

Navigating materials chemical space to discover new battery electrodes using machine learning

Mukhtar Lawan Adam^{a,b}, Oyawale Adetunji Moses^a, Jonathan Pradana Mailoa^c, Chang-Yu Hsieh^{d,*}, Xue-Feng Yu^{a,e,*}, Hao Li^{f,*}, Haitao Zhao^{a,#,*}

^a Center for Intelligent and Biomimetic Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 440305, Guangdong, PR China

^b Physics Department, Bayero University, Kano 700231, Nigeria

^c Tencent Quantum Laboratory, Shenzhen 518057, Guangdong, China

^d Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China

^e Shenzhen Key Laboratory of Micro/Nano Biosensing, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, China

^f Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan

ARTICLE INFO

Keywords:

Machine learning
Active learning
Electrode materials
Voltage
Electronegativity
Physiochemical properties
All-solid-state battery

ABSTRACT

Investigating the role of electrodes' physiochemical properties on their output voltage can be beneficial in developing high-performance batteries. To this end, this study uses a two-step machine learning (ML) approach to predict new electrodes and analyze the effects of their physiochemical properties on the voltage. The first step utilizes an ML model to curate an informative feature space that elucidates the relationship between physiochemical properties and voltage output. The second step trains an active learning model on the informative feature space using Bayesian optimization to screen potential battery electrodes from a dataset of 3656 materials. This strategy successfully identified 41 electrode materials that exhibit good electronic conductivity and host highly electronegative anions. This work provides an efficient strategy to discover novel electrode materials while integrating domain knowledge of chemistry and material science with ML in materials research.

1. Introduction

The quest for clean energy, coupled with the increasing usage of portable devices and electric vehicles, has stimulated a high demand for energy storage. Electrochemical energy storage devices such as batteries and supercapacitors store electricity through an electrochemical process.[1] Battery has three essential components: electrode (cathode/anode), electrolyte, and separator.[1,2] The energy storage performance of a battery largely depends on the electrodes, which dictate the battery's high energy density, overall capacity, and average voltage.[1]

Lithium-ion batteries (LIB) have revolutionized and enabled transformative advances in energy storage.[3,4] They are currently the most reliable energy storage systems due to their high energy density, excellent cycling stability, high working voltage, and relatively good rate capability.[5–7] However, despite the demonstrated technological prowess of LIB in energy storage. They currently face challenges ranging

from production cost to safety issues, coupled with limited global Lithium reserves, a complex distribution network, and high demand that makes producing the electrode for LIBs expensive.[8–13] Moreover, LIBs are known to have poor low-temperature performance,[14] issues with dendrite growth,[15] high tendencies to be damaged under high working voltage, and overheating under extreme temperatures,[16] which might lead to thermal runaway and consequent explosion.[17,18] Therefore, searching for the next-generation post-LIB will go a long way in providing a substantive alternative.

Recently, post-lithium metal-ion batteries based on monovalent ions have been proposed as promising alternatives to LIB.[19–29] The monovalent metal ions (Na^+ and K^+) possess similar electrochemical working principles to LIB, utilizing similar electrolytes and manufacturing technology.[21] Monovalent metal ions batteries do not need scarce and expensive transition metals to enhance their energy density as opposed to the LIB.[30]

However, monovalent metal battery electrodes still face the

* Corresponding authors.

E-mail addresses: kimhsieh@zju.edu.cn (C.-Y. Hsieh), xf.yu@siat.ac.cn (X.-F. Yu), li.hao.b8@tohoku.ac.jp (H. Li), ht.zhao@siat.ac.cn (H. Zhao).

Lead Contact.

challenges of finding suitable negative electrodes and considerable volume changes during charging and discharging.[31] Interestingly, several different multivalent metal ions (Ca^{2+} , Zn^{2+} , Mg^{2+} , and Al^{3+}) have also been suggested as alternatives to the LIB.[30–38] Early research has indicated that most of these metals do not suffer from dendrite formation in the electrolytes compared to the LIB.[39] Their lower reactivity (than alkali) also makes their pure metal state suitable to be utilized as an anode.[40] They are known to achieve greater theoretical energy densities than the monovalent metal ions and use much safer electrolytes than the LIBs.[15] Due to their multivalent chemistry, these elements tend to have an edge in volumetric energy density.[30] Multivalent ions host multiple charges, providing higher capacity than when a similar number of ions of a monovalent element are stored in a given electrode volume,[35] providing an edge over the monovalent metal-ion batteries in terms of better potential volumetric and specific capacity.[15,28] Noteworthy, multivalent ion batteries suffers slow ion conduction through the electrolyte occurs due to electrostatic trapping, worsened by higher cation charges.[41] Although multivalent ion batteries have not yet been poised to overtake LIBs. However, understanding the role of electrodes' physiochemical properties on their voltage can be vital to sustaining the pace in the development of next-generation commercially viable battery systems.

The Edisonian approach has been the traditional way for the search/discovery of new electrode materials.[42,43] Discovery through this path is routinely guided by studying materials having similar compositional and structural motifs to known electrodes. However, given this route's time-, resource-consuming, and serendipitous nature, there arises a need for an alternative paradigm that can give an intuitive picture of how physiochemical properties influence the voltage and identify promising candidate materials more efficiently. However, from an electrochemical perspective, general theories based on materials' atomistic properties exist, but their quantitative and predictive influence on voltage performance is lacking.[44–47] Identifying these salient features (from the vast materials physiochemical space) and their quantitative influence on the voltage is essential to accelerate the pace of electrode design and discovery.

Recently, the development of high-performance computational platforms, high-throughput density functional theory calculations, and their utilization in materials science have helped create scores of materials repositories.[48–54] Several groups have demonstrated the potential of ML in predicting novel organic electrodes and understanding battery electrolytes.[55–58] Therefore, identifying and discovering new electrodes can be accelerated by employing data-driven approaches such as interpretable machine learning (ML) and active learning on these repositories. Interpretable ML [59–64] has the potential to go beyond the conventional ML approaches that merely provide voltage predictions [65–68] in a black-box fashion. The interpretable ML can also elucidate the quantitative influence/importance of the input features in a model and reveal scientific insights. Furthermore, identifying important atomistic features from the interpretable ML can curate an informative feature space that can be exploited by Bayesian optimization to discover new electrodes.

Herein, we aim to use ML to predict new electrode materials and unravel the influence of electrodes' physiochemical properties on their voltage. Hence, a two-step approach was implemented. At the first stage, an interpretable ML was employed to study the physiochemical feature space of 2986 electrode compounds with their DFT-calculated voltage curated from the Materials Project database. The electrodes' voltage was significantly influenced by their electronegativity, polarizability, average atomic radius, average quantum number, atomic density, the heat of atomization, the material's melting point, and the average number of valence electrons. Secondly, riding on the as-acquired knowledge (from the interpretable ML model), a reduced 10-dimensional informative feature space is formed to be utilized by Bayesian optimization (BO) in an active learning (AL) module. A reduced informative of less than twenty feature spaces is essential for attaining an

optimal performance by a BO method.[69–71] The AL module is employed to query a dataset of 3656 stable compounds for potential electrode materials and DFT calculations to verify their voltages.

Going through this two-step screening pipeline, 41 potential electrode materials, mostly ternary metal fluorides and oxide compounds, were predicted. On analyzing their physiochemical properties, the predicted materials host atoms with relatively low atomic radius, polarizability, and highly electronegative atoms (the fluorides and oxides). As attributed by the interpretable ML, these features contribute the most to the voltage. Employing chemical intuition to explain the observations from the AL predictions and the interpretable ML model corroborates with the inductive effect theory in which electrode redox couples are stabilized. Also, the voltage is enhanced due to charge transfer from the anions. The theory explains that high electronegative anions in electrodes can enhance the voltage.

Notwithstanding, the presence of highly electronegative elements in materials can result in a wide band gap, which is considered counter-productive to electron flow in electrodes. Hence, we calculated the predicted electrode materials' band gap and electronic conductivities. The materials were found to have sizeable bandgaps and good electronic conductivity. Through our in-silico pipeline, we integrated domain knowledge in chemistry and material science and corroborated vital physiochemical traits (highly electronegativity anions, narrow bandgap) to predict 41 new potential electrode materials that can identify potent electrodes from a pool of materials.

2. Results and discussion

2.1. Design strategy

Fig. 1 summarizes the schematics of our overall workflow. In the first step, we train a classical ML model that predicts the electrode voltage (Fig. 1B) based on a dataset of 2986 electrode materials curated from the Materials Projects battery electrodes database (Fig. 1A). The features used in the model generation are based on the material's structural and physiochemical properties, which are non-DFT features. The classical ML model (Fig. 1B) is then used to construct an interpretable ML model using the Tree SHAP to understand the roles of individual physiochemical features in voltage predictions and extract the most informative features. To provide us with information to form an informative feature space that will optimize the performance of the BO algorithm in the AL module (Fig. 1D). The AL module starts with the BO algorithm (Fig. 1D) querying the exploration database of 3656 materials (Fig. 1F) for potential electrode materials with the best voltage (the query criteria set to 6 V). The queried best material from Fig. 1F by the BO in Fig. 1D is then passed for DFT calculations to verify the voltage. The predicted material is labeled and deposited into the dataset (Fig. 1F). The AL loop continues until the voltage of the selected materials falls significantly below the selection criteria (Fig. 1C).

2.2. Dataset construction

The electrodes in the initial dataset (4351 entries) containing DFT-calculated voltages, capacity, and discharge formula of the electrodes based on various metal ions (Li, Na, K, Rb, Zn, Al, K, Rb, Y, and Cs) batteries, were curated from the battery explorer of the Materials Project (MP) database.[48] On inspection of duplicate entries based on the discharge formula, 1365 duplicate entries were removed, leaving a final dataset of 2986 electrode materials. Fig. 2A shows the distribution of the elements in the dataset, where Mg and Na predominantly follow Li metal ion-containing materials, Ca, K, Zn, Rb, and Cs. From the scatter plot of the data set in Fig. 2B, the maximum electrode average voltage in the dataset ranges around 7.5 V, and the maximum electrode capacity is around 740 mAh/g. Furthermore, the histogram of the average voltage with the frequency plotted in Fig. 2C shows a normal distribution of the average voltages in the data set.

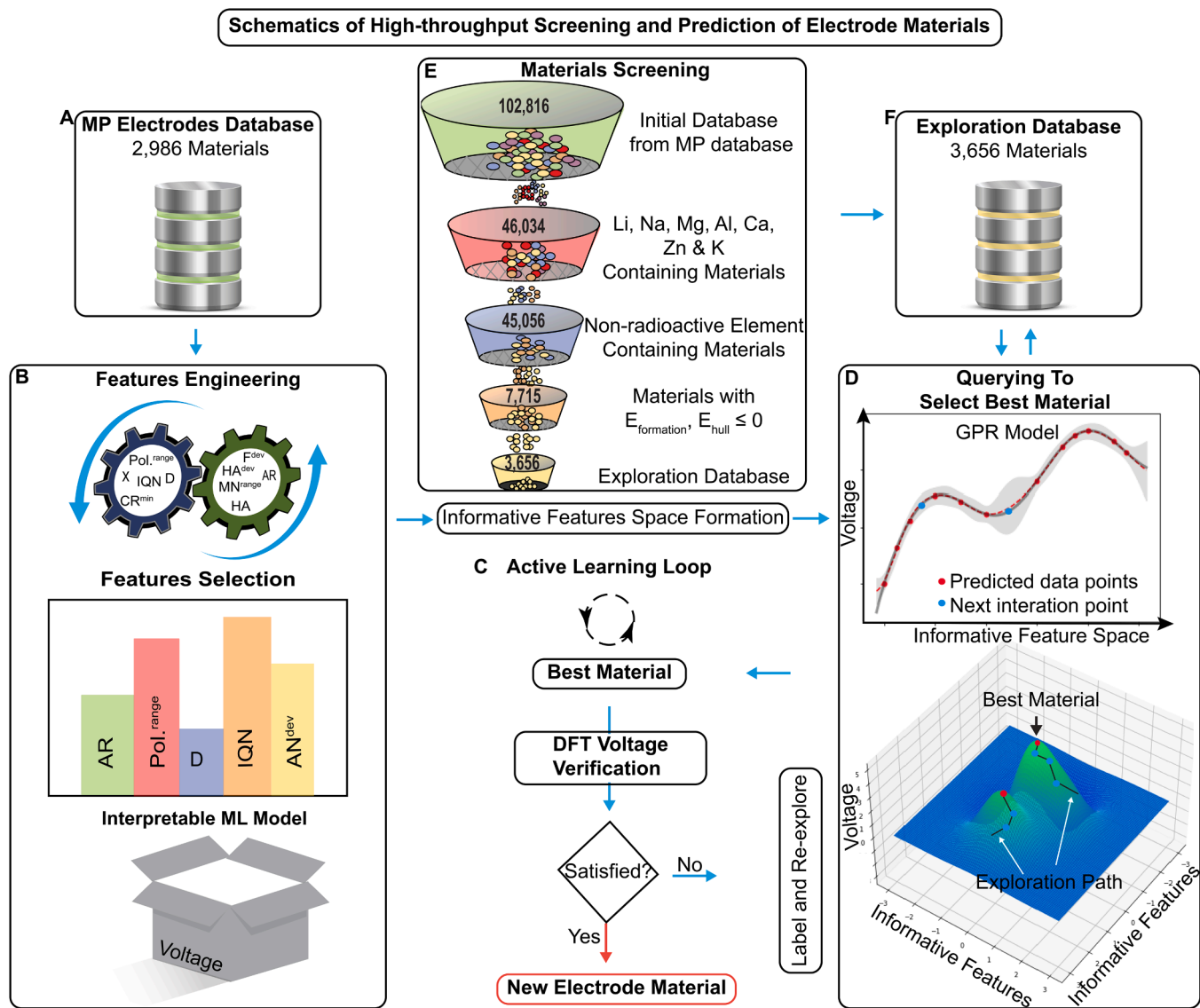


Fig. 1. Workflow and Schematics for the High-throughput Screening and Prediction of Promising Electrode Materials from a Material Database (A) A curated database of electrode materials with their DFT-calculated voltages from the battery explorer of the Materials Project. **(B)** Physiochemical features engineering, generation of an interpretable ML model for the prediction of electrode voltage, and selection of informative features space for use in the AL module. **(C)** The AL loop utilizes Gaussian process regression as a surrogate to explore and exploit the chemical space of materials to select the best possible material, and we utilize DFT for further verification. **(D)** The Gaussian process regression model to query the new database. **(E)** The Materials Project database was curated to include 102,819 materials, of which 3656 were identified as potential electrode materials. **(F)** The active learner will query the exploration database for potent electrode materials.

2.3. Feature engineering and machine learning

Physiochemical features based on electrode composition and structure were generated using the Oliynyk featurization.[72] A total of 324 features were extracted for each potential electrode material in the dataset. Some extra features, such as the crystal density, packing fractions, and volume per atom of the electrodes, were included to capture the structural information of the crystals adequately. The final selected optimal features with their Pearson correlation are shown in Fig. 3A (See Table S1 for the meaning of each symbol).

Furthermore, utilizing the PyCaret package, 8 ML algorithms: Light Gradient Boost Machine (LGBM), Extreme Gradient boosting (XGB), Gradient boosting regressor (GBR), CatBoost regressor (Cat), Random Forest regressor (RFR), Extra trees regressors (ETR), Kernel ridge (Ridge), and Support vector regressor (SVR) were tested on our dataset to find the most suitable algorithm. The performance indices of the ML algorithms, as shown in Fig. 3B, indicate that the LGBM performs the

best on our dataset.

Additionally, each of the 24 optimal feature importance scores is calculated, and then the mean is taken, as shown in Fig. 3C. All 24 features influence the models - signaling a successful feature engineering. The range of polarizability, the average deviation of the atomic radius, the average deviation of the angular quantum number, deviation families, the average deviation of the Pauling electronegativity, and the standard deviation of the heat of atomization are the features with the highest mean scores.

Further on, to generate an optimal ML model, we carried out a hyperparameter tuning using Optuna (Figure S1) of the best-performing ML algorithm (the LGBM) on the 24 feature sets in predicting the voltage. The performance of the final LGBM model is shown in Fig. 3D.

2.4. Interpretable machine learning

Here, we utilized the model-agnostic approach to create an

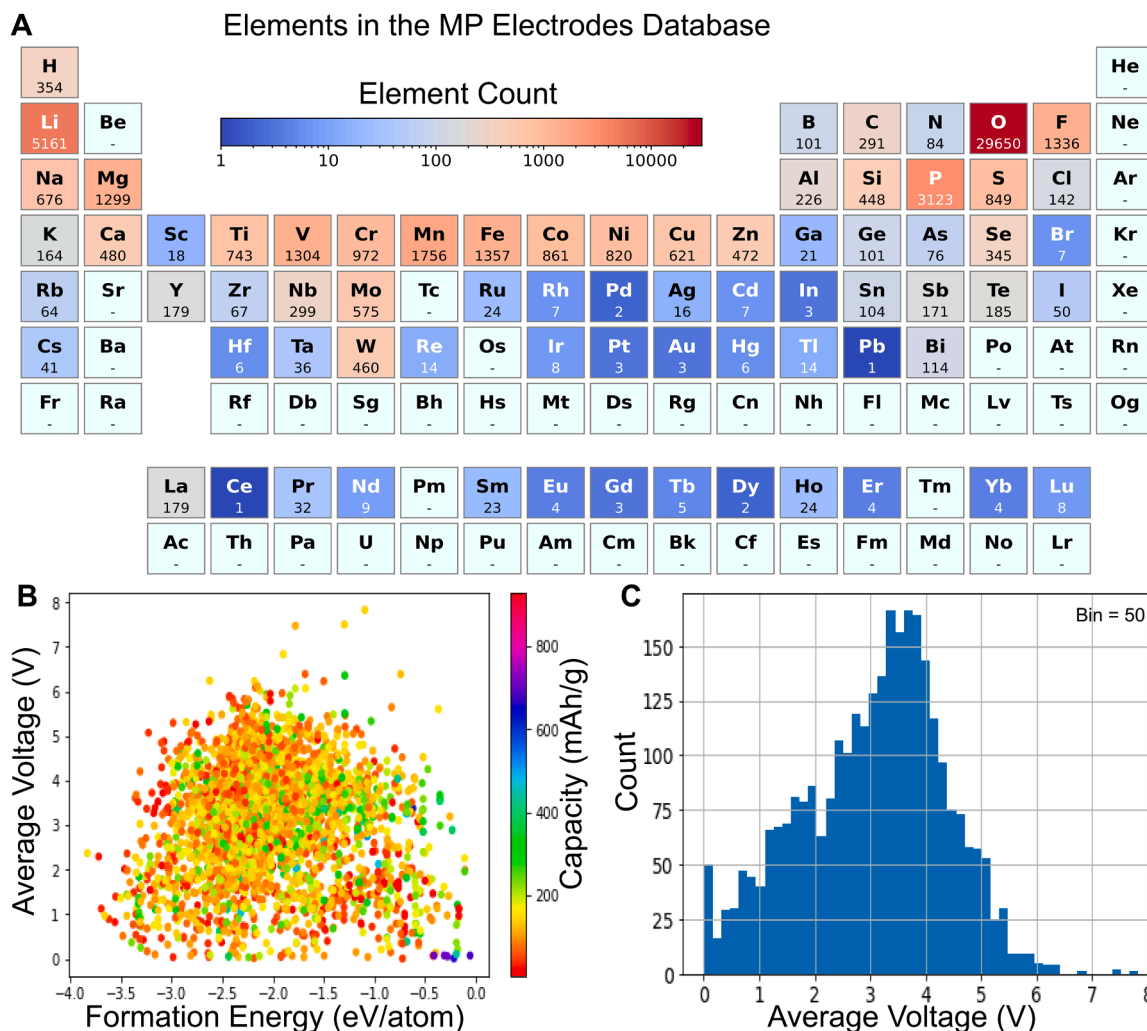


Fig. 2. Statistical Analysis of the Materials Project (Initial) Electrodes Database (A) Distribution and count of the elemental constituents in the Materials Project electrode data set. (B) Scatter plot of the electrodes' average voltage, capacity, and formation energies. (C) Histogram of the average voltage.

interpretable ML model to describe the physicochemical features' influence on electrode voltage.[64] This approach, in principle, uses a surrogate algorithm (i.e., Tree SHAP or Lime) as the front end to learn and extract the global and local interpretability of the features influenced by an already-trained black-box ML model. Hence, utilizing the final LGBM model, we can use Tree Shapley tools to make our black-box LGBM model interpretable.[73] These will provide insights into the contribution of the material's physicochemical features to the resultant electrode voltage and thus might even lead to the formulation of theories that can guide electrode discovery. With the Tree SHAP, global and local interpretability of physicochemical features contribution is simulated. From the Tree SHAP feature importance plot (Fig. 4A), the material's polarizability (Pol^{range}) and average atomic radius (AR) are found to be the most contributing features. From the results in Fig. 4B, average values (~ 20) of the Pol^{range} positively impact the resultant voltage and negatively impact the voltage with lower values (< 20). This trend generally indicates that for materials having an overall combination of lower/higher (AR) and higher/lower (Pol^{range}) values, a higher/lower voltage is to be expected. Similar to the trend that can be corroborated in the $CaCoPO_3$ and $Al(CoS_2)_2$ materials (Figs. 4C and 4D).

Further on, we plot the Tree SHAP Pol^{range} and AR dependency plots to observe if there is possible intractability between them. According to Fig. 5A, the Pol^{range} and AR exhibit a complex interdependency, which can be attributed to the inherent relationship between atomic radius and polarizability. From a classical mechanics viewpoint, the dependency

can be described using the equation:[74]

$$r = \sqrt{\alpha \frac{(I - A)}{2}} \quad (1)$$

where r is the atomic radius, α is the polarizability, I is the ionization potential, and A is the electron affinity. In the periodic table, a consistent trend exists in the elements' polarizability, electronegativity, atomic radius, and electron affinity. As atoms' atomic radius and polarizability decrease, their electronegativity and ionization energy increase. Therefore, elements' atomic radius and polarizability tend to have an inverse relationship with their electronegativity and electron affinity. These physicochemical features dictate the nature of bonding and, thus, materials' overall electronic properties, as described by the Slater rules and Phillips ionicity theory.[75–78]

The output voltage of an electrode is intricately linked to its atomic constituents, which influences the underlying chemical mechanisms, such as polarizability and atomic radius, as seen in Eq. (1). Polarizability, a measure of electron cloud distortion in response to external electric fields, facilitates charge transfer between electrode and electrolyte.[79] This is a fundamental aspect of battery operation, when ions migrate during discharge and charge cycles, leading to charge imbalances within the material and the creation of an electric field. The degree of polarization is determined by factors like ionic conductivity and resistance, which are influenced by intrinsic material properties and the

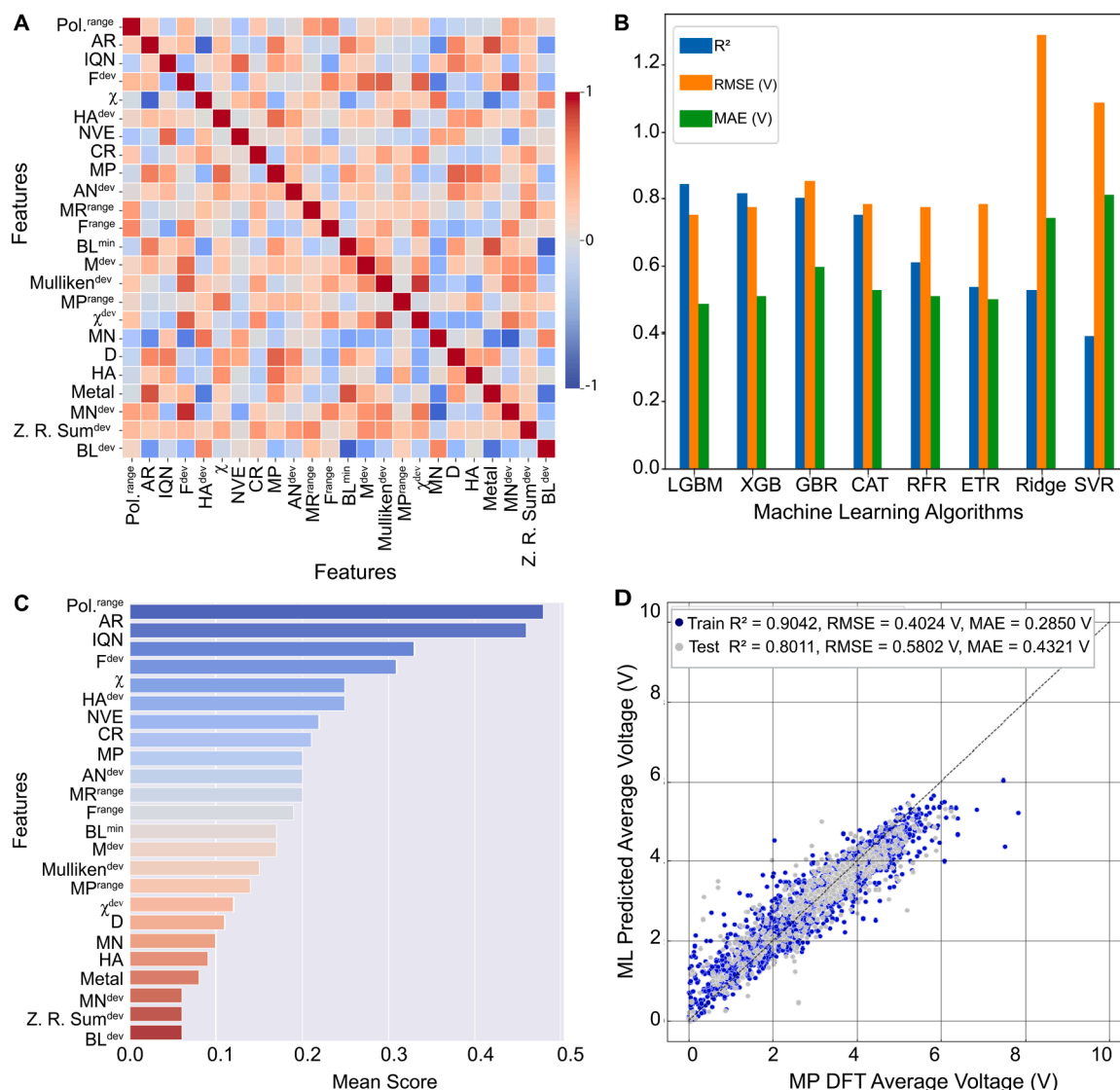


Fig. 3. ML Modeling of the Initial MP Electrode Database (A) Pearson correlation between the best-selected features, after dropping features with low mutual information (MI) scores, and those that are highly correlated ($r > 0.95$). (B) Comparison of various classical ML algorithms' performance on the dataset. (C) The contribution score of the physiochemical features ranked by taking the mean feature importance score from the Light Gradient Boosting Machine (LGBM), Extreme Gradient Boosting Machine (XGB), Gradient Boosting Regressor (GBR), CatBoost regressor (CAT), Random Forest regressor (RFR), Extra Trees regressors (ETR), Kernel Ridge (Ridge), and Support vector regressor (SVR). (D) The parity plot of the MP-DFT-calculated voltage and the LGBM model predicted the voltage after optimizing the model parameters using Optuna.

electrode-electrolyte interface. Materials with higher polarizability exhibit enhanced stability in maintaining charge distributions during the charge/discharge processes, which can ultimately lead to a more consistent output voltage.[80,81] From Figs. 4C and 4D, a trend in which materials with high polarizability can host enhanced voltages. Additionally, atomic radius plays a crucial role in the output voltage, especially in intercalation-type batteries such as lithium-ion batteries. The atomic size of the host material directly impacts its ability to accommodate ions to enable them to store more charge, thereby yielding a higher output voltage. Thus, optimizing electrode materials involves a delicate balance between factors like atomic radius and stability, which can be deduced from the values of SHAP results (Figs. 4C and 4D). These chemical properties are central to rational design and performance optimization of electrode materials.

Consequently, we reaffirm this observation using the uniform manifold approximation and projection (UMAP), an unsupervised clustering model. UMAP helps to identify the local trend in a physiochemical space of material. The physiochemical feature spaces were

reduced to 2-dimensional input space for the UMAP algorithm. The results from the UMAP clustering model are in Fig. 5B, with the voltage of the electrodes projected. In comparison, we also project the electrode's average atomic radius and range polarizability (Figs. 5C and 5D). A trend can be observed in which materials with average voltage from 2 to 6.5 V have an average atomic radius of ≤ 1 Å and range polarizability of ≤ 30 . This trend corroborates that of the Tree SHAP interpretability results.

2.5. Exploration dataset of potential electrode materials

At first, we curated a new dataset of 102,816 materials from the materials project database (Fig. 1E). We screened and selected materials hosting the active metals (Li, Na, Mg, Al, Ca, Zn, and K) to curate potential electrode materials not containing radioactive elements. Also, considering the role of thermodynamic stability in electrodes, only materials with the formation energy and energy above the hull of ≤ 0 are considered, and a total of 3656 potential electrode materials were finally

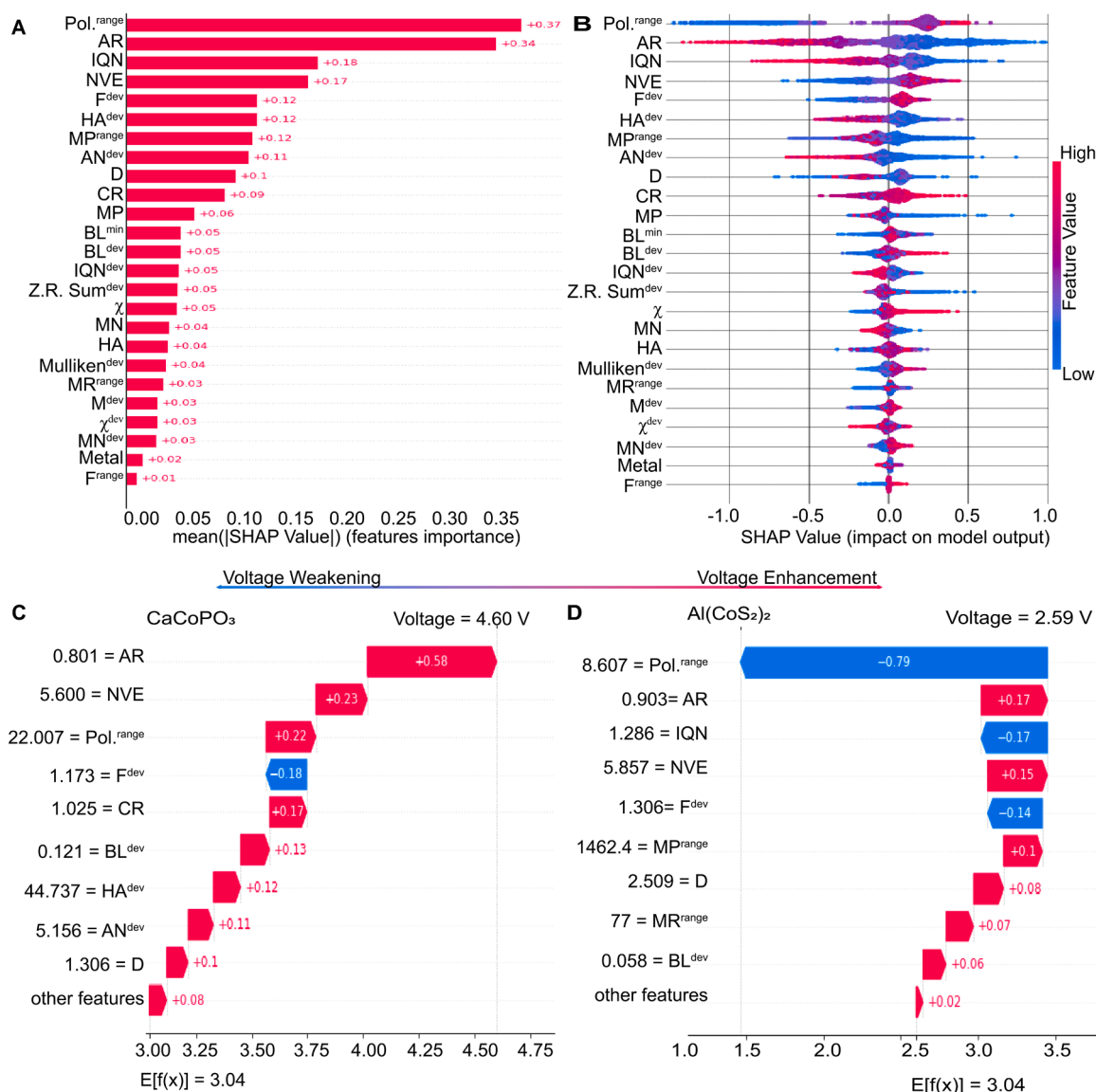


Fig. 4. Interpretability of the ML Model (A) Tree SHAP global interpretability of feature's importance, ranking them among the most significant. (B) The model's output impacts the Tree SHAP global interpretability of the material's elemental/structural features. Tree SHAP local interpretability of the feature's performance on voltage prediction of (C) CaCoPO_3 and (D) $\text{Al(CoS}_2)_2$.

obtained (Fig. 1F).

2.6. Active learning workflow

Based on the previously acquired knowledge of how electrode physiochemical features influence voltage from our interpretable ML model, we then curate an informative feature space to be explored by BO in the active learning (AL) module to efficiently suggest potent electrode materials from the exploration database while minimizing DFT computational costs.

Fig. 6A shows the performance of various functions in optimizing and suggesting the most potent physiochemical features point with a target voltage of 7.5 V. It can be noted that the mixed acquisition function, which uses the maximum uncertainty (MU) acquisition function to reduce the overall uncertainty of the model and the maximum expected improvement (MEI) acquisition function to optimize the voltage prediction, is the only function to identify the material with the target voltage in fewer steps compared to other functions.

As described elsewhere,[69–71] Gaussian process regression (GPR) models' predictive performance and accuracy depend on the feature

space size. We used the feature importance ranking from the interpretable ML model to test the GPR performance on the top 20, 15, and 10 feature spaces (Figure S2). The 10-dimensional feature space is more accurate (Fig. 6B). Therefore, we adopted an informative 10-dimensional physiochemical feature space in the AL module.

Further on, based on the above, we adopt and build an AL module using the mixed acquisition function in the BO and an informative 10-dimensional physiochemical feature space in the exploration dataset. Based on this module, we queried the new exploration dataset of 3656 materials for potent electrode materials (Fig. 1E). Ab initio-before the BO process, we randomly selected five materials in the dataset and labeled their corresponding voltages (obtained through DFT calculations) to be used in training the surrogate model. Then, based on the predicted mean and uncertainties, the as-chosen mixed acquisition function optimizes the objective function to suggest the most potent material for evaluation. The suggested point is then evaluated using DFT calculations, and then the materials labeled (the voltage) back into the exploration dataset to improve the model for the next BO iteration.

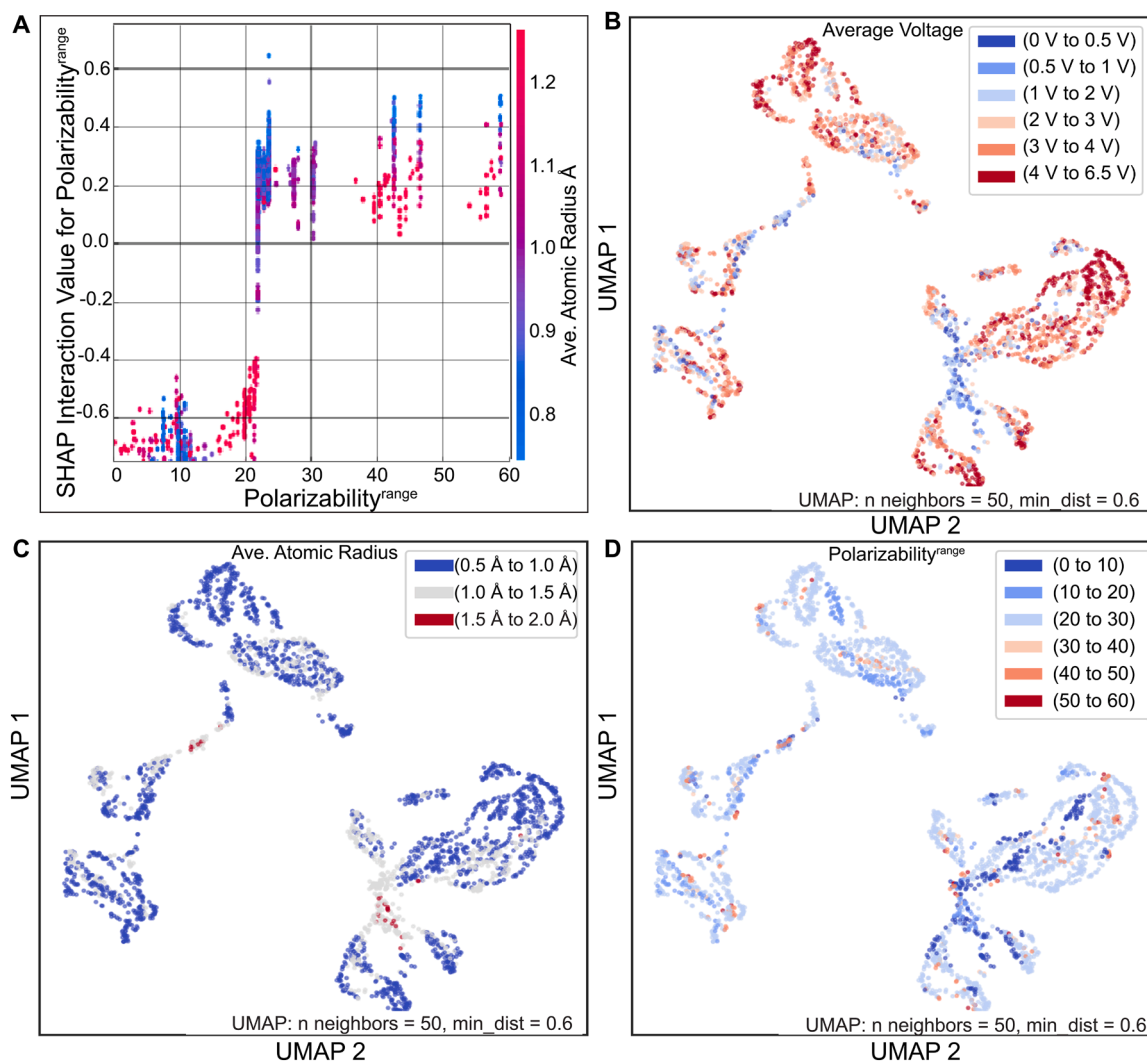


Fig. 5. SHAP Interpretability and Unsupervised ML Clustering Results of the MP Database (A) SHAP dependence plot of $Pol. range$ and AR, showing the interaction in-between. Unsupervised ML clustering of the initial MP database using UMAP (B) with projected voltages. (C) with the average atomic radius. (D) with the projected range polarizability.

2.7. The active learning module results

To evaluate the reliability of our machine learning model's predictions and DFT computed voltages, we conducted a comparative analysis between the computed voltages of the machine learning predicted materials and the corresponding experimental results from literature for some of the predicted electrode materials, including $NaCuSO_4F$, $CaFeO_3$, $KCoPO_4$, $Na_3Fe(SO_4)_3$, and $Li_2Mn_3CoO_8$. The comparison revealed varying degrees of agreement between the predicted and experimental voltages for these materials. Specifically, for $NaCuSO_4F$ (an experimental voltage of 2–5 V and 0–3 V as a cathode and anode, respectively), and $CaFeO_3$ (4.07 V), $KCoPO_4$ (4.02 V), $Na_3Fe(SO_4)_3$ (4.70 V), and $Li_2Mn_3CoO_8$ (4.18 V) [82–86]. All the calculated voltages closely matched the experimental values, demonstrating a solid alignment. In contrast, $KCoPO_4$ showed a noticeable discrepancy between the predicted, computed, and experimental voltages. However, the computed (3 V) values are within the merging errors of the experimental values, indicating that the model may benefit from further refinement for this material.

Further on, to understand the underlying mechanism of the material's voltage, we plot the observed materials with their corresponding average atomic radius and range polarizability (Fig. 6C). Most materials with voltage >4.5 V host a combination of atoms with mostly low ionic

radius and high electronegativity. Notably, this can be attributed to the materials' highly electronegative atoms. A mechanism described by Brent et al. [87] and Goodenough et al. [88,89] shows that electronegativity is a vital factor in the voltage of electrodes. This mechanism involves the ionic-covalent bond dynamics driven by an inductive effect during the active-metal ion intercalation, in which the anions (or polyanions) transfer charges to the transition metal. The transferred charges reduce the covalency between the transition metal and other atoms. Therefore, it enhances the stability and energy level of the transition-metal redox couple and, in essence, increases the electrodes' potential. [90,91] Hence, having/introducing a high electronegativity anion in an electrode material can substantially enhance the voltage of an electrode material.

Notably, compounds with highly electronegative atoms (like F and O) tend to suffer from low electronic conductivity due to their wide band gap caused by the highly ionic bond formation with other metals. On the other hand, the ionic nature tends to make the materials good ionic conductors. A balance of the ionic and electronic conductivity in electrodes is needed to allow the diffusion of the active-metal ion while aiding electron motion. [92] Hence, we calculated the materials' band gap and electronic conductivity. As shown in Fig. 6D, most materials host a sizeable bandgap and an averagely good electronic conductivity comparable to some electrodes. [93]

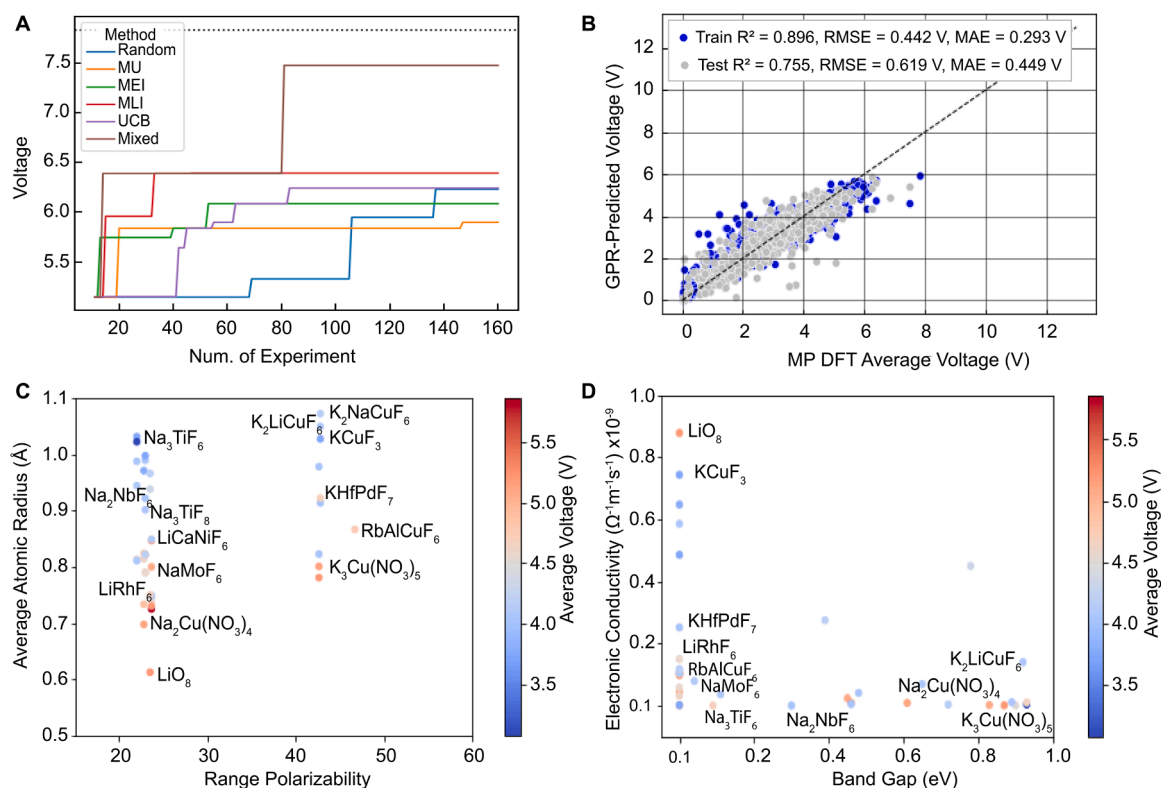


Fig. 6. Performance of the AL Model on the New Database and the Properties of the Predicted Best Electrode Materials (A) Performance of the acquisition functions: Maximum uncertainty (MU), maximum expected improvement (MEI), maximum likelihood improvement (MLI), upper confidence bound (UCB), and mixed (MU and MEI) in the MP dataset. (B) Gaussian progress regression performance on the materials project dataset, with a 10-dimensional feature space. (C) The average atomic radius and range polarizability for electrode materials were predicted through the AL loop. (D) The calculated electronic conductivity and band gap of the predicted electrode materials.

3. Conclusion

The accuracy of AL in the explorations of a new database to predict new materials generally depends on the initial labeled space's quality, dimensionality, and how informative the exploration features space is to the target property. Through an interpretable machine learning model, we could narrow down to 24 out of the 324 electrode materials' physiochemical features significantly influencing the voltage. Further, we adopted a 10-dimensional informative physiochemical feature space to explore a new dataset of materials, using an AL protocol to query the best potent material for electrode applications. The final predicted potent electrode material's physiochemical properties were analyzed using ML, chemistry, and materials science knowledge. From the analysis, we could deduce and observe some physiochemical traits that most of our predicted potent electrode materials manifest. (a) The presence of highly electronegative anions in the materials. That can be associated with enhancing the voltages of electrodes through inductive charge transfer. (b) The materials are narrow bandgap semiconductors with sizable bandgaps and relatively good electronic conductivity. In this work, we leverage and integrate machine learning approaches with domain knowledge of chemistry and material sciences to discover new potent electrode materials and understand the influence of their physiochemical properties on voltage.

4. Experimental procedures

4.1. Machine learning

The ML model was created to predict electrode voltage using the materials project battery electrodes dataset based on features created using Oliynyk featurization.[72] All the ML models were based on

algorithms implemented in Scikit-learn;[94] the PyCaret package was to rank various ML algorithms' performance. The hyperparameter optimization of the Light gradient boost ML algorithm was carried out using the Optuna package. The classical ML model was interpreted using the SHAP package.

To generate an optimal ML model, a multi-step approach is carried out from feature engineering (Fig. 3A and 3C), suitable ML algorithms selection (Fig. 3B) to hyperparameters tuning (Figure S1) of the selected ML algorithm for model generation (Fig. 3D). Features engineering is carried out to determine optimal features, dropping the redundant and highly correlated features. These can effectively improve the model performance and prevent overfitting. The individual feature's mutual information scores (MI) contribution to the voltage was calculated (Figure S3). Uninformative with deficient MI scores were eliminated, indicating a weak contribution of information to the target. Furthermore, using Pearson's correlation coefficient (r) as a criterion, if two features have a high correlation (i.e., $|r| > 0.95$ in this study), one of the features is discarded. The final selected optimal features with their Pearson correlation are shown in Fig. 3A and listed in Table S1.

The Pearson correlation coefficient is defined as

$$r = \frac{\text{cov}(x_i, x_j)}{\sigma_{x_i} \sigma_{x_j}} \quad (2)$$

where $\text{cov}(x_i, x_j)$ and $\sigma_{x_{i/j}}$ are the covariance of features x_i and x_j and standard deviation of the feature $x_{i/j}$, respectively.

Furthermore, utilizing the PyCaret package, 8 ML algorithms: Light Gradient Boost Machine (LGBM), Extreme Gradient boosting (XGB), Gradient boosting regressor (GBR), CatBoost regressor (Cat), Random Forest regressor (RFR), Extra trees regressors (ETR), Kernel ridge (Ridge), and Support vector regressor (SVR) were tested on our datasets

to find the most suitable algorithm.

Consequently, after selecting a suitable ML algorithm - the LGBM, its optimal hyperparameters should be generated for a reliable ML model. Therefore, we carried out a Bayesian hyperparameter optimization of the LGBM model based on the Optuna optimization framework. Hyperparameter optimization was conducted in >100 random trials using the RMSE test as the objective function. The hyperparameter optimization contour plot, as shown in **Figure S1**, indicates that the maximum number of trials yielding the lowest RMSE has a learning rate of around 0.1 for several leaves between 20-to-40, feature fraction (0.4 to 1) and a bagging fraction (0.4 to 1). Furthermore, the optimized hyperparameters were used to create the best ML model with the lowest possible RMSE and the highest R^2 described elsewhere.[95,96] An accurate LGBM regressor model capable of predicting the voltages of materials (using our feature set) is obtained with an MAE of 0.2850 V (**Fig. 3D**). A split ratio of 80/20 was used throughout the study, where random 80 % of the data was allocated to the training set and 20 % to the training set.

4.2. Active learning

Bayesian Optimization (BO), a quintessential classical active AL module, has accelerated the discovery of materials of interest from large databases.[97–100] This scheme minimizes the number of experiments or high-throughput DFT calculations required to identify the most potent candidate from a stream of materials. In BO, the choice of acquisition functions determines the minimal number of experiments needed to find an optimal material. At the same time, BO utilizes the Gaussian process regression (GPR) as a surrogate to predict values and uncertainties while exploring the physiochemical feature space of materials. For the uncertainty quantification, the following acquisition functions: the maximum uncertainty (MU), maximum expected improvement (MEI), maximum likelihood improvement (MLI), upper confidence bound (UCB), and mixed (MU and MEI) were tested on the materials project battery electrode datasets.

The AL model is created based on BO using Gaussian process regression in the algorithm to estimate the uncertainty and functional relationship between the material features and voltage.[94] The following acquisition function: maximum uncertainty (MU), maximum expected improvement (MEI), maximum likelihood improvement (MLI), and upper confidence boundary (UCB) were tested. A mixed acquisition method of MU and MIE was utilized in this work. In the mixed acquisition method, MU is utilized to reduce the uncertainty of the model, while the MEI is used to optimize the voltage of the material. The functions are defined as follows;

$$\text{MU}, x^* = \sigma[f(x_i)] \quad (3)$$

$$\text{MEI}, x^* = E[f(x_i)] \quad (4)$$

$$\text{MLI}, x^* = \frac{E[f(x_i)] - E[f(x_b)]}{\sigma[f(x_i)]} \quad (5)$$

$$\text{UCB}, x^* = [E[f(x_i)] + \sigma[f(x_i)]] \quad (6)$$

where x^* is the new sample obtained by the acquisition function, σ is the standard deviation, $f(x_i)$ is the model, x_i is the material's feature vector, and E is the expected value.

A total of 5 BO iterations were performed and stopped when materials satisfying the selection criterion could not be identified in the exploration dataset. Forty-one materials were extracted with voltages $\in [3.5 \text{ V}, 5.5 \text{ V}]$. The materials found after the final round of the selection are RbAlCuF_6 , NaMoF_6 , $\text{Li}(\text{NiO}_2)_2$, $\text{NaMg}_2(\text{SO}_4)_3$, NaCuSO_4F , Na_2NbF_6 , Na_2CuF_4 , $\text{Na}_2\text{Cu}(\text{NO}_3)_4$, LiRhF_6 , LiMnF_6 , LiCaNiF_6 , Li_2RuF_6 , $\text{KCo}(\text{PO}_3)_3$, NaMoF_6 , LiAgF_6 , CaRuO_3 , CaFeO_3 , K_2NaCuF_6 , CaP_2PdO_7 , LiCuF_4 , KCoPO_4 , KHfPdF_7 , $\text{Na}_3\text{Fe}(\text{SO}_4)_3$, $\text{Ca}(\text{AgO}_2)_2$, KCuF_3 , $\text{Na}(\text{NiO}_2)_2$, Na_3TiF_6 , $\text{NaMn}_2\text{P}_2\text{O}_9$, LiO_8 , KF_3 , LiIrF_6 , CaCrP_2O_7 , KMnPCO_7 , K_2LiCuF_6 ,

Na_3TiF_6 , $\text{Ca}_2\text{Co}_3\text{O}_8$, $\text{Li}_2\text{Mn}_3\text{CoO}_8$, $\text{NaZr}_2\text{PdF}_{11}$, $\text{K}_3\text{Cu}(\text{NO}_3)_5$, and $\text{LiMn}(\text{CO}_3)_2$. Interestingly, it is noteworthy that some of these materials, e.g., NaCuSO_4F , [86,101] CaFeO_3 , [85] KCoPO_4 , [84] $\text{Na}_3\text{Fe}(\text{SO}_4)_3$, [83] and $\text{Li}_2\text{Mn}_3\text{CoO}_8$, were reported in experimental works as suitable insertion electrodes for battery applications, in good agreement with the predictions of this work.

4.3. Density functional theory calculations

All the DFT calculations are performed using the Vienna *ab initio* simulation package (VASP). The generalized gradient approximation (GGA) in the form of Perdew-Burke-Ernzerhof (PBE) functional, hybrid functional calculation using the HSE06 functional, and the projected augmented wave (PAW) were employed. All calculations converged with the residual forces less than 0.001 eV/\AA , and the convergence threshold of the total energy was set to 10^{-6} eV . The voltage (V) of the material is calculated from the total energy under different active metal-ion (Li, Na, Mg, Al, Ca, Zn, and K) concentrations, and MX is the material:

$$V = \frac{E_{A_1MX} - E_{A_2MX} + (y_2 - y_1)E_A}{(y_2 - y_1)} \quad (7)$$

The dependency between the material band gap and its conductivity

From Eq. (8), calculating the material's band gap can also be used to derive their electronic conductivity (σ_e).

$$\sigma_e = \frac{A}{T^{3/2}} e^{E_g/2KT} \quad (8)$$

where A is the constant, T is the temperature, E_g is the band gap, and K is the Boltzmann constant.

The diffusivity D_i is a critical parameter that characterizes the mobility of charge carriers, including electrons, holes, or ions, within materials. A vital determinant of electrical conductivity, with higher values indicating better conductivity, is temperature-dependent due to the influence of thermal motion. Hence making it an essential factor in understanding the electrochemical properties of materials. Moreso, it provides insights into how charge carriers spread out within a material due to random scattering, complementing the concept of carrier drift under electric fields. [102,103] Also, as described by the Nernst-Einstein equation, [104] the ionic conductivity by the active metal ion is given by:

$$\sigma_i = n_i e^2 D_i / k_b T \quad (9)$$

Here, n_i is the concentration of the active metal ion, e is the elementary charge, k_b is the Boltzmann constant, T is the absolute temperature, and D_i is the diffusivity of the ion in the comparing results of the active learning predictions. From a semi-quantitative relationship, [105] atoms' polarizability and electronegative nature can be determined using Eq. (10).

$$\chi = 1.66(n/\alpha)^{1/3} + 0.37 \quad (10)$$

where χ is the electronegativity and α is the polarizability.

Author Contributions

M.L.A. and H. Z. conceived the idea and designed the project. M.L.A. performed the machine learning modeling, DFT calculations and prepared the draft. H.Z., H.L., C.H., and X.Y. acquired the funding and supervised the project. All the authors participated in discussing the results and revising the manuscript. All authors read and approved the final manuscript.

CRediT authorship contribution statement

Mukhtar Lawan Adam: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation,

Visualization, Writing – original draft, Writing – review & editing. **Oyawale Adetunji Moses**: Writing – review & editing. **Jonathan Pradana Mailoa**: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Chang-Yu Hsieh**: Funding acquisition, Project administration, Supervision, Writing – review & editing. **Xue-Feng Yu**: Funding acquisition, Supervision. **Hao Li**: Funding acquisition, Project administration, Supervision, Writing – review & editing. **Haitao Zhao**: Formal analysis, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare no competing financial interests.

Data availability

Data will be made available on request.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (no. 52173234), Natural Science Foundation of Guangdong Province (2022A1515010554), Shenzhen Science and Technology Program (ZDSYS20220606100606013), Shenzhen Science and Technology Program (JCY20210324102008023), Shenzhen-Hong Kong-Macau Technology Research Program (Type C, SGDX2020110309300301), Shenzhen Excellent Science and Technology Innovation Talent Training Project - Outstanding Youth Project (RCJC20200714114435061), CCF-Tencent Open Fund, and the Iwatani Naoji Foundation.

References

- [1] C. Liu, F. Li, L.-P. Ma, H.-M. Cheng, Advanced materials for energy storage, *Adv. Mater.* 22 (2010) E28–E62, <https://doi.org/10.1002/adma.200903328>.
- [2] X. Li, Z. Huang, C.E. Shuck, G. Liang, Y. Gogotsi, C. Zhi, MXene chemistry, electrochemistry and energy storage applications, *Nat Rev Chem* 6 (2022) 389–404, <https://doi.org/10.1038/s41570-022-00384-8>.
- [3] Z. Yang, H. Huang, F. Lin, Sustainable electric vehicle batteries for a sustainable world: perspectives on battery cathodes, environment, supply chain, manufacturing, life cycle, and policy, *Adv. Energy Mater.* 12 (2022), 2200383, <https://doi.org/10.1002/aenm.202200383>.
- [4] X. Zeng, M. Li, D. Abd El-Hady, W. Alshihri, A.S. Al-Bogami, J. Lu, K. Amine, Commercialization of lithium battery technologies for electric vehicles, *Adv. Energy Mater.* 9 (2019), 1900161, <https://doi.org/10.1002/aenm.201900161>.
- [5] J.B. Goodenough, K.-S. Park, The li-ion rechargeable battery: a perspective, *J. Am. Chem. Soc.* 135 (2013) 1167–1176, <https://doi.org/10.1021/ja3091438>.
- [6] D.P. Dubal, O. Ayyad, V. Ruiz, P. Gómez-Romero, Hybrid energy storage: the merging of battery and supercapacitor chemistries, *Chem. Soc. Rev.* 44 (2015) 1777–1790, <https://doi.org/10.1039/C4CS00266K>.
- [7] W. Li, B. Song, A. Manthiram, High-voltage positive electrode materials for lithium-ion batteries, *Chem. Soc. Rev.* 46 (2017) 3006–3059, <https://doi.org/10.1039/C6CS00875E>.
- [8] G. Bridge, E. Faigen, Towards the lithium-ion battery production network: thinking beyond mineral supply chains, *Energy Res. Soc. Sci.* 89 (2022), 102659, <https://doi.org/10.1016/j.erss.2022.102659>.
- [9] A.-M. Desautry, D. Monfort Climent, G. Lefebvre, A. Cristiano-Tassi, D. Peralta, S. Perret, A. Urban, C. Guerrot, Tracing the origin of lithium in Li-ion batteries using lithium isotopes, *Nat. Commun.* 13 (2022) 4172, <https://doi.org/10.1038/s41467-022-31850-y>.
- [10] C.B. Tabein, J. Dallas, S. Casanova, T. Pelech, G. Bournival, S. Saydam, I. Canbulat, Towards a low-carbon society: a review of lithium resource availability, challenges and innovations in mining, extraction and recycling, and future perspectives, *Miner. Eng.* 163 (2021), 106743, <https://doi.org/10.1016/j.mineng.2020.106743>.
- [11] P. Greim, A.A. Solomon, C. Breyer, Assessment of lithium criticality in the global energy transition and addressing policy gaps in transportation, *Nat. Commun.* 11 (2020) 4570, <https://doi.org/10.1038/s41467-020-18402-y>.
- [12] Y. Luo, H. Wei, L. Tang, Y. Huang, Z. Wang, Z. He, C. Yan, J. Mao, K. Dai, J. Zheng, Nickel-rich and cobalt-free layered oxide cathode materials for lithium ion batteries, *Energy Storage Mater.* 50 (2022) 274–307, <https://doi.org/10.1016/j.ensm.2022.05.019>.
- [13] B. Chu, Y.-J. Guo, J.-L. Shi, Y.-X. Yin, T. Huang, H. Su, A. Yu, Y.-G. Guo, Y. Li, Cobalt in high-energy-density layered cathode materials for lithium ion batteries, *J. Power Sources* 544 (2022), 231873, <https://doi.org/10.1016/j.jpowsour.2022.231873>.
- [14] T.M. Bandhauer, S. Garimella, T.F. Fuller, A critical review of thermal issues in lithium-ion batteries, *J. Electrochem. Soc.* 158 (2011) R1, <https://doi.org/10.1149/1.3515880>.
- [15] S. Ponnada, M.S. Kiai, R. Krishnapriya, R. Singhal, R.K. Sharma, Lithium-free batteries: needs and challenges, *Energy Fuels* 36 (2022) 6013–6026, <https://doi.org/10.1021/acs.energyfuels.2c00569>.
- [16] J. Liu, Z. Huang, J. Sun, Q. Wang, Heat generation and thermal runaway of lithium-ion battery induced by slight overcharging cycling, *J. Power Sources* 526 (2022), 231136, <https://doi.org/10.1016/j.jpowsour.2022.231136>.
- [17] Q. Wang, P. Ping, X. Zhao, G. Chu, J. Sun, C. Chen, Thermal runaway caused fire and explosion of lithium ion battery, *J. Power Sources* 208 (2012) 210–224, <https://doi.org/10.1016/j.jpowsour.2012.02.038>.
- [18] K. Liu, Y. Liu, D. Lin, A. Pei, Y. Cui, Materials for lithium-ion battery safety, *Sci. Adv.* 4 (2018), <https://doi.org/10.1126/sciadv.aas9820>.
- [19] C. Eames, M.S. Islam, Ion intercalation into two-dimensional transition-metal carbides: global screening for new high-capacity battery materials, *J. Am. Chem. Soc.* 136 (2014) 16270–16276, <https://doi.org/10.1021/ja508154e>.
- [20] P. Bhauriyal, A. Mahata, B. Pathak, Hexagonal BC₃ electrode for a high-voltage al-ion battery, *J. Phys. Chem. C* 121 (2017) 9748–9756, <https://doi.org/10.1021/acs.jpcc.7b02290>.
- [21] K. Kubota, M. Dahbi, T. Hosaka, S. Kumakura, S. Komaba, Towards K-Ion and Na-ion batteries as "beyond li-ion", *Chem. Record* 18 (2018) 459–479, <https://doi.org/10.1002/tcr.201700057>.
- [22] P.K. Nayak, L. Yang, W. Brehm, P. Adelhelm, Von Lithium- zu natriumionenbatterien: vorteile, herausforderungen und überraschendes, *Angewandte Chem.* 130 (2018) 106–126, <https://doi.org/10.1002/ange.201703772>.
- [23] K.-Y. Zhang, Z.-Y. Gu, E.H. Ang, J.-Z. Guo, X.-T. Wang, Y. Wang, X.-L. Wu, Advanced polyanionic electrode materials for potassium-ion batteries: progresses, challenges and application prospects, *Mater. Today* 54 (2022) 189–201, <https://doi.org/10.1016/j.mattod.2022.02.013>.
- [24] P. Hu, Z. Zou, X. Sun, D. Wang, J. Ma, Q. Kong, D. Xiao, L. Gu, X. Zhou, J. Zhao, et al., Uncovering the potential of M1-site-activated nasicon cathodes for Zn-Ion batteries, *Adv. Mater.* 32 (2020), 1907526, <https://doi.org/10.1002/adma.201907526>.
- [25] M. Song, H. Tan, D. Chao, H.J. Fan, Recent advances in zn-ion batteries, *Adv. Funct. Mater.* 28 (2018), 1802564, <https://doi.org/10.1002/adfm.201802564>.
- [26] J. Shin, D.S. Choi, H.J. Lee, Y. Jung, J.W. Choi, Hydrated intercalation for high-performance aqueous zinc ion batteries, *Adv. Energy Mater.* 9 (2019), 1900083, <https://doi.org/10.1002/aenm.201900083>.
- [27] J. Miranda, E. le Calvez, R. Retoux, O. Crosnier, T. Brousse, Revisiting Rb₂TiNb₆O₁₈ as electrode materials for energy storage devices, *Electrochem. commun.* 137 (2022), 107249, <https://doi.org/10.1016/j.elecom.2022.107249>.
- [28] N. Fu, Y.-T. Xu, S. Zhang, Q. Deng, J. Liu, C.-J. Zhou, X.-W. Wu, Y.-G. Guo, X.-X. Zeng, Electrode materials for aqueous multivalent metal-ion batteries: current status and future prospect, *J. Energy Chem.* 67 (2022) 563–584, <https://doi.org/10.1016/j.jechem.2021.08.057>.
- [29] X. Zhang, T. Xiong, B. He, S. Feng, X. Wang, L. Wei, L. Mai, Recent advances and perspectives in aqueous potassium-ion batteries, *Energy Environ. Sci.* 15 (2022) 3750–3774, <https://doi.org/10.1039/D2EE01573K>.
- [30] Y. Tian, G. Zeng, A. Rutt, T. Shi, H. Kim, J. Wang, J. Koettgen, Y. Sun, B. Ouyang, T. Chen, et al., Promises and challenges of next-generation "beyond li-ion" batteries for electric vehicles and grid decarbonization, *Chem. Rev.* 121 (2021) 1623–1669, <https://doi.org/10.1021/acs.chemrev.0c00767>.
- [31] M.D. Slater, D. Kim, E. Lee, C.S. Johnson, Sodium-ion batteries, *Adv. Funct. Mater.* 23 (2013) 947–958, <https://doi.org/10.1002/adfm.201200691>.
- [32] H. Cui, L. Ma, Z. Huang, Z. Chen, C. Zhi, Organic materials-based cathode for zinc ion battery, *SmartMat* (2022), <https://doi.org/10.1002/smm2.1110>.
- [33] A.L. Lipson, B. Pan, S.H. Lapidus, C. Liao, J.T. Vaughey, B.J. Ingram, Rechargeable Ca-Ion batteries: a new energy storage system, *Chem. Mater.* 27 (2015) 8442–8447, <https://doi.org/10.1021/acs.chemmater.5b04027>.
- [34] Y. Yang, J. Zhou, L. Wang, Z. Jiao, M. Xiao, Q. Huang, M. Liu, Q. Shao, X. Sun, J. Zhang, Prussian blue and its analogues as cathode materials for Na-, K-, Mg-, Ca-, Zn- and Al-ion batteries, *Nano Energy* 99 (2022), 107424, <https://doi.org/10.1016/j.nanoen.2022.107424>.
- [35] L. Peng, Y. Zhu, D. Chen, R.S. Ruoff, G. Yu, Two-dimensional materials for beyond-lithium-ion batteries, *Adv. Energy Mater.* 6 (2016), 1600025, <https://doi.org/10.1002/aenm.201600025>.
- [36] F. Yang, X. Feng, P.-A. Glans, J. Guo, MoS₂ for beyond lithium-ion batteries, *APL Mater.* 9 (2021), 050903, <https://doi.org/10.1063/5.0050118>.
- [37] Y. Zhou, Q. Han, Y. Liu, Y. Wang, F. Jiang, N. Wang, Z. Bai, S. Dou, Molybdenum chalcogenides based anode materials for alkali metal ions batteries: beyond lithium ion batteries, *Energy Storage Mater.* 50 (2022) 308–333, <https://doi.org/10.1016/j.ensm.2022.05.031>.
- [38] J. Ni, A. Dai, Y. Yuan, L. Li, J. Lu, Three-dimensional microbatteries beyond Lithium Ion, *Matter* 2 (2020) 1366–1376, <https://doi.org/10.1016/j.matt.2020.04.020>.
- [39] R.J. Gummow, G. Vamvounis, M.B. Kannan, Y. He, Calcium-Ion Batteries: current State-of-the-Art and future perspectives, *Adv. Mater.* 30 (2018), 1801702, <https://doi.org/10.1002/adma.201801702>.
- [40] J. Biemolt, P. Jungbaker, T. van Teijlingen, N. Yan, G. Rothenberg, Beyond lithium-based batteries, *Materials (Basel)* 13 (2020) 425, <https://doi.org/10.3390/ma13020425>.
- [41] S. Chen, D. Zhao, L. Chen, G. Liu, Y. Ding, Y. Cao, Z. Chen, Emerging intercalation cathode materials for multivalent metal-ion batteries: status and challenges, *Small Struct.* 2 (2021), 2100082, <https://doi.org/10.1002/ssr.202100082>.

- [42] A.J.Y. Wong, X. Zhou, Y. Lum, Z. Yao, Y.C. Chua, Y. Wen, Z.W. Seh, Battery materials discovery and smart grid management using machine learning, *Batter. Supercaps* 5 (2022), <https://doi.org/10.1002/batt.202200309>.
- [43] Z. Wang, Z. Sun, H. Yin, X. Liu, J. Wang, H. Zhao, C.H. Pang, T. Wu, S. Li, Z. Yin, et al., Data-Driven Materials Innovation and Applications. Preprint at, John Wiley and Sons Inc, 2022, <https://doi.org/10.1002/adma.202104113>.
- [44] C. Liu, Z.G. Neale, G. Cao, Understanding electrochemical potentials of cathode materials in rechargeable batteries, *Mater. Today* 19 (2016) 109–123, <https://doi.org/10.1016/j.mattod.2015.10.009>.
- [45] R. Tian, S.-H. Park, P.J. King, G. Cunningham, J. Coelho, V. Nicolosi, J. N. Coleman, Quantifying the factors limiting rate performance in battery electrodes, *Nat. Commun.* 10 (2019) 1933, <https://doi.org/10.1038/s41467-019-09792-9>.
- [46] M. Sotoudeh, A. Groß, Descriptor and scaling relations for ion mobility in crystalline solids, *JACS Au* 2 (2022) 463–471, <https://doi.org/10.1021/jacsau.1c00505>.
- [47] T. Zhang, D. Li, Z. Tao, J. Chen, Understanding electrode materials of rechargeable lithium batteries via DFT calculations, *Prog. Nat. Sci.: Mater. Int.* 23 (2013) 256–272, <https://doi.org/10.1016/j.pnsci.2013.04.005>.
- [48] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, et al., Commentary: the materials project: a materials genome approach to accelerating materials innovation, *APL Mater.* 1 (2013), 011002, <https://doi.org/10.1063/1.4812323>.
- [49] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, et al., AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations, *Comput. Mater. Sci.* 58 (2012) 227–235, <https://doi.org/10.1016/j.compmatsci.2012.02.002>.
- [50] J. Jie, M. Weng, S. Li, D. Chen, S. Li, W. Xiao, J. Zheng, F. Pan, L. Wang, A new MaterialGo database and its comparison with other high-throughput electronic structure databases for their predicted energy band gaps, *Sci. China Technol. Sci.* 62 (2019) 1423–1430, <https://doi.org/10.1007/s11431-019-9514-5>.
- [51] A. Belsky, M. Hellenbrandt, V.L. Karen, P. Luksch, New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design, *Acta Crystallogr. B* 58 (2002) 364–369, <https://doi.org/10.1107/S0108768102006948>.
- [52] C.R. Groom, I.J. Bruno, M.P. Lightfoot, S.C. Ward, The cambridge structural database, *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater.* 72 (2016) 171–179, <https://doi.org/10.1107/S2052520616003954>.
- [53] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD), *JOM* 65 (2013) 1501–1509, <https://doi.org/10.1007/s11837-013-0755-4>.
- [54] H. Zhao, W. Chen, H. Huang, Z. Sun, Z. Chen, L. Wu, B. Zhang, F. Lai, Z. Wang, M. L. Adam, et al., A robotic platform for the synthesis of colloidal nanocrystals, *Nat. Synth.* 2 (2023) 505–514, <https://doi.org/10.1038/s44160-023-00250-5>.
- [55] R.P. Carvalho, C.F.N. Marchiori, D. Brandell, C.M. Araujo, Artificial intelligence driven in-silico discovery of novel organic lithium-ion battery cathodes, *Energy Storage Mater.* 44 (2022) 313–325, <https://doi.org/10.1016/j.ensm.2021.10.029>.
- [56] Y.-C. Gao, N. Yao, X. Chen, L. Yu, R. Zhang, Q. Zhang, Data-Driven insight into the reductive stability of ion-solvent complexes in lithium battery electrolytes, *J. Am. Chem. Soc.* (2023), [10.1021/jacs.3c08346](https://doi.org/10.1021/jacs.3c08346).
- [57] X. Chen, X. Liu, X. Shen, Q. Zhang, Applying Machine Learning to Rechargeable Batteries: From the Microscale to the Macroscale, Preprint at John Wiley and Sons Inc, 2021, <https://doi.org/10.1002/anie.202107369>.
- [58] X. Zhang, Z. Wang, A.M. Lawan, J. Wang, C.Y. Hsieh, C. Duan, C.H. Pang, P. K. Chu, X.F. Yu, H. Zhao, Data-driven structural descriptor for predicting platinum-based alloys as oxygen reduction electrocatalysts, *InfoMat* 5 (2023), <https://doi.org/10.1002/inf2.12406>.
- [59] F. Oviedo, J.L. Ferres, T. Buonassisi, K.T. Butler, Interpretable and explainable machine learning for materials science and chemistry, *Acc. Mater. Res.* 3 (2022) 597–607, <https://doi.org/10.1021/accounts.1c00244>.
- [60] Dean, J., Scheffler, M., Purcell, T.A.R., Barabash, S.V., Bhowmik, R., and Bazhiron, T. (2021). Interpretable machine learning for materials design.
- [61] J.A. Esterhuizen, B.R. Goldsmith, S. Linic, Interpretable machine learning for knowledge generation in heterogeneous catalysis, *Nat. Catal.* 5 (2022) 175–184, <https://doi.org/10.1038/s41929-022-00744-z>.
- [62] S. Axelrod, D. Schwalbe-Koda, S. Mohapatra, J. Damewood, K.P. Greenman, R. Gómez-Bombarelli, Learning Matter: materials design with machine learning and atomistic simulations, *Acc. Mater. Res.* 3 (2022) 343–357, [10.1021/accounts.1c00238](https://doi.org/10.1021/accounts.1c00238).
- [63] A.E.A. Allen, A. Tkatchenko, Machine learning of material properties: predictive and interpretable multilinear models, *Sci. Adv.* 8 (2022) 7185, <https://doi.org/10.1126/sciadv.abm7185>.
- [64] R. Dybowski, Interpretable machine learning as a tool for scientific discovery in chemistry, *N. J. Chem.* 44 (2020) 20914–20920, <https://doi.org/10.1039/D0NJ02592E>.
- [65] I.A. Moses, R.P. Joshi, B. Ozdemir, N. Kumar, J. Eickholt, V. Barone, Machine learning screening of metal-ion battery electrode materials, *ACS Appl. Mater. Interfaces* 13 (2021) 53355–53362, [10.1021/acsami.1c04627](https://doi.org/10.1021/acsami.1c04627).
- [66] S.-Y. Louis, E.M.D. Siriwardane, R.P. Joshi, S.S. Omei, N. Kumar, J. Hu, Accurate prediction of voltage of battery electrode materials using attention-based graph neural networks, *ACS Appl. Mater. Interfaces* 14 (2022) 26587–26594, [10.1021/acsami.2c00029](https://doi.org/10.1021/acsami.2c00029).
- [67] R.P. Joshi, J. Eickholt, L. Li, M. Fornari, V. Barone, J.E. Peralta, Machine learning the voltage of electrode materials in metal-ion batteries, *ACS Appl. Mater. Interfaces* 11 (2019) 18494–18503, <https://doi.org/10.1021/acsami.9b04933>.
- [68] O. Allam, B.W. Cho, K.C. Kim, S.S. Jang, Application of DFT-based machine learning for developing molecular electrode materials in Li-ion batteries, *RSC Adv.* 8 (2018) 39414–39420, <https://doi.org/10.1039/C8RA07112H>.
- [69] E. Siivola, A. Paleyes, J. González, A. Veltari, Good practices for bayesian optimization of high dimensional structured spaces, *Appl. AI Lett.* 2 (2021), <https://doi.org/10.1002/ail2.24>.
- [70] Q. Liang, A.E. Gongora, Z. Ren, A. Tiihonen, Z. Liu, S. Sun, J.R. Deneault, D. Bash, F. Mekki-Berrada, S.A. Khan, et al., Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains, *NPJ Comput. Mater.* 7 (2021), <https://doi.org/10.1038/s41524-021-00656-9>.
- [71] C. Gambella, B. Ghaddar, J. Naoum-Sawaya, Optimization problems for machine learning: a survey, *Eur. J. Oper. Res.* 290 (2021) 807–828, <https://doi.org/10.1016/j.ejor.2020.08.045>.
- [72] A.O. Olynyk, E. Antono, T.D. Sparks, L. Ghadbeigi, M.W. Gaultois, B. Meredig, A. Mar, High-throughput machine-learning-driven synthesis of full-heusler compounds, *Chem. Mater.* 28 (2016) 7324–7331, [10.1021/acs.chemmater.6b02724](https://doi.org/10.1021/acs.chemmater.6b02724).
- [73] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.* 2 (2020) 56–67, <https://doi.org/10.1038/s42256-019-0138-9>.
- [74] P. Szarek, W. Grochala, Most probable distance between the nucleus and homo electron: the latent meaning of atomic radius from the product of chemical hardness and polarizability, *J. Phys. Chem. A* 118 (2014) 10281–10287, <https://doi.org/10.1021/jp507423p>.
- [75] K. Li, Y. Li, D. Xue, Band gap engineering of crystal materials: band gap estimation of semiconductors via electronegativity, *Funct. Mater. Lett.* (2012), <https://doi.org/10.1142/S1793604712600028>.
- [76] M. Rahm, P. Erhart, R. Cammi, Relating atomic energy, radius and electronegativity through compression, *Chem. Sci.* 12 (2021) 2397–2403, <https://doi.org/10.1039/d0sc06675c>.
- [77] J. Liang, X. Zhu, Phillips-inspired machine learning for band gap and exciton binding energy prediction, *J. Phys. Chem. Lett.* 10 (2019) 5640–5646, [10.1021/acs.jpclett.9b02232](https://doi.org/10.1021/acs.jpclett.9b02232).
- [78] J.C. Phillips, Ionicity of the chemical bond in crystals, *Rev. Mod. Phys.* 42 (1970) 317–356, <https://doi.org/10.1103/RevModPhys.42.317>.
- [79] Y.A. Budkov, S.V. Zavarzin, A.L. Kolesnikov, Theory of ionic liquids with polarizable ions on a charged electrode, *J. Phys. Chem. C* 125 (2021) 21151–21159, [10.1021/acs.jpcc.1c05548](https://doi.org/10.1021/acs.jpcc.1c05548).
- [80] J.G. McDaniel, A. Yethiraj, Influence of electronic polarization on the structure of ionic liquids, *J. Phys. Chem. Lett.* 9 (2018) 4765–4770, [10.1021/acs.jpclett.8b02120](https://doi.org/10.1021/acs.jpclett.8b02120).
- [81] D. Frydel, Polarizable poisson-boltzmann equation: the study of polarizability effects on the structure of a double layer, *J. Chem. Phys.* 134 (2011), <https://doi.org/10.1063/1.3598476>.
- [82] P. Strobel, A.I. Palos, M. Anne, F. Le Cras, Materials Structural, magnetic and lithium insertion properties of spinel-type $\text{Li}_2\text{Mn}_3\text{MO}_8$ oxides (M=Mg, Co, Ni, Cu), *J. Mater. Chem.* 10 (2020) 429–436, <https://doi.org/10.1039/A905962H>.
- [83] P.H. Salame, Synthesis and electrical studies of $\text{Na}_3\text{Fe}(\text{SO}_4)_3$ cathode material for sodium ion batteries, in: AIP Conference Proceedings, American Institute of Physics Inc, 2019, <https://doi.org/10.1063/1.5113459>.
- [84] N. Priyadarsini, S. Surendran, B. Senthikumar, L. Vasylychko, R.K. Selvan, Synthesis and electrochemical performances of $\gamma\text{-KCoPO}_4$ nanocrystals as promising electrode for aqueous supercapacitors, *ChemElectroChem* 6 (2019) 369–377, <https://doi.org/10.1002/celec.201801440>.
- [85] M. Hibino, R. Harimoto, Y. Ogasawara, R. Kido, A. Sugahara, T. Kudo, E. Tochigi, N. Shibata, Y. Ikahara, N. Mizuno, A new rechargeable sodium battery utilizing reversible topotactic oxygen extraction/insertion of CaFeO_z ($2.5 \leq z \leq 3$) in an organic electrolyte, *J. Am. Chem. Soc.* 136 (2014) 488–494, <https://doi.org/10.1021/ja411365z>.
- [86] M. Reynaud, P. Barpanda, G. Rouse, J.N. Chotard, B.C. Melot, N. Recham, J. M. Tarascon, Synthesis and crystal chemistry of the NaMSO_4F family (M = Mg, Fe, Co, Cu, Zn), *Solid State Sci.* 14 (2012) 15–20, <https://doi.org/10.1016/j.solidstatesciences.2011.09.004>.
- [87] B.C. Melot, J.M. Tarascon, Design and preparation of materials for advanced electrochemical storage, *Acc. Chem. Res.* 46 (2013) 1226–1238, <https://doi.org/10.1021/ar300088q>.
- [88] A.K. Padhi, K.S. Nanjundaswamy, J.B. Goodenough, Phospho-olivines as positive-electrode materials for rechargeable lithium batteries, *J. Electrochem. Soc.* 144 (1997) 1188–1194, <https://doi.org/10.1149/1.1837571>.
- [89] A.K. Padhi, K.S. Nanjundaswamy, C. Masquelier, S. Okada, J.B. Goodenough, Effect of Structure on the $\text{Fe}_3 + \text{Fe}_2 + \text{Redox Couple}$ in Iron Phosphates, *J. Electrochem. Soc.* 144 (1997) 1609–1613, <https://doi.org/10.1149/1.1837649>.
- [90] M.E.A.y de Dompablo, U. Amador, J.-M. Tarascon, A computational investigation on fluorinated-polyanionic compounds as positive electrode for lithium batteries, *J. Power Sources* 174 (2007) 1251–1257, [10.1016/j.jpowsour.2007.06.178](https://doi.org/10.1016/j.jpowsour.2007.06.178).
- [91] R.C. Massé, C. Liu, Y. Li, L. Mai, G. Cao, Energy storage through intercalation reactions: electrodes for rechargeable batteries, *Natl. Sci. Rev.* 4 (2017) 26–53, [10.1093/nsr/nw093](https://doi.org/10.1093/nsr/nw093).
- [92] M. Greaves, S. Barg, M.A. Bissett, MXene-based anodes for metal-ion batteries, *Batter. Supercaps* 3 (2020) 214–235, <https://doi.org/10.1002/batt.201900165>.

- [93] X. Cui, K. Tuo, Y. Xie, C. Li, D. Zhao, L. Yang, X. Fu, S. Li, Investigation on electrochemical performance at the low temperature of LFP/C-P composite based on phosphorus doping carbon network, *Ionics* 26 (2020) 3795–3808, <https://doi.org/10.1007/s11581-020-03567-9>.
- [94] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, A. Mueller, Scikit-learn, *GetMobile: Mob. Comput. Commun.* 19 (2015) 29–33, <https://doi.org/10.1145/2786984.2786995>.
- [95] M. Mukherjee, S. Satsangi, A.K. Singh, A Statistical approach for the rapid prediction of electron relaxation time using elemental representatives, *Chem. Mater.* 32 (2020) 6507–6514. [10.1021/acs.chemmater.0c01778](https://doi.org/10.1021/acs.chemmater.0c01778).
- [96] R. Kumar, A.K. Singh, Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts, *NPJ Comput. Mater.* 7 (2021) 197, <https://doi.org/10.1038/s41524-021-00669-4>.
- [97] R. Jaleem, K. Kanamori, I. Takeuchi, M. Nakayama, H. Yamasaki, T. Saito, Bayesian-driven first-principles calculations for accelerating exploration of fast ion conductors for rechargeable battery application, *Sci. Rep.* 8 (2018) 5845, <https://doi.org/10.1038/s41598-018-23852-y>.
- [98] Y. Sheng, Y. Wu, J. Yang, W. Lu, P. Villars, W. Zhang, Active learning for the power factor prediction in diamond-like thermoelectric materials, *NPJ Comput. Mater.* 6 (2020) 171, <https://doi.org/10.1038/s41524-020-00439-8>.
- [99] Y.K. Wakabayashi, T. Otsuka, Y. Krockenberger, H. Sawada, Y. Taniyasu, H. Yamamoto, Bayesian optimization with experimental failure for high-throughput materials growth, *NPJ Comput. Mater.* 8 (2022) 180, <https://doi.org/10.1038/s41524-022-00859-8>.
- [100] L. Bassman, P. Rajak, R.K. Kalia, A. Nakano, F. Sha, J. Sun, D.J. Singh, M. Aykol, P. Huck, K. Persson, et al., Active learning for accelerated design of layered materials, *NPJ Comput. Mater.* 4 (2018) 74, <https://doi.org/10.1038/s41524-018-0129-0>.
- [101] P. Barpanda, Sulfate Chemistry For High-Voltage Insertion materials: Synthetic, Structural and Electrochemical Insights, Preprint at Wiley-VCH Verlag, 2015, <https://doi.org/10.1002/ijch.201400157>.
- [102] N. Yao, X. Chen, Z.H. Fu, Q. Zhang, Applying Classical, Ab Initio, and Machine-Learning Molecular Dynamics Simulations to the Liquid Electrolyte For Rechargeable Batteries, Preprint at American Chemical Society, 2022, <https://doi.org/10.1021/acs.chemrev.1c00904>.
- [103] A.D. Sendek, E.D. Cubuk, E.R. Antoniak, G. Cheon, Y. Cui, E.J. Reed, Machine learning-assisted discovery of Solid Li-Ion conducting materials, *Chem. Mater.* 31 (2019) 342–352. [10.1021/acs.chemmater.8b03272](https://doi.org/10.1021/acs.chemmater.8b03272).
- [104] A. France-Lanord, J.C. Grossman, Correlations from Ion Pairing and the nernst-einstein equation, *Phys. Rev. Lett.* 122 (2019), 136001, <https://doi.org/10.1103/PhysRevLett.122.136001>.
- [105] S.A. Blair, A.J. Thakkar, Relating polarizability to volume, ionization energy, electronegativity, hardness, moments of momentum, and other molecular properties, *J. Chem. Phys.* 141 (2014), 074306, <https://doi.org/10.1063/1.4893178>.