Predicting Toxicity of Ionic Liquid Compounds Using Random Forest Approach

Safa SADAGHIYANFAM[1], Hiqmet KAMBERAJ[2,3], Yalcin ISLER[4, *]
[1]Cozumcul Technology, Department of Software Development, Menderes, Izmir, Turkey
[2]International Balkan University, Department of Computer Engineering, Skopje, North Macedonia
[3]National Institute of Physics, Fan Noli Square, Academy of Sciences of Albania, Tirana, Albania
[4]Alanya Alaaddin Keykubat University, Department of Electrical and Electronics Engineering, Alanya, Antalya, Turkey
ORCIDs: 0000-0002-6487-7501, 0000-0001-7357-7490, 0000-0002-2150-4756
E-mails: safa.sd1993@gmail.com, h.kamberaj@gmail.com, islerya@yahoo.com
*Corresponding author.

***Abstract*— Ionic liquids have applications across various scientific fields, in part, due to their interesting physical and chemical properties. Comprehensive assessments of their toxicological profiles are necessary to allow their safe and suitable applications. Unlike prior ionic liquid toxicity predictions which rely on small descriptor sets, we integrate joint 2D and 3D descriptors with random forest which explains the most important descriptors to address these limitations. This study aims to predict the toxicity of ionic liquids using 2D and 3D molecular descriptors by utilizing machine learning. In particular, we propose the random forest regression model to uncover molecular descriptors and toxicity patterns. Additionally, GridSearchCV is used to tune the hyperparameters to ensure optimal model performance. Several metrics were calculated to evaluate the model's accuracy. The model achieved R^2=0.879 which indicates strong predictive performance. Our study demonstrates the benefits of 2D and 3D descriptors for predicting the toxicity of ionic liquids, showing strong correlations between experimental and predicted toxicities. Our analysis of features using 2D and 3D descriptors highlighted those descriptors that are strongly associated with toxicity predictions. Feature importance highlights that physicochemical factors effect toxicity which provides interpretation for ionic liquid design. This study demonstrates the effectiveness of predicting the toxicity of ionic liquids by integrating molecular descriptors and machine learning, thereby facilitating the safer production and application of ionic liquids.**

***Keywords*— Ionic Liquids, Random Forest, Explainable ML, Toxicity Prediction, GridSearchCV, hyperparameter tuning, Molecular Descriptors, 2D and 3D descriptors**

I.INTRODUCTION

ILs are a unique and new class of salts consisting entirely of cations and anions that maintain a liquid form below 100˚C. ILs have gained attention for their applications in bio-catalysis and electrochemistry due to their high thermal stability and versatility, also known as green solvents [1,2,3]. However, recent studies highlight their environmental impact due to their varying level of toxicity [4,5,6,7].

The ILs pose a hurdle because the cation-anion combinations limit the ability to test every compound for toxicity. To address this gap, new methods, such as QSAR (Quantitative Structure-Activity Relationship), are being implemented [8,9]. QSAR model uses computational power to predict toxicity using molecular descriptors, resulting in faster and cheaper results than laboratory work without sacrificing quality.

Moreover, random forests, support vector machines, and gradient boosting approaches have demonstrated the ability to accurately describe complex nonlinear correlations between the structural characteristics of a molecular entity and toxicity. Hence, the use of machine learning techniques in QSAR has never been higher [10,11,12]. Moreover, molecular data in SMILES format can be more effectively described by deep models using convolutional neural networks (CNNs), which automatically perform feature extraction and enhance the model's performance [13,14].

This study focuses on utilizing a random forest regression model, implemented in Python[1], to examine the predictive capability of molecular descriptors for estimating the toxicity (logEC50) of ionic liquids, accomplishing the objectives of this study. This work aims to develop predictive models from two-dimensional (2D) and three-dimensional (3D) molecular descriptors. Additionally, this study seeks to utilize feature selection tools to identify key molecular descriptors, along with hyperparameter tuning, to enhance model reliability. The

[1] https:// https://www.geeksforgeeks.org/machine-learning/random-forest-regression-in-python/

proposed approach focuses on predicting the toxicity of ILs and improving the interpretability of descriptors corresponding to toxicity. Consequently, it stimulates the creation of more environmentally friendly and safer chemical alternatives.

This work offers a framework that is both interpretable and grounded in existing literature, connecting prevalent QSAR descriptors to chemically reasonable toxicity mechanisms and provides screening recommendations for ionic liquid candidates with reduced toxicity (see Figure 5).

II.*MATERIALS and METHODS*

A. Dataset acquisition

For this research, the dataset containing information on the toxicity of ILs was obtained from literature [15]. The dataset comprises toxicological data for 355 ILs represented as logEC50 values, which are measured on a logarithmic scale. Each record represents an ionic liquid in SMILES (Simplified Molecular Input Line Entry System) format. Each entry's experimentally obtained logEC50 value provides a quantitative assessment of toxicity, enabling analysis based on empirical data. Some samples of the dataset are provided below in Table 1, to show its structure and the type of information that are included.

Table 1: Data for Ionic Liquids and their Experimental logEC50

IL No.	SMILES	Experimental logEC50
1	[N+](C)(C)(CC)COCC.[Cl-]	3.59
2	O1c4c(O[B-]12Oc3c(O2)cccc3)cccc4.CC[N+](CC)(CC)CC	1.17
3	[N+](C)(C)(Cc1ccccc1)CCCCCCCCCC.[Cl-]	0.64

In the current study, ILs of the dataset are characterized by their molecular structure and associated toxicity values. The molecular structures are represented in SMILES notation, and their toxicity is expressed by experimental logEC50 values, which represent half-maximal effective concentration. We determined the 2D and 3D descriptors using the RDKit toolkit in Python (in particular, the functions in the Chem module)[2], representing the input vectors of the ionic liquid geometrical and chemical structure composition, as well as their physical/chemical molecular features. The final matrix included all computed 2d and 3d descriptors which were successfully calculated.

B. Dataset Preprocessing Tasks

We preprocessed the input data to enhance its quality and to make it suitable for use in machine learning.

• Imputation of Missing Values: Missing values were imputed with column-wise mean imputation using Scikit-learn's [16] SimpleImputer module in Python . Handling missing values was applied to the full descriptor matrix before training the model.

• Feature Integration and Normalization [17,18]: The toxicity dataset and molecular descriptor dataset were combined to yield an integrated feature matrix. Normalization was not applied either for calculated descriptors or random forest. Standardization was implemented only for plots to be diagnosed.

C. Model development

Since the RF repressor [19] is effective with high-dimensional datasets, resistant to overfitting, and interpretable using feature importance metrics, it was selected as the prediction model for this study. The dataset was split into an 80:20 ratios (80% for training, and 20% for testing). The test dataset was used to evaluate the model's prediction accuracy on new and unseen data [20].

Hyperparameter tuning was performed using *GridSearchCV* [21], targeting parameter adjustments of values such as the maximum tree depth (the so-called *max_depth* parameter in Python) and the number of estimators (the parameter *n_estimators* in Python). Hyperparameters were selected with GridSearchCV using 3-fold cross-validation on the training split with R^2 optimizing. The grid spanned *n_estimators {100, 200, 300}, max_depth {None, 10, 20, 30}, min_samples_split {2, 5, 10}, and min_samples_leaf{1, 2, 4} with n_jobs = −1 and random_state = 42*. The selected configuration was n_estimators = 300, max_depth = 10, min_samples_split = 2, and min_samples_leaf = 1.

D. Performance Metrics

The model's prediction accuracy was assessed using the following standard metrics.

- Mean Absolute Error (MAE): A measure of the average magnitude of prediction errors [22].
- Root Mean Squared Error (RMSE): It gives more emphasis to prediction errors for greater differences [23].
- Pearson correlation coefficient (*r*): It measures the scale of linear relationship between the predicted and observed toxicity (logEC50) values [24].
- Coefficient of Determination (R^2): It gives a proportion of variance in observed values that explained by the model [25].

[2] https://www.rdkit.org/docs/GettingStartedInPython.html

ml

E. Visualization and analysis

- *Importance of features:* The relative contribution of each molecular descriptor towards the prediction of IL toxicity was calculated based on the values of feature importance derived using the RF model [26]. A bar plot was created to identify the most critical features, providing information about the significant molecular properties accountable for toxicity predictions.
- *Correlation analysis:* Predictive validity of the model was assessed by comparing the experimental and predicted logEC50 values in a plot based on scatter. The Pearson correlation coefficient was also calculated, giving a numerical measure of the capacity of the model to approximate the linear relationship between these values [27].

III.RESULTS

A. Model Performance

RF regression model demonstrated good prediction capability for toxicity values of ionic liquids from 2D and 3D molecular descriptors. The following parameters were used to evaluate the model's performance on the test dataset, as shown in Table 2.

Table 2: Performance parameters of the model on the test dataset.

Evaluation Metrics	**Value**
R^2 Score	0.8790
MAE	0.2776
RMSE	0.3833
Pearson Correlation Coefficient	0.9379

In particular, our results demonstrate the model's accuracy in predicting toxicity using these molecular descriptors through the ML approach. This is evident from the high linear correlation between predicted and actual values, as shown by the high Pearson correlation coefficient (see Table 2). The first goal of accurate toxicity prediction of ionic liquid is supported in Table 2. The second goal which is identifying key descriptors is addressed by explaining the most influential descriptors below.

B. Feature importance

The feature importance analysis showed the molecular descriptors most important to the predictive accuracy of the model. Among the 20 top descriptors, *SMR_VSA5* and *VSA_EState7* had significantly higher importance scores compared to the others. *SMR_VSA5* calculates van-der-Waals areas for atoms that are contained within a mid-range molar-refractivity bin, which is chosen here to positively track exposed hydrophobic surface.

Higher values, therefore, indicate bulkier or more hydrophobic fragments, a trend that is also seen at higher IL toxicity (due to better penetration into the membrane).

VSA_EState7 is the sum of atom contributions for a mid-high area bin; it connects local electronic environment with available surface. In application, higher VSA_EState7 values mean more electronically active surface exposed atoms as would be expected for more intense intermolecular interactions that can raise observed toxicity.

When combined, these descriptors suggest a hydrophobicity/accessible-surface and local-electronics process, suggesting that bulk and charge-distribution properties are the main sources of the toxicity estimates in the model. A relative importance bar chart of the 20 most influential descriptors is provided in Fig. 1.

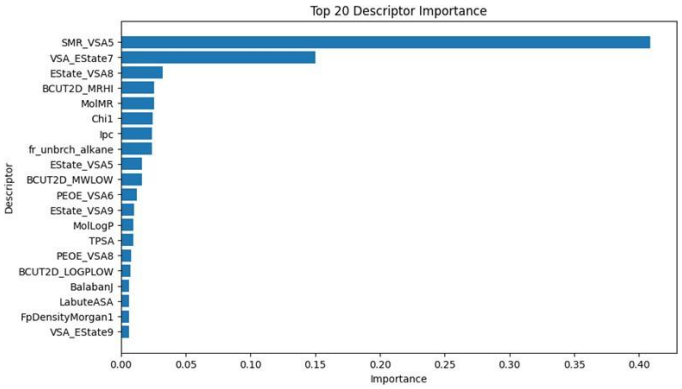

Figure 1: Bar plot of the top 20 molecular descriptors ranked by importance.

C. Correlation between actual and predicted values

Fig. 2 is a scatter plot between observed and expected logEC50 values. The scatter plot is highly linear along the diagonal line (y = x) and exhibits good predictability. The random outliers could have been a result of dataset limitations, even though almost all the predictions, except for a few, were close to the experimental value.

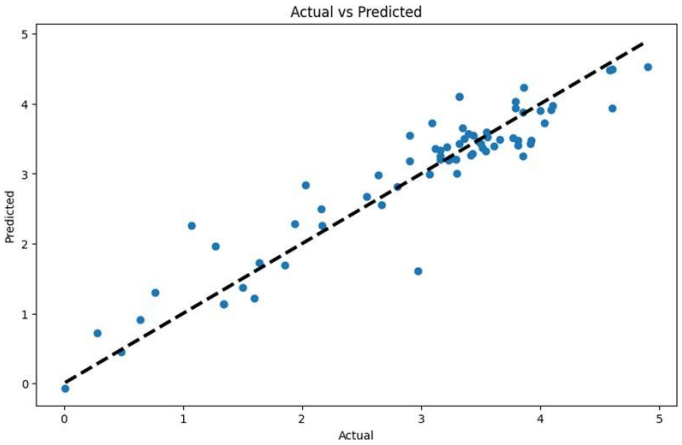

Figure 2: Scatter plot of the observed versus predicted logEC50 values. The diagonal line represents the baseline.

D. Model evaluation with confidence intervals

In this study, we investigated the RF model's performance on training and test datasets, with a focus on confidence intervals, which can provide insight into the uncertainty and variability of model predictions for individual data points.

E. Confidence interval for test data

Fig. 3 shows the mean of predicted and actual points for the test set, along with a 95% confidence interval for each point. The same analysis also highlights the spread in predictions and indicates where the model's performance might be less uniform.

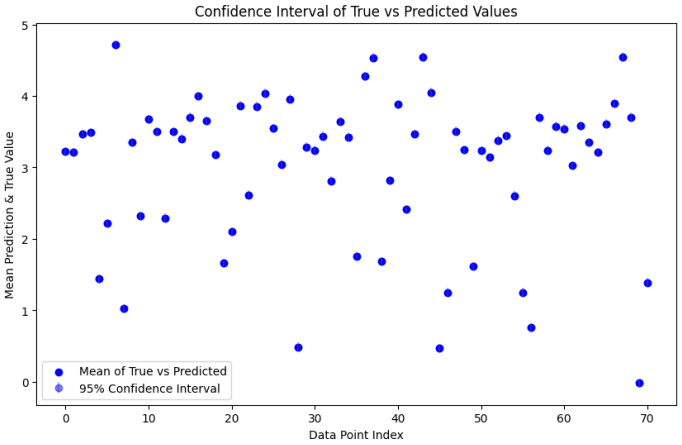

Figure 3: The 95% confidence interval of observed versus predicted toxicities measured for the test dataset.

The critical observations are as follows:

- The ones with broad intervals indicate instances where the model doesn't predict well Most of the data points have tight confidence intervals, showing that the predictions are reasonable.

F. Comparison across training and datasets

The mean of actual and predicted values for the train and test sets, along with a 95% confidence interval for each point, is depicted in Fig. 4.

The critical observations are as follows:

- The model is a good fit to the training data, as is evident from the lower confidence intervals of the training dataset.
- As is the case when extrapolating to new data, the test data intervals are marginally wider.

Confidence interval analysis identifies areas for optimization in the outliers and verifies that the model performs well on training data and reasonably well on test data.

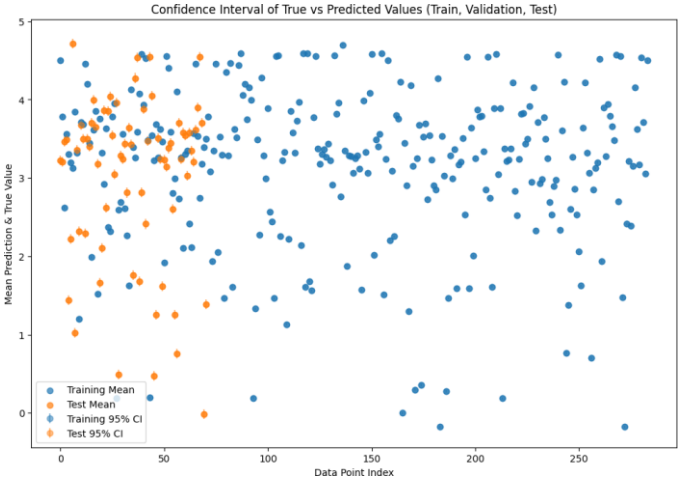

Figure 4: Confidence Interval of True vs. Predicted Values for Training and Test Datasets.

length beyond a certain point may negatively impact the model's performance due to insufficient data.

IV.Discussion

The correlation coefficient obtained between all molecular descriptors and IL toxicity was 0.8790, and the Pearson correlation coefficient was 0.9379, demonstrating the high capability of the model to predict IL toxicity from their molecular descriptors. In addition, compared with other models (See Table 3), the low MAE (0.2776) and RMSE (0.3833) verify the reliability and universality of the model.

Table 3: comparison table that benchmarks results against representative IL toxicity QSAR studies.

Study	Endpoint / cell	n	Algorithm(s)	Reported R^2 (test unless noted)
Zhao et al., 2014 [28].	IPC-81, logEC50	304	MLR, SVM	0.892 (MLR), 0.958 (SVM)
Wang et al., 2020 [2].	IPC-81, logEC50	355	FNN, SVM	0.8917 (FNN), 0.9202 (SVM)
Ahmadi et al., 2022 [29].	IPC-81, logEC50	304	CORAL (Monte Carlo QSTR)	0.85 (validation)
Proposed Model	IPC-81, logEC50	355	Random Forest	0.879 (test)

The feature importance analysis identified molecular descriptors critical to the predictive ability of the model. The two highest-ranked descriptors, *SMR_VSA5 and VSA_EState7*, are associated with electronic and spatial molecular properties. They are likely to describe meaningful physicochemical interactions involved in IL toxicity, such as lipophilicity, size, electron density, reactivity, and polarity. The prevalence of these traits aligns with previous research, which has also portrayed the prominence of electronic and spatial characteristics in IL interactions with biological systems. Our test set performance ($R^2 = 0.879$) is consistent with the reported range for ionic-liquid toxicity QSAR models for comparable endpoints and assessment schemes [30,31]. Consistent with the QSAR IL literature in general, tree-based approaches are rivaling conventional baselines: previous work shows SVM and MLR giving good results on similar IL toxicity tasks, against which our Random Forest result is competitive [12,28].

In order to investigate the chemical reasonableness of salience of SMR_VSA5 and VSA_EState7; RDKit classifies them as surface-area-hybrid descriptors summarizing molar-refractivity-binned van-der-Waals surface (SMR_VSA) or electrotopological-state values over surface-area bins (VSA_EState) with hydrophobic surface-exposure and local electronic environment tied to interaction potential [32]. The underlying electrotopological-state indices of VSA_EState are long-established QSAR indices of historical precedent and represent atom-level electronic and topological effects pertinent to activity and toxicity.

For clarity, we say that cross-paper metric comparison should be treated with caution since endpoints, descriptor sets, and split protocols vary, thus our quantitative benchmarking is stated as "within-range" and not superiority assertions [33].

The RF model performed as well as other recent machine learning models, such as Meta-Ensemble for IL toxicity prediction [34]. It is as effective as complex models in providing a satisfactory linear correlation between experimental and predicted toxicity values. Specifically, RF's interpretability, as revealed through feature importance analysis, gives it a significant advantage compared with more sophisticated models. The appropriateness of RF for model interpretability predictive tasks comes to the front as a result of this trade-off between accuracy and transparency [3,4].

Though the model functioned reasonably well overall, there were outlying data points in the scatter plot that resulted in discrepancies in experimental and calculated values. These are induced by missing data points in the given dataset; for example, present noise, diversity of the structures, and physical-chemical properties of ionic liquid molecules, or lack of significant descriptors that could play a crucial role in predicting toxicity. Furthermore, overreliance on the descriptors necessitates a closer examination of their influence on the prediction of ionic liquids' toxicity using the machine learning approaches. Below is a schematic overview summarizing the study's workflow.

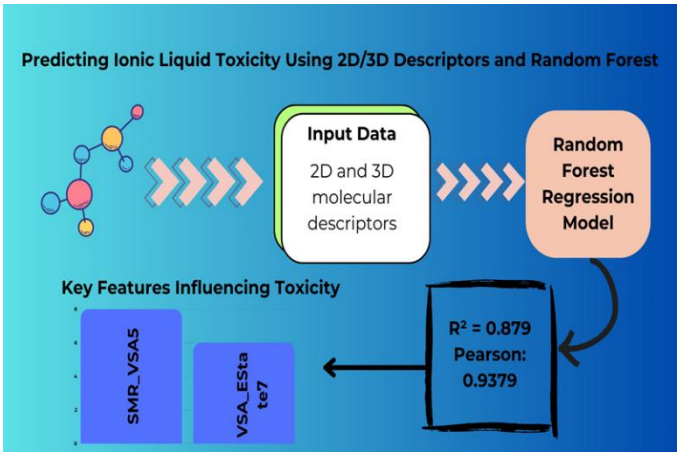

Figure 5: Conceptual summary of the QSAR workflow for ionic-liquid toxicity prediction. Curated IL structures were encoded as 2D and 3D RDKit descriptors, modeled with a RF regression. Feature-importance analysis highlights SMR_VSA5 and VSA_EState7 among the top contributors to the model's estimates.

V.Implications and Future Directions

The findings of this study hold tremendous implications for the design of cleaner and more secure ionic liquids. This work provides a structure-based optimization framework for ILs that reduce environmental and biological risks by defining molecular descriptors that have the most influence on predicting ionic liquid toxicity.

Furthermore, the efforts may include expanding the diversity of molecules in the dataset with ionic liquids of different physical and chemical properties, as well as geometrical structures, including various chemical functional groups. Furthermore, including other descriptors, such as quantum-chemical descriptors, could enhance the predictive performance of machine learning approaches. Exploring ensemble or hybrid models could further improve prediction power by improving the algorithms. The work demonstrates the potential of applying 2D and 3D molecular descriptors, along with machine learning, for more effective prediction of IL toxicity. The creation of computational facilities in green chemistry facilitates the development of environmentally friendly and safer chemical substitutes.

VI.Conclusions

In this study, we examined the use of the RF regression model to predict the toxicity (in logEC50 scale) of ionic liquid molecules. The molecules were represented by their 2D and 3D molecular descriptors.

The model was found to be predictive, as indicated by an R^2 value of 0.8790, a Pearson correlation coefficient of 0.9379, and low error values (MAE: 0.2776, RMSE: 0.3833). These findings suggest the potential use of machine learning approaches to model even complex chemical and physical relationships between molecular descriptors and IL toxicity at a desired level of accuracy.

Furthermore, feature selection analysis identified crucial molecular descriptors (e.g., SMR_VSA5 and VSA_EState7) as the most critical features, based on their scores, in predicting ILs toxicity. This is crucial in providing insights into the electronic and structural properties that govern IL toxicity, thereby facilitating the development of safer and greener ionic liquids. The high correlation between predicted and experimentally obtained values further assures the robustness of the Random Forest model and its applicability in cheminformatics.

Despite these results, the study has some limitations, including missing descriptors and outliers, which can affect toxicity. Future studies could explore the use of more sophisticated models, such as ensemble or hybrid models, to improve predictive accuracy and generalizability of the molecular descriptors.

In short, this study is a contribution to green chemistry as it presents a computational method of estimating IL toxicity. The results opened the door to the rational design of less harmful ionic liquids, securing their safe and sustainable application in industry and the environment.

Acknowledgement

This study was also supported by the Izmir Katip Çelebi University Scientific Research Council Agency, as project number 2024-TDR-FEBE-0024, for the doctoral thesis studied by Safa Sadaghiyanfam.

CONFLICTS

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Data and Software Availability

The following supplementary files are available upon request to the corresponding author.

- DescriptorsCalculation.ipynb: Python notebook file for Calculating 2D and 3D descriptors by using RDkit first.
- Descriptors+RandomForest.ipynb: Python notebook file for Random Forest model by using GridSearch for hyperparameter optimization after calculating descriptors.
- Book1.csv: SMILES and corresponding toxicity values.

REFERENCES

[1] Earle, M. J.; Esperança, J. M.; Gilea, M. A.; Canongia Lopes, J. N.; Rebelo, L. P.; Magee, J. W.; Seddon, K. R.; Widegren, J. A. The distillation and volatility of ionic liquids. *Nature* **2006**, *439*, 831–834.

[2] Wang, Z.; Song, Z.; Zhou, T. Machine learning for ionic liquid toxicity prediction. *Processes* **2020**, *9*, 65.

[3] Mousavi, S. P.; Nakhaei-Kohani, R.; Atashrouz, S.; Hadavimoghaddam, F.; Abedi, A.; Hemmati-Sarapardeh, A.; Mohaddespour, A. Modeling of H2S solubility in ionic liquids: comparison of white-box machine learning, deep learning and ensemble learning approaches. *Scientific Reports* **2023**, *13*, 7946.

[4] Hemmerich, J.; Ecker, G. F. In silico toxicology: From structure–activity relationships towards deep learning and adverse outcome pathways. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2020**, *10*, e1475.

[5] Goh, G. B.; Siegel, C.; Vishnu, A.; Hodas, N. O.; Baker, N. Chemception: a deep neural network with minimal chemistry knowledge matches the performance of expert-developed QSAR/QSPR models. *arXiv preprint arXiv:1706.06689* **2017**,

[6] Pretti, C.; Chiappe, C.; Pieraccini, D.; Gregori, M.; Abramo, F.; Monni, G.; Intorre, L. Acute toxicity of ionic

liquids to the zebrafish (Danio rerio). Green Chemistry 2006, 8, 238–240.

[7] Pawłowska, B.; Telesiński, A.; Biczak, R. Phytotoxicity of ionic liquids. *Chemosphere* **2019**, *237*, 124436.

[8] Tabaaza, G. A.; Tackie-Otoo, B. N.; Zaini, D. B.; Otchere, D. A.; Lal, B. Application of machine learning models to predict cytotoxicity of ionic liquids using VolSurf principal properties. *Computational Toxicology* **2023**, *26*, 100266.

[9] Yuan, Q.; Wei, Z.; Guan, X.; Jiang, M.; Wang, S.; Zhang, S.; Li, Z. Toxicity prediction method based on multi-channel convolutional neural network. *Molecules* **2019**, *24*, 3383.

[10] Baskin, I.; Epshtein, A.; Ein-Eli, Y. Benchmarking machine learning methods for modeling physical properties of ionic liquids. *Journal of Molecular Liquids* **2022**, *351*, 118616.

[11] Fan, D.; Xue, K.; Zhang, R.; Zhu, W.; Zhang, H.; Qi, J.; Zhu, Z.; Wang, Y.; Cui, P. Application of interpretable machine learning models to improve the prediction performance of ionic liquids toxicity. *Science of The Total Environment* **2024**, *908*, 168168.

[12] Zhu, P.; Kang, X.; Zhao, Y.; Latif, U.; Zhang, H. Predicting the toxicity of ionic liquids toward acetylcholinesterase enzymes using novel QSAR models. *International journal of molecular sciences* **2019**, *20*, 2186.

[13] Zhou, Z. H. *Ensemble methods: foundations and algorithms*; CRC Press, 2012.

[14] Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences* **1988**, *28*, 31–36.

[15] Chen, G.; Song, Z.; Qi, Z.; Sundmacher, K. Generalizing property prediction of ionic Liquids from limited labeled data: a one-step framework empowered by transfer learning. *Digital Discovery* 2023, 2, 591-601.

[16] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

[17] Mining, W. I. Data mining: Concepts and techniques. *Morgan Kaufmann* **2006**, *10*, 4.

[18] Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Machine learning* **2002**, *46*, 389–422.

[19] Schonlau, M.; Zou, R. Y. The random forest algorithm for statistical learning. *The Stata Journal* **2020**, *20*, 3–29.

[20] Bishop, C. M. Pattern recognition and machine learning. *Springer Google Scholar* **2006**, *2*, 1122–1128.

[21] Bergstra, J.; Bengio, Y. Random search for hyperparameter optimization. *Journal of machine learning research* **2012**, *13*.

[22] Willmott, C. J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research* **2005**, *30*, 79–82.

[23] Chai, T.; Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, **2014**, *7*, 1247–1250.

[24] Cohen, I.; Huang, Y.; Chen, J.; Benesty, J.; Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. *Noise reduction in speech processing* **2009**, 1–4.

[25] Kvålseth, T. O. (1985). Cautionary Note about R^2. *The American Statistician*, *39*(4), 279–285.

[26] Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.

[27] Heuillet, A.; Couthouis, F.; Díaz-Rodríguez, N. Explainability in deep reinforcement learning. *Knowledge-Based Systems* **2021**, *214*, 106685.

[28] Zhao, Yongsheng, et al. "Toxicity of ionic liquids: database and prediction via quantitative structure–activity relationship method." *Journal of hazardous materials* 278 (2014): 320-329.

[29] Ahmadi, Shahin, Shahram Lotfi, and Parvin Kumar. "Quantitative structure–toxicity relationship models for predication of toxicity of ionic liquids toward leukemia rat cell line IPC-81 based on index of ideality of correlation." *Toxicology Mechanisms and Methods* 32.4 (2022): 302-312.

[30] Peng, Daili, and Francesco Picchioni. "Prediction of toxicity of Ionic Liquids based on GC-COSMO method." Journal of hazardous materials 398 (2020): 122964.

[31] Chipofya, Mapopa, Hilal Tayara, and Kil To Chong. "Deep probabilistic learning model for prediction of ionic liquids toxicity." *International Journal of Molecular Sciences* 23.9 (2022): 5258.

[32] Landrum G. What are the VSA descriptors? RDKit blog. 2023 Apr 17. Available from: https://greglandrum.github.io/rdkit-blog/posts/2023-04-17-what-are-the-vsa-descriptors.html

[33] Alexander DL, Tropsha A, Winkler DA. Beware of R(2): Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. J Chem Inf Model. 2015 Jul 27;55(7):1316-22.

[34] Sadaghiyanfam, Safa, Hiqmet Kamberaj, and Yalcin Isler. "Enhanced prediction of ionic liquid toxicity using a meta-ensemble learning framework with data augmentation." *Artificial Intelligence Chemistry* 3.1 (2025): 100087.