QSPR Models for Predicting Density and Refractive Index of Ionic Liquid Binary Mixtures

Shamkhal Baybekov[1], Durna Aghamaliyeva[2,3], Kamran Heydarov[4], Orkhan Abdullayev[5], Khayala Gaybaliyeva[6], Avtandil Talybov[2], Rufat Azizov[3], Vagif Abbasov[2]

[1] *Chemoinformatics and Computational Chemistry Research Group, Azerbaijan State Oil and Industry University, 34 Azadliq ave., AZ 1010, Baku, Azerbaijan.*

[2] *Institute of Petrochemical Processes of Azerbaijan National Academy of Sciences, Hojaly ave. 30, AZ1025, Baku, Azerbaijan.*

[3] *Azerbaijan State Oil and Industry University, 34 Azadliq ave., AZ 1010, Baku, Azerbaijan.*

[4] *French-Azerbaijani University, Azerbaijan State Oil and Industry University, AZ1010, Baku, Azerbaijan.*

[5] *University of Strasbourg, Blaise Pascal str. 4, 67000, Strasbourg, France.*

[6] *Independent Researcher, 17 Auguste Kervern str., 29200, Brest, France.*

Ionic liquids (ILs) are versatile tunable salts with emerging applications in catalysis, separations, and electrochemistry. Despite growing interest, predictive models for their physicochemical properties in solvent mixtures remain scarce. In this work, we developed quantitative structure–property relationship (QSPR) models for predicting the density and refractive index of IL binary mixtures with water, ethanol, and isopropanol. Six solvent-specific XGBoost ensemble models were trained on curated data from the ILThermo database using 2D Mordred molecular descriptors, temperature, and mole fraction as input features. A GroupKFold cross-validation strategy based on IL identity ensured that performance metrics reflect genuine generalization to unseen structures. Models trained exclusively on pure IL data failed to predict mixture properties, confirming the necessity of mixture-specific training data that explicitly incorporates composition. The dedicated mixture models achieved 5-fold cross-validation R^2 values of 0.90–0.94 for density and 0.90–0.93 for refractive index. External validation was performed on a newly synthesized set of 31 ammonium–carboxylate ILs, characterized at the Institute of Petrochemical Processes using an Anton Paar DMA 4500 M densimeter and an Abbemat 500 refractometer. SHAP feature importance analysis identified mole fraction as the dominant predictor across all six models, followed by temperature and structural descriptors encoding molecular size, polarizability, and halogen content. The dominance of mole fraction over molecular structure at low IL concentrations explains the models' maintained prediction accuracy on structurally diverse external test compounds. Chemical space analysis via functional group profiling and UMAP visualization confirmed substantial structural differences between the ILThermo training data and the ammonium–carboxylate test set, further supporting the robustness of the composition-anchored predictions. All trained models, curated datasets, and the pyIonics data retrieval tool are freely available on GitHub (figure 1).

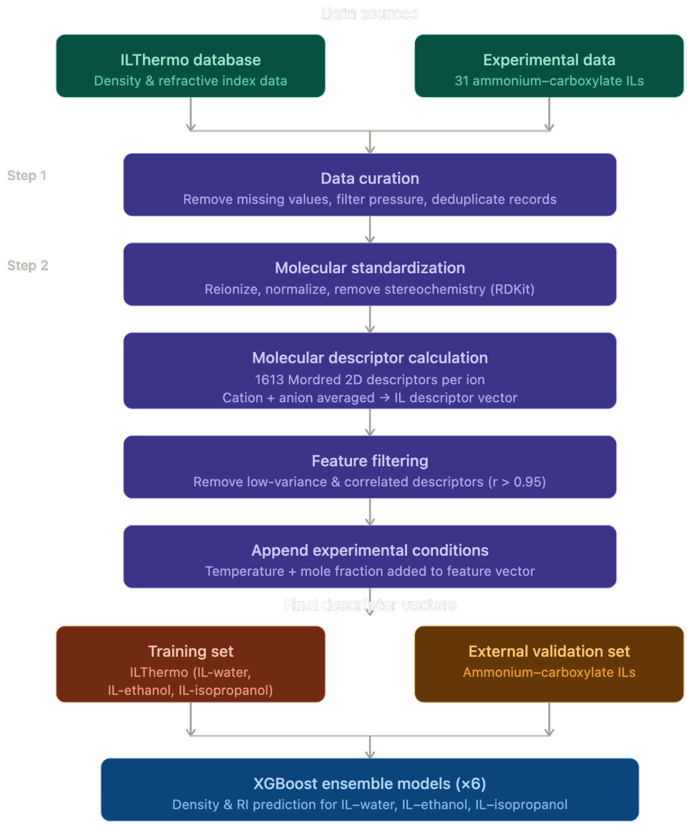

Figure 1. Overview of the modeling workflow. Density and refractive index data for pure ILs and their binary mixtures were retrieved from ILThermo using pyIonics, curated, and used to train six solvent-specific XGBoost ensemble models. External validation was performed on ammonium–carboxylate ILs synthesized at the Institute of Petrochemical Processes.

Bibliography:

[1] Baskin, I.I. et al. J. Chem. Inf. Model. 62 (2022) 5974–5985.

[2] Liu, Z. et al. Chem. Eng. J. 455 (2023) 140579.

[3] Moriwaki, H. et al. J. Cheminform. 10 (2018) 4.

[4] Dong, Q. et al. J. Chem. Eng. Data 52 (2007) 1151–1159.