Fall 2021

Searching Extreme Mechanical Properties Using Active Machine Learning and Density Functional Theory

Joshua Ojih

SEARCHING EXTREME MECHANICAL PROPERTIES USING ACTIVE MACHINE LEARNING AND DENSITY FUNCTIONAL THEORY

by

Joshua Ojih

Bachelor of Engineering
University of Benin, 2015

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in

Mechanical Engineering

College of Engineering and Computing

University of South Carolina

2021

Accepted by:

Ming Hu, Director of Thesis

Jianjun Hu, Reader

Chen Li, Reader

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

DEDICATION

This Thesis is dedicated to God almighty for his endless mercies and provisions for me, all knowledge and understanding concerning this Thesis comes from God. I would also thank my parents and my siblings for their love and encouragement in completing this research.

ACKNOWLEDGEMENTS

First and foremost, I offer my deepest gratitude to my Advisor, Dr. Ming Hu, for his constructive criticism and direction throughout my thesis. I would also love to extend my appreciation to other members of my research team, Alejandro Rodriguez and Mohammed Al-fahdi for their support in completing the research.

ABSTRACT

Materials with extreme mechanical properties leads to future technological advancements. However, discovery of these materials is non-trivial. The use of machine learning (ML) techniques and density functional theory (DFT) calculation for structure properties prediction has helped to the discovery of novel materials over the past decade. ML techniques are highly efficient, but less accurate and density functional theory (DFT) calculation is highly accurate, but less efficient. We proposed a technique to combine ML methods and DFT calculations in discovering new materials with desired properties. This combination improves the search for materials because it combines the efficiency of ML and the accuracy of DFT. Here, we train a ML algorithm, the algorithm is used to make prediction. We use stein novelty (SN) score which recommends potential candidates from the ML prediction. We then verify the potential candidates using DFT calculation. In our demonstration, we use three property space for our search: Bulk Modulus vs Shear Modulus, Shear Modulus vs Hardness and Pugh's ratio vs Poisson's ratio. In exploring our property space, we found 30 crystal structures with high bulk and shear moduli, 21 crystal structures with ultrahigh hardness, and 11 crystal structures with negative Poisson's ratio from original 85,707 crystal structures taking from material project database. The method deployed herein can be extended to push other material properties to the limit.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF SYMBOLS

G	Shear Modulus
k	Pugh's Ratio
K	Bulk Modulus
p	Predicted Uncheck Point by ML
v	Poisson's Ratio

LIST OF ABBREVIATIONS

AI ..... Artificial Intelligence

BLOX..... Boundless Objective-free eXploration

CFID ..... Classical Force-field Inspired Descriptors

CNN ..... Convolutional Neural Network

CGCNN..... Crystal Graph Convolutional Neural Network

DFT ..... Density Functional Theory

GPa..... gigapascal

HTS ..... High Throughput Screening

MAE..... Mean Absolute Error

ML..... Machine Learning

PES..... Potential Energy Surface

PAW..... Projector Augmented Wave

RF..... Random Forest

SD ..... Stein Discrepancy

SN ..... Stein Novelty

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

The term extreme is defined as something farthest or highest from one another, that is materials with unusual mechanical properties such as super hard materials, extremely negative Poisson's ratio[1]. In the past decade, material scientists have been favorably using high throughput screening (HTS) for structure property prediction with high accuracy in searching for new materials[2]. However, HTS prediction at the quantum level (first principles) is, although highly accurate, less efficient and hence time consuming and computationally expensive[2,3]. In contrast prediction at the classical level (such as classical molecular dynamics) is highly efficient but less accurate since they usually scale linear with the number of atoms[4,5]. Because of the computation cost of DFT[6] and the less accuracy of classical potential, an intuitive idea is to bridge the gap between DFT-level accuracy and classical-level efficiency.

ML technique offers the possibility of bridging this gap[7], and the application of ML has already help in speeding the process for novel material discovery[8]. These discovered materials are important to the Engineer because they enable future technological developments in area such as sustainable energies technologies, energy efficient processes[9]. ML techniques have been used extensively for material properties prediction over the past decades, because ML has high efficiency, and accuracy closed to

that of DFT[10,11]. The accuracy of a ML technique depends on the effective input representation of the crystal structures since the atomic positions are not rotationally and translationally invariant[12]. Such input representation is known as descriptors or features. The idea behind the use of ML technique for structure properties prediction is to analyze and map the relationship between the properties of materials and their characteristics by extracting information from existing data without given any explicit knowledge on how to draw conclusion from data[13]. With given data, ML algorithm learn the rules and relationship that underlie a dataset by assessing the data and build a model to make prediction[13], for example ML model have been used for the prediction of mechanical properties of metal alloy[14,15], band gap energies of crystal[16,17], the formation energies of crystal[18–20], melting temperature of binary inorganic compounds[21]. ML techniques have been used to introduced a new kind of representation of DFT potential energy surface (PES), which provides the energy and forces as a function of all atomic position in system of arbitrary size[22].

Though ML is highly efficient, it has some limitation which tends to reduce its accuracy in predicting properties. Such limitation include measurement error[23], lack of generality and precision, reliance on high-quality data[24], inability to determine high level concept[25], prone to artifact[26], good in interpolation but poor in extrapolation[21,27]. Another critic in ML techniques is the lack of novel laws, understanding , and knowledge from their use because ML techniques are treated as black box[5].

In our research, we combined ML efficiency with DFT accuracy for structure property prediction, we used mechanical properties as the property search space (bulk

modulus (K) vs shear modulus (G), shear modulus vs hardness and Pugh's ratio (k) vs Poisson's ratio(v), after thoroughly screening of 85707 crystal structures gotten from material project[28] database, we found 30 crystal structures with high bulk and shear moduli, 21 crystal structures with high shear modulus and hardness and 11 crystal structures with negative Poisson's ratio. We compare our result with ML techniques: Crystal graph convolution neural network (CGCNN)[10], Lasso regression[29] and Ridge regression[30].

1.2 RESEARCH GOAL

Most empirical potential used in classical molecular dynamics simulation are efficient but not accurate, on the other hand, accurate electronic structure calculation (DFT) are limited to several hundreds of atoms making modelling beyond nanoscale inadequate and they are also computational expensive. The goal of this research is to bridge the gap between DFT-level accuracy and classical-level efficiency by combine ML techniques and DFT calculations.

Objective of the thesis include:

1.) To investigate material properties using boundless objective-free eXploration (BLOX) and DFT
2.) To search extreme mechanical properties of materials
3.) To compare some ML techniques with DFT results

The result of this study shows how we can successfully combine ML techniques with DFT to discover materials with unusual properties.

1.3 THESIS STRUCTURE

This thesis is organized into five main chapters. The first chapter present the introduction, background, research goal and objectives. The second chapter present a literature review on CGCNN, BLOX and Descriptors. The Third chapter present the methodology used in this study. The fourth chapter present results and discussion from the study. The final chapter present conclusion and future work to be done to improve the existing method used in this study.

CHAPTER 2

LITERATURE REVIEW

In this chapter, an overview of research work related to combining ML and DFT is presented. This chapter will discuss the general machine learning approach, BLOX algorithm in selecting of potential candidates and descriptors used to obtain accurate result in our ML techniques.

2.1 GENERAL MACHINE LEARNING APPROACH FOR MATERIAL SCIENCE

The material scientist relies on experiments and simulation-based model for material characterization[31]. However, in recent years ML techniques have been used for material properties prediction and material design[32]. ML bypasses the computational cost of solving Schrodinger equation[33] there by making ML techniques highly efficient.

ML is the study of computer algorithm that uses data to carry out specific task, and improves automatically through experience, it is a subset of artificial intelligence (AI). ML algorithm build a model based on input data (training data), to make prediction without being explicitly programmed[34]. Traditionally, ML is broadly divided into three categories:

1) Supervised learning: In supervise learning, the algorithm is presented with inputs and the desire output in other to learn and establish a relationship between the input and output.
2) Unsupervised learning: In unsupervised learning, only the input is presented to the algorithm, the algorithm itself finds a pattern within the data.
3) Semi-supervised learning: Semi-supervise learning fall between supervised and unsupervised learning, some of the training data do not have a label.

ML algorithm can be used for regression, Classification, or active learning.

2.2 DESCRIPTORS

In ML techniques, controlling the performance to enhance its accuracy is based on how compound/structures are represented in dataset[35]. For the material scientist, the atomic positions in crystal structure are not suitable for direct input representation because they are not rotationally and translationally invariant[12]. The transform input data are called descriptors/features. Selecting a good descriptor is a very important step because a good descriptor can explain a target property well and this led to a robust prediction model of a target property[35], combining descriptors with ML methods led to a model capable of accurate structure properties prediction. Chemical descriptors based on elemental properties have been successfully applied for various computational discovery[36], nonetheless, this is not suitable for modelling crystal structures with the same composition since they ignore structural information[37]. Here, we introduce a new descriptor that combine both structural and elemental descriptors, this is called classical force-field inspired descriptors (CFID)[37,38], this is because to cover a wide range of crystal structures, it is good to combined elemental and structural representation as

descriptors[35,39], these combine descriptors can also be applied to molecular system. Elemental representations include atomic number, atomic mass, period, and group in the period table, first ionization energy, second ionization energy, electron affinity, Pauling electronegativity, Allen electronegativity, Van der Walls radius, covalent radius, atomic radius, melting and boiling point, density, molar volume, heat of fusion, heat of vaporization, thermal conductivity, and specific heat. These helps to captures essential information about compounds. Structural representations include simple coordination number, Voronoi polyhedron of central atom, angular distribution function, radial distribution function, bond-orientational order parameter[40] and angular Fourier series[41]. The CFID consist of 1557 descriptors for each crystal structures: 438 average chemical, 4 simulation-box-size, 378 radial charge-distribution, 100 radial distribution, 179 angle-distribution up to the first neighbor, 179 angle-distribution up to the second neighbor, 179 dihedral angle up to the first neighbor and 100 nearest neighbor descriptors.

2.3 CGCNN

CGCNN is a ML technique used for structure properties prediction, the aim is to see if we can rely on ML model in finding extreme mechanical properties, this will in turn reduce computational as cost. The CGCNN model combines the descriptor and the learning model into one inseparable step i.e. the model learn material properties directly from the connection of atom in the crystal[10]. The CGCNN framework represent periodic crystal that provides material property prediction with DFT accuracy[3,10]. Here the crystal structure is represented by a crystal graph that encodes both atomic information and bonding interaction between atoms, and then it builds a convolutional

neural network (CNN) on top of the graph to automatically extract features that are used for predicting target properties by training with DFT calculated data. The CNN consist of two major components: convolutional layer and pooling layers. After convolutions, the network automatically learns the features vectors for each atom. The pooling layer is then used for producing an overall features vector for each atom. The pooling layer is then used for producing an overall features vector for the crystal.
Implementation of CGCNN is available at https://github.com/txie-93/cgcnn[10].

2.4 BLOX ALGORITHM

Implementation of BLOX is available at https://github.com/tsudalab/BLOX[42]. In BLOX implementation, a ML model is built to predict the properties of materials for which current data on calculated properties is available. In searching a property space, BLOX search outside the boundary to capture properties of materials that lies at the edge of the boundary. BLOX employs stein discrepancy which boundless evaluate a kind of distance between any two distribution in any dimensional space[43]. After the initial preparation by selecting materials from materials project database[28], the search is performed by repeating the following steps. Step1, construct a property prediction model. Step2, selection of candidate using SN score based on stein discrepancy. Step3, evaluation of selected candidate by DFT.

CHAPTER 3

METHODOLOGY

In this chapter, the methodology in which the research is conducted is presented as show in Figure 3.1 below. This includes training a CGCNN model, running BLOX algorithm, and DFT calculations.

First, we used the CFID to transform our crystal structures to ML input, we used BLOX coupled with Random Forest (RF) ML algorithm to screen 85707 crystal structures (unchecked data) downloaded from material project database. We split our unchecked data into 10 different jobs, so we can run the jobs in parallel. RF algorithm coupled in BLOX learns the features input and the RF model predict properties of the unchecked data. The BLOX algorithm uses the SN score to recommend potential candidate base on the highest score given (SN scores measures a deviation between the observed property and the predicted property of the unchecked data by using stein

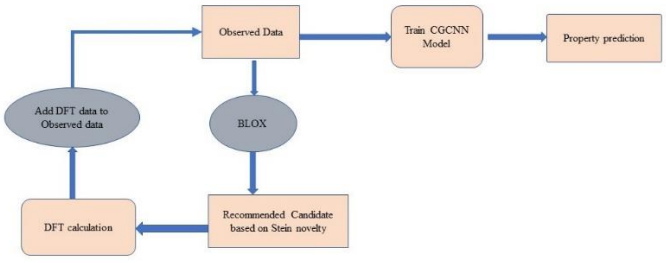

Figure 3.1 Schematic of workflow performed in this study

discrepancy), as shown in the equation below.

$$SN(V\ U\ \{p\}) = SD(V) - SD(V\ U\ \{p\}) \quad (1)$$

Where p is the predicted unchecked point by ML, we select the candidate with the largest SN.

We selected 25 crystal structures each from the 10 jobs, making a total of 250 recommended candidate. We perform DFT calculation using plane-wave basis projector augmented wave (PAW) method[44], with the Perdew-Burke-Ernzerhof exchange-correlation functional[45], as implemented in the VASP code[46–48], the cutoff energy is set to 500eV for the recommended crystal structures to find extreme mechanical properties. DFT was calculated for these recommended structures because the ML prediction was not accurate enough due to the descriptors used and the number of input data. We build a CGCNN, Lasso regression and Ridge regression model and compare our prediction with DFT for the recommended structures.

After each round of DFT, we add the DFT verified data to the observed data and repeat the process, we found out that BLOX explore out of trend material in every direction, so we must set a constraint in other to control the direction of the search i.e., if we are looking for ultrahigh mechanical properties, we have to clean our data by removing very low mechanical properties from our input data, so that BLOX can search the upper boundary and vice versa. Figure 3.2 below shows the scatter plot of the observed properties for bulk vs shear moduli property space. Here, we used 2000

structure as the input to train the model for the property prediction. Figure 3.3 below show the scatter plot for the shear vs hardness property space. Here we used 1980 structures as the input to train the model for the property prediction, while figure 3.4 below show the scatter plot for the Pugh's ratio vs Poisson's ratio property space. Here, we used 2000 structures as the input to train the model for the property prediction.

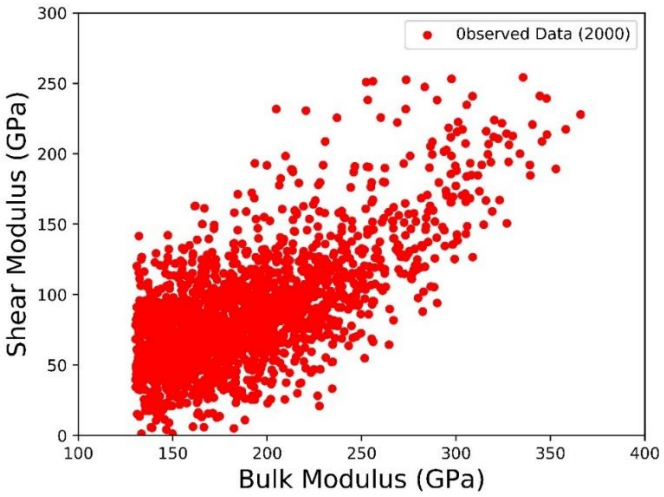

Figure 3.2 Properties of Observed data for Bulk Modulus vs Shear Modulus property space

In each round we train a CGCNN model with the observed data and make prediction for our recommended materials in other to compare with DFT calculations. The loop of DFT/BLOX/recommendations were performed at least 4 rounds until there is no significant amount of new interested material properties obtained.

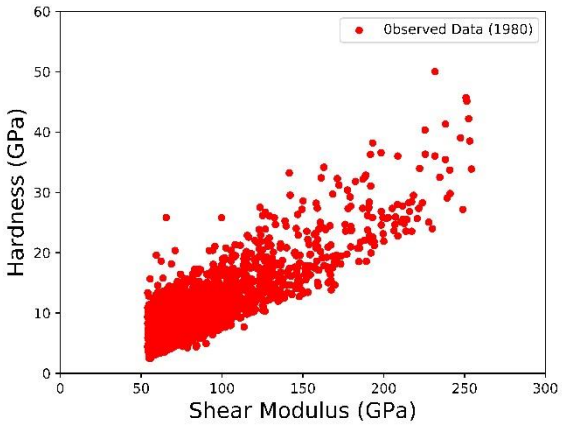

Figure 3.3 Properties of Observed data for Shear Modulus vs Hardness property space.

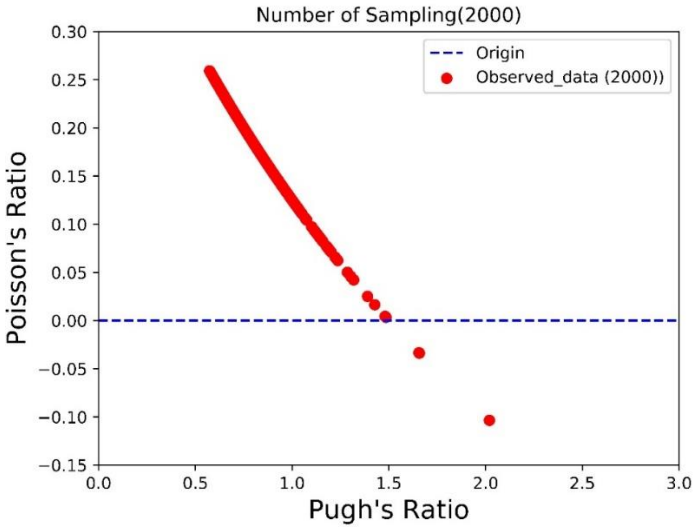

Figure 3.4 Properties of Observed data for Pugh's ratio vs Poisson's ratio property space

CHAPTER 4

RESULT AND DISCUSSION

The result of this research is described in three major subsections based on the property space used.

4.1 BULK VS SHEAR MODULI

The first property space that was explore for searching extreme mechanical properties was the bulk modulus vs shear modulus property space. Here, we are interested in structure with high bulk modulus and shear modulus, because structure with high bulk and shear modulus has the tendency of being super hard material. Super hard materials are materials with hardness exceeding 40 GPa and they are of great importance because of their industrial application such as abrasives, polishing, disc brakes, proactive coating and cutting tools. Here, we show the results for different rounds of BLOX and DFT calculations to find materials with high bulk and shear moduli from exploring the material project database. Figure 4.1 below shows the recommended materials by BLOX.

From Figure 4.1, BLOX using SN score recommend some materials which have higher chance to be out of trend for the first round. Figure 4.2 below show the DFT calculation for the first round of recommended materials.

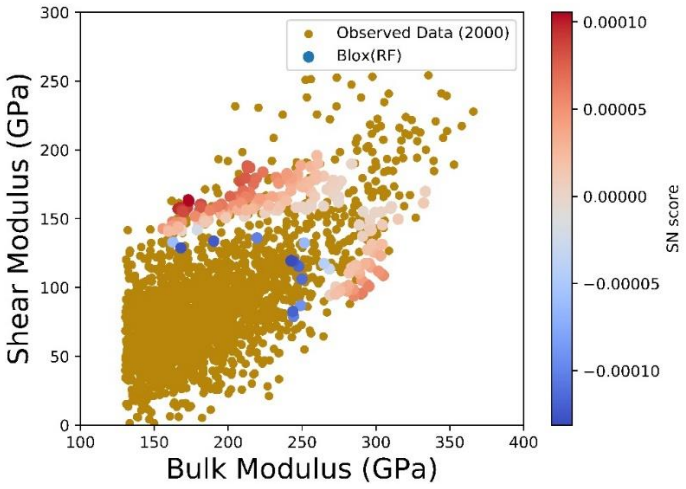

Figure 4.1 Observed data and BLOX prediction for first round for recommended structures.

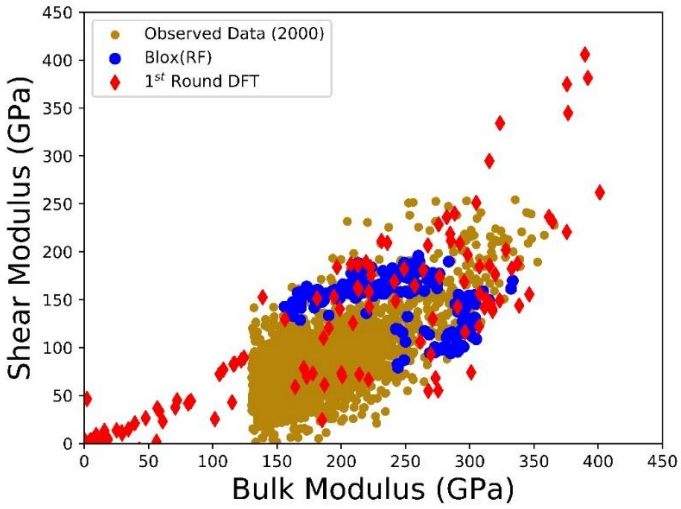

Figure 4.2 Observed data and first round DFT calculation for recommended structure without boundary.

The DFT calculated values pushed the properties to the limit in every direction, as we found materials with extremely high and low bulk modulus and shear modulus. This is because BLOX algorithm search for out of trend materials in all direction. Figure 4.3 below show that BLOX algorithm for this second round recommend materials that are

not in the direction of interest, because BLOX is not smart to know the direction of interested since it searches in all direction whether the upper limit or the lower limit, so in other to get the right direction of interest we set a boundary to help guide BLOX in the direction needed.

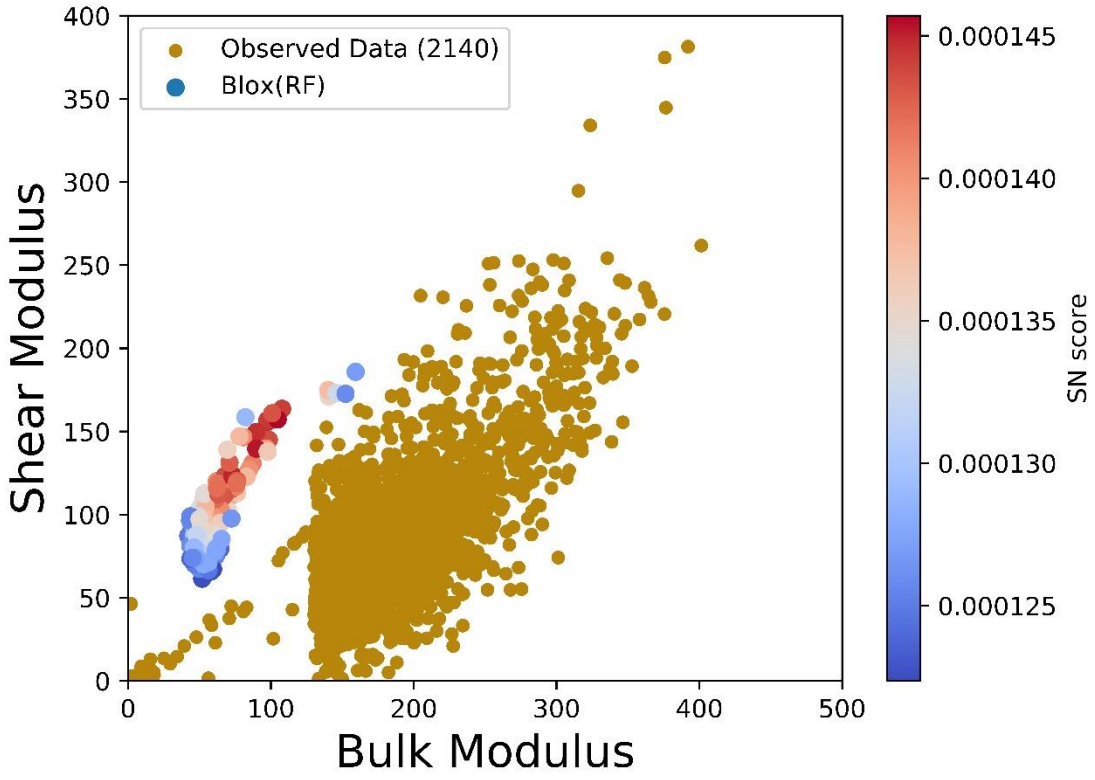

Figure 4.3 Observed data and BLOX prediction for second round without boundary.

For this property search we remove any materials with bulk modulus less than 130 GPa, since our target is to find materials with extremely high mechanical strength. Figure 4.4 below show the first round DFT for the recommended structures after the boundary has been applied.

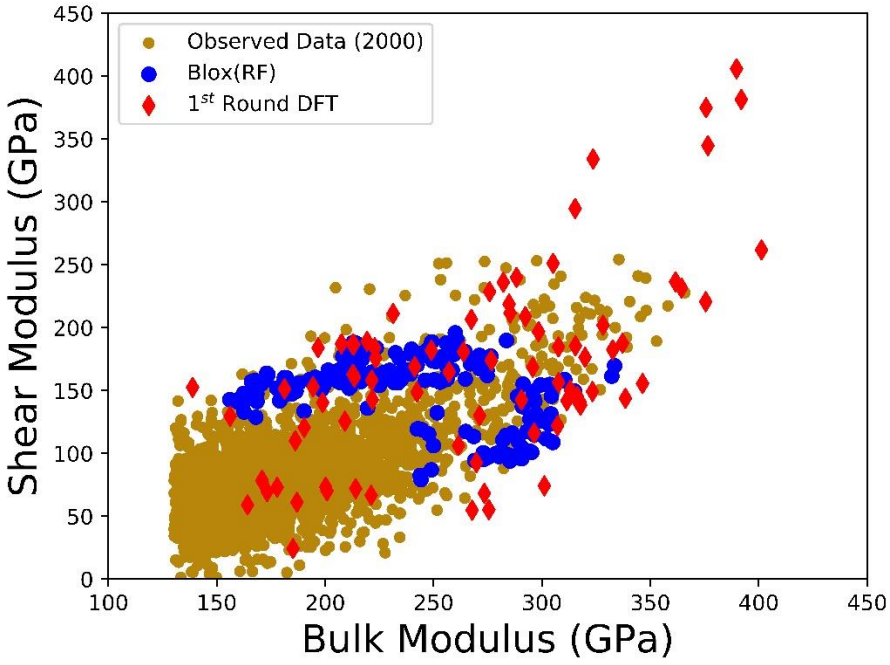

Figure 4.4 Observed data and first round DFT calculation for recommended structure with boundary.

Figure 4.5 below show the second round recommended material by BLOX after boundary has been applied, here we can see that the BLOX prediction are now in the direction of interest. BLOX was able to recommend out of trend material towards the upper limit, these recommended materials have a high probability of having extremely high bulk and shear moduli. Figure 4.6 below, we observed the DFT calculations for the recommended candidate from BLOX, we found about 3 crystal structures with extremely high bulk and shear moduli. Figure 4.7 and figure 4.8 show the recommended material from BLOX, here we can observe that BLOX recommended material in the upper limit after the boundary has been applied.

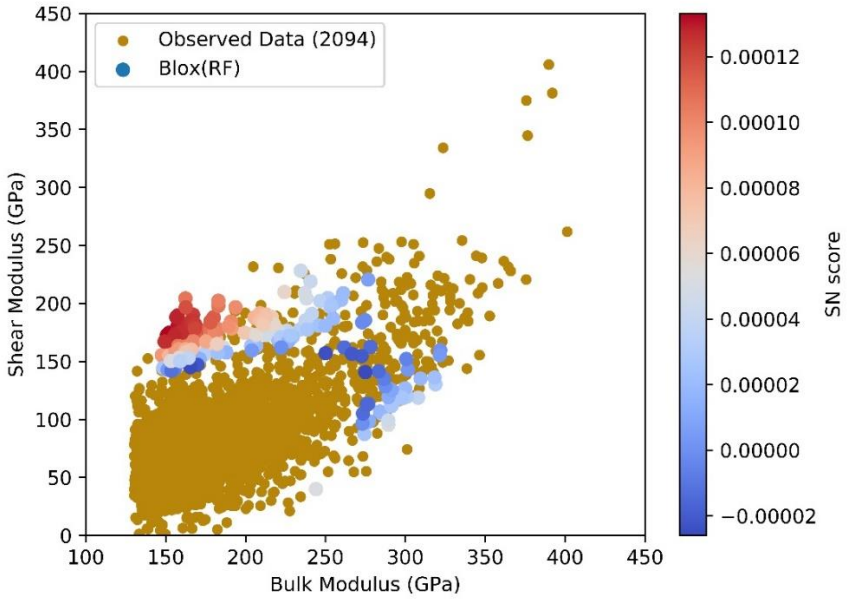

Figure 4.5 Observed data and BLOX prediction for second round with boundary.

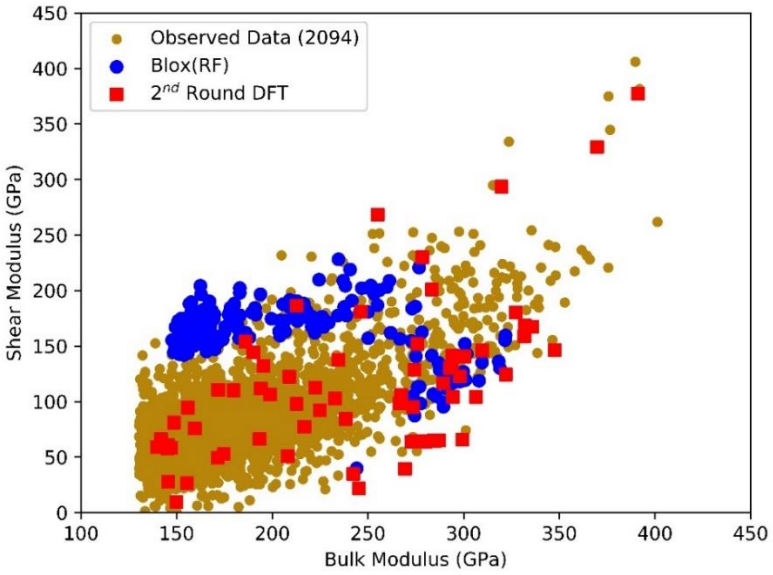

Figure 4.6 Observed data and second round DFT for recommended structure.

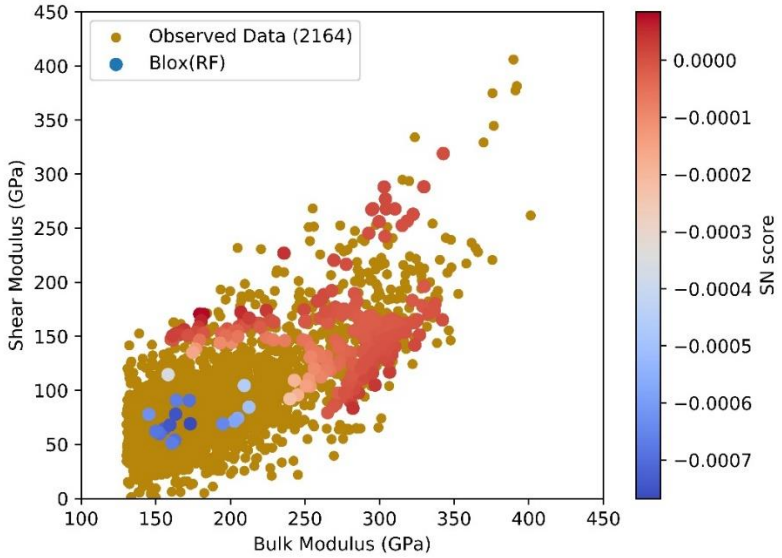

Figure 4.7 Observed data and BLOX prediction for third round.

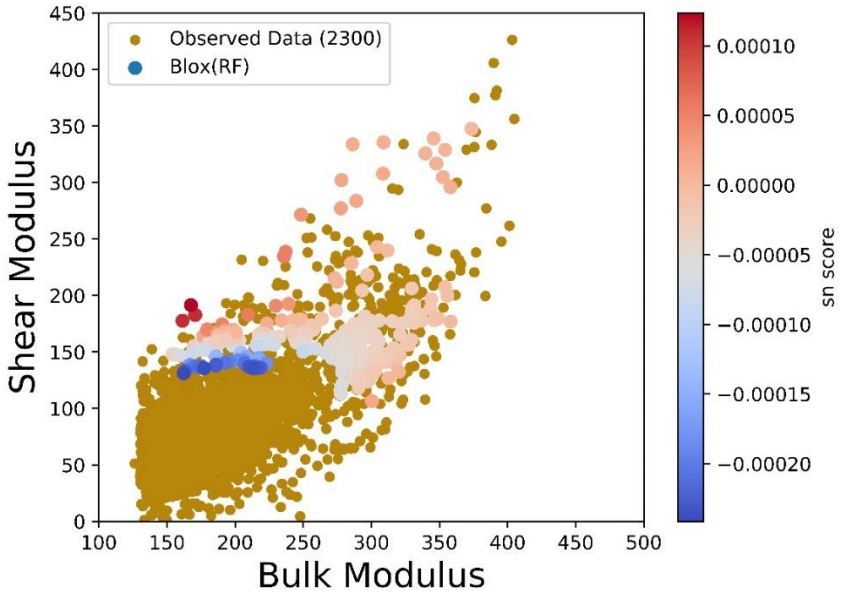

Figure 4.8 Observed data and BLOX prediction for fourth round.

Figure 4.9 and figure 4.10 below show the third and fourth rounds DFT calculation for the recommended material respectively. From figure 4.9, we observed the DFT calculations for the recommended materials from BLOX, we found about 10 structures with high bulk and shear modulus. Similarly, we found about 12 crystal structure from the fourth round of BLOX recommended material as shown in figure 4.10.

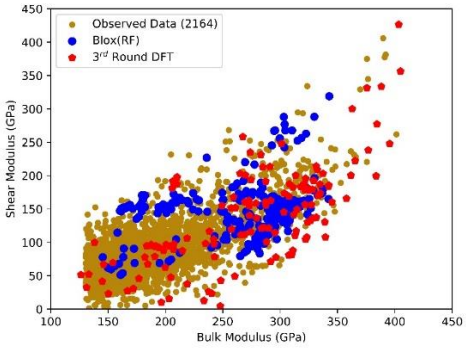

Figure 4.9 Observed data and fourth round DFT for recommended structure.

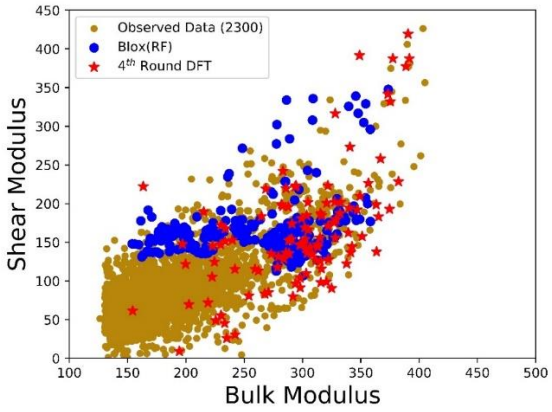

Figure 4.10 Observed data and fourth round DFT for recommended structure.

Figure 4.11 below show the initial observed data with all rounds of DFT calculation, here we can see that after carefully exploring 85707 crystal structures from the material project database, we found 30 structures with high bulk and low moduli.

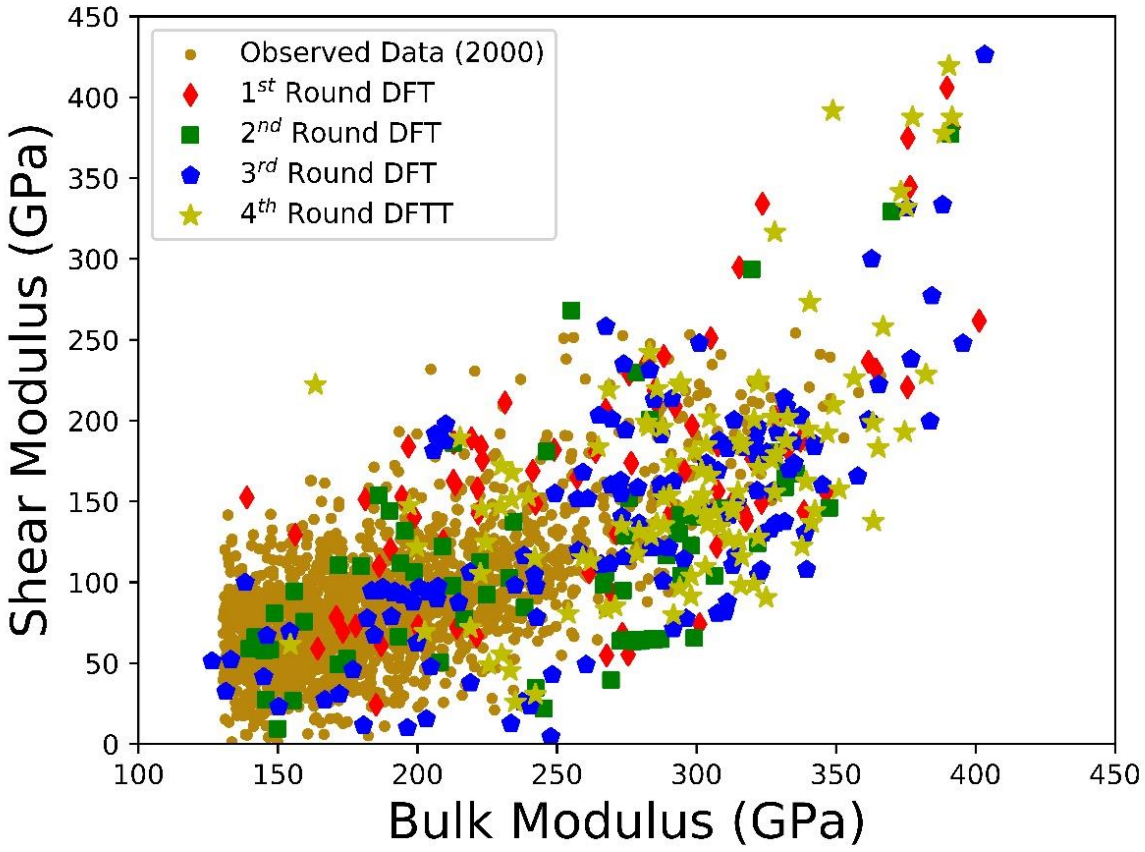

Figure 4.11 Observed data and all rounds of DFT for recommended structures.

Figure 4.12 below show the comparison between the machine learning model (CGCNN, Lasso regression and Ridge regression) used for property prediction and DFT calculation for the first round. We can see that the ML model could not push the property to the limit, because the accuracy of the ML model was reduced due to the small number of input data available, and descriptor used.

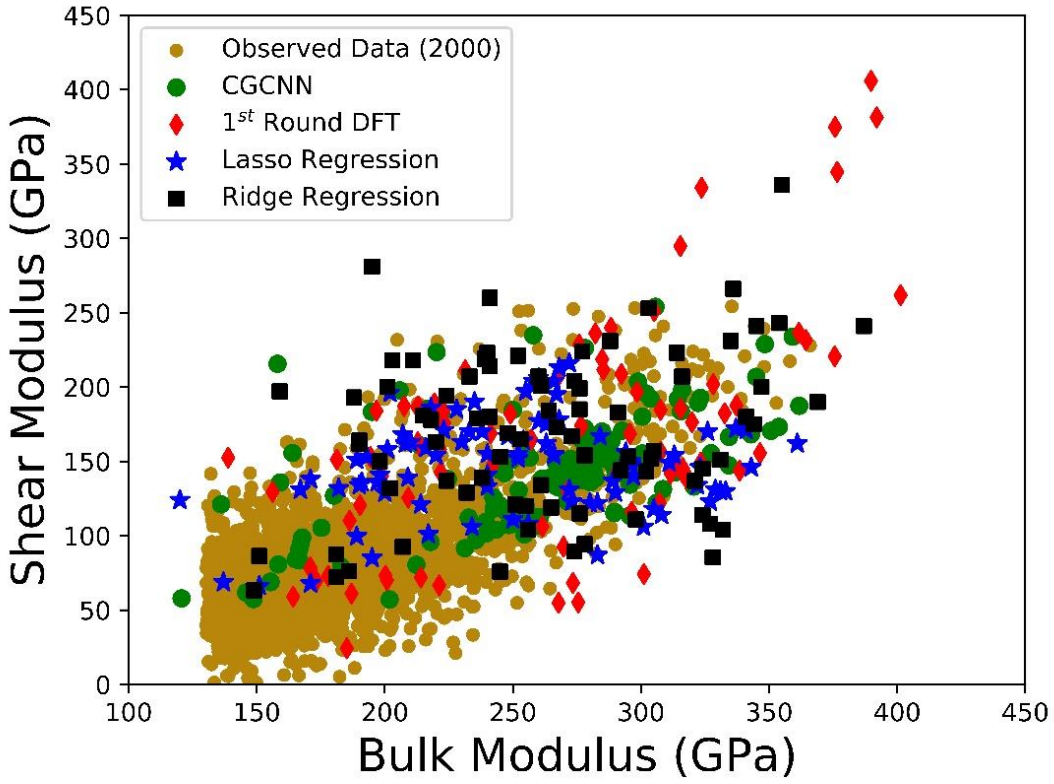

Figure 4.12 Comparison between DFT calculation and prediction from ML model for first round recommended structures.

Figure 4.13 below show the maximum and average distance between CGCNN prediction and DFT calculations for the recommended structure with high bulk and shear moduli, we observed that as the number of rounds increases, the distance between the CGCNN prediction and DFT calculation decreases.

Figure 4.14 below show the mean absolute error for CGCNN, Lasso regression and Ridge regression for bulk and shear moduli respectively, we observed that the ML model do not necessarily improve as we add few hundreds to the initial 2000 observed data.

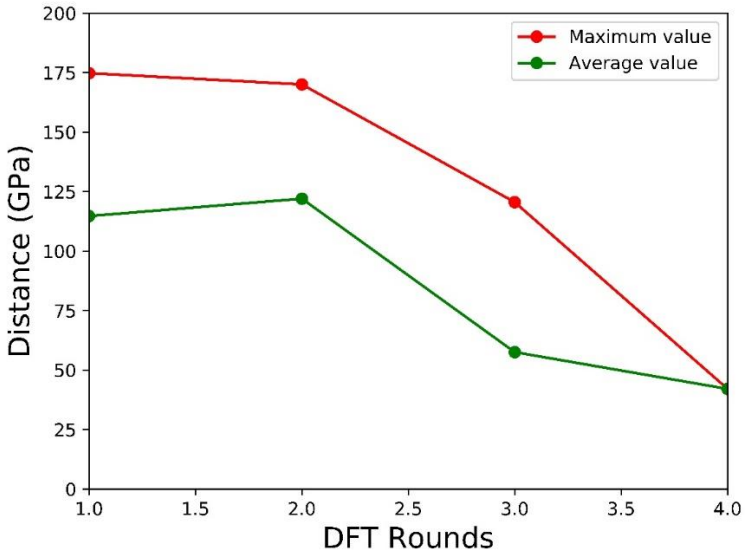

Figure 4.13 Maximum and average distance between outlier of CGCNN and DFT calculation with high bulk and shear moduli.

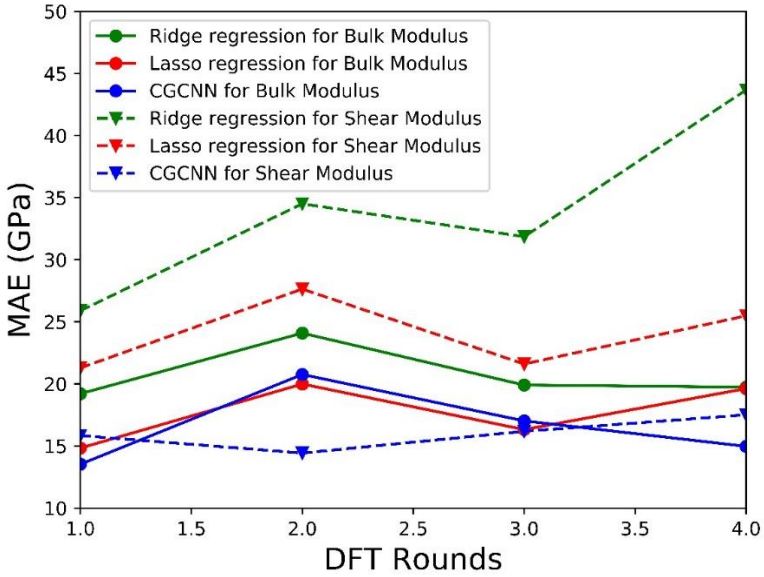

Figure 4.14 MAE for CGCNN, Lasso regression and Ridge regression for bulk modulus and shear modulus.

4.2 ULTRAHIGH HARDNESS

Our recent high throughput on ultrahigh carbon allotropes illustrates that the hardness is strongly correlated with the shear modulus[49]. Using the shear modulus and hardness property space, we were able to find some materials with ultrahigh hardness. A high shear modulus is essential to high hardness because they are highly correlated, and it show more substantial relationship with hardness[50]. Hardness is the measure of the resistance to localized plastic deformation, materials with a hardness value exceeding 40 GPa when measured by the Vickers hardness test are super hard materials[51]. While hard materials usually have high bulk modulus, high bulk modulus does not necessarily mean a material is hard. Shear modulus provides a better correlation with hardness than bulk modulus[52]. Figure 4.15 below show the BLOX prediction for each of recommended materials using SN scores, these materials have the tendency of having ultrahigh hardness.

Figure 4.16 below show the DFT calculations for each round of DFT. We observed from the DFT calculations that BLOX was able to recommend structures with high hardness.

Figure 4.17 below show the initial observed data and all rounds of DFT calculation

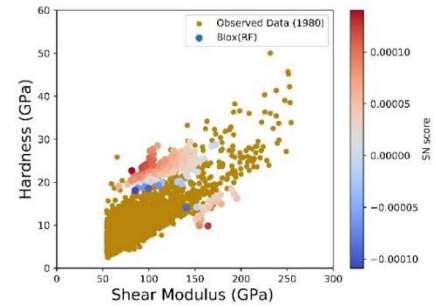

Figure 4.15a

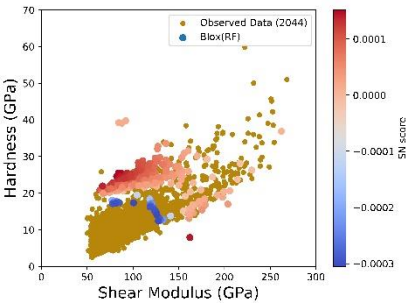

Figure 4.15b

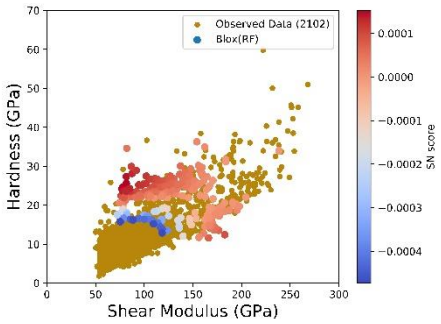

Figure 4.15c

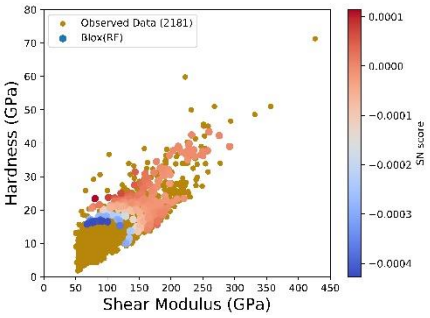

Figure 4.15d

Figure 4.15 Observed data and BLOX prediction for each round

We were able to find 21 structures with high shear modulus and high hardness as observed from figure 4.17. Figure 4.18 below show the maximum and average distance between CGCNN prediction and DFT calculations for the recommended structure with high shear modulus and hardness, we observed that as the number of rounds increases, the distance between the CGCNN prediction and DFT calculation decreases.

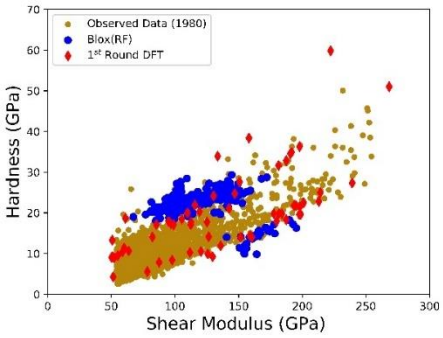

Figure 4.16a

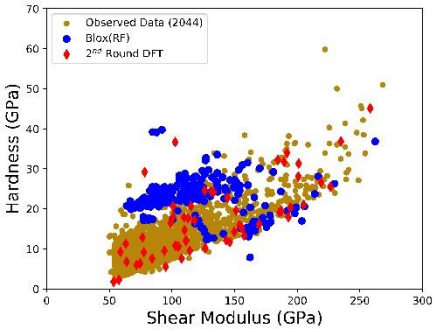

Figure 4.16b

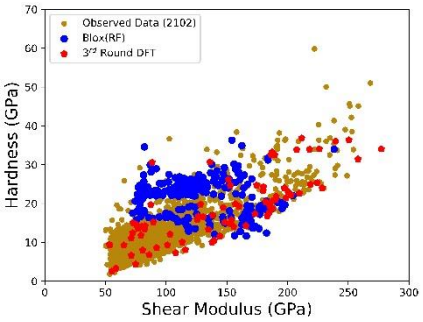

Figure4.16c

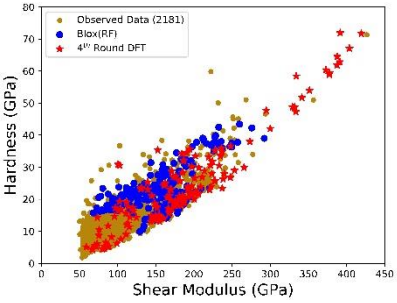

Figure 1.16d

Figure 4.16 Observed data and DFT calculations for each round

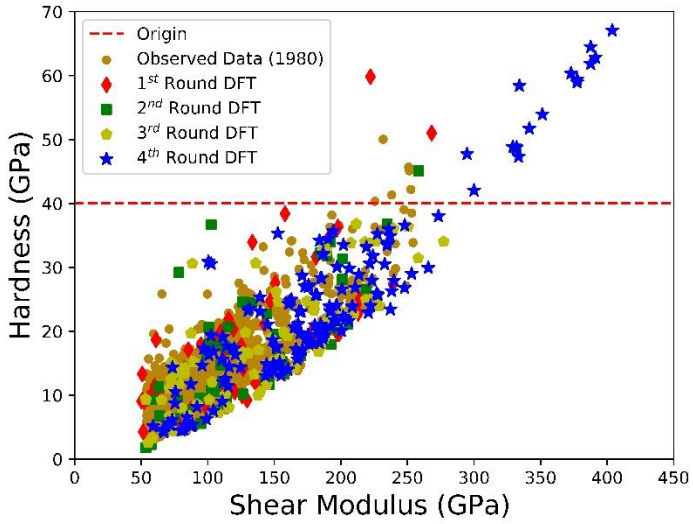

Figure 4.17 Observed data and all rounds of DFT calculation for the recommended structures

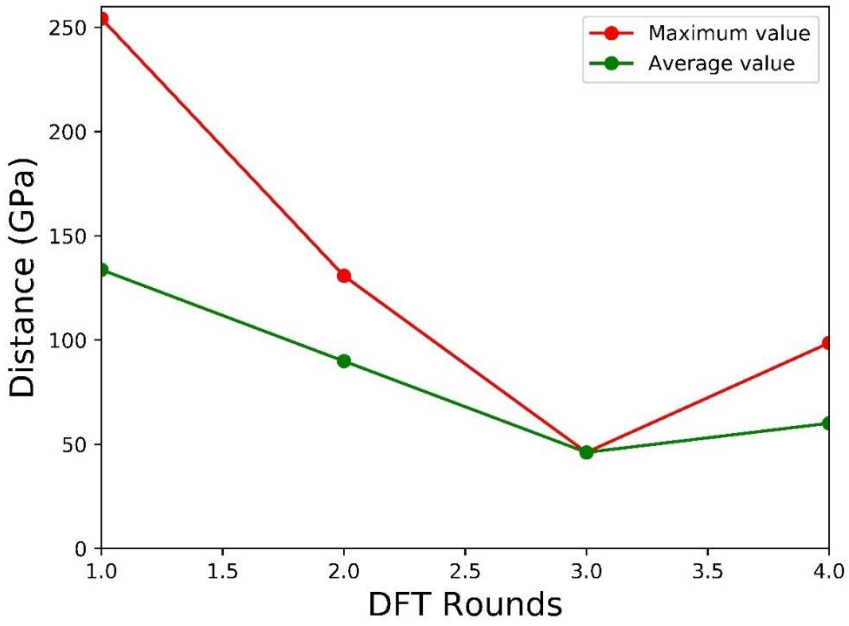

Figure 4.18 Maximum and average distance between outlier of CGCNN and DFT calculation with high and shear moduli

4.3 NEGATIVE POISSON'S RATIO

Poisson's ratio is defined as the ratio of lateral strain in solid over the longitudinal strain measured in a simple tension experiment[53]. Most materials have positive Poisson's ratio, but a portion of solid materials has negative Poisson's ratio, known as auxetic materials. The materials with negative Poisson's ratio have unique properties such as high energy absorption, high fracture resistance, difficult to shear, enhance toughness[54] and very difficult to cause indentation. We use the Pugh's ratio (defined as the ratio between the shear modulus and the bulk modulus to distinguish the ductile/brittle behavior of material[55,56]) for this search because it is negatively correlated with Poisson's ratio. BLOX prediction of recommended material using SN scores is shown in figure 4.19.

Figure 4.20 shows below shows the DFT calculations for each round. Here, we were able to find some structures from each of BLOX recommendation. Figure 4.21 shows the initial observed data and all rounds of DFT calculation. We were able to find 11 new structures with negative Poisson's ratio.

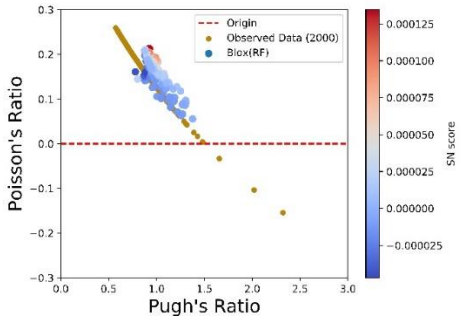

Figure 4.19a

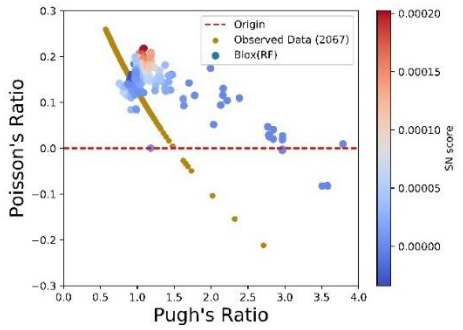

Figure 4.19b

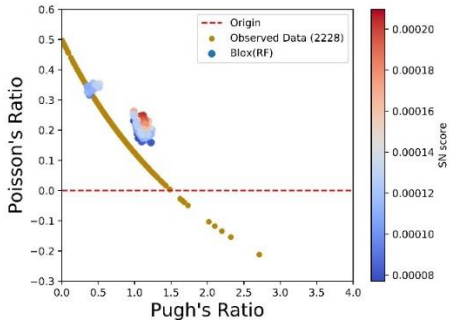

Figure 2.19c

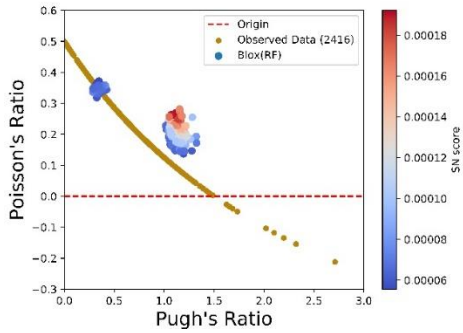

Figure 4.19d

Figure 4.19 Observed data and BLOX prediction for each round

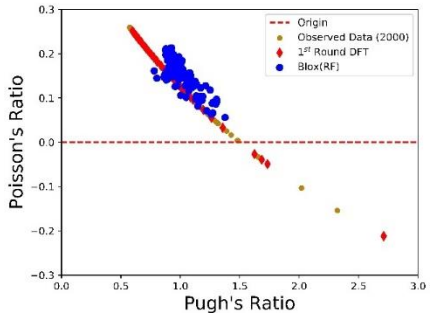

Figure 4.20a

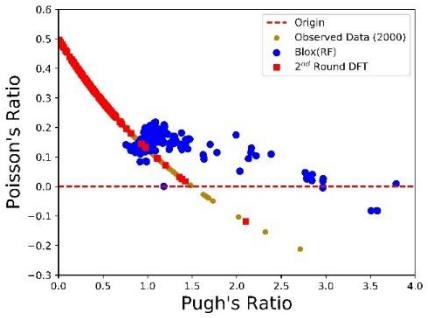

Figure 3.20b

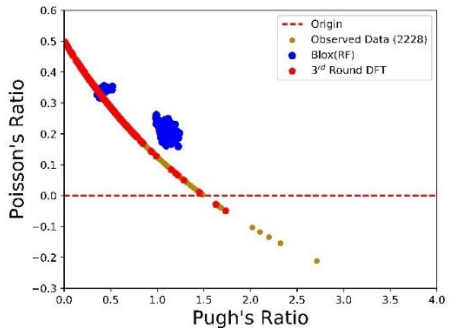

Figure 4.20c

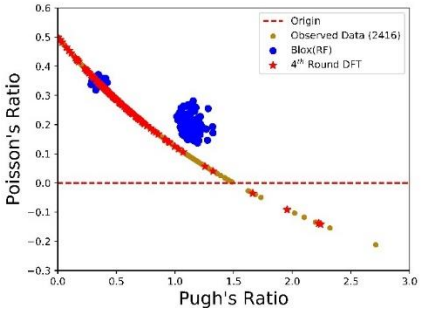

Figure 4.20d

Figure 4.20 Observed data and DFT calculations for each round

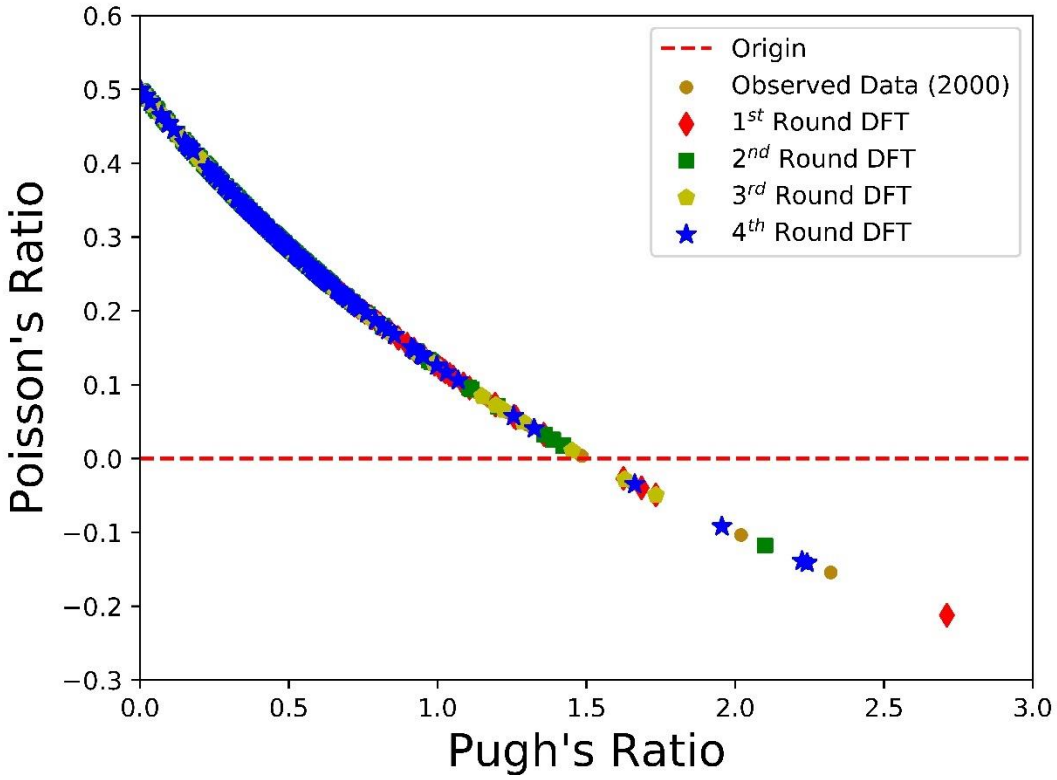

Figure 4.21 Observed data and all rounds DFT calculation for the recommended structures.

CHAPTER 5

CONCLUSION

In summary, we have presented the application of machine learning method and DFT calculation for structure property search. Here, we used BLOX algorithm coupled with RF ML model to recommend out of trend materials using SN score. We verify the recommended structures with DFT calculations. We were able to bypass the heavy computational cost of searching database with full DFT calculations. The efficiency of ML method helps to improves the search while verifying the prediction with DFT improve the accuracy of the search. We employ the use of mechanical properties for the property space. After exploring 85,707 crystal structures using bulk modulus, shear modulus, hardness, Pugh's ratio, and Poisson's ratio as the property space for the search, we found 30 structures with high bulk and shear moduli, 21 structures with high hardness and 11 structures with negative Poisson's ratio.

We also compare our DFT result and ML model (CGCNN, Lasso regression and Ridge regression) for the recommended candidate by BLOX, and we found out that the ML models could not push the properties to the limit which is one of the main drawbacks for almost all ML models. This is understandable considering that most ML models can only predict properties within the original range of training data, while they can hardly predict properties outside.

5.1 FUTURE WORK

1) BLOX algorithm recommends out of trend materials based on SN score. These materials have the tendency of having ultrahigh or low mechanical properties. We want to improve the success rate of our recommended materials such that most of the recommended materials will go in the direction of interest.

2) We need improve our ML algorithm by generate different descriptors and using other ML model to increase the efficiency of the model in other to push properties to the limit.

PUBLICATION

Authors Names: Joshua Ojih, Mohammed Al-Fahdi, Kamal Choudhary, Alejandro David Rodriguez, and Ming Hu

Title: Questing for Extreme Mechanical Properties via Boundless Objective-Free Exploration and Density Functional Theory

Under Preparation

REFERENCES

[1] Sabouni-Zawadzka, A. Al, 2020, "Extreme Mechanical Properties of Regular Tensegrity Unit Cells in 3D Lattice Metamaterials," Materials (Basel)., **13**(21), pp. 1–17.

[2] Chibani, S., and Coudert, F. X., 2020, "Machine Learning Approaches for the Prediction of Materials Properties," APL Mater., **8**(8).

[3] Noh, J., Gu, G. H., Kim, S., and Jung, Y., 2020, "Uncertainty-Quantified Hybrid Machine Learning/Density Functional Theory High Throughput Screening Method for Crystals," J. Chem. Inf. Model., **60**(4), pp. 1996–2003.

[4] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M., 1998, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins," J. Phys. Chem. B, **102**(18), pp. 3586–3616.

[5] Schmidt, J., Marques, M. R. G., Botti, S., and Marques, M. A. L., 2019, "Recent Advances and Applications of Machine Learning in Solid-State Materials Science," npj Comput. Mater., **5**(1).

[6] Zinola, C. F., 2010, "Production, Storage, Use, and Delivery of Hydrogen in the

Electrochemical Conversion of Energy," Electrocatal. Comput. Exp. Ind. Asp., pp. 589–631.

[7] Schütt, K. T., Glawe, H., Brockherde, F., Sanna, A., Müller, K. R., and Gross, E. K. U., 2014, "How to Represent Crystal Structures for Machine Learning: Towards Fast Prediction of Electronic Properties," Phys. Rev. B - Condens. Matter Mater. Phys., **89**(20), pp. 1–5.

[8] Callaghan, S., 2021, "Toward Machine Learning-Enhanced High-Throughput Experimentation for Chemistry," Patterns, **2**(3), p. 100221.

[9] Ludwig, A., 2019, "Discovery of New Materials Using Combinatorial Synthesis and High-Throughput Characterization of Thin-Film Materials Libraries Combined with Computational Methods," npj Comput. Mater., **5**(1).

[10] Xie, T., and Grossman, J. C., 2018, "Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties," Phys. Rev. Lett., **120**(14), pp. 1–14.

[11] Seko, A., Hayashi, H., Tsuda, K., Chaput, L., and Tanaka, I., 2015, "Prediction of Low-Thermal-Conductivity Compounds with First-Principles Anharmonic Lattice-Dynamics Calculations and Bayesian Optimization," **205901**(November), pp. 1–5.

[12] Himanen, L., Jäger, M. O. J., Morooka, E. V., Federici Canova, F., Ranawat, Y. S., Gao, D. Z., Rinke, P., and Foster, A. S., 2020, "DScribe: Library of Descriptors for Machine Learning in Materials Science," Comput. Phys. Commun., **247**, p. 106949.

[13] Lu, L., Dao, M., Kumar, P., Ramamurty, U., Karniadakis, G. E., and Suresh, S., 2020, “Extraction of Mechanical Properties of Materials through Deep Learning from Instrumented Indentation,” Proc. Natl. Acad. Sci. U. S. A., **117**(13), pp. 7052–7062.

[14] Chatterjee, S., Murugananth, M., and Bhadeshia, H. K. D. H., 2007, “δ TRIP Steel,” Mater. Sci. Technol., **23**(7), pp. 819–827.

[15] Bhadeshia, H. K. D. H., Dimitriu, R. C., Forsik, S., Pak, J. H., and Ryu, J. H., 2009, “Performance of Neural Networks in Materials Science,” Mater. Sci. Technol., **25**(4), pp. 504–510.

[16] Pilania, G., Mannodi-Kanakkithodi, A., Uberuaga, B. P., Ramprasad, R., Gubernatis, J. E., and Lookman, T., 2016, “Machine Learning Bandgaps of Double Perovskites,” Sci. Rep., **6**(October 2015), pp. 1–10.

[17] Dey, P., Bible, J., Datta, S., Broderick, S., Jasinski, J., Sunkara, M., Menon, M., and Rajan, K., 2014, “Informatics-Aided Bandgap Engineering for Solar Materials,” Comput. Mater. Sci., **83**, pp. 185–195.

[18] Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C., and Scheffler, M., 2015, “Big Data of Materials Science: Critical Role of the Descriptor,” Phys. Rev. Lett., **114**(10), pp. 1–5.

[19] Meredig, B., Agrawal, A., Kirklin, S., Saal, J. E., Doak, J. W., Thompson, A., Zhang, K., Choudhary, A., and Wolverton, C., 2014, “Combinatorial Screening for New Materials in Unconstrained Composition Space with Machine Learning,”

Phys. Rev. B - Condens. Matter Mater. Phys., **89**(9), pp. 1–7.

[20] Curtarolo, S., Morgan, D., Persson, K., Rodgers, J., and Ceder, G., 2003, "Predicting Crystal Structures with Data Mining of Quantum Calculations," Phys. Rev. Lett., **91**(13), pp. 1–4.

[21] Seko, A., Maekawa, T., Tsuda, K., and Tanaka, I., 2014, "Machine Learning with Systematic Density-Functional Theory Calculations: Application to Melting Temperatures of Single- and Binary-Component Solids," Phys. Rev. B - Condens. Matter Mater. Phys., **89**(5), pp. 1–9.

[22] Behler, J., and Parrinello, M., 2007, "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces," Phys. Rev. Lett., **98**(14), pp. 1–4.

[23] Fan, J., Han, F., and Liu, H., 2014, "Challenges of Big Data Analysis," Natl. Sci. Rev., **1**(2), pp. 293–314.

[24] Keith, J. A., Vassilev-Galindo, V., Cheng, B., Chmiela, S., Gastegger, M., Müller, K. R., and Tkatchenko, A., 2021, "Combining Machine Learning and Computational Chemistry for Predictive Insights into Chemical Systems," Chem. Rev., **121**(16), pp. 9816–9872.

[25] Bietti, A., and Mairal, J., 2019, "On the Inductive Bias of Neural Tangent Kernels," Adv. Neural Inf. Process. Syst., **32**(NeurIPS).

[26] Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., and Müller, K. R., 2019, "Unmasking Clever Hans Predictors and Assessing What Machines

Really Learn," Nat. Commun., **10**(1), pp. 1–8.

[27] Pun, G. P. P., Batra, R., Ramprasad, R., and Mishin, Y., 2019, "Physically Informed Artificial Neural Networks for Atomistic Modeling of Materials," Nat. Commun., **10**(1), pp. 1–10.

[28] Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and Persson, K. A., 2013, "Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation," APL Mater., **1**(1), pp. 1–11.

[29] Kayser, K., Jacinto, S. D., Böhm, G., Fritz, P., Kunze, W. P., Nehrlich, A., and Gabius, H. J., 1997, "Application of Computer-Assisted Morphometry to the Analysis of Prenatal Development of Human Lung," Anat. Histol. Embryol., **26**(2), pp. 135–139.

[30] Hoerl, R. W., 2020, "Ridge Regression: A Historical Context," Technometrics, **62**(4), pp. 420–425.

[31] Agrawal, A., and Choudhary, A., 2016, "Perspective: Materials Informatics and Big Data: Realization of the 'Fourth Paradigm' of Science in Materials Science," APL Mater., **4**(5).

[32] Dimiduk, D. M., Holm, E. A., and Niezgoda, S. R., 2018, "Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering," Integr. Mater. Manuf. Innov., **7**(3), pp. 157–172.

[33] Takahashi, K., and Tanaka, Y., 2016, “Materials Informatics: A Journey towards Material Design and Synthesis,” Dalt. Trans., **45**(26), pp. 10497–10499.

[34] Kubat, M., 2017, *An Introduction to Machine Learning*.

[35] Seko, A., Hayashi, H., Nakayama, K., Takahashi, A., and Tanaka, I., 2017, “Representation of Compounds for Machine-Learning Prediction of Physical Properties,” Phys. Rev. B, **95**(14), pp. 1–11.

[36] Ward, L., Agrawal, A., Choudhary, A., and Wolverton, C., 2016, “A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials,” npj Comput. Mater., **2**(June), pp. 1–7.

[37] Choudhary, K., Decost, B., and Tavazza, F., 2018, “Machine Learning with Force-Field-Inspired Descriptors for Materials: Fast Screening and Mapping Energy Landscape,” Phys. Rev. Mater., **2**(8).

[38] Choudhary, K., Garrity, K. F., Reid, A. C. E., DeCost, B., Biacchi, A. J., Hight Walker, A. R., Trautt, Z., Hattrick-Simpers, J., Kusne, A. G., Centrone, A., Davydov, A., Jiang, J., Pachter, R., Cheon, G., Reed, E., Agrawal, A., Qian, X., Sharma, V., Zhuang, H., Kalinin, S. V., Sumpter, B. G., Pilania, G., Acar, P., Mandal, S., Haule, K., Vanderbilt, D., Rabe, K., and Tavazza, F., 2020, “The Joint Automated Repository for Various Integrated Simulations (JARVIS) for Data-Driven Materials Design,” npj Comput. Mater., **6**(1).

[39] Tanaka, I., 2018, *Nanoinformatics*.

[40] Steinhardt, P. J., Nelson, D. R., and Ronchetti, M., 1983, “Bond-Orientational

Order in Liquids and Glasses," Phys. Rev. B, **28**(2), pp. 784–805.

[41] Bart, A. P., 2013, "On Representing Chemical Environments," **184115**, pp. 1–16.

[42] Terayama, K., Terayama, K., Terayama, K., Terayama, K., Sumita, M., Sumita, M., Tamura, R., Tamura, R., Tamura, R., Payne, D. T., Chahal, M. K., Ishihara, S., Tsuda, K., Tsuda, K., and Tsuda, K., 2020, "Pushing Property Limits in Materials Discovery: Via Boundless Objective-Free Exploration," Chem. Sci., **11**(23), pp. 5959–5968.

[43] Liu, Q., Lee, J. D., and Jordan, M. I., 2016, "A Kernelized Stein Discrepancy for Goodness-of-Fit Tests and Model Evaluation," (1).

[44] Blöchl, P. E., 1994, "Projector Augmented-Wave Method," Phys. Rev. B, **50**(24), pp. 17953–17979.

[45] Perdew, J. P., Burke, K., and Ernzerhof, M., 1996, "Generalized Gradient Approximation Made Simple," Phys. Rev. Lett., **77**(18), pp. 3865–3868.

[46] Kresse, G., and Hafner, J., 1993, "Ab Initio Molecular Dynamics for Liquid Metals," Phys. Rev. B, **47**(1), pp. 558–561.

[47] Joubert, D., 1999, "From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method," Phys. Rev. B - Condens. Matter Mater. Phys., **59**(3), pp. 1758–1775.

[48] Vargas-Hernández, R. A., 2020, "Bayesian Optimization for Calibrating and Selecting Hybrid-Density Functional Models," J. Phys. Chem. A, **124**(20), pp. 4053–4061.

[49] Hybridization, D., and Hardness, U., 2021, “Diverse Hybridization and Ultrahigh Hardness,” pp. 1–15.

[50] Mansouri Tehrani, A., Oliynyk, A. O., Parry, M., Rizvi, Z., Couper, S., Lin, F., Miyagi, L., Sparks, T. D., and Brgoch, J., 2018, “Machine Learning Directed Search for Ultraincompressible, Superhard Materials,” J. Am. Chem. Soc., **140**(31), pp. 9844–9853.

[51] Chung, H. Y., Weinberger, M. B., Yang, J. M., Tolbert, S. H., and Kaner, R. B., 2008, “Correlation between Hardness and Elastic Moduli of the Ultraincompressible Transition Metal Diborides RuB#!2, OsB2, and ReB2,” Appl. Phys. Lett., **92**(26), pp. 2008–2010.

[52] Levine, B. J. B., Tolbert, S. H., and Kaner, R. B., 2009, “Advancements in the Search for Superhard Ultra-Incompressible Metal Borides,” pp. 3519–3533.

[53] Lakes, R., 1993, “Advances in Negative Poisson’s Ratio Materials,” Adv. Mater., **5**(4), pp. 293–296.

[54] Jiang, J. W., and Park, H. S., 2014, “Negative Poisson’s Ratio in Single-Layer Black Phosphorus,” Nat. Commun., **5**, pp. 1–7.

[55] Wen, Y., Wang, L., Liu, H., and Song, L., 2017, “Properties of B19 TiAl,” pp. 1–11.

[56] Liu, Z. T. Y., Zhou, X., Gall, D., and Khare, S. V, 2014, “First-Principles Investigation of the Structural , Mechanical and Electronic Properties of the NbO-Structured 3d , 4d and 5d Transition Metal Nitrides First-Principles Investigation

of the Structural , Mechanical and Electronic Properties of the NbO-Structured 3d , 4d and 5d Transition Metal Nitrides," Comput. Mater. Sci., **84**(December 2017), pp. 365–373.