Rechargeable Zn batteries with aqueous electrolytes have been considered as promising alternative energy storage technology, with various advantages such as low cost, high volumetric capacity, environmentally friendly, and high safety. However, a lack of reliable cathode materials has largely pledged their applications. Herein, a machine learning (ML)-based approach to predict cathodes with high capacity (>100 mAh g\(^{-1}\)) and high voltage (>0.5 V) is developed. Over \(\approx 130 \,000\) inorganic materials from the materials project database are screened and the crystal graph convolutional neural network based ML approach is applied with data from the AFLLOW database, the combination of these two gives rise to \(\approx 80\) predicted cathode materials. Among them, \(\approx 10\) cathode materials have been experimentally discovered previously, which agrees remarkably well with experimental measurements, while \(\approx 70\) new promising candidates have been predicted for further experimental validations. The authors hope this study could spur further interests in ML-based advanced theoretical tools for battery materials discovery.

1. Introduction

Rechargeable Zn-ion batteries with aqueous electrolytes have been considered as promising alternatives and supplements for the commercialized rechargeable Li-ion batteries, with many advantages such as low cost, high volumetric energy density, environmentally benign, and high safety.\(^{[1–6]}\) They are promising for device applications in large-scale grid energy storage and renewable green energy electrochemical storage systems, which are sensitive to cost, space, and environmental effects. However, one of the key challenges facing the Zn-ion battery systems is the lack of proper cathode materials with high structure stability, high capacity, and high voltage.\(^{[7]}\) To date, manganese oxides,\(^{[8–11]}\) vanadium oxides,\(^{[12–17]}\) Prussian blue analogs,\(^{[18–20]}\) Mo\(_6\)S\(_8\)-based Chevrel phase compounds,\(^{[21]}\) and some other organic materials,\(^{[22,23]}\) etc., have been widely investigated as cathode materials for Zn batteries. In particular, so far, vanadium oxides (V\(_x\)O\(_y\)) and MnO\(_2\) are the most extensively studied Zn battery cathode systems.\(^{[24]}\) Meanwhile, they still suffer from low capacity, low voltage, high cost, toxicity, low cycle life, and so on, which severely hindered their practical applications. Discovering and designing new reliable cathode materials are essential for the future development, application, and commercialization of the Zn battery systems.

Recently, the data-driven machine learning (ML) approach has largely renovated the paradigm for materials design and discovery, which opened a new era for the computational materials science community.\(^{[25,26]}\) The combination of large data sets in the materials databases, data-mining technique, structure featureization, and ML approach could not only expand the materials screening scope but also gain unique insights into the complex structure-property and composition-property relationships, using the wealth of the existing computational and experimental materials data.\(^{[27,28]}\) It can achieve a desirable prediction accuracy for materials properties that is comparable to the conventional computational materials tools such as density functional theory (DFT) with a lower computational cost, larger computational length scale or timescale. One particular example is the recent development of the crystal graph convolutional neural network (CGCNN),\(^{[29]}\) an ML tool that encodes the crystal structure as a “graph,” has enabled the accelerated design of crystalline materials with desired properties, while also providing fundamental insights on the elementary/structural contributions to the energetics and properties of the materials. It has been successfully applied to the design of various crystalline materials for specific properties and applications, including the screening of solid electrolytes for dendrite suppression in lithium battery,\(^{[30]}\) prediction
of thermodynamic, mechanical, and electronic properties for 2-D materials, the formation energy of compounds, and the methane adsorption properties of metal-organic frameworks, etc.

Herein, we have developed a CGCNN-based ML tool to predict the high voltage cathode materials for Zn-ion batteries. We have screened over 130,000 inorganic materials from two representative materials databases, namely the materials project (MP) and AFLOW, which has predicted approximately 80 candidates that could exhibit high voltage (over 0.5 V vs Zn/Zn²⁺), low toxicity, high abundant, and high capacity (>100 mAh g⁻¹). It was discovered that the mixing prediction potentials which combined both the MP-derived and ML-predicted voltages agree remarkably well with the experimental measurements for 10 well-known Zn battery cathode systems. While other ≈70 candidates that have not been reported were also predicted and tabulated for future experimental validations. We hope this study could spur further interest in the experimental discovery of cathode materials for Zn batteries and theoretical screening of other battery materials with highly efficient ML-based tools.

2. Methods

From classic electrochemical theory, the electrochemical reaction for a standard Zn full cell with Zn metal anode can be written as:

\[ a\text{Zn} + b\text{X} = \text{Zn}_a\text{X}_b \]  

where \( X/X \) is the cathode material pair before and after intercalation of Zn. The electrochemical potential for the Zn//X full cell can be derived from the Gibbs free energy of the reaction (1)

\[ \Delta G_1, \] namely:

\[ \Delta G_1 = G_{\text{Zn}_a\text{X}_b} - (aG_{\text{Zn}} + bG_X) \]  

\[ \varphi = \Delta G_1 / (2aF) \]  

where \( a, b \) are the balance factors, \( G_{\text{Zn}_a\text{X}_b} \), \( G_{\text{Zn}} \), and \( G_X \) are the Gibbs free energy of \( \text{Zn}_a\text{X}_b \), Zn and X, respectively. \( \varphi \), 2, and \( F \) are the equilibrium electrochemical potential of the cathode pair, number of electrons involved per Zn atom for the electrochemical reaction (1), and Faraday constants, respectively. Combining Equations (2) and (3), it is indicated that the accurate prediction of electrochemical potential \( \varphi \) requires reliable calculations of the Gibbs free energy for each material. Traditionally, the free energies of materials in the materials database are obtained from high-throughput calculations, which could give rise to a systematic error that led to large deviations for potential predictions. The systematic error from high-throughput calculations could arise from the chosen of simulation system size, cut-off energy, convergence criteria, initial magnetization setup, k-points, and U value, etc. To overcome this problem, we employ a combined approach with the CGCNN-based ML tool and multiple databases to improve the prediction accuracy while also provide uncertainty qualifications for the predictions.

The schematics of the battery cathode prediction protocol used in this study are given in Figure 1. Using the Pymatgen application programming interface (API), we search through the existing ≈130,000 inorganic materials in the MP database and obtain ≈19,000 pairs of electrodes with Zn-containing compounds.
(Zn\(_a\)X\(_b\)) and the corresponding material before intercalation of Zn (X). Among them, compounds with toxic elements (e.g., Pb, Cd, Tl, Cr, Hg) and radioactive elements (U, Fr, Ra, etc.) are excluded which reduces the candidates to \(\approx\)9000 pairs. The electrochemical potential of these cathode pairs versus Zn/Zn\(^{2+}\) can be obtained by combining Equations (2) and (3), with the Gibbs free energy data extracted from the MP database (denoted as \(\varphi_{\text{MP}}\)). Meanwhile, subsequently, a DFT-Solver based on the CGCNN algorithm was trained to predict the Gibbs free energy for all the \(\approx\)130,000 inorganic material pool. Then, a total of \(\approx\)60,000 data instances are taken from the AFLOW database, where \(\approx\)42,000 data instances are used as the training set (70%), \(\approx\)9000 data are chosen as the validation set (15%) and \(\approx\)9000 data are employed as the testing set (15%). Similarly, the electrochemical potential for the \(\approx\)9000 cathode pairs can also be calculated with the Gibbs free energy predicted from the CGCNN-based ML approach, which is denoted as \(\varphi_{\text{ML}}\). A step further, after filtering out all the positive potential values for the two electrochemical potentials (\(\varphi_{\text{MP}}\) and \(\varphi_{\text{ML}}\)), we define a mixing prediction potential by calculating the square root of the two predicted potentials, that is,

\[ \varphi_{\text{mix}} = \sqrt{\varphi_{\text{MP}}\varphi_{\text{ML}}} \]  

(4)

Here the geometric means \(\varphi_{\text{mix}}\) is used instead of simple average (i.e., \(\varphi_{\text{a}} = \frac{\varphi_{\text{MP}} + \varphi_{\text{ML}}}{2}\)) because we want to choose candidates that have high voltage from both databases, while \(\varphi_{\text{a}}\) could be dominated by the highest value of the two potentials, which could be strongly affected by the accuracy of the data in one database. Further filtering is applied to screen all the candidates that have both high mixing prediction potential (>0.5 V) and high capacity (>100 mAh g\(^{-1}\)). The electrode candidates can be obtained after combining all the compounds that have the same elements and similar \(a\) to \(b\) ratio for further experimental materials synthesis and test to validate the prediction.

3. Results and Discussion

To validate the CGCNN-based ML DFT-solver, the parity plot for the ML-predicted energies and high throughput DFT calculated energies taken from the two materials databases is given in Figure 2. It can be observed that the CGCNN-DFT-solver with training data taken from both MP and AFLOW shows good linearity with the energy data extracted directly from the two materials databases. The coefficients of determination \(R^2\) for both cases are close to 1, indicating very small standard deviations, while the slopes of the lines for both MP and AFLOW are 0.98 and 0.99, respectively, showing excellent correlations between the ML-predicted energies and the energies extracted directly from the databases. The mean absolute error (MAE) is plotted against the machine learning epoch for the training and validation set, showing the saturation of MAE after 100 epochs for both databases, giving rise to an MAE value of \(\approx\)0.1 eV.
atom$^{-1}$, consistent with the previous report. This MAE value also indicates that the standard error for the CGCNN-ML model is within the DFT model uncertainty.

The statistics for this study are further given in Figure 3. It can be seen that a total of $\approx 9000$ electrode pairs have been screened from the MP database out of $\approx 130,000$ inorganic compounds (Figure 3a). Around $3/4$ of these electrode pairs have a positive electrochemical potential. Notably, the number of candidates decays exponentially with the increasing of electrochemical potential, the majority of candidates show a calculated electrochemical equilibrium potential between 0 and 2 V. A similar trend is observed from the ML-learned potentials with training data from AFLOW (e.g., $\varphi_{ML}$), where most of the electrode pairs have a relatively low voltage between 0 and 1.5 V. The low predicted voltage can be understood since the electrochemical potential for Zn//Zn$^{2+}$ ($\approx 0.76$ V vs S.H.E.) is much higher than Li//Li$^+$ ($\approx 3.06$ V vs S.H.E.). Meanwhile, it is shown that the actual statistical distributions for the two cases are not identical. To quantify the agreements between the two potentials, the correlation between $\varphi_{MP}$ and $\varphi_{ML}$ is further plotted in Figure 3b. It can be observed that these two potentials show a wide dispersion, weak and positive correlation with R-squared of $\approx 0.4$ and $\rho_{xy}$ of 0.6, while the slope is 0.41. The weak correlation can be understood since the materials data in the two databases are obtained with different high throughput calculation methods. Meanwhile, the fundamental elemental and structural contributions to the energies should be similar throughout the databases. The distribution of $\varphi_{mix}$ is demonstrated in Figure 3c, which gives rise to an almost equal number of pairs in between 0 and 1.5 V while decays exponentially above 1.5 V. The total number of pairs that exhibit a potential above 0.5 V reduces to $\approx 3000$.

The theoretical specific capacity was further calculated, using the molecular weight of the cathode material X and the charge transfer in electrochemical reaction Equation (1). It can be observed in Figure 3d that among all the pairs with a predicted potential greater than 0.5 V, $\approx 500$ pairs have shown a theoretical specific capacity $>100$ mAh g$^{-1}$. Among them, some candidates share the same elements and atomic ratio with different total atom numbers (e.g., for Mn peroxide, they may exist in the form of MnO$_2$, Mn$_2$O$_3$, and Mn$_3$O$_6$, etc.). In this study, we choose only the lowest total atom numbers and the polymorph with the lowest energy as the ground state. Some representative predicted cathode materials are listed in Figure 4, after all the screening and filtering. It can be discovered that the majority of experimentally discovered inorganic cathode materials are captured, including MnO$_2$, VO$_2$, V$_2$O$_3$, V$_3$O$_7$, V$_2$O$_5$, VS$_2$, V$_2$O$_5$, LiV$_3$O$_8$, O$_2$, etc., showing the fidelity of this model, while the experimentally measured
Figure 4. The theoretical prediction with uncertainty for representative candidates from the list of finalists. The results on the left shows the mixing prediction is in consistence with the experiment data for the experimentally discovered inorganic cathodes. Meanwhile, some predicted cathode materials are listed on the right.

Table 1. Comparison of MAE and RMSE between statistic mean potentials, MP-predicted potentials, and machine learning potentials.

|      | ϕ_{MP} | ϕ_{ML} | ϕ_{H} | ϕ_{G} | ϕ_{A} | ϕ_{Q} |
|------|--------|--------|-------|-------|-------|-------|
| MAE (V) | 0.399  | 0.334  | 0.174 | 0.111 | 0.131 | 0.184 |
| RMSE (V) | 0.482  | 0.372  | 0.214 | 0.172 | 0.182 | 0.237 |

values for these materials agree remarkably well with the predicted potentials by mixing prediction. Note that interestingly, most of these inorganic cathode materials are vanadium oxides (VₓOᵧ) and MnO₂, which is consistent with the experimental observations.[24] To further rationalize the chosen of geometric means as the predictions, we compared the performance matrix for the different statistic means, namely, the harmonic mean ϕ_{H}, geometric mean ϕ_{GM}, arithmetic mean ϕ_{A}, quadratic mean ϕ_{Q} with ϕ_{MP} and ϕ_{ML}. As can be seen in Table 1, the geometric mean shows the best performance with the lowest MAE and root mean square error (RMSE). In particular, the MAE for ϕ_{G} (0.111 V) is way smaller than ϕ_{MP} (0.399 V) and ϕ_{ML} (0.334 V), showing the necessity to use mix prediction instead of simply extracting the data from databases.

Meanwhile, some other new promising cathode materials are predicted with a potential value between 0.5 and 2.0 V and a reasonably high capacity, including SnPO₄, MnPO₄, AgPS₂, Co₃O₄, Mn₃PO₄, BaMnF₇, BaNi₂O₆, Li₂Fe₂(PO₄)₂, etc. A full list of all the 78 finalists and their materials project identity number (MP-id) for ZnₓXₙ and X are given in Table S2, Supporting Information. It is worthwhile to note that there are only a few vanadium-based materials (VF₄, TiV(PO₄)₂, BaV₂O₇, V₂P₂O₁₁, and VPO₅) for the other predicted candidates, most of which have nothing in common with the experimentally verified datasets shown on the left side of Figure 4. This can be understood since in this model the 10 experimentally measured voltages are not used as the training sets in the machine learning model, so the predicted materials don’t necessarily have to be similar to the 10 experimentally verified materials. Besides, it can be seen that the high voltage candidates are mainly in four categories, including oxides, fluorides, sulfides, and phosphates, which shows some similarity with the state-of-art high voltage cathode materials for Li-batteries, showing the reliability of the predictions.

4. Conclusions

In conclusion, we have developed a CGCNN-based ML tool to predict high voltage cathode materials for Zn-ion batteries, which mix the potential data calculated from the MP database and predicted from ML-learned potential with the AFLOW database. We have screened over ≈130 000 inorganic materials and extracted tens of candidates that are predicted to have high voltage (over 0.5 V vs Zn/Zn²⁺) and high capacity (over 100 mAh g⁻¹). The mixing predictions agree remarkably well with experimental data, while the two potentials can provide an estimate for the uncertainty of the predictions. We further provide possible candidates for high potential Zn battery cathode materials for future experimental exploration. This approach can be widely extended to other battery systems including Li-, Na-, K-, Al-, and Mg-based batteries. We hope this study could spur further interest in the applications of the ML-based approach for battery materials discovery.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Z.H. was supported by a start-up grant from Zhejiang University. This work was supported by the Fundamental Research Funds for the Central Universities No. 2021FZZX001-08.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author (hongzijian100@zju.edu.cn) upon reasonable request.

Keywords

AFLOW, crystal graph convolutional neural network, machine learning, Zn batteries
