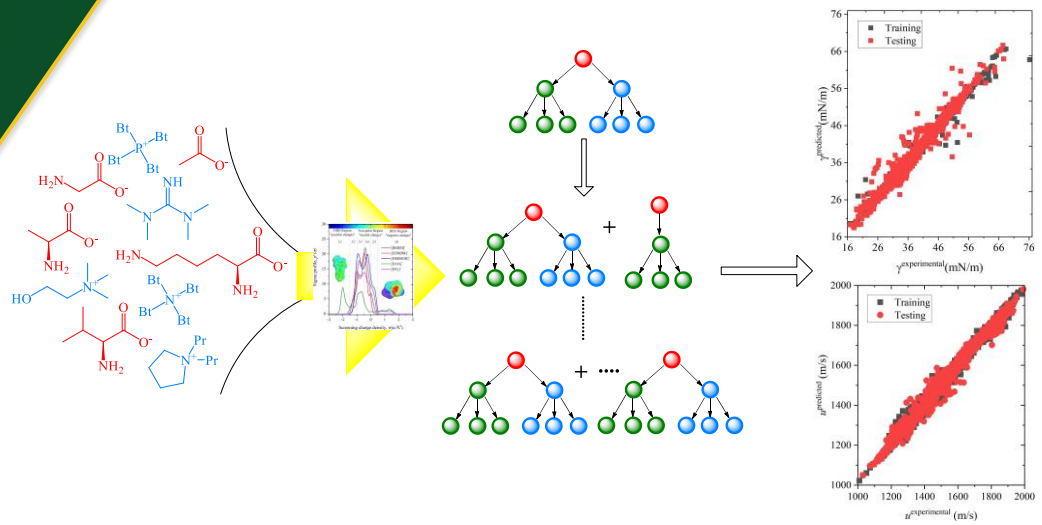
Artificial Intelligence Predicts Key Physical Properties of Ionic Liquids



Gradient boosting tree machine learning model.

1. Mohan M, Smith MD, Demerdash ON, Simmons BA, Singh S, Kidder MK, Smith JC. (2023) Quantum Chemistry-Driven Machine Learning Approach for the Prediction of the Surface Tension and Speed of Sound in Ionic Liquids; *ACS Sust. Chem. Eng.* DOI: 10.1021/acssuschemeng.3c00624
2. Mohan M, Smith MD, Demerdash ON, Kidder MK, Smith JC. (2023) Predictive Understanding the Surface Tension and Velocity of Sound in Ionic Liquids using Machine Learning. *J. Chem. Phys.* (accepted)

This work was supported and provided by the U. S. DOE Office of Science Genomic Science Program Office of Biological and Environmental Research (contract no. FWP ERKP752). Michelle K. Kidder acknowledges the U. S. DOE Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences (CSGB) Grant Number 3ERKCG25 for partially supporting this research

Scientific Achievement

We demonstrated that descriptive features derived from quantum chemical COnductor-like Screening MOdel for Realistic Solvents (COSMO-RS) surface polarization profiles of ionic liquids (ILs) can be used in training machine-learned (ML) regression models to accurately predict certain physical properties of ionic liquids, i.e. surface tension.

Significance and Impact

Machine-learning approaches to predict physiochemical properties of solvents have the potential to reduce the need for expensive and time-consuming experimental solvent screening campaigns in bioenergy research and elsewhere; however, prior work using traditional cheminformatics features to train models has had mixed success. Here, in an ORNL/JBEI collaboration we demonstrate that using features derived from quantum chemistry it possible to use AI to generate superior predictive models. These should be able to be extended into other important IL solvent properties.

Research Details

- A comprehensive dataset of 2524 data points was collected from the literature of the surface tension of ILs for 360 different ILs.
- For the speed of sound, 5702 data points were collected from the literature for 218 different ILs at a different temperatures and pressures
- We calculated a binned probability of polarized charge at the molecular surface (sigma profile) that we hypothesized is likely to implicitly capture the propensity for critical intermolecular interactions.
- This hypothesis was validated through the ML model's excellent performance as well as the *post hoc* interrogation of the ML model to ascertain the relative importance of features used to train the model.
- Gradient boosting tree (GBT) machine learning models showed excellent predictive power.