supporting Information), so that the progress in the field with "only" 2068 ILs is around 1%. Of course, the models like those presented in this work may help to fill those "gaps". However, I strongly recommend to apply them only for families of ILs (as well as both T and η ranges) shown in Figure 1. In particular, using the models in estimating the viscosity of cation−anion combinations yet unexplored experimentally (see Figure 1a) may lead to results with vague uncertainty. This is an intrinsic feature of each empirical model that one should be always aware of, particularly when the model aims to capture such complex property (like viscosity) of such complex systems (like ILs).

Summing up, I believe that all results and the Supporting Information recapitulated in this paper will be very useful for both academic and industrial communities concerned with ILs. I also hope that this work will convince the researchers working in the field that the proposed models are remarkably advantageous over the other methods reported thus far. Moreover, I strongly encourage both experimentalists and theoreticians to test the models developed herein, respectively by comparing their outcomes with the results of measurements of viscous behavior of novel ILs and simulations/predictions of it using other models.

Finally, future studies on IL viscosity modeling should be concerned with: (1) explaining non-Newtonian viscous behavior of ILs; (2) estimating the viscosity of binary and higher mixtures of ILs as well as the mixtures of ILs with organic solvents and/or water. It is quite clear that treatment of topic (1) with the models like those shown in this work is impossible, mostly due to the fact the problem has not been sufficiently explored yet—indeed, there are not many articles published in the open literature, where apparent viscosity of ILs is reported.[72−74] Extending the modeling to multicomponent systems, that is the topic (2), seems to be more readily feasible. Formally, the models presented herein could be applied to predict viscosity of mixtures of any number of ILs, if fractional GCs were employed to represent both qualitative and quantitative composition of the system. However, such an approach is not recommended, as using noninteger GCs will result in input variables outside the models' AD. The same refers to modeling molecular solvents using the GCs proposed in this work. Such calculations are prohibited as the functional groups adopted in this work can be applied only for ILs, despite the fact that they formally are capable of representing common organic compounds (e.g., alkanes, alcohols, esters). In my opinion, studies on empirical correlations should limit solely to pure fluids, whereas for multicomponent systems, the approaches like UNIFAC-VISCO[54,55] are much more promising and should be systematically revised and extended.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.iecr.9b03150.

Functional groups used to represent ILs; ions and their GC assignments; viscosity database with references; calculated data; model parameters (ZIP)

MATLAB function *ilviscogc* returning viscosity using the methods developed in this work along with a template for input file and instructions (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kpaduszynski@ch.pw.edu.pl. Phone: +48 (22) 234 56 40.

ORCID ⓘ

Kamil Paduszyński: 0000-0003-2489-6983

Notes

The author declares no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Welton, T. Room-Temperature Ionic Liquids. Solvents for Synthesis and Catalysis. *Chem. Rev.* **1999**, *99*, 2071−2084.

(2) Hallett, J. P.; Welton, T. Room-Temperature Ionic Liquids: Solvents for Synthesis and Catalysis. 2. *Chem. Rev.* **2011**, *111*, 3508−3576.

(3) Ghandi, K. Review of Ionic Liquids, Their Limits and Applications. *Green Sustainable Chem.* **2014**, *04*, 44−53.

(4) Holbrey, J. D.; Rogers, R. D. In *Ionic Liquids*; Rogers, R. D., Seddon, K. R., Eds.; American Chemical Society: Washington, DC, 2002; Vol. *818*, pp 2−14.

(5) Plechkova, N. V.; Seddon, K. R. Applications of Ionic Liquids in the Chemical Industry. *Chem. Soc. Rev.* **2008**, *37*, 123−150.

(6) Werner, S.; Haumann, M.; Wasserscheid, P. Ionic Liquids in Chemical Engineering. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 203−230.

(7) Dietz, M. L. Ionic Liquids as Extraction Solvents: Where Do We Stand? *Sep. Sci. Technol.* **2006**, *41*, 2047−2063.

(8) Pereiro, A. B.; Araújo, J. M. M.; Esperança, J. M. S. S.; Marrucho, I. M.; Rebelo, L. P. N. Ionic Liquids in Separations of Azeotropic Systems — A Review. *J. Chem. Thermodyn.* **2012**, *46*, 2−28.

(9) Ventura, S. P. M.; e Silva, F. A.; Quental, M. V.; Mondal, D.; Freire, M. G.; Coutinho, J. A. P. Ionic-Liquid-Mediated Extraction and Separation Processes for Bioactive Compounds: Past, Present, and Future Trends. *Chem. Rev.* **2017**, *117*, 6984−7052.

(10) Paduszyński, K. Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 1. Density. *Ind. Eng. Chem. Res.* **2019**, *58*, 5322−5338.

(11) Krossing, I.; Slattery, J. M. Semi-Empirical Methods to Predict the Physical Properties of Ionic Liquids: An Overview of Recent Developments. *Z. Phys. Chem.* **2006**, *220*, 1343−1359.

(12) Vega, L. F.; Vilaseca, O.; Llovell, F.; Andreu, J. S. Modeling Ionic Liquids and the Solubility of Gases in Them: Recent Advances and Perspectives. *Fluid Phase Equilib.* **2010**, *294*, 15−30.

(13) Maia, F. M.; Tsivintzelis, I.; Rodriguez, O.; Macedo, E. A.; Kontogeorgis, G. M. Equation of State Modelling of Systems with Ionic Liquids: Literature Review and Application with the Cubic Plus Association (CPA) Model. *Fluid Phase Equilib.* **2012**, *332*, 128−143.

(14) Coutinho, J. A. P.; Carvalho, P. J.; Oliveira, N. M. C. Predictive Methods for the Estimation of Thermophysical Properties of Ionic Liquids. *RSC Adv.* **2012**, *2*, 7322.

(15) Das, R. N.; Roy, K. Advances in QSPR/QSTR Models of Ionic Liquids for the Design of Greener Solvents of the Future. *Mol. Diversity* **2013**, *17*, 151−196.

(16) Dong, K.; Liu, X.; Dong, H.; Zhang, X.; Zhang, S. Multiscale Studies on Ionic Liquids. *Chem. Rev.* **2017**, *117*, 6636−6695.

(17) Hosseini, S. M.; Mulero, A.; Alavianmehr, M. M. Predictive Methods and Semi-Classical Equations of State for Pure Ionic Liquids: A Review. *J. Chem. Thermodyn.* **2019**, *130*, 47−94.

(18) Abbott, A. P. Application of Hole Theory to the Viscosity of Ionic and Molecular Liquids. *ChemPhysChem* **2004**, *5*, 1242−1246.

(19) Matsuda, H.; Yamamoto, H.; Kurihara, K.; Tochigi, K. Computer-Aided Reverse Design for Ionic Liquids by QSPR Using Descriptors of Group Contribution Type for Ionic Conductivities and Viscosities. *Fluid Phase Equilib.* **2007**, *261*, 434−443.

(20) Matsuda, H.; Yamamoto, H.; Kurihara, K.; Tochigi, K. Prediction of The Ionic Conductivity and Viscosity of Ionic Liquids by QSPR Using Descriptors of Group Contribution Type. *J. Comput. Aided Chem.* **2007**, *8*, 114−127.

(21) Slattery, J. M.; Daguenet, C.; Dyson, P. J.; Schubert, T. J. S.; Krossing, I. How to Predict the Physical Properties of Ionic Liquids: A Volume-Based Approach. *Angew. Chem., Int. Ed.* **2007**, *46*, 5384−5388.

(22) Tochigi, K.; Yamamoto, H. Estimation of Ionic Conductivity and Viscosity of Ionic Liquids Using a QSPR Model. *J. Phys. Chem. C* **2007**, *111*, 15989−15994.

(23) Bini, R.; Malvaldi, M.; Pitner, W. R.; Chiappe, C. QSPR Correlation for Conductivities and Viscosities of Low-Temperature Melting Ionic Liquids. *J. Phys. Org. Chem.* **2008**, *21*, 622−629.

(24) Gardas, R. L.; Coutinho, J. A. P. A Group Contribution Method for Viscosity Estimation of Ionic Liquids. *Fluid Phase Equilib.* **2008**, *266*, 195−201.

(25) Gardas, R. L.; Coutinho, J. A. P. Group Contribution Methods for the Prediction of Thermophysical and Transport Properties of Ionic Liquids. *AIChE J.* **2009**, *55*, 1274−1290.

(26) Billard, I.; Marcou, G.; Ouadi, A.; Varnek, A. In Silico Design of New Ionic Liquids Based on Quantitative Structure-Property Relation-ship Models of Ionic Liquid Viscosity. *J. Phys. Chem. B* **2011**, *115*, 93−98.

(27) Eiden, P.; Bulut, S.; Köchner, T.; Friedrich, C.; Schubert, T.; Krossing, I. In Silico Predictions of the Temperature-Dependent Viscosities and Electrical Conductivities of Functionalized and Nonfunctionalized Ionic Liquids. *J. Phys. Chem. B* **2011**, *115*, 300−309.

(28) Han, C.; Yu, G.; Wen, L.; Zhao, D.; Asumana, C.; Chen, X. Data and QSPR Study for Viscosity of Imidazolium-Based Ionic Liquids. *Fluid Phase Equilib.* **2011**, *300*, 95−104.

(29) Valderrama, J. O.; Muñoz, J. M.; Rojas, R. E. Viscosity of Ionic Liquids Using the Concept of Mass Connectivity and Artificial Neural Networks. *Korean J. Chem. Eng.* **2011**, *28*, 1451−1457.

(30) Gharagheizi, F.; Ilani-Kashkouli, P.; Mohammadi, A. H.; Ramjugernath, D.; Richon, D. Development of a Group Contribution Method for Determination of Viscosity of Ionic Liquids at Atmospheric Pressure. *Chem. Eng. Sci.* **2012**, *80*, 326−333.

(31) Mirkhani, S. A.; Gharagheizi, F. Predictive Quantitative Structure−Property Relationship Model for the Estimation of Ionic Liquid Viscosity. *Ind. Eng. Chem. Res.* **2012**, *51*, 2470−2477.

(32) Yu, G.; Zhao, D.; Wen, L.; Yang, S.; Chen, X. Viscosity of Ionic Liquids: Database, Observation, and Quantitative Structure-Property Relationship Analysis. *AIChE J.* **2012**, *58*, 2885−2899.

(33) Dutt, N. V. K.; Ravikumar, Y. V. L.; Rani, K. Y. Representation of Ionic Liquid Viscosity-Temperature Data by Generalized Correlations and an Artificial Neural Network (ANN) Model. *Chem. Eng. Commun.* **2013**, *200*, 1600−1622.

(34) Chen, B.-K.; Liang, M.-J.; Wu, T.-Y.; Wang, H. P. A High Correlate and Simplified QSPR for Viscosity of Imidazolium-Based Ionic Liquids. *Fluid Phase Equilib.* **2013**, *350*, 37−42.

(35) Yu, G.; Wen, L.; Zhao, D.; Asumana, C.; Chen, X. QSPR Study on the Viscosity of Bis(Trifluoromethylsulfonyl)Imide-Based Ionic Liquids. *J. Mol. Liq.* **2013**, *184*, 51−59.

(36) Paduszyński, K.; Domańska, U. Viscosity of Ionic Liquids: An Extensive Database and a New Group Contribution Model Based on a Feed-Forward Artificial Neural Network. *J. Chem. Inf. Model.* **2014**, *54*, 1311−1324.

(37) Alcalde, R.; García, G.; Atilhan, M.; Aparicio, S. Systematic Study on the Viscosity of Ionic Liquids: Measurement and Prediction. *Ind. Eng. Chem. Res.* **2015**, *54*, 10918−10924.

(38) Lazzús, J. A.; Pulgar-Villarroel, G. A Group Contribution Method to Estimate the Viscosity of Ionic Liquids at Different Temperatures. *J. Mol. Liq.* **2015**, *209*, 161−168.

(39) Zhao, Y.; Huang, Y.; Zhang, X.; Zhang, S. A Quantitative Prediction of the Viscosity of Ionic Liquids Using S_σ-profile Molecular Descriptors. *Phys. Chem. Chem. Phys.* **2015**, *17*, 3761−3767.

(40) Esmaeili-Jaghdan, Z.; Shariati, A.; Nikou, M. R. K. A Hybrid Smart Modeling Approach for Estimation of Pure Ionic Liquids Viscosity. *J. Mol. Liq.* **2016**, *222*, 14−27.

(41) Barycki, M.; Sosnowska, A.; Gajewicz, A.; Bobrowski, M.; Wileńska, D.; Skurski, P.; Giełdoń, A.; Czaplewski, C.; Uhl, S.; Laux, E.; et al. Temperature-Dependent Structure-Property Modeling of Viscosity for Ionic Liquids. *Fluid Phase Equilib.* **2016**, *427*, 9−17.

(42) Kang, X.; Zhao, Z.; Qian, J.; Muhammad Afzal, R. Predicting the Viscosity of Ionic Liquids by the ELM Intelligence Algorithm. *Ind. Eng. Chem. Res.* **2017**, *56*, 11344−11351.

(43) Baghban, A.; Kardani, M. N.; Habibzadeh, S. Prediction Viscosity of Ionic Liquids Using a Hybrid LSSVM and Group Contribution Method. *J. Mol. Liq.* **2017**, *236*, 452−464.

(44) Fatehi, M.-R.; Raeissi, S.; Mowla, D. Estimation of Viscosities of Pure Ionic Liquids Using an Artificial Neural Network Based on Only Structural Characteristics. *J. Mol. Liq.* **2017**, *227*, 309−317.

(45) Yan, F.; He, W.; Jia, Q.; Wang, Q.; Xia, S.; Ma, P. Prediction of Ionic Liquids Viscosity at Variable Temperatures and Pressures. *Chem. Eng. Sci.* **2018**, *184*, 134−140.

(46) Haghbakhsh, R.; Raeissi, S. Estimation of Viscosities of 1-Alkyl-3-Methylimidazolium Ionic Liquids over a Range of Temperatures Using a Simple Correlation. *Phys. Chem. Liq.* **2019**, *57*, 401−421.

(47) He, M.; Zhu, C.; Liu, X. Estimating the Viscosity of Ionic Liquid at High Pressure Using Eyring's Absolute Rate Theory. *Fluid Phase Equilib.* **2018**, *458*, 170−176.

(48) Macías-Salinas, R. A Viscosity Model for Ionic Liquids Based on the Eyring's Theory and a Cubic EoS. *J. Mol. Liq.* **2018**, *262*, 161−174.

(49) Polishuk, I. Modeling of Viscosities in Extended Pressure Range Using SAFT + Cubic EoS and Modified Yarranton−Satyro Correlation. *Ind. Eng. Chem. Res.* **2012**, *51*, 13527−13537.

(50) Abolala, M.; Peyvandi, K.; Varaminian, F. Modeling the Viscosity of Pure Imidazolium-Based Ionic Liquids Using SAFT-VR-Mie EoS. *Fluid Phase Equilib.* **2015**, *394*, 61−70.

(51) Dowell, N. M.; Llovell, F.; Sun, N.; Hallett, J. P.; George, A.; Hunt, P. A.; Welton, T.; Simmons, B. A.; Vega, L. F. New Experimental Density Data and Soft-SAFT Models of Alkylimidazolium ($[C_nC_1Im]^+$) Chloride (Cl^-), Methylsulfate ($[MeSO_4]^-$), and Dimethylphosphate ($[Me_2PO_4]^-$) Based Ionic Liquids. *J. Phys. Chem. B* **2014**, *118*, 6206−6221.

(52) Llovell, F.; Vega, L. F. Assessing Ionic Liquids Experimental Data Using Molecular Modeling: $[C_n Mim][BF_4]$ Case Study. *J. Chem. Eng. Data* **2014**, *59*, 3220−3231.

(53) Sun, Y.; Shen, G.; Held, C.; Feng, X.; Lu, X.; Ji, X. Modeling Viscosity of Ionic Liquids with Electrolyte Perturbed-Chain Statistical Associating Fluid Theory and Free Volume Theory. *Ind. Eng. Chem. Res.* **2018**, *57*, 8784−8801.

(54) Zhao, N.; Jacquemin, J.; Oozeerally, R.; Degirmenci, V. New Method for the Estimation of Viscosity of Pure and Mixtures of Ionic Liquids Based on the UNIFAC-VISCO Model. *J. Chem. Eng. Data* **2016**, *61*, 2160−2169.

(55) Zhao, N.; Oozeerally, R.; Degirmenci, V.; Wagner, Z.; Bendová, M.; Jacquemin, J. New Method Based on the UNIFAC−VISCO Model for the Estimation of Ionic Liquids Viscosity Using the Experimental Data Recommended by Mathematical Gnostics. *J. Chem. Eng. Data* **2016**, *61*, 3908−3921.

(56) Zhang, J.; Qin, L.; Peng, D.; Zhou, T.; Cheng, H.; Chen, L.; Qi, Z. COSMO-Descriptor Based Computer-Aided Ionic Liquid Design for Separation Processes. *Chem. Eng. Sci.* **2017**, *162*, 364−374.

(57) Song, Z.; Li, X.; Chao, H.; Mo, F.; Zhou, T.; Cheng, H.; Chen, L.; Qi, Z. Computer-Aided Ionic Liquid Design for Alkane/Cycloalkane Extractive Distillation Process. *Green Energy Environ.* **2018**, *4*, 154−165.

(58) Wang, J.; Song, Z.; Cheng, H.; Chen, L.; Deng, L.; Qi, Z. Computer-Aided Design of Ionic Liquids as Absorbent for Gas Separation Exemplified by CO_2 Capture Cases. *ACS Sustainable Chem. Eng.* **2018**, *6*, 12025−12035.

(59) Freire, M. G.; Neves, C. M. S. S.; Marrucho, I. M.; Coutinho, J. A. P.; Fernandes, A. M. Hydrolysis of Tetrafluoroborate and Hexafluorophosphate Counter Ions in Imidazolium-Based Ionic Liquids. *J. Phys. Chem. A* **2010**, *114*, 3744−3749.

(60) Snee, R. D. Validation of Regression Models: Methods and Examples. *Technometrics* **1977**, *19*, 415−428.

(61) Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the Definition of the Q^2 Parameter for QSAR Validation. *J. Chem. Inf. Model.* **2009**, *49*, 1669−1678.

(62) LSSVMlab (version 1.8). http://www.esat.kuleuven.be/sista/lssvmlab/ accessed (April 2, 2019).

(63) Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environ. Health Perspect.* **2003**, *111*, 1361−1375.

(64) Klimasauskas, C. C. In *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real World Performance*; Trippi, R. R., Turban, E., Eds.; Probus: Chicago, 1993; Chapter: Applying Neural Networks, pp 64−65.

(65) Paduszyński, K. In Silico Calculation of Infinite Dilution Activity Coefficients of Molecular Solutes in Ionic Liquids: Critical Review of Current Methods and New Models Based on Three Machine Learning Algorithms. *J. Chem. Inf. Model.* **2016**, *56*, 1420−1437.

(66) Paduszyński, K.; Królikowski, M.; Zawadzki, M.; Orzeł, P. Computer-Aided Molecular Design of New Task-Specific Ionic Liquids for Extractive Desulfurization of Gasoline. *ACS Sustainable Chem. Eng.* **2017**, *5*, 9032−9042.

(67) Marsh, K. N.; Brennecke, J. F.; Chirico, R. D.; Frenkel, M.; Heintz, A.; Magee, J. W.; Peters, C. J.; Rebelo, L. P. N.; Seddon, K. R. Thermodynamic and Thermophysical Properties of the Reference Ionic Liquid: 1-Hexyl-3-methylimidazolium Bis[(trifluoromethyl)-sulfonyl]amide (Including Mixtures). Part 1. Experimental Methods and Results (IUPAC Technical Report). *Pure Appl. Chem.* **2009**, *81*, 781−790.

(68) Netzeva, T. I.; Worth, A. P.; Aldenberg, T.; Benigni, R.; Cronin, M. T. D.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; et al. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships: The Report and Recommendations of ECVAM Workshop 52. *Altern. Lab. Anim.* **2005**, *33*, 155−173.

(69) Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review. *Altern. Lab. Anim.* **2005**, *33*, 445−459.

(70) Roy, K.; Kar, S.; Ambure, P. On a Simple Approach for Determining Applicability Domain of QSAR Models. *Chemom. Intell. Lab. Syst.* **2015**, *145*, 22−29.

(71) Eckert, F.; Klamt, A. *COSMOtherm*, version C3.0, Release 17.01; COSMOlogic GmbH & Co. KG: Leverkusen, Germany, 2017.

(72) Burrell, G. L.; Dunlop, N. F.; Separovic, F. Non-Newtonian Viscous Shear Thinning in Ionic Liquids. *Soft Matter* **2010**, *6*, 2080−2086.

(73) Jacquemin, J.; Anouti, M.; Lemordant, D. Physico-Chemical Properties of Non-Newtonian Shear Thickening Diisopropyl-ethyl-ammonium-Based Protic Ionic Liquids and Their Mixtures with Water and Acetonitrile. *J. Chem. Eng. Data* **2011**, *56*, 556−564.

(74) Smith, J. A.; Webber, G. B.; Warr, G. G.; Atkin, R. Rheology of Protic Ionic Liquids and Their Mixtures. *J. Phys. Chem. B* **2013**, *117*, 13930−13935.