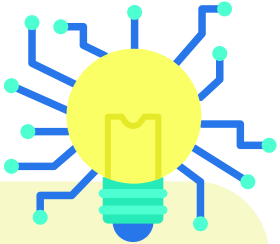
Machine Learning-Enhanced Computational Modeling of Metal-Protein Interactions

Charlie Klawitter
Mentor: Dr. Mengdi Huai
July 11, 2025

Project Role



Train and Refine a Machine Learning Model to Enhance the Modeling of Metal-Protein Interactions

Summary of Work to Date

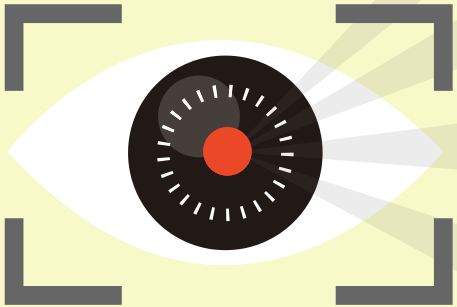
- Develop a program to extract necessary data from EFP .log files
- Create baseline ML model to predict the Final EFP Energies
- Program a neural network model to predict key energies from a range of molecule features on synthetic data

Future Plans

- Introduce real, up to date data for metal systems into the baseline neural network model
- Record discoveries and refine model to increase accuracy and efficiency
- Generate future ambition for the project

Learning Outcomes

Learning Points



Machine Learning — How to develop a machine learning model in PyTorch

Git , Linux, and Python — Git flow, version control, Linux operations, and python skills

Effective Fragment Potential — Basic understanding of EFP and key features

Research Methodology — The proper flow of developing a strong research project

Progress (EFP Energies Model)

Data Extraction Python Script

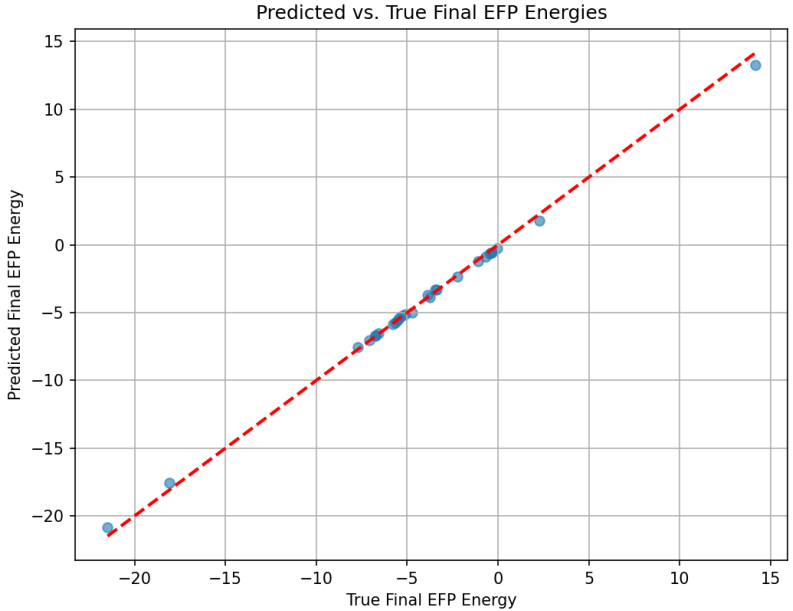
- Extracted data to obtain key energies from molecules
- Converted from Hartree to Kcal/Mol
- Electrostatic, Penetration, Polarization, Final EFP, etc.

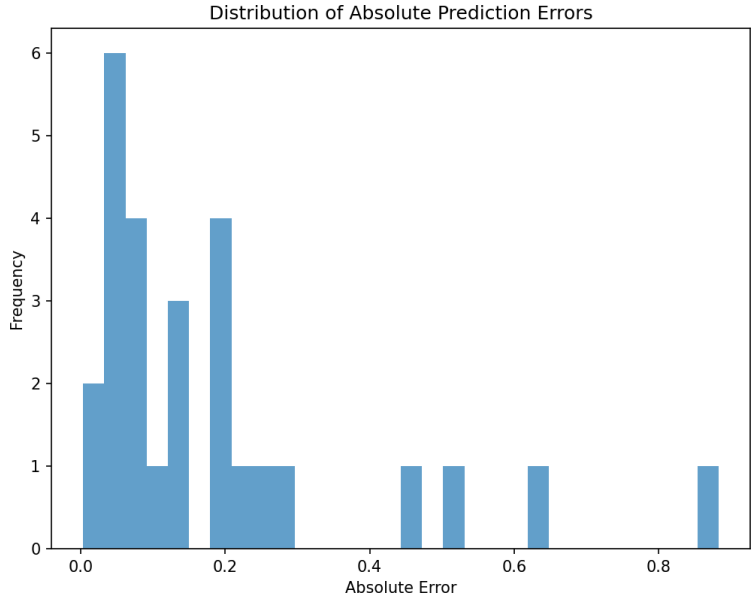
and

Basic Neural Network to Predict Final EFP Energy

- 80/20 Train/Test split
- Test Mean Squared Error: 0.0781

Progress (EFP Energies Model)





Progress (Synthetic Data Neural Network)

01 Input Features

Charge of fragments, distance between fragments, dipole moment of fragments, etc.

02 Target Outputs

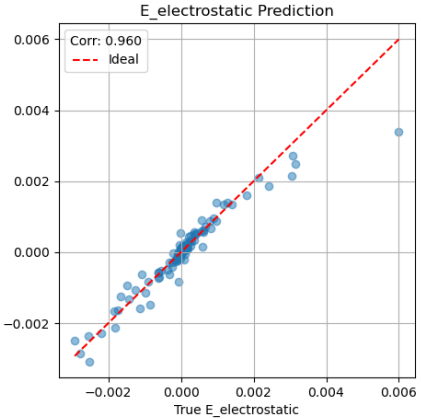
Electrostatic energy, Polarization energy, Dispersion energy, etc.

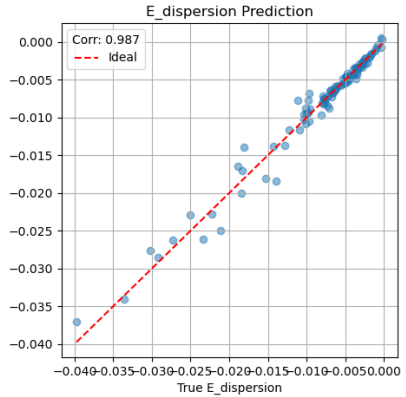
03 Model Specifications

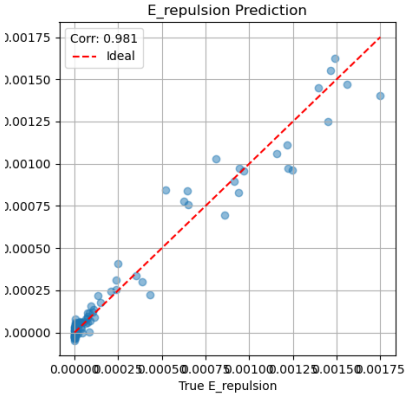
200 epochs, 80/20 split, 3 layers for NN, and 500 data points

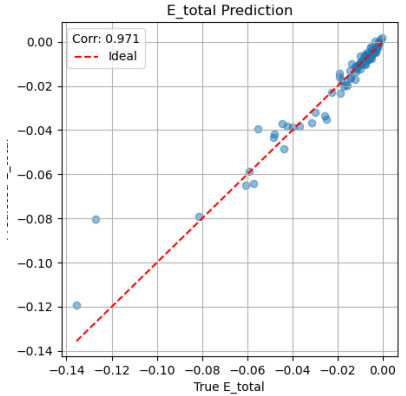
04 Results

Test MSE: 0.1037









Progress (Sij/Rij/undamped electrostatic energy)

$$E_{e-e}^{pen} = -2\left(\frac{1}{-2ln|S_{ij}|}\right)^{\frac{1}{2}} \frac{S_{ij}^2}{R_{ij}}$$

S_{ij}: Intermolecular overlap integral between LMOs i and j

R_{ij}: The distance between centroids of LMOs i and j

- Extracted Sij and Rij values along with Undamped Electrostatic Energy
- These features were used as input data for a neural network model with 1 hidden layer (simple at first)
- The goal is to predict the damped electrostatic energy, therefore solving the problem of creating a more accurate and efficient dampening function

Progress (Sij/Rij/undamped electrostatic energy)

01 Input Features

SIJ/RIJ Values and Undamped Electrostatic Energy

02 Target Outputs

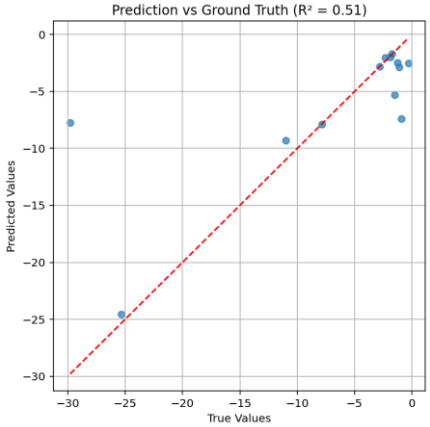
Damped Electrostatic Energy

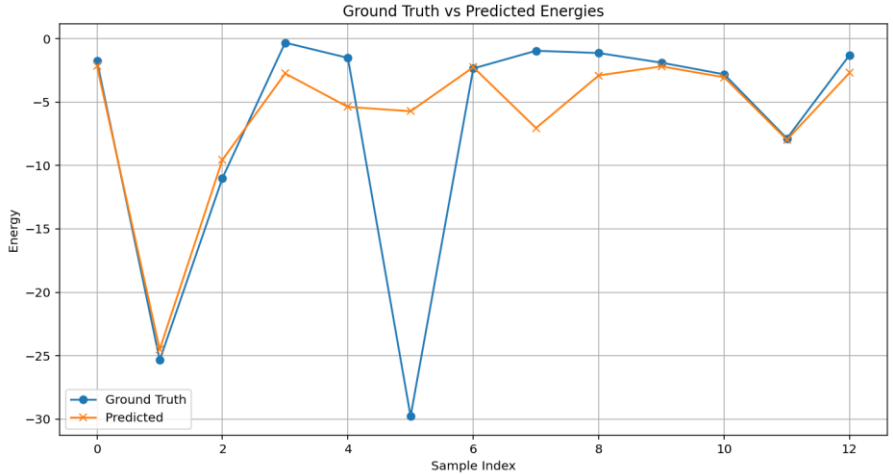
03 Model Specifications

65 samples, 100 epochs, 80/20, 3 layers

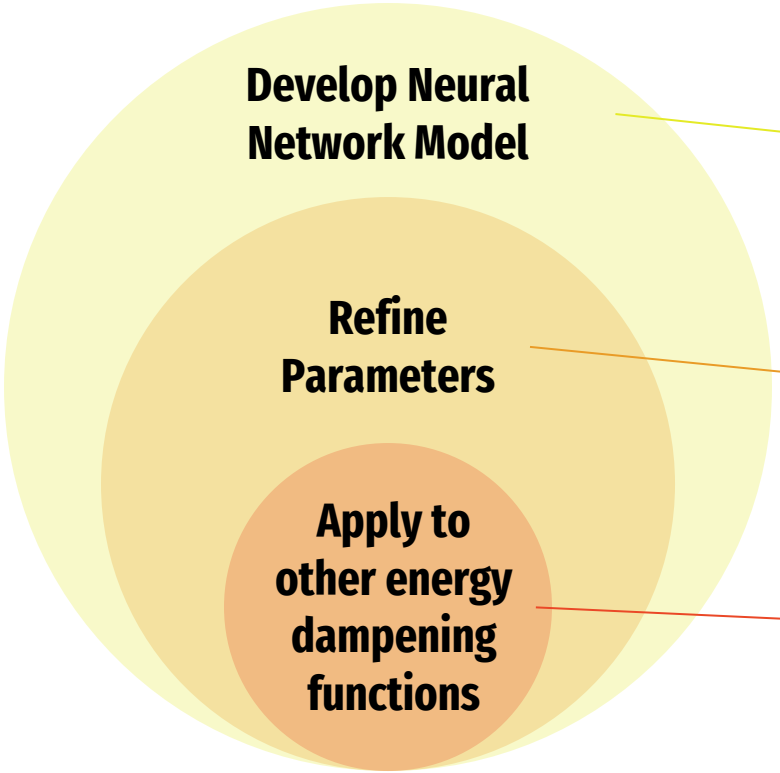
04 Results

R^2: 0.5143





Future Goals



Create a strong and efficient neural netork model to predict dampened energy calculations

Continue to refine the model and increase accuracy

Generate models to predict dampened values for Polarization Energy and Dispersion Energy

Greatest Dificulty

Challenges (Chemistry)

01 Quantum Chemistry

Understanding the Quantum Chemistry aspect of the project

02 Chemistry Knowledge

Lack the intuition behind the input data (Rij/Sij values and energies)

Acknowledgements

- NSF for funding SIMCODES
- SIMCODES and ISU for mentorship