

Article

Not peer-reviewed version

AI-based Nano-scale Material Property Prediction For Li-Ion Batteries

Mohit Anil Lal , Akashdeep Singh , Ryan Mzik , Amirmasoud Lanjan , [Seshasai Srinivasan](#) *

Posted Date: 28 November 2023

doi: 10.20944/preprints202311.1620.v1

Keywords: Quantum Mechanics; Molecular Dynamics; battery materials; Artificial Intelligence; Machine Learning; Nanotechnology







Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

AI-Based Nano-Scale Material Property Prediction For Li-Ion Batteries

Mohit Anil Lal ^{1,†} , Akashdeep Singh ^{1,†} , Ryan Mzik ¹, Amirmasoud Lanjan ²  and Seshasai Srinivasan ^{1,2,*} 

¹ W Booth School of Engineering Practice and Technology, McMaster University, Hamilton, Ontario, Canada

² Department of Mechanical Engineering, McMaster University, Hamilton, Ontario, Canada

* Correspondence: ssriniv@mcmaster.ca

† These authors contributed equally to this work.

Abstract: Molecular dynamics (MD) simulations present a sophisticated nano-scale computational approach that can play a critical role in material design for next generation batteries. One critical piece of information needed for MD simulations is the non-bonded potential parameters, which can be obtained through quantum mechanics (QM) calculations or experimental methods. However, experimental data is not available for exploratory and novel materials to derive the potential parameters. Also, the QM approach needs significant computational power and is too time consuming for large systems. On the other hand, current Artificial Intelligence techniques are faster, but they cannot be generalized across molecules and properties. As a result, there is a significant barrier to discovering and exploring new materials. Overcoming this barrier, in this work, we propose a Machine Learning (ML) based technique that can learn inter-atomic potential parameters for various particle-particle interactions employing QM calculations. This ML model can be used as an alternative for QM calculations for predicting non-bonded interactions in a computationally efficient manner. Using these parameters as input to MD simulations, we can predict a diverse range of properties, enabling researchers to design new and novel materials suitable for various applications in the absence of experimental data. We employ our ML-based technique to learn the Buckingham potential, a non-bonded interatomic potential. Subsequently, we utilize these predicted values to compute the densities of four distinct molecules, achieving an accuracy exceeding 93%. This serves as a strong demonstration of the efficacy of our proposed approach.

Keywords: quantum mechanics; molecular dynamics; Artificial Intelligence; Machine Learning; nanotechnology

1. Introduction

The chemical compound space (CCS) is the theoretical space consisting of every possible compound known (and unknown) to us [1]. Even some of our largest databases consisting of approximately 10^8 known substances are a mere drop in the ocean compared to an estimated 10^{180} substances that possibly make up the CCS [2]. Needless to say, the next big discovery of a compound that can revolutionize energy storage devices of the future is far from trivial.

The status quo for techniques used in the discovery of new and novel materials to enhance battery technologies has progressed from expensive and time-consuming empirical trial and error methods to the more recent first principles approach of using quantum mechanics (QM) [3–6], Monte Carlo simulations and molecular dynamics (MD) [7–10]. QM calculations evaluate electron-electron interactions solving the complex Schrödinger equation, thereby enabling accurate results on a wide variety of properties. However, the computational cost is a bottleneck for molecules larger than a couple hundred atoms. Hence, for multi-component and/or multi-layer structures such as the solid electrolyte interface layer, QM is not a feasible approach. Additionally, many battery components including ionic and polymer electrolytes, crystal structures, and electrode-electrolyte interactions [8,11] are better analyzed on larger length and time scales that are inaccessible with QM. MD simulations

simplify particle-particle interactions to five main types of interactions, namely, Nonbonded, Bonded, Angle, Dihedral, and Improper interactions. These interactions, which can be obtained using a simple algebraic equation, reduce the computational cost significantly and are applicable to systems almost 10^6 times larger. To analyze ion migration in perovskite nickelate with 200 atoms, QM techniques, even using Density Functional Theory (DFT) approximation to reduce computational costs, require about 10^5 core-hours of computational time on a picosecond range simulation. On the other hand, MD simulations with 105 atoms required only 10^4 core-hours of computational time [12]. Thus, MD simulations enable the analysis of a wide variety of properties and behavior of materials at the atomic scale, such as crystal structure, thermal properties, and mechanical properties, often too complex to model using QM calculations.

Though MD simulations are widely used to investigate the properties of materials at the atomic level, these simulations rely on experimentally derived interatomic potential parameters that determine the forces between particles [13]. This dependence on prior experimental data poses a challenge in using MD to design new and novel materials. To address this issue, Lanjan et al. [14] recently proposed a novel computational framework that couples QM calculations with MD simulations. This generates a wide range of crystal structures by varying a single system parameter (ex. bond length) while keeping other parameters relaxed at their minimum energy level. The QM calculations are then used to evaluate the system's energy as a function of these changes, and the resulting data points are used to fit the interaction equations to estimate the potential parameters for each type of particle-particle interaction. Employing this framework enables the study of crystal structures with the accuracy of QM calculations but at the speed and system sizes permissible by MD techniques. While this framework enhances nano-based computational methods, the QM calculations still need massive amounts of computational power which can be significantly reduced with the AI-based technique proposed in this work.

The emergence of ML, deep learning (DL) and Artificial Intelligence (AI) has helped alleviate the bottlenecks posed by QM and MD simulations and has made it possible to expand the scope of our search for novel materials in the CCS. ML and DL algorithms are orders of magnitude faster than ab-initio techniques. Unlike the QM-based simulations, that can take days to complete, ML algorithms can produce results within seconds. The use of AI has brought a paradigm shift in research related to improving battery technology as well as molecular property prediction, and material discovery in general. For example, Sandhu et. al. [15] used DL to examine optimal crystal structures of doped cathode materials in Lithium Manganese Oxide (LMO) batteries. Failed or unsuccessful synthesis data was used to predict the reaction success rate for the crystallization of templated vanadium selenites [16]. Using QM and ML techniques, Lu et al. [17] developed a method to predict undiscovered hybrid organic-inorganic perovskites (HOIPs) for photovoltaics. Their screening technique was able to shortlist six HOIPs with ideal bandgap and thermal stability from 5158 unexplored candidates. To identify material compositions with suitable properties, Meredig et al. [18] built a ML model trained on thousands of ground state crystal structures and used this model to scan roughly 1.6 million candidate compositions of novel ternary compounds to produce a ranked list of 4500 stable ternary compositions that would possibly represent undiscovered materials.

The broad approach employed when using AI-based property prediction models consists of three overarching components: a reference database consisting of relevant quantum-mechanical data which is used to fit the AI model; a mathematical representation that not only uniquely describes the attributes of the reference materials but also enables effective model training; and finally, a suitable AI model that can accomplish the learning task itself. In the ensuing sections, we describe these components in further detail.

1.1. Database

The fundamental premise of AI is the ability to draw inferences from patterns in data and enable an accurate prediction in unknown domains. Hence, the data, that makes up the training examples

for our learning task, becomes a critical aspect for successful prediction. With the introduction of the Materials Genome Initiative in 2011, the United States signaled the importance of unifying the infrastructure for materials innovation and harnessing the power of materials data. In lieu of the same goal, there has been an advent of various materials databases, such as the Inorganic Crystal Structure Database (ICSD) [19], the Open Quantum Materials Database (OQMD) [20], the Cambridge Structural Databases [21], the Harvard Clean Energy Project [22], the Materials Project [23], and the AFLOWLIB [24]. Specifically, the size of the training examples, the diversity of the dataset, and the degrees of freedom, all contribute to how effective the learning task for a specific objective can be [25]. In predicting properties such as band-gap energy and glass-forming ability for crystalline and amorphous materials, Ward et. al. [26] methodically selected a chemically diverse set of attributes taken from the OQMD. Similarly, for electronic-structure problems, Schütt et. al. [27] noted that the density of states at the Fermi energy is the critical property of concern. In predicting this property, around 7000 crystal structures from the ICSD were used, observing higher predicted variance for certain configurations and the need to extend the training set in these specific areas. The process of material discovery is complex and diverse, and it is not surprising that there is no one-size-fits-all database that can accurately predict the properties of all materials. The physical and chemical characteristics of materials vary widely, requiring different methods and techniques for precise analysis and prediction. Moreover, the current methodologies rely on the availability of well-curated data or the ability to manually generate such data, which is a daunting and often infeasible task, especially for new and unexplored materials. Thus, there is a need to develop generalizable and adaptable approaches that can efficiently handle a diverse range of materials, properties, and configurations, without the need for extensive data generation or curation.

1.2. Molecular Representation

ML algorithms draw inferences from data to establish a relationship between atomic structure and the properties of a system. To enable the best possible structure-property approximation, a good representation of the material (also referred to as the 'fingerprint' or 'descriptor') is crucial. The first Hohenberg-Kohn theorem of DFT proves that the electron density of a system contains all the information needed to describe its ground state properties, and it is a "universal descriptor" that can be used to predict these properties without knowledge of the details of the interactions between the electrons [28]. Crucially, for ML, a good molecular representation is invariant to rotation and translation of the system, and to permutation of atomic indices [29]. Therefore, unfortunately, electronic density is not a universally suitable representation of a system. Additionally, a good descriptor must be unique, continuous, compact, and computationally cheap [29]. Often, there are multiple molecular geometries that possess similar values for a property. Hence, there is no single universal representation for all properties leading to hundreds of molecular descriptors that are suitable only for a small subset of the CCS and a small subset of properties [30]. A commonly used molecular representation that satisfies the above-mentioned criteria of a good representation is the 'Coulomb matrix'. It uses the same parameters that constitute the Hamiltonian for any given system, viz., the set of Cartesian coordinates, R_i , and nuclear charges, Z_i [31]. While the coulomb matrix representation has shown tremendous success for property prediction in finite systems, it is unable to do the same for infinite periodic crystal structures [27]. Hansen et. al. [32] proposed a new descriptor called 'bag-of-bonds' that performed better due to incorporating the many-body interactions of a system. In fact, the use of different descriptors in an ML endeavor for material property prediction is so common that there are open-source software packages that provide implementations for a myriad of different descriptors [29]. Unfortunately, a lack of clarity on the right descriptor makes the use of AI inaccessible to researchers that possess domain expertise but lack the needed knowledge of AI. Additionally, the lack of generalizability of a chosen descriptor makes the current AI-based techniques inaccurate and narrow-scoped. Overcoming these challenges, the novel technique proposed in this work makes material discovery and property prediction easier

and more accessible without the time-consuming process of selecting a suitable descriptor. Specifically, our approach leverages a 2-stage process combining AI with MD simulations.

1.3. AI Model

In addition to an appropriate database and the precise molecular representation, a critical aspect in the material property prediction process is the choice of the AI algorithm. AI algorithms can be categorized into supervised learning, unsupervised learning, and reinforcement learning. Supervised learning uses a standard fitting procedure that attempts to determine a mapping function between known input features and the corresponding output labels. The goal is to make accurate predictions for new, unseen data. In contrast, unsupervised learning does not have prior knowledge of the desired output, and goal is to find patterns and structures in this unlabeled data. Reinforcement learning uses an iterative trial-and-error process where the actions are determined based on reinforcement in the form of a reward-penalty system. The goal here is to maximize the cumulative reward over time. Supervised learning is the most widespread category of learning used in materials research. Different models may be better suited for certain types of materials or properties, and the choice of model often depends on the available data and the specific goals of the prediction task. Akbarpour et. al. [33] found that Artificial Neural Networks (ANNs) performed better in predicting the synthesis conditions of nano-porous anodic aluminum oxide at the interpore distance in comparison to both Multiple Linear Regression and experimental studies. On the other hand, for the modelling and synthesis of zeolite synthesis, Manuel Serra et. al. [34] found that Support Vector Regression (SVR) outperformed ANNs and decision trees. Fang et. al. [35] proposed a novel hybrid methodology for forecasting the atmospheric corrosion of metallic materials where the optimal hyperparameters for an SVR model were automatically determined using a Generic algorithm. These examples highlight the need for AI expertise when choosing the right algorithm for a given application, which can be a barrier to making AI methods accessible for materials-based research.

In this work, we have presented a ML model to predict nonbonded potential parameters for conventional elements in the periodic table. We propose a novel approach that uses ML to learn a common empirical non-bonded interatomic potential, the Buckingham potential [36], and we successfully demonstrate the ability of this machine-learned potential to predict a wide range of properties when used as an input to classical MD simulations. We also demonstrate a marked improvement in the time taken to determine such properties compared to a traditional first principles approach.

2. Materials and Methods

Due to the enormity of the CCS, it is impossible to generate exhaustive datasets and consequently difficult for AI models to generalize well beyond the dimensional space of the training data. Additionally, the lack of a general representation that can scale well to very different properties results in AI-based techniques that fail to provide both the accuracy and the generalizability that comes with ab-initio techniques. Therefore, it is essential to use techniques that combine the speed of AI with the accuracy and generalizability of QM and MD simulations [37]. Hence, while most AI-based approaches follow the 3-step process described above to predict a confined set of properties in a narrow subset of the CCS, our approach generalizes well to a large set of properties. We delve into the details of our process in the ensuing paragraphs.

2.1. Database Generation

To train our ML model, we generated a database employing the QM approach and Quantum Espresso (QE) software package [38–40]. QE uses the principles of QM and computational methods to solve the Schrödinger equation using DFT approximation and can predict the electronic structure and properties of a system at the atomic scale. The database consists of QE generated non-bonded atom-pair energies for each element in the periodic table. Self-consistent field calculations (QE configurations in

Table 1) were performed for every possible same atom-pair system. Further, for each atom-pair, various plausible total charges, each with multiple inter-atomic distances were considered. We refer to each such combination of atom-pairs with their respective total charge as a system-charge configuration. The inter-atomic distance in each simulation was chosen to include different energy levels, from extremely close and unstable to far apart. There were more data points around the equilibrium range and fewer at farther distances. By exploring identical atom-pair combinations for each element in the periodic table and introducing a range of charge values for each atom-pair, we conducted simulations for 340 unique system-charge configurations. For each configuration, we evaluated interactions over 20 different distances, totaling nearly 700 hours of computational time. This effort resulted in the creation of a comprehensive database comprising non-bonded atom-pair energies for 6,400 distinct configurations. It must be noted that not all system-charge configurations were simulated with the same number of distance values. Some systems with larger elements may be unstable at close distances, and the QM simulations will not work well for those cases. Similar behavior may be seen for systems with higher charge values owing to the repulsion between atoms.

Table 1. The summary of settings for QM calculations in this work.

| Properties | Value Method |
|-----------------------|--------------------------------|
| XC Functional | PBE |
| Convergence Tolerance | $1.0 \times 10^{-6} Ry$ |
| W.F. Cutoff | 1.0×10^2 |
| Charge Cutoff | $1.0 \times 10^2 Ry$ |
| Maximum Force | $1.0 \times 10^{-3} Ry / Bohr$ |
| Smearing Factor | $1.0 \times 10^{-2} Ry$ |

2.2. Data Preprocessing

2.2.1. Curve Fitting:

For the system-charge configuration of Carbon–Carbon, for instance, the plot of relative non-bonded energy versus distance is shown in Figure 1 for various partial charges. In order to learn an interatomic potential for each configuration, the Buckingham Potential was selected as the appropriate choice.

$$U_{non-bonded} = Ae^{\frac{-r}{B}} - \frac{C}{r^6}.$$

(1)

In the above equation, r is the inter-atomic distance for a non-bonded atom-pair. By fitting each configuration’s energy and distance values to the above equation, the constants A, B, and C in the Buckingham potential were obtained using the "Levenberg-Marquardt" algorithm. This algorithm starts with an initial guess for the parameters (A, B and C) of the function, then calculates the gradient of the residuals with respect to the parameters and iterates until the residuals are minimized, or a maximum number of iterations is reached. To optimize the initial starting values for this algorithm, the grid search technique was utilized. As a result, a set of Buckingham constants was obtained for each system-charge configuration.

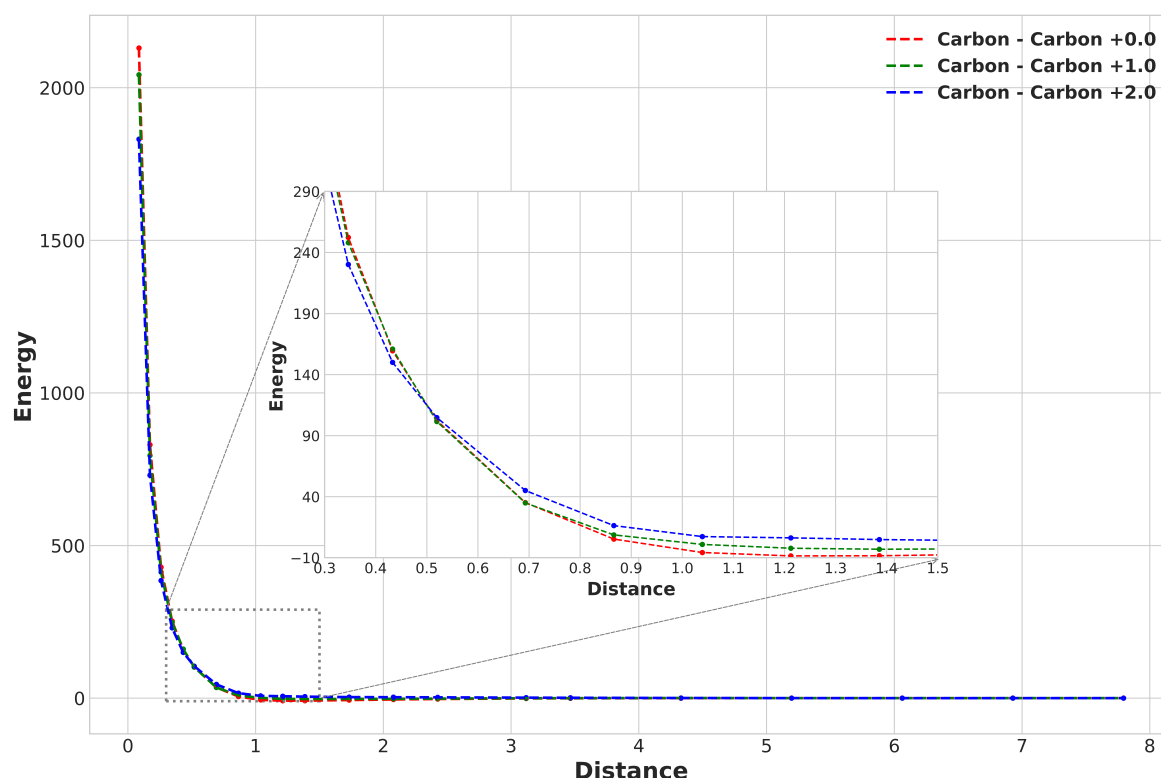


Figure 1. The Buckingham Potential for Carbon atom-pairs for various partial charges

2.2.2. Clustering:

From the Buckingham Potential equation, it is intuitively clear that for a given value of energy and distance, there can be multiple combinations of Buckingham constants. Now, it is prudent to choose a combination that could best enable the learning process of our model. This was done by clustering the system-charge configurations representing the same element. To cluster configurations for the same elements, the configuration with the highest R^2 value was first chosen for each element. All the Buckingham potentials obtained previously were recomputed while providing upper and lower bounds to the algorithm centred around this chosen configuration bearing the highest R^2 . The original computed configurations and the resulting recomputed configurations are shown in Figure 2 and Figure 3, respectively.

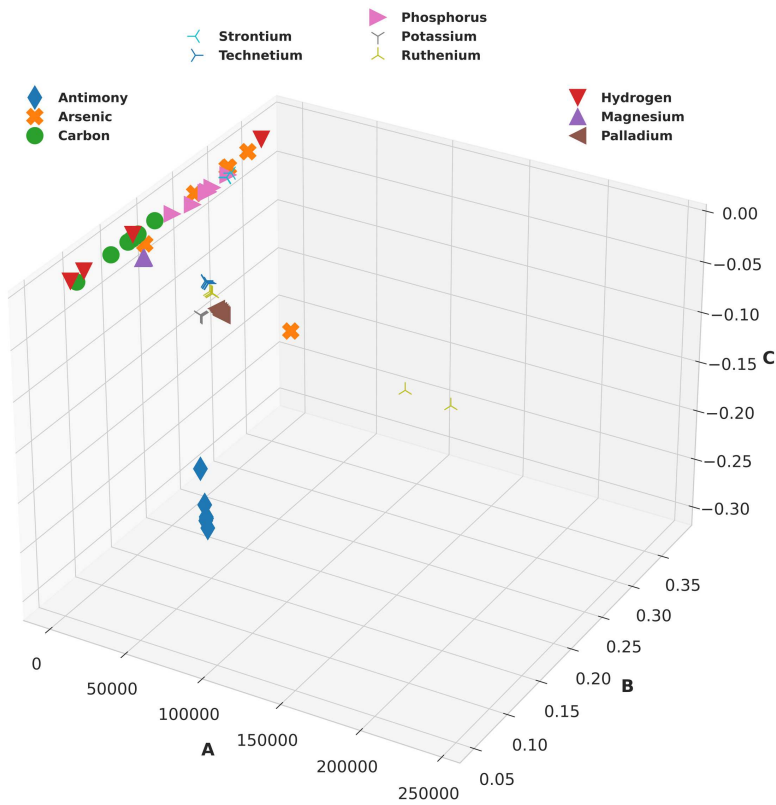


Figure 2. Original unbounded Buckingham Potential constants show a lack of element clustering thereby making model training difficult.

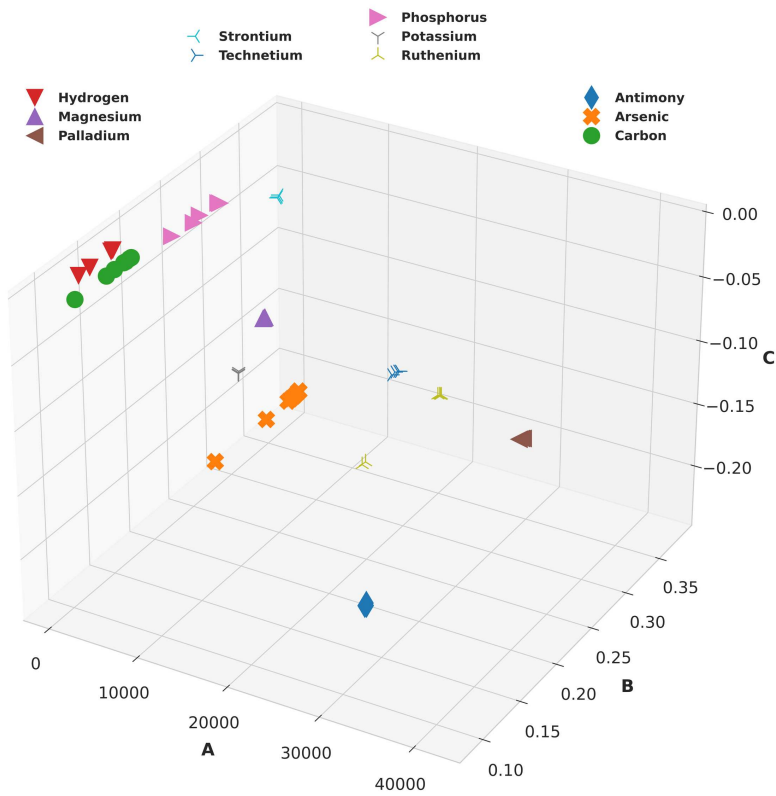


Figure 3. Clustered Buckingham Potential constants show grouped system-charge configurations for same elements.

2.3. AI Model and Training

2.3.1. Fingerprint:

The task of training an AI model for predicting non-bonded interatomic potentials, such as the Buckingham potential, requires a careful selection of both the configuration representation and the algorithm. The goal of this work is to determine a potential that can be used as input for MD simulations, giving researchers the flexibility to adjust additional parameters for their specific applications. To this end, we have chosen the most basic properties of a system-charge configuration as the input for our AI model. These include the atomic mass, atomic radius, atomic number, and partial charge of each atom. The task of selecting the appropriate representation, or "fingerprint," of each configuration is thus simplified, as we are only concerned with modeling the non-bonded interactions between atoms.

2.3.2. Training:

To achieve an optimal model, all labels generated using the Levenberg-Marquardt algorithm with an R^2 value of less than 90% were eliminated. The remaining dataset was then split into training and test sets, with a ratio of 75% to 25%, respectively, due to the small size of the dataset. Instead of further dividing the training data into a training and validation sets, k-fold cross-validation was used to train and evaluate the model. K-fold cross-validation divides data into k subsets and trains a model k times, each time using a different subset as the test set and the remaining ones as the training set. The performance is then averaged across all iterations to estimate the model's performance on unseen data. This technique helps utilize all the data, reduces the impact of sampling bias, and reduces the risk of overfitting, especially in small datasets.

2.3.3. Algorithm:

The use of ML was determined to be the most appropriate choice for this dataset as DL models are often prone to overfitting with smaller datasets. After evaluating several ML algorithms, the Random Forest Regressor was selected as the most suitable candidate due to its enhanced accuracy and robustness compared to traditional decision tree algorithms. The Random Forest Regressor operates by combining multiple decision trees, each of which is trained on a different subset of the training data and a randomly selected subset of features. The final prediction is made by averaging the predictions from all decision trees in the forest. The model was optimized for maximum accuracy with the grid search technique, which focused on tuning three hyperparameters: the number of estimators in the forest, the minimum number of samples required to split a node, and the utilization of bootstrap samples for each tree. The trained model was able to predict the Buckingham potential constants for test data with an accuracy of 93%.

2.3.4. MD Simulations:

With a trained model, we can predict Buckingham potential parameters for same element atom-pairs for any given partial charge. We then use the mixing rule to calculate the Buckingham constants for dissimilar atom pairs using the following equations:

$$A_{mn} = (A_{mm} \times A_{nn})^{0.5} \quad (2)$$

$$B_{mn} = \frac{1}{\left(\frac{1}{B_{mm}} \times \frac{1}{B_{nn}}\right)^{0.5}} \quad (3)$$

$$C_{mn} = -(C_{mm}^6 \times C_{nn}^6)^{1/12} \quad (4)$$

where A, B, and C represent the Buckingham potential parameters. Also, m and n represent the index of the atom-type in the system.

The non-bonded potentials obtained can subsequently be used for computational investigations at the atomic-molecular scale. Potential constants for the other types of interactions apart from non-bonded interactions are taken from the work by Lanjan et. al [14]. In this work, the “LAMMPS” software package [41] is employed with the settings and potentials described in Table 2 and 3, respectively.

Table 2. The summary of the settings for MD simulations in this work.

| Properties | Description/Specification |
|-------------------------|--|
| Energy minimization | Conjugate gradient for 2×10^4 steps |
| Equilibrium | 1ns NVT run and 10ns NPT run |
| Production run | 10ns |
| Motions integrator | Stoermer-Verlet, 1 fs time-step |
| Temperature coupling | 25° C, Nose-Hoover thermostat |
| Pressure coupling | 1 bar, Parrinello-Rahman barostat |
| Constraint solver | Constraining all bonds |
| Periodic boundary | x, y and z directions |
| Long-range interactions | Ewald summation with 1.0×10^{-5} accuracy |
| Trajectory output | Every 1,000 time-step (fs) |
| Neighbor list updating | Every 10 fs |
| Dynamic load balance | Yes |

Table 3. The potentials used in the MD simulations using the LAMMPS software package.

| Interaction Type | Potential Style | Equation |
|------------------|----------------------|---|
| Nonbonded | Buckingham/Coulombic | $E = Ae^{-\frac{r}{B}} - \frac{C}{r^6}$ |
| Bonded | Harmonic | $E = K(r - r_0)^2$ |
| Angle | Harmonic | $E = K(\theta - \theta_0)^2$ |
| Dihedral | Quadratic | $E = K(\phi - \phi_0)^2$ |
| Improper | Harmonic | $E = K(\chi - \chi_0)^2$ |

3. Results and Discussion

To evaluate the effectiveness of our method, we selected four molecules with different levels of complexity: (i) H₂O, a simple molecule (ii) (CH₂O)₂CO Ethylene Carbonate (EC), a relatively complex molecule with a ring section, (iii) C₂H₅OH (ethanol), a short-length hydrocarbon, and (iv) C₈H₁₈ (octane), a long chain molecule. Firstly, we used the partial charges from the literature [14] for all possible unique similar atom-pair combinations for each molecule to predict the corresponding Buckingham potential parameters using our trained ML model. We then computed the Buckingham potential parameters for the dissimilar atom-pair combinations using mixing rules outlined in equations 2-4. The accuracy of the predicted potential parameters is provided in Table 4. The comparison of the predicted values with the experimental values is shown in Figure 4. Next, we used these predicted potential parameters as inputs for MD simulations to predict the density of these molecules.

Table 4. Buckingham potential prediction using the trained ML model for the atoms and their respective partial charges constituting the four molecules viz. Water, Ethylene Carbonate, Ethanol and Octane.

| Element Name | Partial Charge | R ² (A) | R ² (B) | R ² (C) |
|--------------|----------------|--------------------|--------------------|--------------------|
| Carbon | -0.4656 | 100.00% | 97.75% | 94.18% |
| Carbon | -0.0257 | 100.00% | 97.75% | 94.18% |
| Carbon | 0.7305 | 99.15% | 96.15% | 94.18% |
| Carbon | -0.3101 | 100.00% | 97.75% | 94.18% |
| Carbon | -0.0714 | 100.00% | 97.75% | 94.18% |
| Hydrogen | 0.222 | 37.19% | 76.07% | 47.82% |
| Hydrogen | 0.4053 | 31.15% | 99.83% | 11.42% |
| Hydrogen | 0.1899 | 37.19% | 76.07% | 47.82% |
| Hydrogen | 0.1968 | 37.19% | 76.07% | 47.82% |
| Hydrogen | 0.4153 | 31.15% | 99.83% | 11.42% |
| Hydrogen | 0.1783 | 37.19% | 76.07% | 47.82% |
| Oxygen | -0.3745 | 99.75% | 98.94% | 98.04% |
| Oxygen | -0.711 | 98.91% | 98.94% | 98.04% |
| Oxygen | -0.5357 | 98.91% | 98.94% | 98.04% |
| Oxygen | -0.2865 | 99.75% | 98.94% | 98.04% |

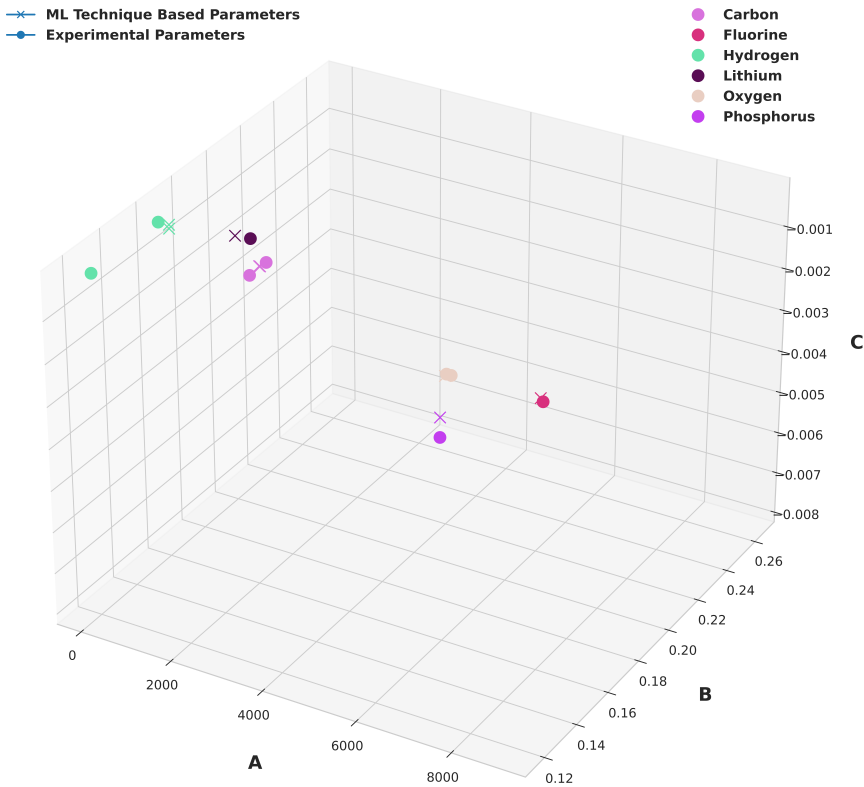


Figure 4. A 3D plot of the Buckingham potential parameter values for different elements and partial charges from Table 4 obtained experimentally and using our trained ML model.

Density is an important property of molecules as it can provide information about their packing and intermolecular forces. An accurate prediction of density requires an accurate modeling of interatomic forces and interactions, including both bonded and non-bonded interactions. Non-bonded interactions are sensitive to temperature and pressure changes and have a significant impact on the density of a molecule. As such, calculating the density precisely is a good indicator that the proposed ML-based technique can be employed to determine other molecular properties such as mechanical properties, thermal properties, and electrochemical properties, that are influenced by

similar interatomic interactions. Furthermore, density is a thermodynamic property that can be measured experimentally and accurately calculated using QM techniques. Hence, comparing the predicted densities of materials with the experimental values is an effective approach to assess the accuracy and reliability of our ML-based method. This comparison is summarized in Table 5, where our predicted densities are shown to have an accuracy greater than 93% with respect to the experimental data. Also, the densities obtained with MD simulations (specifications in Table 2) using our ML-predicted potential parameters are shown as a function of time in Figure 5. The density results in our MD simulations align closely with the expected values for Ethylene Carbonate (EC) and Octane, with slight deviations well within the permissible range for computational models. The dynamic density fluctuations observed in H₂O and Ethanol simulations are characteristic of the inherent complexities of molecular dynamics. Such variations are anticipated in MD simulations, reflecting the system’s responsiveness to changing conditions and interactions, while the overall trends remain consistent with experimental expectations, demonstrating the reliability of our computational approach.

Table 5. Comparison of density results from this work with experimental values

| Density | Experimental | This Work | Error |
|------------------|--------------|-----------|-------|
| H ₂ O | 0.99 | 0.95 | 4.04% |
| Octane | 0.7 | 0.73 | 4.29% |
| Ethanol | 0.79 | 0.78 | 1.27% |
| EC | 1.33 | 1.42 | 6.77% |

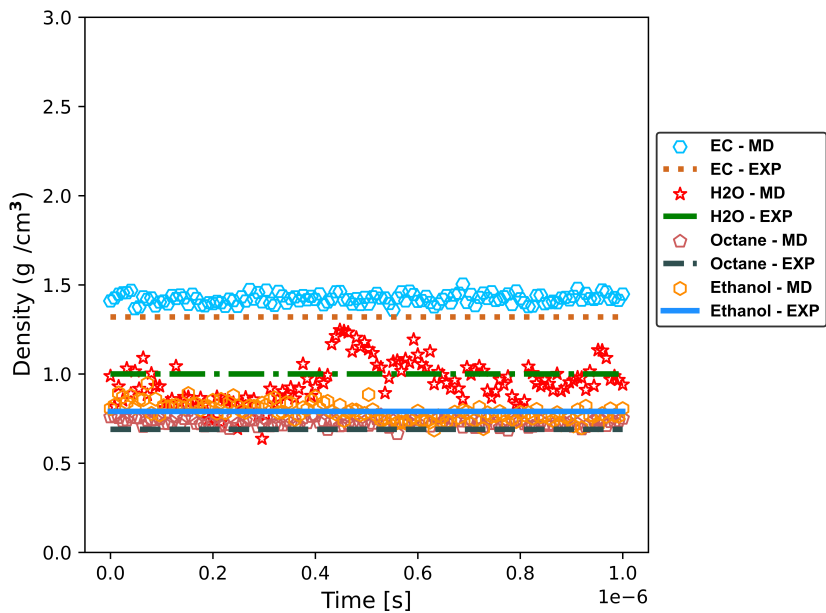


Figure 5. Density as a function of time: Comparison of density values of EC, H₂O, Octane and Ethanol molecules obtained from MD simulations using the non-bonded potential parameters from the trained ML model with the corresponding experimental data.

4. Conclusions

Nano-based computational techniques, such as QM and MD, have been the status-quo for material discovery and material property prediction. However, larger length and time scale analysis, vital for understanding important battery phenomena, is often infeasible with QM simulations due to their computational limitations. On the other hand, MD simulations require potential parameters as inputs, which are often difficult to obtain for new and novel materials. The need for accurate material property prediction is critical for the development of better, more efficient battery technologies. Addressing

these limitations, this work presents a novel ML-based technique that can learn inter-atomic potential parameters for various particle-particle interactions with the accuracy of conventional computational techniques like QM. When used as input to MD simulations, these learned potential parameters can predict a diverse range of properties, enabling the rapid screening and comparison of large databases of materials properties for battery applications.

In this study, we demonstrate the efficacy and validity of our proposed technique by learning a non-bonded interatomic potential, the Buckingham potential. We use the non-bonded potential parameters predicted in this work in conjunction with the potential parameters obtained from the literature for other types of interactions to predict the density of four different complex molecules. The obtained values were in close agreement with the experimental values for all four molecules, establishing the accuracy and efficacy of our proposed technique for the nano-scale evaluation of new and novel materials. Our technique can help quickly eliminate materials that are unlikely to meet the desired criteria, narrowing down the list of potential candidates for further evaluation. By identifying the most promising battery compositions and materials for further testing and development, this technique can accelerate the discovery of novel materials and the improvement of existing battery technologies.

In conclusion, the proposed ML-based technique provides a promising path towards discovering and developing novel materials with enhanced properties for applications such as next-generation batteries with superior electrochemical performance. Our technique can accelerate the search for new materials with desirable properties, allowing for the rapid screening and comparison of large databases of materials properties for such applications.

Author Contributions: Conceptualization: M.L., A.S., A.L. and S.S.; methodology, M.L. and A.S.; software: A.S.; validation: M.L., A.S. and R.M.; formal analysis: M.L., A.S., A.L; investigation: M.L. and A.S.; computational resources: S.S.; data curation: M.L., R.M. and A.S.; writing—original draft preparation: M.L.; writing—review and editing: A.L. and S.S.; supervision: A.L. and S.S.; project administration: S.S.

Funding: This research was funded by NSERC Discovery grants program. The grant number is RGPIN-2022-04988.

Acknowledgments: The authors would like to thank NSERC Canada for funding this research through their Discovery Grants program. The authors would also like to thank the reviewers for their constructive suggestions that helped improve the quality of this manuscript.

Data Availability Statement: The dataset generated and analyzed as part of this study can be found here: https://github.com/sudo-singh/predmodel_microproperties/tree/main/sim_data. The source code can be found here: https://github.com/sudo-singh/predmodel_microproperties

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-------|--------------------------------------|
| MD | Molecular Dynamics |
| QM | Quantum Mechanics |
| ML | Machine Learning |
| CCS | Chemical Compound Space |
| DFT | Density Functional Theory |
| DL | Deep Learning |
| AI | Artificial Intelligence |
| LMO | Lithium Manganese Oxide |
| HOIPs | Hybrid Organic-Inorganic Perovskites |
| OQMD | Open Quantum Materials Database |
| ICSD | Inorganic Crystal Structure Database |
| ANNs | Artificial Neural Networks |
| SVR | Support Vector Regression |
| QE | Quantum Espresso |
| EC | Ethylene Carbonate |

References

1. von Lilienfeld, O.A. First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties. *International Journal of Quantum Chemistry* **2013**, *113*, 1676–1689. doi:10.1002/qua.24375.
2. Lemonick, S. Exploring chemical space: Can AI take us where no human has gone before? *Chemical and Engineering News* **2020**, *98*.
3. Moradi, Z.; Heydarinasab, A.; Shariati, F.P. First-principle study of doping effects (Ti, Cu, and Zn) on electrochemical performance of Li_2MnO_3 cathode materials for lithium-ion batteries. *International Journal of Quantum Chemistry* **2021**, *121*. doi:10.1002/qua.26458.
4. Moradi, Z.; Lanjan, A.; Srinivasan, S. Multiscale Investigation into the Co-Doping Strategy on the Electrochemical Properties of Li_2RuO_3 Cathodes for Li-Ion Batteries. *ChemElectroChem* **2021**, *8*, 112–124. doi:10.1002/celc.202001206.
5. Tyagi, R.; Lanjan, A.; Srinivasan, S. Co-Doping Strategies to Improve the Electrochemical Properties of $\text{Li}_x\text{Mn}_2\text{O}_4$ Cathodes for Li-Ion Batteries. *ChemElectroChem* **2022**, *9*. doi:10.1002/celc.202101626.
6. Moradi, Z.; Lanjan, A.; Srinivasan, S. Enhancement of Electrochemical Properties of Lithium Rich Li_2RuO_3 Cathode Material. *Journal of The Electrochemical Society* **2020**, *167*, 110537. doi:10.1149/1945-7111/aba44d.
7. Lanjan, A.; Srinivasan, S. An Enhanced Battery Aging Model Based on a Detailed Diffusing Mechanism in the SEI Layer. *ECS Advances* **2022**, *1*, 030504. doi:10.1149/2754-2734/ac8e84.
8. Lanjan, A.; Moradi, Z.; Srinivasan, S. Multiscale Investigation of the Diffusion Mechanism within the Solid–Electrolyte Interface Layer: Coupling Quantum Mechanics, Molecular Dynamics, and Macroscale Mathematical Modeling. *ACS Applied Materials & Interfaces* **2021**, *13*, 42220–42229. doi:10.1021/acsami.1c12322.
9. Lanjan, A.; Choobar, B.G.; Amjad-Iranagh, S. First principle study on the application of crystalline cathodes $\text{Li}_2\text{Mn}_0.5\text{TM}_0.5\text{O}_3$ for promoting the performance of lithium-ion batteries. *Computational Materials Science* **2020**, *173*, 109417. doi:10.1016/j.commatsci.2019.109417.
10. Lanjan, A.; Choobar, B.G.; Amjad-Iranagh, S. Promoting lithium-ion battery performance by application of crystalline cathodes $\text{Li}_x\text{Mn}_{1-z}\text{Fe}_z\text{PO}_4$. *Journal of Solid State Electrochemistry* **2020**, *24*, 157–171. doi:10.1007/s10008-019-04480-6.
11. Aykol, M.; Herring, P.; Anapolsky, A. Machine learning for continuous innovation in battery technologies. *Nature Reviews Materials* **2020**, *5*, 725–727. doi:10.1038/s41578-020-0216-y.
12. Chan, H.; Narayanan, B.; Cherukara, M.J.; Sen, F.G.; Sasikumar, K.; Gray, S.K.; Chan, M.K.Y.; Sankaranarayanan, S.K.R.S. Machine Learning Classical Interatomic Potentials for Molecular Dynamics from First-Principles Training Data. *The Journal of Physical Chemistry C* **2019**, *123*, 6941–6957. doi:10.1021/acs.jpcc.8b09917.
13. Behler, J. Perspective: Machine learning potentials for atomistic simulations. *The Journal of Chemical Physics* **2016**, *145*, 170901. doi:10.1063/1.4966192.
14. Lanjan, A.; Moradi, Z.; Srinivasan, S. A computational framework for evaluating molecular dynamics potential parameters employing quantum mechanics. *Molecular Systems Design & Engineering* **2023**. doi:10.1039/D3ME00007A.
15. Sandhu, S.; Tyagi, R.; Talaie, E.; Srinivasan, S. Using neurocomputing techniques to determine microstructural properties in a Li-ion battery. *Neural Computing and Applications* **2022**, *34*, 9983–9999. doi:10.1007/s00521-022-06985-0.
16. Raccuglia, P.; Elbert, K.C.; Adler, P.D.F.; Falk, C.; Wenny, M.B.; Mollo, A.; Zeller, M.; Friedler, S.A.; Schrier, J.; Norquist, A.J. Machine-learning-assisted materials discovery using failed experiments. *Nature* **2016**, *533*, 73–76. doi:10.1038/nature17439.
17. Lu, S.; Zhou, Q.; Ouyang, Y.; Guo, Y.; Li, Q.; Wang, J. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nature Communications* **2018**, *9*, 3405. doi:10.1038/s41467-018-05761-w.
18. Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J.E.; Doak, J.W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Physical Review B* **2014**, *89*, 094104. doi:10.1103/PhysRevB.89.094104.

19. Belsky, A.; Hellenbrandt, M.; Karen, V.L.; Luksch, P. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallographica Section B Structural Science* **2002**, *58*, 364–369. doi:10.1107/S0108768102006948.
20. Kirklin, S.; Saal, J.E.; Meredig, B.; Thompson, A.; Doak, J.W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials* **2015**, *1*, 15010. doi:10.1038/npjcompumats.2015.10.
21. Groom, C.R.; Bruno, I.J.; Lightfoot, M.P.; Ward, S.C. The Cambridge Structural Database. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials* **2016**, *72*, 171–179. doi:10.1107/S2052520616003954.
22. Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R.S.; Gold-Parker, A.; Vogt, L.; Brockway, A.M.; Aspuru-Guzik, A. The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid. *The Journal of Physical Chemistry Letters* **2011**, *2*, 2241–2251. doi:10.1021/jz200866s.
23. Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K.A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*, 011002. doi:10.1063/1.4812323.
24. Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R.H.; Nelson, L.J.; Hart, G.L.; Sanvito, S.; Buongiorno-Nardelli, M.; Mingo, N.; Levy, O. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science* **2012**, *58*, 227–235. doi:10.1016/j.commatsci.2012.02.002.
25. Zhang, Y.; Ling, C. A strategy to apply machine learning to small datasets in materials science. *npj Computational Materials* **2018**, *4*, 25. doi:10.1038/s41524-018-0081-z.
26. Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials* **2016**, *2*, 16028. doi:10.1038/npjcompumats.2016.28.
27. Schütt, K.T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K.R.; Gross, E.K.U. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Physical Review B* **2014**, *89*, 205118. doi:10.1103/PhysRevB.89.205118.
28. Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, *136*, B864–B871. doi:10.1103/PhysRev.136.B864.
29. Himanen, L.; Jäger, M.O.; Morooka, E.V.; Canova, F.F.; Ranawat, Y.S.; Gao, D.Z.; Rinke, P.; Foster, A.S. DScribe: Library of descriptors for machine learning in materials science. *Computer Physics Communications* **2020**, *247*, 106949. doi:10.1016/j.cpc.2019.106949.
30. Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360–365. doi:10.1126/science.aat2663.
31. Rupp, M.; Tkatchenko, A.; Müller, K.R.; von Lilienfeld, O.A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Physical Review Letters* **2012**, *108*, 058301. doi:10.1103/PhysRevLett.108.058301.
32. Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O.A.; Müller, K.R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *The Journal of Physical Chemistry Letters* **2015**, *6*, 2326–2331. doi:10.1021/acs.jpclett.5b00831.
33. Akbarpour, H.; Mohajeri, M.; Moradi, M. Investigation on the synthesis conditions at the interpore distance of nanoporous anodic aluminum oxide: A comparison of experimental study, artificial neural network, and multiple linear regression. *Computational Materials Science* **2013**, *79*, 75–81. doi:10.1016/j.commatsci.2013.05.048.
34. Serra, J.M.; Baumes, L.A.; Moliner, M.; Serna, P.; Corma, A. Zeolite Synthesis Modelling with Support Vector Machines: A Combinatorial Approach. *Combinatorial Chemistry & High Throughput Screening* **2007**, *10*, 13–24. doi:10.2174/138620707779802779.
35. Fang, S.; Wang, M.; Qi, W.; Zheng, F. Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials. *Computational Materials Science* **2008**, *44*, 647–655. doi:10.1016/j.commatsci.2008.05.010.
36. The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1938**, *168*, 264–283. doi:10.1098/rspa.1938.0173.

37. Wei, J.; Chu, X.; Sun, X.; Xu, K.; Deng, H.; Chen, J.; Wei, Z.; Lei, M. Machine learning in materials science. *InfoMat* **2019**, *1*, 338–358. doi:10.1002/inf2.12028.
38. Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G.L.; Cococcioni, M.; Dabo, I.; Corso, A.D.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A.P.; Smogunov, A.; Umari, P.; Wentzcovitch, R.M. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter* **2009**, *21*, 395502. doi:10.1088/0953-8984/21/39/395502.
39. Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M.B.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; Colonna, N.; Carnimeo, I.; Corso, A.D.; de Gironcoli, S.; Delugas, P.; DiStasio, R.A.; Ferretti, A.; Floris, A.; Fratesi, G.; Fugallo, G.; Gebauer, R.; Gerstmann, U.; Giustino, F.; Gorni, T.; Jia, J.; Kawamura, M.; Ko, H.Y.; Kokalj, A.; Küçükbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N.L.; Nguyen, H.V.; de-la Roza, A.O.; Paulatto, L.; Poncé, S.; Rocca, D.; Sabatini, R.; Santra, B.; Schlipf, M.; Seitsonen, A.P.; Smogunov, A.; Timrov, I.; Thonhauser, T.; Umari, P.; Vast, N.; Wu, X.; Baroni, S. Advanced capabilities for materials modelling with Quantum ESPRESSO. *Journal of Physics: Condensed Matter* **2017**, *29*, 465901. doi:10.1088/1361-648X/aa8f79.
40. Giannozzi, P.; Baseggio, O.; Bonfà, P.; Brunato, D.; Car, R.; Carnimeo, I.; Cavazzoni, C.; de Gironcoli, S.; Delugas, P.; Ruffino, F.F.; Ferretti, A.; Marzari, N.; Timrov, I.; Urru, A.; Baroni, S. Quantum ESPRESSO toward the exascale. *The Journal of Chemical Physics* **2020**, *152*, 154105. doi:10.1063/5.0005082.
41. Thompson, A.P.; Aktulga, H.M.; Berger, R.; Bolintineanu, D.S.; Brown, W.M.; Crozier, P.S.; in 't Veld, P.J.; Kohlmeyer, A.; Moore, S.G.; Nguyen, T.D.; Shan, R.; Stevens, M.J.; Tranchida, J.; Trott, C.; Plimpton, S.J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications* **2022**, *271*, 108171. doi:10.1016/j.cpc.2021.108171.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.