JOHN T. BARBER & RAYMOND W. HOLSAPPLE, Department of Physical Sciences and Mathematics, West Liberty University, West Liberty, WV, 26074. Using neural networks for molecular energy regression.

The PubChem library houses structural data for more than 100 million molecules. Simulations of molecular properties are both computationally and financially expensive. In the first phase of this research, we demonstrated how supervised learning could be used to predict molecular energy using only a molecule's structure. Specifically, we used a vectorized Coulomb matrix concatenated with molecular shape multipoles as our regression training feature. Various models were compared, with our two primary metrics being root mean squared error (RMSE) and the coefficient of determination (R2). A random forest (RF) model performed best on these metrics; however, model training time was also taken into consideration. A bagging regression model (BG) was consistently within 1% of RF in the primary metrics, but was faster by an order of magnitude. In the second phase of this research, we extended our regression modeling to include various neural network architectures using the same training feature from phase one. We have created and tested models with single and multiple hidden layers. In this phase we have also created a new evaluation metric, K-Accuracy, which computes the proportion of predictions that fall within a specified percentage error. After fine tuning various model parameters in our neural network, results were comparable to RF and BG. Currently, we are in phase three of this research where we are investigating two key improvements. We are coding an algorithm that creates a training feature called Bag-of-Bonds, which accounts for bond-type, and we are extending our neural network to a deep neural network.