ACS APPLIED MATERIALS & INTERFACES

www.acsami.org

Forum Article

Machine Learning Screening of Metal-Ion Battery Electrode Materials

Isaiah A. Moses, Rajendra P. Joshi, Burak Ozdemir, Neeraj Kumar, Jesse Eickholt, and Veronica Barone*

Cite This: https://doi.org/10.1021/acsami.1c04627

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Rechargeable batteries provide crucial energy storage systems for renewable energy sources, as well as consumer electronics and electrical vehicles. There are a number of important parameters that determine the suitability of electrode materials for battery applications, such as the average voltage and the maximum specific capacity which contribute to the overall energy density. Another important performance criterion for battery electrode materials is their volume change upon charging and discharging, which contributes to determine the cyclability, Coulombic efficiency, and safety of a battery. In this work, we present deep neural network regression machine learning models (ML), trained on data obtained from the Materials Project database, for predicting average voltages and volume change upon charging and discharging of electrode materials for metal-ion batteries. Our models exhibit good performance as measured by the average mean absolute error obtained from a 10-fold cross-validation, as well as on independent test sets. We further assess the robustness of our ML models by investigating their screening potential beyond the training database. We produce Na-ion electrodes by systematically replacing Li-ions in the original database by Na-ions and, then, selecting a set of 22 electrodes that exhibit a good performance in energy density, as well as small volume variations upon charging and discharging, as predicted by the machine learning model. The ML predictions for these materials are then compared to quantum-mechanics based calculations. Our results reaffirm the significant role of machine learning techniques in the exploration of materials for battery applications.

KEYWORDS: *metal-ion batteries, machine learning, deep learning, deep neural networks, electrode voltage, electrode volume change*

INTRODUCTION

The quest for environmentally friendly energy sources has resulted in a growing interest in the search of novel materials for sustainable energy.[1] Renewable energy sources are mostly intermittent, and hence, there is a need for efficient energy storage systems. Rechargeable batteries provide appropriate storage technologies for renewable energy sources and are also heavily used in consumer electronics and electrical vehicles.[2,3]

A leading technology in the battery market is the Li-ion battery (LIB). These batteries provide high efficiencies, high energy densities and long cycle life.[4,5] However, the limited abundance of Li and the associated high cost of LIBs is a challenge, especially for large scale applications.[6−8] Alternative battery technologies to meet current and future energy storage needs are being explored. Two other alkali metals, Na and K, have relatively low ionic radius and are abundant in the earth's crust.[9] These properties have sparked research in appropriate electrode materials for efficient Na- or K-ion batteries.[10−12] Unfortunately, materials that work well for Li ions do not necessarily work for Na and K,[13] which makes finding suitable electrode materials for these larger ions a challenging task. The exploration of the chemical space for new materials using quantum mechanical methods is computationally demanding and time-consuming. However, over the past decade, robust databases have been made available to the scientific community. These databases, generated over time partly from combinatorial experiments[14,15] and, majorly, from high-throughput calculations that utilize quantum mechanical methods, such as density functional theory (DFT),[16] combined with improved ML algorithms, are accelerating the discovery of materials for energy conversion and storage.[17−19]

ML applications to materials science range from the exploration of microscopic properties of materials, such as the band structure,[20−22] density of states,[23] and formation energy,[24] to applications in specific areas that include solar cells,[25,26] catalysis,[27,28] and batteries.[29−36] Regarding battery technologies, a number of studies have been reported on the

Special Issue: Artificial Intelligence/Machine Learning for Design and Development of Applied Materials

Received: March 11, 2021
Accepted: June 8, 2021

use of ML to identify promising candidates as electrolytes for LIBs.[30−36] Relatively fewer studies have been reported on electrodes.[17,37] These reports include the prediction of three types of crystal symmetries (monoclinic, orthorhombic, and triclinic) of cathode materials with Li−Si−(Mn, Fe, Co)−O composition,[38] machine learning potential to predict voltage of graphite-based LIBs,[37] a utilization of ridge regression model by Eremin et al.[39] to predict the energy of $LiNiO_2$ and $LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$ as cathode materials and a combination of DFT and ML to predict redox potentials of carbon-based electrode materials.[40] Recently, we presented a more general study on the voltage of electrode materials in metal-ion batteries[41] in which we show that deep learning models can be used to predict the voltages of diverse electrode materials with reasonable accuracy.

Important metrics to measure battery performance include specific capacity, voltage, energy density, thermal stability, Coulombic efficiency, safety, cyclability, electrical conductivity of electrodes, and lifetime.[5,42] Many of these metrics, including the Coulombic efficiency, as well as the cyclability and safety, largely depend on the volume change of the electrode upon charging and discharging.[43,44] So far, the volume change of electrode materials has been studied only through either demanding DFT simulations or experiments. In this work, we present the first machine learning model to quantify the volume change in electrode materials upon charging and discharging. The model is obtained employing regression based deep neural networks (DNN). In addition, we extend and improve our previous work[41] for predicting the voltages of electrode materials for metal-ion batteries by including a larger and more balanced training data set. Using this ML approach, we propose top candidates with small changes in volume and large energy density as potential cathode materials for Na-ion batteries. These materials are further validated through first-principles simulations.

Data and Methods. *Machine Learning.* The data for training our models was retrieved from the Materials Project (MP) database[16,45] using the Materials Application Programming Interface (MAPI)[46] and the Python Materials Genomics (*pymatgen*)[47] material analysis package. We extracted nearly 4860 unique instances of computed data for training our ML models from the MP database. The distribution of the data set, based on the working intercalation ion, is shown in Figure 1. For each instance in the database, we gathered the chemical formula of the electrodes with low and high concentration of the working ion, the working ion (Li, Na, K, Rb, Cs, Mg, Ca, Al, Zn, or Y), the type of the electrode (either intercalation or conversion), the Bravais lattice type, the space group, the average voltage (V_{av}), and the percentage change in the volume of the electrodes with low and high concentration of working ions ($\Delta V_{\%}$).

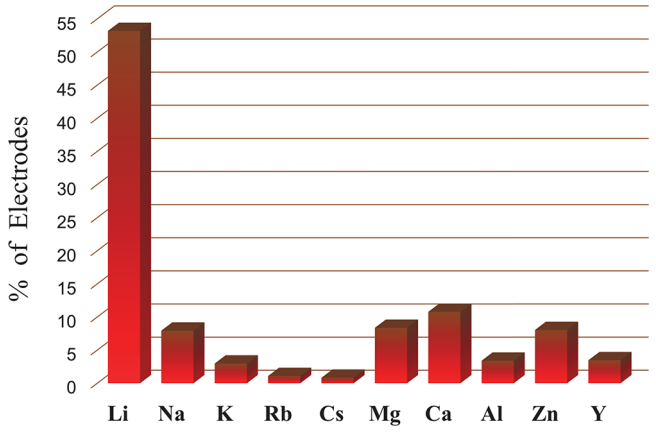

Figure 1. Distribution of instances in the training data set based on the active metal ion.

By utilizing the stoichiometries of both charged and discharged electrodes and the *matminer* software[48] we generated additional features to form a set of 306 features to uniquely represent each reaction in our ML models. These additional atomic features are derived from the structural and chemical compositions of the electrodes and have been previously used to predict several properties of crystalline materials, including voltages.[41,49] A more detailed discussion on featurization is presented in Supporting Information. Once features are generated, feature-wise normalization is carried out such that the value of each feature ranges from −1 to 1 to make learning easier for our neural network.[50] To avoid the dimensionality curse associated with a large dimension of the feature vector, we use principal component analysis (PCA) to optimize the total number of features.[51] This analysis shrinks the dimension of the feature vector by 65%. We assessed the performance of our models with and without PCA optimization and we found that while model performance for $\Delta V_{\%}$ is not significantly affected by PCA, the mean absolute errors for the V_{av} models are about 20% smaller when PCA is used. Therefore, all results presented in this work are obtained employing the PCA optimized dimension of the feature vector.

Our previous work on voltages of battery electrodes, trained with about 22% less data than the present work, has shown that the deep learning algorithm performed significantly better than shallow learning.[41] Therefore, in this work, we have used regression based deep neural network (DNN) as our working model. We train models with three different sets of data: (i) on the entire data set consisting of electrode materials for 10 different active metal-ions (labeled All), (ii) on all alkali-ion based electrode materials (labeled Alkali), and (iii) on only Li-ion based electrode materials (labeled Li). We utilize the *Keras* deep learning library[50] to build our models where hyperparameters are optimized to find the optimal model architecture. The architecture that exhibits the best performance consists of an input layer, three hidden layers and an output layer. For the three different data sets (All, Alkali, and Li) and for both targets, V_{av} and $\Delta V_{\%}$, the number of nodes for the input layer is 106 as obtained by PCA, while the number of nodes for the first, second and third hidden layers is 360, 90, and 30, respectively.

We used the *relu* activation function and the *adam* optimizer in our models.[10] Batch normalization was applied to the V_{av} models, while dropout[52] and weight decay[53] were utilized to combat overfitting in both V_{av} and $\Delta V_{\%}$ models. The mean squared error is used as the loss function while the performance of each model is measured employing the mean absolute error (MAE).

Density Functional Theory. To assess the performance of our ML models, we carried out additional DFT calculations using the Quantum-Espresso software.[54,55] Projector-augmented wave pseudopotentials (PAW) are used to replace core electrons and the generalized gradient approximation exchange-correlation density functional for solids by Perdew, Burke, and Ernzerhof (PBEsol) with Hubbard U correction is employed in all calculations[56,57] and dubbed DFT+U

https://doi.org/10.1021/acsami.1c04627
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

Table 1. Distribution of MAE for V_{av} (in V) for Each of the 10 Rounds (R_i) in Models Trained with the Entire Data Set (All), the Alkali-Only Data Set (Alkali), and the Li-Only Data Set (Li)[a]

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	SD	$\overline{MAE}$	MAE_T
All	0.42	0.39	0.38	0.38	0.43	0.39	0.38	0.39	0.39	0.37	0.02	0.39	0.35
Alkali	0.37	0.38	0.43	0.41	0.36	0.39	0.36	0.41	0.40	0.39	0.02	0.39	0.45
Li	0.42	0.43	0.42	0.43	0.38	0.44	0.42	0.40	0.41	0.38	0.02	0.41	0.43

[a]The standard deviation (SD), the average MAE across the 10 rounds ($\overline{MAE}$), and the MAE for the final model in the test set (MAE_T) for each training set (All, Alkali, and Li) are shown in the last three columns, respectively.

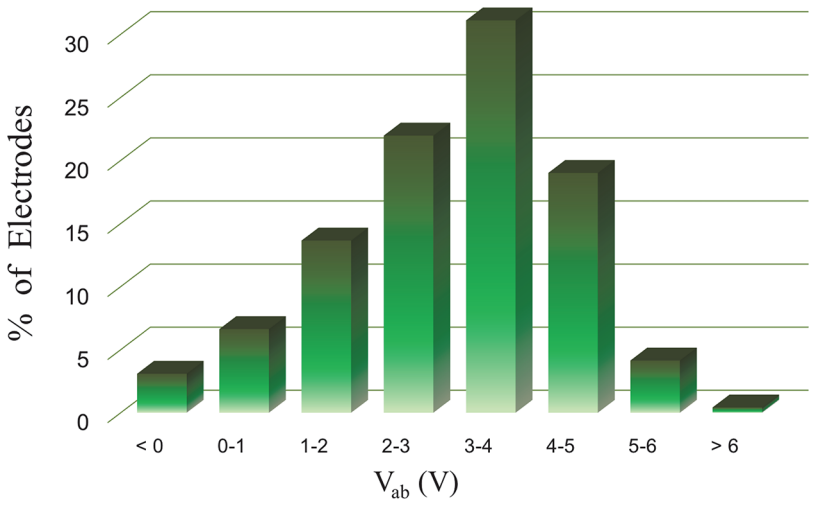

Figure 2. Distribution of the average voltage, V_{av}, in the entire data set.

Table 2. Distribution of MAE for V_{av} (in V) for Each of the 10 Rounds (R_i) in Models Trained after Removing the Underrepresented Instances, with the Entire Data Set (All), the Alkali-Only Data Set (Alkali), and the Li-Only Data Set (Li)[a]

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	SD	$\overline{MAE}$	MAE_T
All	0.36	0.39	0.41	0.37	0.35	0.36	0.34	0.36	0.35	0.36	0.02	0.37	0.37
Alkali	0.38	0.40	0.42	0.46	0.38	0.38	0.38	0.40	0.42	0.36	0.03	0.40	0.38
Li	0.37	0.40	0.38	0.40	0.36	0.40	0.40	0.40	0.33	0.33	0.03	0.38	0.38

[a]The standard deviation (SD), the average MAE across the 10 rounds ($\overline{MAE}$), and the MAE for the final model in the test set (MAE_T) for each training set (All, Alkali, and Li) are shown in the last three columns, respectively.

throughout this work. In order to compare ML results with DFT+U, we utilized the same U values documented in MP[58] from which our training data was retrieved. The value of the kinetic energy cutoffs for the wave functions and charge densities are 140 and 1400 Ry, respectively. Appropriate Monkhorst−Pack[59] meshes are used depending on the structure by ensuring that the total energy is converged to 1 mRy. Structural optimizations were carried out with convergence thresholds in the total energy, the atomic forces, and the pressure of 10^{-4} Ry, 10^{-3} Ry/Bohr, and 0.5 kbar, respectively.

RESULTS AND DISCUSSION

Average Voltage Models. We begin our analysis by targeting the average voltage, V_{av}. Different models were built, trained, and assessed independently utilizing the three different sets of data: All, Alkali, and Li. In each data set, the data is randomly divided into two parts, with 90% of instances used for training and validation (training set), and 10% of instances saved exclusively for testing final models (test set). Constrained by the relatively small amount of training data, we employed a 10-fold cross-validation procedure on the training set in order to guide the hyperparameter tuning as well as to better assess the predictive capacity of our models. In addition to hyperparameter tuning, batch normalization was observed to improve the performance of the V_{av} models. The values of the MAEs for the 10 different rounds in the 10-fold cross-validation are shown in Table 1, where R_i (with i = 1, 2, 3, ..., 10) represents the different rounds in the cross-validation (additional details presented in Supporting Information). For the model trained on the All data set (Table 1-all), we obtain a distribution of MAEs in the range of 0.37−0.43 V among the 10 rounds of cross-validation with an average MAE ($\overline{MAE}$) of 0.39 V. Similar results are obtained for the models trained on the Alkali and Li training sets (Table 1-Alkali, Li). With the optimum hyperparameters obtained from the cross-validation, we build the final models (for each training data set, All, Alkali, and Li) using the corresponding entire training set (90% of the training data). These final models are then tested in the corresponding test sets held specifically for this purpose. The MAE results for the final models (MAE_T) are 0.35, 0.45, and 0.43 V for the All, Alkali, and Li data sets, respectively.

We further investigated how the performance of the models for V_{av} is impacted by the presence of statistical underrepresented instances which are defined as those instances in the training set with V_{av} values that are either negative or too large and are scarcely represented in the data set. These

https://doi.org/10.1021/acsami.1c04627
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

Table 3. Distribution of MAE for $\Delta V_{\%}$ for the 10 Rounds of Training (R_i) in Models Trained with the Entire Data Set (All), the Alkali-Only Data Set (Alkali), and the Li-Only Data Set (Li)[a]

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	SD	$\overline{\text{MAE}}$	MAE_T
All	8.6	9.4	16.7	11.0	7.4	49.4	6.7	56.3	16.8	49.5	19.1	23.2	10.6
Alkali	3.5	7.2	6.6	5.0	4.2	3.6	5.5	5.8	11.5	4.1	2.3	5.7	7.8
Li	3.9	5.4	2.5	6.0	5.0	2.6	2.1	2.2	2.6	7.0	1.7	3.9	3.8

[a]The standard deviation (SD), the average MAE across the 10 rounds ($\overline{\text{MAE}}$), and the MAE for the final model in the test set (MAE_T) for each training set (All, Alkali, and Li) are shown in the last three columns, respectively.

instances could impact the learning and, hence, the performance of our models. Negative V_{av} and those with values greater than 6 V constitute about 3.5% of the data (Figure 2). Considering these instances as underrepresented we retrained our models with the remaining data. Our results show no significant differences with the results shown in Table 2. However, even if significant differences were obtained by removing statistical underrepresented instances, we believe that including these instances is important for the predictive capacity of our models, especially when considering electrode materials with low voltages or those that will not work as battery electrodes ($V_{av} < 0$). The scarcity of data in these voltage ranges presents a problem for ML models such as, for instance, the poor performance recently reported for graphite based electrodes.[37]

Overall, our DNN models for V_{av} exhibit a moderate increase in accuracy compared to our previous work[41] in which we obtained a $\overline{\text{MAE}}$ of 0.43 and 0.47 V for an All data set (consisting of only 6 metal-ion electrodes) and a Li-only data set, respectively. There are about 22% more data instances in our database than in the one used previously, with most of the additions being active metals other than Li. Therefore, the slight improvement in performance of the current models could be attributed, in part, to the more robust data set used for training. However, in the case of the Li-only models, we obtain a $\overline{\text{MAE}}$ of 0.41 V compared to the 0.47 V obtained previously, despite using roughly the same data set. Therefore, the moderate improvement observed in this work can be attributed to the more robust architecture of the current DNN models that are deeper and wider than the ones employed in the previous work.[41]

An additional assessment of the performance of our V_{av} models is presented in Supporting Information Table S4.

Change in Volume Models. Utilizing the same optimum architecture employed in the V_{av} models, we set out to train ML models for predicting the percentage of their volume variation upon charging and discharging, $\Delta V_{\%}$. The performance of our models for $\Delta V_{\%}$ in each data set is presented in Table 3. For the model trained on the All data set, MAE values in each round vary in the range of 6.7% to 56.3% depending on the fold considered, with a $\overline{\text{MAE}}$ of 23% and a SD of 19% in the 10-fold set. For the model trained on alkali-ion based electrode materials, MAE values for $\Delta V_{\%}$ drop to the range of 3.6% to 11.5%, with a SD of 2.3% and a $\overline{\text{MAE}}$ of 5.7%. Further improvement is observed for the model trained on Li-only data where we obtain a $\overline{\text{MAE}}$ of 3.9%, a range of 5% among the 10 fold set. For the final model trained with all 90% of the data in each corresponding database (All, Alkali, and Li), the mean absolute errors on the corresponding holdout test set (MAE_T) are 10.6%, 7.8%, and 3.8% for the All, Alkali, and Li models, respectively.

While there are no significant variations in the performance of the models targeting V_{av} trained on the All, Alkali, and Li data sets, this trend contrasts sharply with the one obtained for the $\Delta V_{\%}$ models. The $\overline{\text{MAE}}$ and the SD observed for models trained on the All data set are too large for its use in predicting the $\Delta V_{\%}$ of new electrode materials. However, these differences in performance are to be expected if we consider the relatively small range in the values of V_{av} and the extremely large range of the $\Delta V_{\%}$ values. The distribution of the $\Delta V_{\%}$ in the All data set is shown in Figure 3. This distribution shows a heavily imbalanced data set toward small volume variations. The values of the $\Delta V_{\%}$ in our data set range from about 0% to as much as a staggering 26 000%. However, more than 80% of the data present small values of $\Delta V_{\%}$ of less than 10%. Just to understand the influence of the low frequency data on the performance of our models, we retrained our models after removing instances with $\Delta V_{\%}$ greater than 30%.

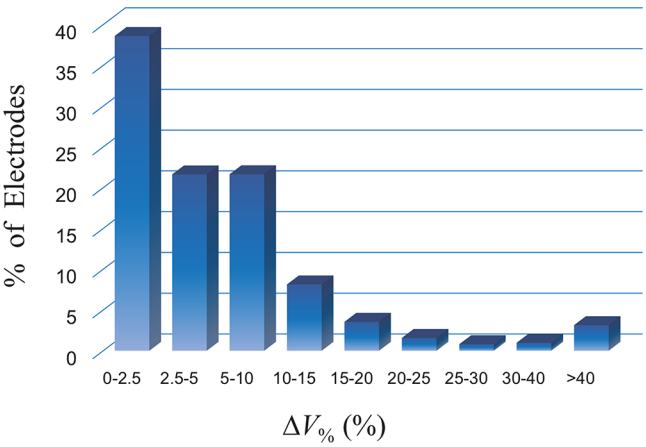

Figure 3. Distribution of $\Delta V_{\%}$ in the entire data set.

Not surprisingly, the $\overline{\text{MAE}}$ for models trained on the All, Alkali, and Li databases is significantly reduced to 2.7%, 2.4%, and 2.1%, respectively, accompanied by a substantial reduction in the corresponding SDs (Table 4). These results indicate that our models will perform best for materials that are not expected to expand or contract significantly upon charging and discharging. On the other hand, it is clear that more data is needed to expand current databases in the literature to account for materials with large volume changes upon charging or discharging in order to produce more robust ML models.

Li to Na Knowledge Transfer. To further assess the potential of the models built in the previous sections, we set out to investigate the robustness of these models by screening novel electrodes for Na-ion batteries. We are particularly interested in finding new Na-based electrodes with large values of voltages and large changes in active-ion stoichiometries to target high-energy density electrodes. We also considered candidate electrodes with minimal $\Delta V_{\%}$ on charging and discharging since, on one hand, small changes in volume

https://doi.org/10.1021/acsami.1c04627
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

Table 4. Distribution of MAE for $\Delta V_{\%}$ for the 10 Rounds of Training (R_i) in Models Trained after Removing Instances with $\Delta V_{\%}$ Greater than 30% with the Entire Data Set (All), the Alkali Only Data Set (Alkali), and the Li Only Data Set (Li)[a]

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	SD	$\overline{\text{MAE}}$	MAE_T
All	2.7	2.4	2.5	2.4	2.7	2.9	2.8	2.8	2.9	2.7	0.17	2.7	2.7
Alkali	2.2	2.6	2.1	2.6	2.5	2.3	2.5	2.4	2.5	2.4	0.15	2.4	2.2
Li	1.9	2.3	2.5	2.2	2.2	2.2	1.9	2.3	1.9	2.2	0.20	2.1	2.0

[a]The standard deviation (SD), the average MAE across the 10 rounds ($\overline{\text{MAE}}$), and the MAE for the final model in the test set (MAE_T) for each training set (All, Alkali, and Li) are shown in the last three columns, respectively.

Table 5. Absolute Deviation in $\Delta V_{\%}$ and V_{av} between Values Predicted by the ML Models Trained with the Alkali Data Set and First-Principles DFT+U Calculations[a]

	electrode		absolute deviation			
			$\Delta V_{\%}$		V_{av} (V)	
	charged	discharged	reduced	whole	reduced	whole
1	$NaMn_3OF_8$	$Na_4Mn_3OF_8$	10.0	14.6	0.43	0.12
2	$Na(CoO_2)_2$	Na_2CoO_2	30.0	19.1	2.31	0.72
3	$NaNbTe_2WO_{12}$	$Na_4NbTe_2WO_{12}$	11.7	6.8	0.40	0.31
4	$NaCuF_4$	Na_2CuF_4	10.1	10.6	0.02	0.19
5	Na_2CrO_4	Na_4CrO_4	1.4	2.6	2.12	1.38
6	$NaMn_2F_9$	$Na_3Mn_2F_9$	0.9	3.2	0.63	0.51
7	$Mn_3P_6WO_{24}$	$Na_6Mn_3P_6WO_{24}$	24.4	19.1	0.37	0.13
8	$CrWO_6$	Na_2CrWO_6	1.8	2.4	0.31	0.34
9	$NaNbF_6$	Na_3NbF_6	1.1	9.7	1.91	0.18
10	$NaVTe(WO_6)_2$	$Na_4VTe(WO_6)_2$	8.2	6.0	0.01	0.09
11	$NaSbF_6$	Na_3SbF_6	6.3	0.4	0.87	0.81
12	$NaVOF_{11}$	Na_4VOF_{11}	15.8	5.7	0.20	0.50
13	NaV_2O_4	Na_2VO_2	20.1	29.3	2.78	1.62
14	NaV_2OF_7	$Na_3V_2OF_7$	6.4	14.4	0.06	0.42
15	CrPOF	$Na_3Cr_2P_2(O_4F)_2$	18.9	11.5	0.67	1.05
16	NbO_2F	$NaNbO_2F$	2.4	1.6	0.47	0.84
17	$TiCrO_4$	Na_2TiCrO_4	12.1	17.5	0.72	0.53
18	$NaTiV_3O_{10}$	$Na_4TiV_3O_{10}$	16.6	7.3	0.52	0.04
19	Mn_2CrO_6	$Na_3Mn_2CrO_6$	5.3	3.8	0.53	0.16
20	$NaMn(PO_4)_2$	$Na_3Mn(PO_4)_2$	18.3	15.4	1.24	0.66
21	$NaMn_8O_{16}$	$Na_3Mn_4O_8$	8.3	14.6	2.12	0.07
22	VF_5	Na_2VF_5	1.8	1.0	0.43	1.32
MAE			**10.5**	**9.8**	**0.87**	**0.55**

[a]Reduced and whole represent the models trained on data excluding underrepresented instances and including underrepresented instances, respectively.

produce better cyclability of electrodes and, on the other hand, our models perform best for small volume expansions. We therefore replaced Li with Na in a wide range of Li-ion electrodes already present in our database and used our ML models trained on the Alkali data set to predict V_{av} and $\Delta V_{\%}$. We then selected those materials with large energy density and small volume variations. With the new electrodes identified, we performed additional DFT+U calculations to assess the performance of the ML models in electrodes outside the training set. Absolute deviations between our ML results in this set of 22 new materials are presented in Table 5 (individual DFT+U and ML results are provided in Supporting Information). As seen in Table 5, deviations between results from ML models and DFT+U calculations are somewhat larger than the $\overline{\text{MAE}}$ obtained for training. However, this trend is expected since we are testing in novel Na-ion electrodes outside the training data set. Additionally, we are assessing the performance of ML models in Na-ion battery electrodes which are severely underrepresented in our training data set (Figure 1).

We note that the results in Table 5 indicate that, on average, deviations obtained with the unfiltered data set (i.e., without removing underrepresented instances) performs slightly better for both targets, $\Delta V_{\%}$ and V_{av}, indicating the importance of including all instances in the training of ML models. Overall, we observe that while there is a clear need to significantly increase the number of instances with working ions other than Li through systematic high-throughput first-principles calculations, our current models provide a valuable tool for the initial screening of a large number of electrodes.

CONCLUSIONS

We have built deep neural network regression machine learning models for predicting the percentage change in volume ($\Delta V_{\%}$) and the average voltage (V_{av}) of battery electrodes upon charging and discharging with metal ions. Our models exhibit a good performance as measured by the $\overline{\text{MAE}}$ obtained by using 10-fold cross-validation, as well as the MAE_T obtained on the independent test sets. Further assessment is carried out by investigating the screening potential of ML models outside instances included in the database. Novel Na-

https://doi.org/10.1021/acsami.1c04627
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

ion electrodes are produced by systematically replacing Li-ions in the original database by Na-ions, and selecting those electrodes that exhibit good performance in energy density, as well as small volume variations upon charging and discharging, as predicted by ML models. ML results for the selected electrodes is then compared to DFT+U calculations. As expected, we find that on this new evaluation set, MAEs for V_{av} and $\Delta V_{\%}$ are relatively larger than the ones obtained in the 10-fold cross-validation. Nonetheless, the results are encouraging considering the negligible amount of time involved in the ML screening compared to the quantum-mechanics based methods. We also investigated the effect of removing underrepresented instances from the training database. We find that while removing these instances decreases the $\overline{\text{MAE}}$ in the 10-fold cross-validation, the predictive performance of the models in electrode materials outside the database diminishes when these instances are removed from the training. Therefore, the predictive capacity of these models is expected to improve further as more data is made available to the scientific community in public databases, especially if new data instances produce more balanced data sets for model generation.

In summary, our work underscores the fundamental role of deep learning in the exploration of materials for battery applications.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsami.1c04627.

Details of featurization, 10-fold cross-validation method, data instances used in the ML models, Comparison of experimental and ML average voltages, V_{av}, scatter plots showing the performance of ML models on the test sets, individual ML and DFT+U results for a set of 22 new Na-ion electrodes, and individual ML and DFT+U results for a set of 7 new Li-ion electrodes (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Veronica Barone − *Department of Physics, Central Michigan University, Mount Pleasant, Michigan 48859, United States; Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, Michigan 48859, United States;* orcid.org/0000-0002-5881-3640; Email: v.barone@cmich.edu

Authors

Isaiah A. Moses − *Science of Advanced Materials Program, Central Michigan University, Mount Pleasant, Michigan 48859, United States*

Rajendra P. Joshi − *Pacific Northwest National Laboratory, Richland, Washington 99352, United States*

Burak Ozdemir − *Department of Physics, Faculty of Science, University of Ostrava, 70103 Ostrava, Czech Republic;* orcid.org/0000-0001-7265-4490

Neeraj Kumar − *Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* orcid.org/0000-0001-6713-2129

Jesse Eickholt − *Department of Computer Science, Central Michigan University, Mount Pleasant, Michigan 48859, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsami.1c04627

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

V.B. and I.A.M. acknowledge the computational resources and services provided by the Institute for Cyber-Enabled Research at Michigan State University. N.K. and R.P.J. acknowledge the Laboratory Directed Research and Development Program at the Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy under Contract DE-AC06- 76RLO.

■ REFERENCES

(1) Dresselhaus, M.; Thomas, I. Alternative Energy Technologies. *Nature* **2001**, *414*, 332−337.

(2) Dunn, B.; Kamath, H.; Tarascon, J.-M. Electrical Energy Storage for the Grid: A Battery of Choices. *Science* **2011**, *334*, 928−935.

(3) Cheng, F.; Liang, J.; Tao, Z.; Chen, J. Functional Materials for Rechargeable Batteries. *Adv. Mater.* **2011**, *23*, 1695−1715.

(4) Tarascon, J.-M.; Armand, M. *Materials for Sustainable Energy: A Collection of Peer-reviewed Research and Review Articles from Nature Publishing Group*; World Scientific, 2011; pp 171−179.

(5) Nitta, N.; Wu, F.; Lee, J. T.; Yushin, G. Li-ion Battery Materials: Present and Future. *Mater. Mater. Today* **2015**, *18*, 252−264.

(6) Larcher, D.; Tarascon, J.-M. Towards Greener and More Sustainable Batteries for Electrical Energy Storage. *Nat. Chem.* **2015**, *7*, 19−29.

(7) Nithya, C.; Gopukumar, S. Sodium Ion Batteries: A Newer Electrochemical Storage. *Wiley Interdiscip. Rev. Energy Environ.* **2015**, *4*, 253−278.

(8) Scrosati, B.; Garche, J. Lithium Batteries: Status, Prospects and Future. *J. Power Sources* **2010**, *195*, 2419−2430.

(9) Taylor, S. Abundance of Chemical Elements in the Continental Crust: A New Table. *Geochim. Cosmochim. Acta* **1964**, *28*, 1273−1285.

(10) Zhang, Q.; Wang, Z.; Zhang, S.; Zhou, T.; Mao, J.; Guo, Z. Cathode Materials for Potassium-ion Batteries: Current Status and Perspective. *Electrochem. Energy Rev.* **2018**, *1*, 625−658.

(11) Han, M. H.; Gonzalo, E.; Singh, G.; Rojo, T. A Comprehensive Review of Sodium Layered Oxides: Powerful Cathodes for Na-ion Batteries. *Energy Environ. Sci.* **2015**, *8*, 81−102.

(12) Kubota, K.; Dahbi, M.; Hosaka, T.; Kumakura, S.; Komaba, S. Towards K-ion and Na-ion Batteries as "Beyond Li-Ion. *Chem. Rec.* **2018**, *18*, 459−479.

(13) Liu, Y.; Merinov, B. V.; Goddard, W. A. Origin of Low Sodium Capacity in Graphite and Generally Weak Substrate Binding of Na and Mg among Alkali and Alkaline Earth Metals. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 3735−3739.

(14) Chan, E. M. Combinatorial Approaches for Developing Upconverting Nanomaterials: High-throughput Screening, Modeling, and Applications. *Chem. Soc. Rev.* **2015**, *44*, 1653−1679.

(15) Xiang, C.; Suram, S. K.; Haber, J. A.; Guevarra, D. W.; Soedarmadji, E.; Jin, J.; Gregoire, J. M. High-throughput Bubble Screening Method for Combinatorial Discovery of Electrocatalysts for Water Splitting. *ACS Comb. Sci.* **2014**, *16*, 47−52.

(16) 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 Mater.* **2013**, *1*, 011002.

(17) Chen, A.; Zhang, X.; Zhou, Z. Machine Learning: Accelerating Materials Development for Energy Storage and Conversion. *InfoMat* **2020**, *2*, 553−576.

(18) Suh, C.; Fare, C.; Warren, J. A.; Pyzer-Knapp, E. O. Evolving the Materials Genome: How Machine Learning Is Fueling the Next Generation of Materials Discovery. *Annu. Rev. Mater. Res.* **2020**, *50* (1), 1−25.
(19) Schmidt, J.; Marques, M. R.; Botti, S.; Marques, M. A. Recent Advances and Applications of Machine Learning in Solid-state Materials Science. *npj Comput. Mater.* **2019**, *5*, 83.
(20) Dong, Y.; Wu, C.; Zhang, C.; Liu, Y.; Cheng, J.; Lin, J. Bandgap Prediction by Deep Learning in Configurationally Hybridized Graphene and Boron Nitride. *npj Comput. Mater.* **2019**, *5*, 26.
(21) Zhuo, Y.; Mansouri Tehrani, A.; Brgoch, J. Predicting the Band Gaps of Inorganic Solids by Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 1668−1673.
(22) Kolb, B.; Lentz, L. C.; Kolpak, A. M. Discovering Charge Density Functionals and Structure-property Relationships with PROPhet: A General Framework for Coupling Machine Learning and First-principles Methods. *Sci. Rep.* **2017**, *7*, 1192.
(23) Schütt, K. T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K.-R.; Gross, E. K. How to Represent Crystal Structures for Machine Learning: Towards Fast Prediction of Electronic Properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 205118.
(24) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating Materials Property Predictions Using Machine Learning. *Sci. Rep.* **2013**, *3*, 2810.
(25) Pilania, G.; Mannodi-Kanakkithodi, A.; Uberuaga, B.; Ramprasad, R.; Gubernatis, J.; Lookman, T. Machine Learning Bandgaps of Double Perovskites. *Sci. Rep.* **2016**, *6*, 19375.
(26) Takahashi, K.; Takahashi, L.; Miyazato, I.; Tanaka, Y. Searching for Hidden Perovskite Materials for Photovoltaic Systems by Combining Data Science and First Principle Calculations. *ACS Photonics* **2018**, *5*, 771−775.
(27) Kitchin, J. R. Machine Learning in Catalysis. *Nat. Catal.* **2018**, *1*, 230−232.
(28) Goldsmith, B. R.; Esterhuizen, J.; Liu, J.-X.; Bartel, C. J.; Sutton, C. Machine Learning for Heterogeneous Catalyst Design and Discovery. *AIChE J.* **2018**, *64*, 2311−2323.
(29) Ng, M.-F.; Zhao, J.; Yan, Q.; Conduit, G. J.; Seh, Z. W. Predicting the State of Charge and Health of Batteries Using Data-driven Machine Learning. *Nat. Mach. Intell.* **2020**, *2*, 161−170.
(30) Sodeyama, K.; Igarashi, Y.; Nakayama, T.; Tateyama, Y.; Okada, M. Liquid Electrolyte Informatics Using an Exhaustive Search with Linear Regression. *Phys. Chem. Chem. Phys.* **2018**, *20*, 22585−22591.
(31) Okamoto, Y.; Kubo, Y. Ab Initio Calculations of the Redox Potentials of Additives for Lithium-ion Batteries and their Prediction through Machine Learning. *ACS Omega* **2018**, *3*, 7868−7874.
(32) Jalem, R.; Nakayama, M.; Kasuga, T. An Efficient Rule-based Screening Approach for Discovering Fast Lithium Ion Conductors Using Density Functional Theory and Artificial Neural Networks. *J. Mater. Chem. A* **2014**, *2*, 720−734.
(33) Fujimura, K.; Seko, A.; Koyama, Y.; Kuwabara, A.; Kishida, I.; Shitara, K.; Fisher, C. A.; Moriwake, H.; Tanaka, I. Accelerated Materials Design of Lithium Superionic Conductors Based on First-principles Calculations and Machine Learning Algorithms. *Adv. Energy Mater.* **2013**, *3*, 980−985.
(34) Kireeva, N.; Pervov, V. S. Materials Space of Solid-state Electrolytes: Unraveling Chemical Composition−structure−ionic Conductivity Relationships in Garnet-type Metal Oxides Using Cheminformatics Virtual Screening Approaches. *Phys. Chem. Chem. Phys.* **2017**, *19*, 20904−20918.
(35) Cubuk, E. D.; Sendek, A. D.; Reed, E. J. Screening Billions of Candidates for Solid Lithium-ion Conductors: A Transfer Learning Approach for Small Data. *J. Chem. Phys.* **2019**, *150*, 214701.
(36) Sendek, A. D.; Yang, Q.; Cubuk, E. D.; Duerloo, K.-A. N.; Cui, Y.; Reed, E. J. Holistic Computational Structure Screening of more than 12000 Candidates for Solid Lithium-ion Conductor Materials. *Energy Environ. Sci.* **2017**, *10*, 306−320.
(37) Babar, M.; Parks, H. L.; Houchins, G.; Viswanathan, V. An Accurate Machine Learning Calculator for the Lithium-graphite System. *JPhys. Energy* **2021**, *3*, 014005.
(38) Attarian Shandiz, M.; Gauvin, R. Application of Machine Learning Methods for the Prediction of Crystal System of Cathode Materials in Lithium-ion Batteries. *Comput. Mater. Sci.* **2016**, *117*, 270−278.
(39) Eremin, R. A.; Zolotarev, P. N.; Ivanshina, O. Y.; Bobrikov, I. A. Li(Ni, Co, Al)O_2 Cathode Delithiation: A Combination of Topological Analysis, Density Functional Theory, Neutron Diffraction, and Machine Learning Techniques. *J. Phys. Chem. C* **2017**, *121*, 28293−28305.
(40) Allam, O.; Cho, B. W.; Kim, K. C.; Jang, S. S. Application of DFT-based Machine Learning for Developing Molecular Electrode Materials in Li-ion Batteries. *RSC Adv.* **2018**, *8*, 39414−39420.
(41) Joshi, R. P.; Eickholt, J.; Li, L.; Fornari, M.; Barone, V.; Peralta, J. E. Machine Learning the Voltage of Electrode Materials in Metal-ion Batteries. *ACS Appl. Mater. Interfaces* **2019**, *11*, 18494−18503.
(42) Wang, L.; Maxisch, T.; Ceder, G. A First-principles Approach to Studying the Thermal Stability of Oxide Cathode Materials. *Chem. Mater.* **2007**, *19*, 543−552.
(43) Chiang, Y.-M. Building a Better Battery. *Science* **2010**, *330*, 1485−1486.
(44) Ye, H.; Zhang, Y.; Yin, Y.-X.; Cao, F.-F.; Guo, Y.-G. An Outlook on Low-Volume-Change Lithium Metal Anodes for Long-Life Batteries. *ACS Cent. Sci.* **2020**, *6*, 661−671.
(45) Zhou, F.; Cococcioni, M.; Marianetti, C. A.; Morgan, D.; Ceder, G. First-principles Prediction of Redox Potentials in Transition-metal Compounds with LDA+U. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2004**, *70*, 235121.
(46) Ong, S. P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder, G.; Persson, K. A. The Materials Application Programming Interface (API): A Simple, Flexible and Efficient API for Materials Data Based on REpresentational State Transfer (REST) Principles. *Comput. Mater. Sci.* **2015**, *97*, 209−215.
(47) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A Robust, Open-source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.
(48) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J. H.; Chen, J.; Bystrom, K.; Dylla, M.; Chard, K.; Asta, P. K.M.; Snyder, G. J.; Foster, I.; Jain, A.; et al. Matminer: An Open Source Toolkit for Materials Data Mining. *Comput. Mater. Sci.* **2018**, *152*, 60−69.
(49) Siriwardane, E. M. D.; Joshi, R. P.; Kumar, N.; Çakır, D. Revealing the Formation Energy−Exfoliation Energy−Structure Correlation of MAB Phases Using Machine Learning and DFT. *ACS Appl. Mater. Interfaces* **2020**, *12*, 29424−29431.
(50) Chollet, F. *Keras*, 2015. https://github.com/fchollet/keras.
(51) Jolliffe, I. T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc., A* **2016**, *374*, 20150202.
(52) Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929−1958.
(53) Weigend, A. S.; Rumelhart, D. E.; Huberman, B. A. Generalization by Weight-elimination with Application to Forecasting. *Advances in Neural Information Processing Systems* **1991**, 875−882.
(54) Giannozzi, P.; et al. QUANTUM ESPRESSO: A Modular and Open-source Software Project for Quantum Simulations of Materials. *J. Phys.: Condens. Matter* **2009**, *21*, 395502.
(55) Giannozzi, P.; et al. Advanced Capabilities for Materials Modelling with Quantum ESPRESSO. *J. Phys.: Condens. Matter* **2017**, *29*, 465901.
(56) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2008**, *100*, 136406.

https://doi.org/10.1021/acsami.1c04627
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

(57) Wang, L.; Maxisch, T.; Ceder, G. Oxidation Energies of Transition Metal Oxides within the GGA+ U Framework. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *73*, 195107.
(58) https://wiki.materialsproject.org/GGA%2BU_calculations.
(59) Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-zone Integrations. *Phys. Rev. B Condens. Matter* **1976**, *13*, 5188.

https://doi.org/10.1021/acsami.1c04627
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX