Description of Research Project for Engineering Physics Undergraduate Honors Program

William L Cai

One of the largest outstanding problems in computational material science is elucidating the full spectrum of materials that can be synthesized. Traditional approaches are either computationally exhausting for screening through millions of potential materials or heavily rely on the experience and intuition of human experts, which are costly and constrain the variations of the predicted new materials. This project sets out to make reliable predictions on novel materials through a data-driven approach by using state-of-the-art machine learning architecture and by leveraging the work I have done during my 2020 Material Science REU Program and my final project of CS361 Engineering Design Optimization.

During the 2020 Material Science REU Program, I constructed an automated pipeline to extract desired data from The Materials Project database and Inorganic Crystal Structure Database (ICSD) and applied unweighted and weighted Elkanoto Classifiers to predict the synthesizability of two-dimensional inorganic materials. The pipeline functioned perfectly. The classifiers, however, had a performance below expectations. Then in the later CS361 Engineering Design Optimization class, I attempted to optimize the Weighted Elkanoto Classifier in the inorganic material synthesizability prediction problem through hyperparameter tuning. The tuning process involves a proper sampling plan selection, surrogate model construction, and implementation of the off-the-shelf optimization method. The tuning process enables us to find the approximated optimized frontier in C-Gamma space. The optimized classifier showed improvement in its performance but still cannot compete with human experts.

This project will utilize multi-task learning and meta-learning, which are machine learning algorithms that can learn how to learn across a collection of related tasks and datasets and then make more robust predictions. By building on the previous pipeline I have constructed and my hyperparameter tuning techniques, this project will incorporate physics-based reasoning into multi-task learning and meta-learning techniques to learn from multiple material science datasets beyond The Materials Project database and ICSD I have been working with. The end goal of this project is to have a pipeline that can make accurate predictions of the synthesizability for the unlabeled dataset using a physics-based machine learning approach that is powered by multi-task learning and meta-learning. The baseline accuracy of this pipeline will be those of traditional supervised, semi-supervised, and unsupervised models.