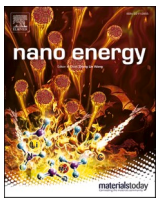
Nano Energy 98 (2022) 107214

Contents lists available at ScienceDirect

Nano Energy

journal homepage: www.elsevier.com/locate/nanoen



Machine learning assisted synthesis of lithium-ion batteries cathode materials

Chi Hao Liow [a], Hyeonmuk Kang [a], Seunggu Kim [b], Moony Na [b], Yongju Lee [a], Arthur Baucour [a], Kihoon Bang [a], Yoonsu Shim [a], Jacob Choe [a], Gyuseong Hwang [a], Seongwoo Cho [a], Gun Park [a], Jiwon Yeom [a], Joshua C. Agar [d], Jong Min Yuk [a], Jonghwa Shin [a], Hyuck Mo Lee [a], Hye Ryung Byon [b], EunAe Cho [a], Seungbum Hong [a,c,*]

[a] Department of Materials Science and Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 3414, Republic of Korea
[b] Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
[c] KAIST Institute for the NanoCentury, KAIST, Daejeon 34141, Republic of Korea
[d] Department of Materials Science and Engineering, Lehigh University, Bethlehem, PA 18015, USA

ARTICLE INFO

Keywords:
Lithium-ion batteries
NCM cathode
Inverse design
Machine learning
Design-to-device pipeline

ABSTRACT

Optimizing synthesis parameters is crucial in fabricating an ideal cathode material; however, the design space is too vast to be fully explored using an Edisonian approach. Here, by clustering eleven domain-expert-derived-descriptors from literature, we use an inverse design surrogate model to build up the experimental parameters-property relationship. Without struggling with the trial-and-error method, the model enables design variables prediction that serves as an effective strategy for cathode retrosynthesis. More importantly, not only did we overcome the data scarcity problem, but the machine learning model has guided us to achieve cathode with high discharge capacity and Coulombic efficiency of 209.5 mAh/g and 86%, respectively. This work demonstrates an inverse design-to-device pipeline with unprecedented potential to accelerate the discovery of high-energy-density cathodes.

1. Introduction

Machine intelligence's ability to approximate correlation on high-dimensional parameter spaces can provide physical insight that accelerates materials discovery [1–4]. Today, Lithium-ion batteries (LiB) is one of the most important technology that has revolutionized portable electronic and electric vehicle industries. However, the traditional approach of developing higher performance battery materials usually requires long-time study and complicated experimental endeavors, because the potential combination of compounds that could intercalate metal ions (Li, Na, K, Mg, Ca, and Al) is easily in the order of thousands.

As such, machine learning (ML) has garnered immense interest to circumvent untenable Edisonian-type experiments aiming at reducing human effort and research costs [5]. For example, Sendek et al. used logistic regression to screen ~12,000 new solid Li superionic conductors for all-solid-state LiB [6]. Similarly, by utilizing the feature vectors derived from chemical properties and their elemental constituents into ML algorithms such as deep neural networks, support vector machine, and kernel ridge regression, Joshi et al. proposed ~5000 electrode candidates for Na- and K-ion batteries [3].

Despite current progress of ML for new materials discovery based on public databases (e.g., Materials projects [7], Novel Materials Discovery (NOMAD) [8], Automatic FLOW for Materials Discovery (AFLOW) [9], and Open Quantum Materials Database (OQMD)) [10]), its application in the guided experimental procedure for LiB electrode is still in the early stage pioneered by a few groups. For example, Ceder and his colleagues have demonstrated the capability of using supervised and unsupervised natural language processing techniques to extract "codified recipes" for synthesis routes from millions of unstructured scientific publications [11–14]. Chueh et al. had employed both the early-predictive model and Bayesian optimization algorithms. They had reduced the number and the duration of experiments required to maximize the battery cycle life [5]. Aspuru-Guzik and his colleagues further advanced in integrating ML into a self-driving robotic laboratory

* Corresponding author at: Department of Materials Science and Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 3414, Republic of Korea.
E-mail address: seungbum@kaist.ac.kr (S. Hong).

https://doi.org/10.1016/j.nanoen.2022.107214
Received 2 December 2021; Received in revised form 5 March 2022; Accepted 27 March 2022
Available online 1 April 2022

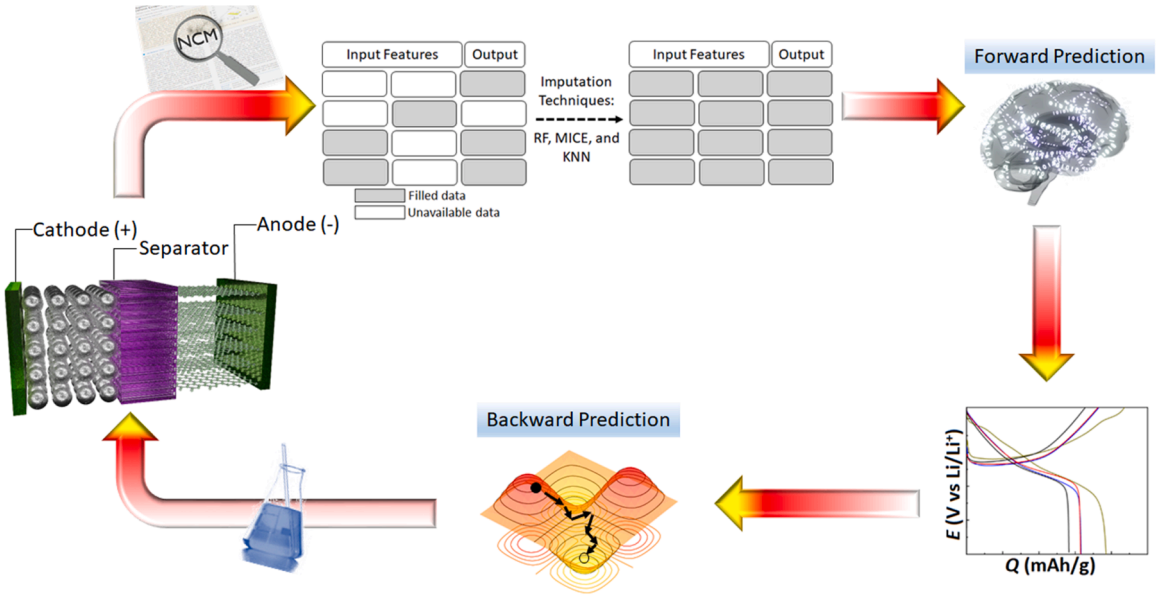


Fig. 1. Schematic illustration of three steps of the design-to-device pipeline. First, we curated data from published literature on the basis of Nickel (Ni), Cobalt (Co), and Manganese (Mn) at different compositional ratio and their latent experimental descriptors as input features, while discharge capacity as the target properties. From the literation curation, there are ~ 16% of missing data. Here, we utilized statistical techniques to "fill" the missing values. Subsequently, the completed data are used as input for forward modeling. Second, we demonstrate inverse design using particle-swarm optimization (PSO) to predict optimized design variables that of the defined target property, the discharge capacity. Finally, we realize the design-to-device pipeline via experimental validation.

that is crucial in reducing human effort in research and development [15].

Nevertheless, ML-guided optimized cathode material's design variables are rarely found, because of the challenge of preparing feature vectors that represent the entire experimental procedure for the notoriously data-hungry ML in cathode synthesis. Although one can create the training dataset from his/her experimental result or computer simulation [16], this approach may not be universal because of the different individual's experience. Consequently, one may find it difficult to replicate the optimized design variables because of the specific synthesis parameters. To address this problem, we propose building the input features based on the frequently mentioned parameters from published scientific records. Subsequently, we combine inverse design and realize the design-to-device pipeline of the fundamental experimental guidelines for a cathode material with high discharge capacity.

Herein, we employed an inverse design framework to predict the optimal experimental design of $LiNi_xCo_{1-x-y}Mn_{1-x-y-z}O_2$ (NCM) cathode with $x + y + z = 1$. This framework (Fig. 1) contains three parts: (1) statistical imputation techniques to approximate the missing data from literature curation; (2) machine learning and hyperparameter search to maximize the generalizability; (3) experimental validation of optimized design variable prediction. From the inverse design, we observed that the composition of Ni, sintering temperature, cut-off voltage, and charge-rate (C-rate) are strongly correlated with the performance of Li-ion batteries, which agrees with the previous studies [17,18,39]. Subsequently, based on the inverse-design guideline, we carefully validated the prediction experimentally. We found that our prediction achieved high accuracy, with root mean square error, RMSE of 8.17 mAh/g. We expect that our study will lay the foundation for a rapid, reliable, and rational design of cathode materials.

2. Results and discussion

2.1. Data imputation

Our objective is to provide highly viable and quantitative guidelines for the experimental design variables of the optimal NCM cathode. Here, we established the input variables that describe the physical descriptors of the NCM cathode, such as compositional ratio and the secondary particle size, to predict the cathode's discharge capacity. Other factors include the sintering temperature and duration because of their influence on crystallinity and structural stability. Lastly, the measurement conditions, such as cut-off voltage and C-rate that control the potential window and rate of electrochemical, were selected as the features. These feature descriptors were our input variables, while the discharge capacity was the target output (as shown in Table 1).

Table 1
Summary of the inverse design guided NCM synthesis.

Target Discharge Capacity		Ni	Co	Mn	Size	1st Sin. Temp	1st Sin. Time	2nd Sin. Temp	2nd Sin. Time	Cut-off Volt.	C-rate
150	MICE_ML	0.564	0.261	0.167	13.4	687	9.30	838	19.38	4.35	1.73
	MICE_syn	0.512	0.276	0.168	6.5	686	9.00	838	19.00	4.35	1.73
175	MICE_ML	0.653	0.093	0.244	15.1	483	6.78	797	14.25	4.45	0.70
	MICE_syn	0.660	0.094	0.247	1.5	482	7.00	797	14.00	4.50	0.70
200	MICE_ML	0.794	0.129	0.063	11.9	690	4.34	789	17.24	4.70	0.22
	MICE_syn	0.757	0.144	0.066	3.9	690	4.00	789	17.00	4.70	0.22

*Note: _ML and _syn denote machine learning predicted and experiment parameters, respectively. The abbreviation of Ni, Co, Mn, Size, 1st Sin. Temp, 1st Sin. Time, 2nd Sin. Temp, 2nd Sin. Time, Cut-off volt., and C-rate represent Nickel, Cobalt, Manganese, Secondary particle size, first sintering temperature, first sintering time, second sintering temperature, second sintering time, cut-off voltage, and charge rate.

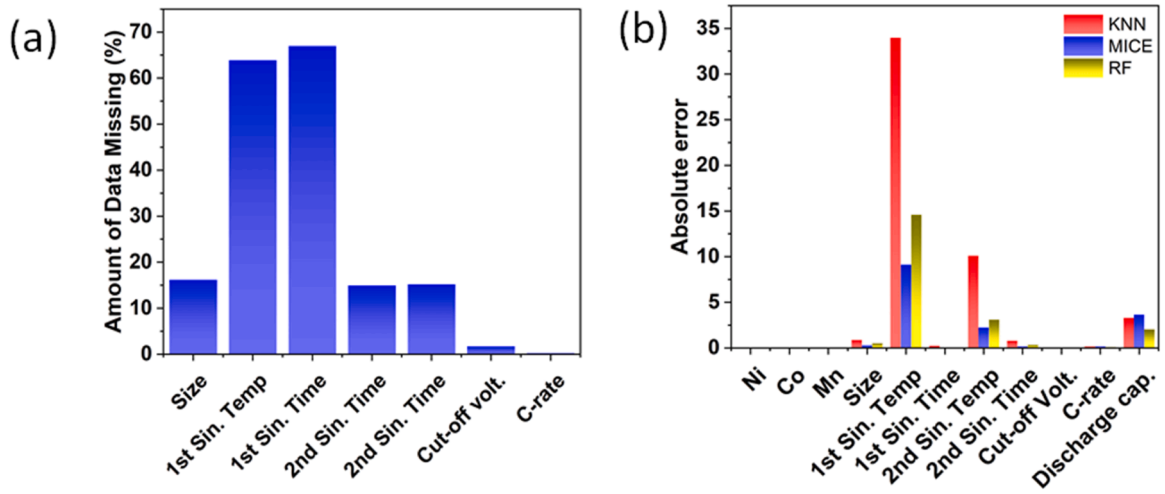


Fig. 2. (a) Percentage of missing data in each feature. (b) Comparison of different imputation methods, namely, KNN (k-neighbour), multiple imputations by chained equations (MICE), and random forest (RF).

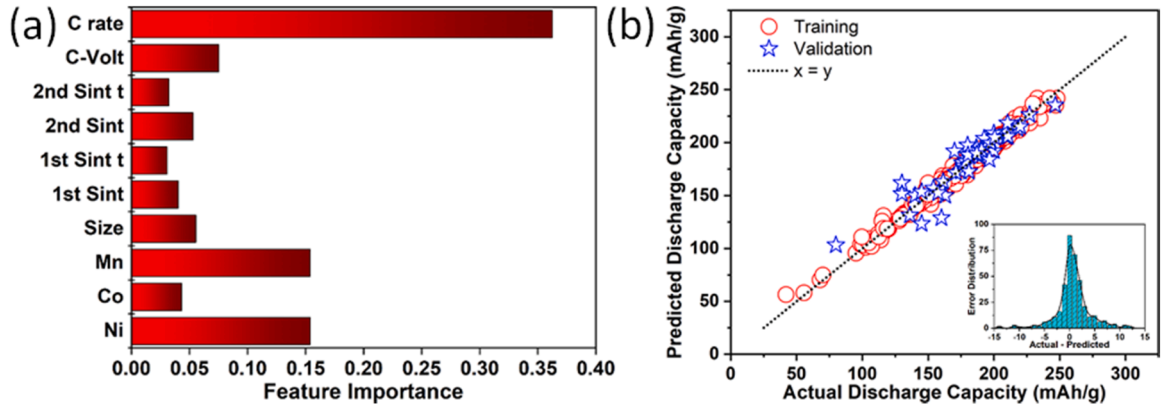


Fig. 3. (a) Feature importance of input descriptors. (b) Performance of forward modeling between predicted and target values using MICE. Inset shows the histogram of the training set mean absolute error (MAE). The dashed line (x = y) is representing the best fit.

We found that the inconsistent report of synthesis and measurement parameters from individual publications poses a great challenge in ML-driven predictions of materials science. Once we standardize the input features, we found that almost 16% of the total 415 data are not available. To overcome this problem, we utilized several imputation techniques, namely k-nearest neighbors (KNN) [19], random forest (RF) [20,21], and multiple imputations by chained equations (MICE) [22], to "fill" in the missing values. Although the data imputation is not often used in materials science, their application in biomedical research is widely used to preserve valuable information [23,24].

To evaluate the reliability of each imputation method, we first removed any row of data that consisted of missing/null values. This process reduced our data size from 415 to 112. We observed that majority of the missing information is related to the first sintering temperature and duration, followed by the second sintering and particle size, as shown in Fig. 2(a).

Before we verified the reliability of the respective mathematical technique, we randomly removed 16% of the data from the 112 data. Ideally, the difference between the "true" value and the calculated one should be close to 0. Here, we spot that the "first sintering temperature" feature exhibits the largest variance between the imputed and the original datasets. KNN shows the poorest performance out of the three methods. The reason is that KNN uses the "feature similarity" from the k's nearest neighbors to predict the missing value. In the "first sintering temperature," more than 80% of the temperature used by researchers is 500 °C (Fig. S1). The lack of diversity can mislead the KNN prediction, while the outlier might be strongly suppressed. In contrast, MICE produced the smallest variance compared to KNN and RF and has become the obvious choice to deal with the missing data. With this, we carried out the MICE imputation for 415 data for further analysis and ML prediction.

2.2. Forward prediction

The input variables' feature importance is interpreted using Gini importance [25], as illustrated in Fig. 3(a). It was found that variables such as Ni and second sintering temperature, C-rate, and cut-off voltage are the four important features that were highly correlated to the discharge capacity. A similar trend was also captured using Pearson correlation, as shown in Fig. S2.

Surprisingly, component "Co" was not ranked as an important feature as their compositional counterparts (Ni and Mn). We believe this might be due to the research trend moving away from Co-rich to Ni-rich composition between 2004 and 2019 [39]. Despite its high theoretical specific capacity of 274 mAh/g, it has several crucial drawbacks, including high raw material cost, toxicity, low thermal stability, and rapid capacity fading at high C-rate [26].

Another reason is the discovery of Ni-rich NCM materials that exhibit high capacity with low production costs. Hence, a concerted effort has been made for the past 15 years to minimize the Co component while

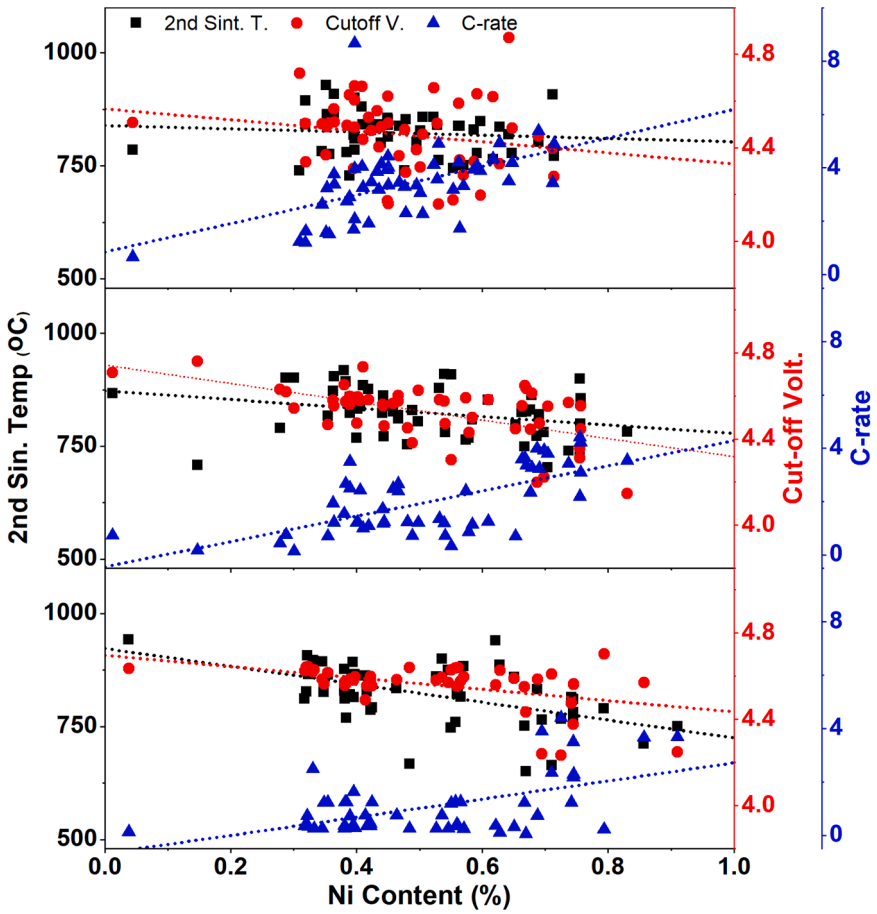


Fig. 4. Predicted inverse design variable based on the different imputed datasets with different desired discharge capacity. Target property (top) 150 mAh/g, (middle) 175 mAh/g, (bottom) 200 mAh/g. Dotted lines serve as a guide to the eye.

maintaining cathode materials' superior properties. Ni and Mn are among many active components widely used to overcome this issue within that period, resulting in a low feature importance ranking for Co.

We employed gradient boosting regression with decision trees (GBR) as the forward modeling algorithm [27], for its robustness on small datasets (< 1000 data) prediction [28,29]. Hyperparameters of GBR were tuned by a 10-fold cross-validation grid search based on 90% of the training set and 10% of the validated set. In this work, we evaluated the model's performance based on the coefficient of the determinant (Fig. 3 (b)), R^2 and mean absolute error ($MAE = 1/N \sum_{i=1}^{N} |y_i^p - y_i^t|$), where N is the number of samples; y_i^p and $y_i^t$are the predicted and target value of sample *i*, respectively. We found that the training of GBR fit well with an R^2 of 0.99 with an MAE of less than 2.2 mAh/g while the validation R^2 is 0.85 with an MAE of 8.9 mAh/g. Although we notice that there might be some degree of overfitting, considering the reasonable R^2 of validation data, we believe that the complexity of descriptors and electrochemical properties is well captured via the GBR algorithm.

2.3. Backward prediction

While forward modeling is commonly used to describe the underlying process-structure-property relationship in materials science, we adopt an inverse design model based on the particle swarm optimization (PSO) algorithm to search for the optimum cathode NCM experimental conditions. Using the trained model, we predicted new design variables that satisfy the electrochemical specification of different target discharge capacities as a guideline for the design-to-device pipeline.

Next, we backward predicted the design variables according to the desired target properties of 150, 175, and 200 mAh/g, respectively. Although PSO can produce many design candidates [27], the quality of the predicted design variables may vary due to the possibility of extrapolation that lies beyond their training set (Fig. S3). Since experimentally validating all the predicted variables is inefficient, a neural network method to rank the predicted variables' reliability was employed. In this method, the predicted variables are fed into the neural network. We then compared the neural network predicted discharge capacity with the GBR prediction. The variance of the comparison will serve as one of the confident guides for our design-to-device realization.

2.4. Rationalization of inverse design

From the feature importance, we selected the four most important parameters that describe the discharge capacity, as depicted in Fig. 4. Subsequently, we established a design rule for cathode materials that meet the desired electrochemical property from the predicted inverse design. It is clear that for the Ni-rich cathode ($LiNi_xCo_{1-x-y}Mn_{1-x-y-z}O_2$, $x > 0.6$), the sintering temperature generally decreases as a function of the Ni content. The reason is twofold; first, due to the comparable ionic radii of Li^+ (0.69 Å) and Ni^+ (0.76 Å), high sintering temperature causes Li^+ to be replaced by Ni^{2+} and hinders the Li-ion diffusion pathway during the electrochemical process. Second, due to the nature of low bonding energy between Ni^{3+} and O^{2-}, Ni-rich cathode's sintering leads to the loss of oxygen atoms and creates surface defects [30,31]. As a result, more oxygen deficiency sites will be built on the cathode's surface at the higher sintering temperature, leading to structural degradation and retarding the interface's electrochemical reaction. As such, researchers lowered the sintering temperature as they increased the Ni-content in the cathode.

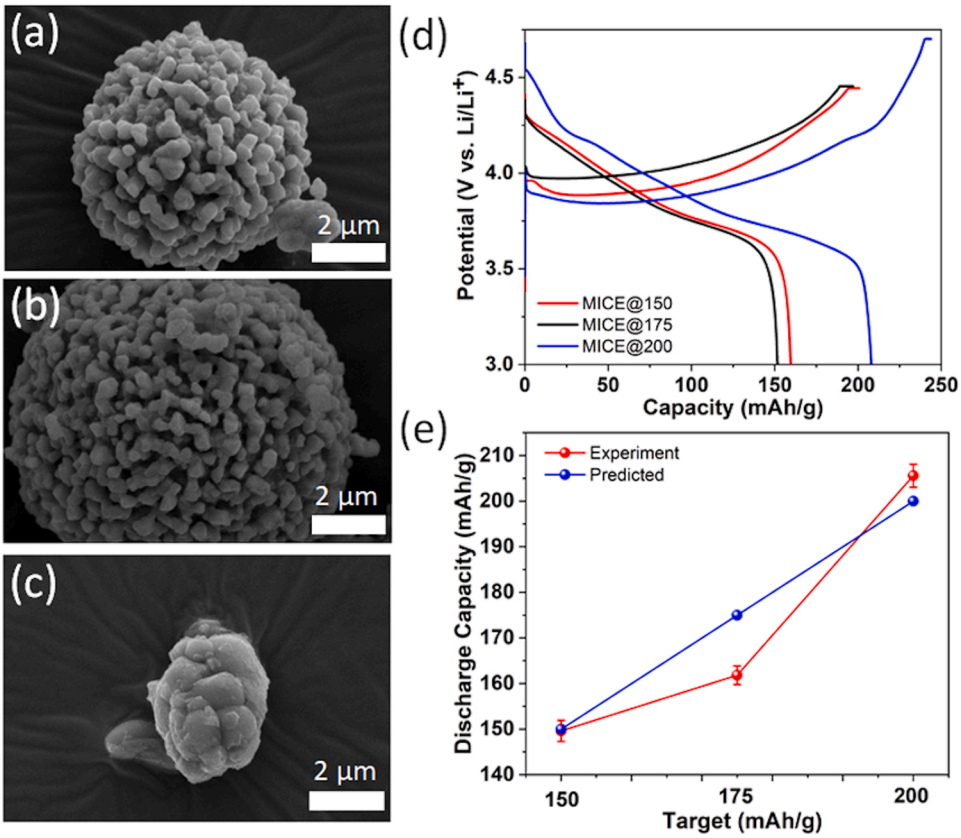


Fig. 5. Design-to-device pipeline based on inverse design prediction. (a), (b), and (c) are scanning electron micrographs based on the prediction of imputed datasets MICE@150, MICE@175, and MICE@200, respectively. (d) Charge-discharge curves of MICE@150, MICE@175, and MICE@200 for the first cycle with 1 M $LiPF_6$ in EC/DEC. The C-rate and cut-off potential follow model trained on MICE imputed dataset (Table 1), (e) Comparison between the predicted and experimental target discharge capacity. The error bar is calculated from 3 set of experiments. The inset table shows Coulombic efficiency measured from each cell. Note: the design variable prediction of various target discharge capacities 150, 175, and 200 mAh/g based on MICE database are denoted as MICE@150, MICE@175, and MICE@200, respectively.

Next, C-rate is another crucial variable that determines the battery performance. At a high C-rate, the electrochemical reaction is faster than the diffusion of sluggish Li^+; resulting in the Li-ion diffusion predominately occurring near the interface between the electrolyte solution and the cathode surface, while bulk NCM far from the separator is severely underutilized [32].

In general, the average predicted cut-off voltage increases as the value of target discharge capacity increases; this agrees with the previous experimental results [33,34]. However, without specifying the Ni content and cut-off voltage in our dataset, surprisingly, the model was able to accurately predict the correlation between Ni content and cut-off voltage, where high Ni content always results in low cut-off voltage. Moreover, since NCM is a ternary system, increasing one component decreases either or both components, which means increasing Ni leads to decreasing the Co and Mn content ratio over Ni. Lacking Mn^{4+}, the structure stabilizer, Ni-rich cathodes are more vulnerable to the phase transition, leading to low cut-off voltage [35]. To prove our model's validity, we carried out the design-to-device pipeline approach.

2.5. Experimental validation

Table 1 summarizes the predicted design variables and the experimentally used parameters to validate the prediction. The full set of predicted design variables is available in Tables S1, S2, and S3. In this prediction, we also considered the processing cost, which was imposed as a constraint to predict the design variables. By doing so, we found the processing condition with the lowest possible dependency on the Co component while achieving the desired target discharge capacity.

In order to verify the crystal structure of the synthesized NCM samples, the peaks of each sample were indexed by the Miller indices based on layered hexagonal α-$NaFeO_2$ structure (space group *R-3 m, No. 160*), with no obvious impurity peak as shown in Fig. S4 [31,36]. We further refined XRD data using the Rietveld method (Fig. S5(a)-(c)), of which results are summarized in Fig. S5(d). From the calculated values of unit cell parameters in Fig. S5(d), the values *c/a* are greater than 4.9362 (required *c/a* ratio for oxygen lattice distortion), indicative of a well-developed layered structure.

We also notice the $R(I_{003}/I_{104}) > 1.2$ [37,38], corresponding to the lowest degree of desired cation mixing. From three synthesized cathodes, NCM synthesized with a predicted target discharge capacity of 150 mAh/g exhibits the best-layered structure and the lowest degree of Li^+/Ni^{2+} mixing, which is beneficial for Li-ion transport and good electrochemical performance. We further characterized and verified the compositions of the as-synthesized NCM cathodes by elemental mapping using a scanning transmission electron microscopy combined with energy dispersive X-ray spectroscopy (STEM-EDS) (Fig. S6(a)-(i)). The dark mode-field STEM images revealed the spherical structures of each NCM cathode materials. From the EDS elemental mapping, we confirmed the existence of Ni, Co, and Mn that are homogenously mixed. The composition ratios obtained from EDS are compared to those of the predicted data by machine learning model and the measured data by inductively coupled plasma (ICP) (Fig. S7). In general, the element ratios obtained from EDS are in a good agreement with ICP results.

Fig. 5 shows the experimental validation of the inverse design prediction, where the discharge capacities of different designs are summarized from the voltage-capacity profile. According to the inverse design prediction from the target discharge capacities, the synthesized NCM powders were designated as MICE@150, MICE@175, and MICE@200, respectively.

The scanning electron microscopy (SEM) image shows the as-prepared NCM powders (Fig. 5(a)-(c)). The secondary particles are densely packed with smaller primary particles. The size of the primary particles decreased as the Mn's ratio increased, which agrees with the previous study [34]. We observed that MICE-based NCM electrodes delivered discharge capacity of ~150.0, ~161.0, and ~209.5 mAh/g for MICE@150, MICE@175, and MICE@200, respectively. Their respective

coulombic efficiencies are 91.7%, 80.0%, and 85.9%, for MICE@150, MICE@175, and MICE@200, respectively. With the chosen sets, we further compared the ML prediction and that of experimental measurement in a series of varied C-rate and cut-off voltage (Table S4), the average percentage error between prediction and experiment is 11%, which indicates the discharge capacity trend is reasonably captured by the ML model.

Finally, we confirm that simple imputation method can be beneficial for our inverse prediction. Without reducing the size of curated dataset, we demonstrated our method can reverse engineer an optimized cathode material with high viability. Nevertheless, to further improve the ML prediction, we suggest to include additional information such as internal cell pressure and density or loading mass because these parameters may impact the end product. Furthermore, if the data are collected from the published journals, filtering out low-quality data can improve the prediction. As ML is sensitive to the input data, improper input values might influence the predictability.

3. Conclusion

In summary, we demonstrated a robust data-driven inverse design to rationally explore the design parameters without prior experimental knowledge. We performed data imputation to "fill" out the missing values using several imputation techniques and obtained a trained model with a remarkable predictive capability ($R^2 = 0.99$). Subsequently, we used the inverse design technique to predict the design variables for desired target discharge capacities of 150, 175, and 200 mAh/g. Lastly, we validated the predicted design rules from the inverse design. The low RMSE of 8.17 mAh/g between experimental measurement and prediction indicates that this approach is highly reliable and able to accelerate the pace of research for Li-ion batteries.

Moreover, the prediction of high discharge capacity based on the dataset imputed via MICE agrees well with the experimental measurement. We expect that the data curation from the literature potentially can eliminate the barrier of knowledge transferability, which slows down the progress in achieving high-performance Li-ion batteries. Broadly speaking, this work facilitates a promising approach that complements the conventional Edisonian methods to accelerate materials development.

CRediT authorship contribution statement

C.L. and S.H. conceived and designed the machine learning model and experiments. C.L. conducted the inverse design, forward modeling, and data imputation. H. K., M. N., S. K., Y. L., G. H. synthesized and characterized the cathode materials. A.B., K. B., Y. S., J. C., S. C., G. P., J. Y. conducted the data curation. All the authors participated in the discussion of the data analysis and manuscript writing. C.L. and S. H. wrote the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the KAIST-funded Global Singularity Research Program for 2020, 2021 and 2022 under award number 1711100689. J.C.A. acknowledges support from the National Science Foundation under Grant TRIPODS + X:RES-1839234 and the Nano/Human Interfaces Presidential Initiative. We also thank the valuable collaboration of the M3I3 initiative board members for their constructive suggestions.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.nanoen.2022.107214.

References

[1] K.A. Severson, P.M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M.H. Chen, M. Aykol, P.K. Herring, D. Fraggedakis, M.Z. Bazan, S.J. Harris, W.C. Chueh, R. D. Braatz, Data-driven prediction of battery cycle life before capacity degradation, Nat. Energy 5 (2019) 383–391.
[2] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, Phys. Rev. B 9 (2014), 094104.
[3] R.P. Joshi, J. Eickholt, L.L. Li, M. Fornari, V. Barone, J.E. Peraltata, Machine learning the voltage of electrode materials in metal-ion batteries, ACS Appl. Mater. Interfaces 20 (2019) 18494–18503.
[4] G. Hautier, C.C. Fischer, A. Jain, T. Mueller, G. Ceder, Finding nature's missing ternary oxide compounds using machine learning and density functional theory, Chem. Mater. 12 (2010) 3762–3767.
[5] P.M. Attia, A. Grover, N. Jin, K.A. Severson, T.M. Markov, Y.-H. Liao, M.H. Chen, B. Cheong, N. Perkins, Z. Yang, P.K. Herring, M. Aykol, S.J. Harris, R.D. Braatz, S. Ermon, W.C. Chueh, Closed-loop optimization of fast-charging protocols for batteries with machine learning, Nature 7795 (2020) 397–402.
[6] A.D. Sendek, E.D. Cubuk, E.R. Antoniuk, G. Cheon, Y. Cui, E.J. Reed, Machine learning-assisted discovery of solid Li-ion conducting materials, Chem. Mater. 2 (2019) 342–352.
[7] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: the materials project: a materials genome approach to accelerating materials innovation, APL Mater. 1 (2013), 011002.
[8] C. Draxl, M. Scheffler, NOMAD: the FAIR concept for big data-driven materials science, MRS Bull. 9 (2018) 676–682.
[9] S. Curtarolo, W. Setyawan, G.L.W. Hart, M. Jahnatek, R.V. Chepulskii, R.H. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M.J. Mehl, H.T. Stokes, D.O. Demchenko, D. Morgan, AFLOW: an automatic framework for high-throughput materials discovery, Comput. Mater. Sci. 58 (2012) 218–226.
[10] 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), J. Oper. Manag. 11 (2013) 1501–1509.
[11] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z.Q. Rong, O. Kononova, K. A. Persson, G. Ceder, A. Jain, Unsupervised word embeddings capture latent knowledge from materials science literature, Nature 7763 (2019) 95–98.
[12] O. Kononova, H. Huo, T. He, Z. Rong, T. Botari, W. Sun, V. Tshitoyan, G. Ceder, Text-mined dataset of inorganic materials synthesis recipes, Sci. Data 1 (2019) 203.
[13] H. Huo, Z. Rong, O. Kononova, W. Sun, T. Botari, T. He, V. Tshitoyan, G. Ceder, Semi-supervised machine-learning classification of materials synthesis procedures, npj Comput. Mater. 1 (2019) 62.
[14] E. Kim, K. Huang, A. Saunders, A. McCallum, G. Ceder, E. Olivetti, Materials synthesis insights from scientific literature via text extraction and machine learning, Chem. Mater. 21 (2017) 9436–9444.
[15] B.P. MacLeod, F.G.L. Parlane, T.D. Morrissey, F. Hase, L.M. Roch, K.E. Dettelbach, R. Moreira, L.P.E. Yunker, M.B. Rooney, J.R. Deeth, V. Lai, G.J. Ng, H. Situ, R. H. Zhang, M.S. Elliott, T.H. Haley, D.J. Dvorak, A. Aspuru-Guzik, J.E. Hein, C. P. Berlinguette, Self-driving laboratory for accelerated discovery of thin-film materials, Sci. Adv. 20 (2020) eaaz8867.
[16] K. Min, B. Choi, K. Park, E. Cho, Machine learning assisted optimization of electrochemical properties for Ni-rich cathode materials, Sci. Rep. 1 (2018) 15778.
[17] J.H. Kim, K.J. Park, S.J. Kim, C.S. Yoon, Y.K. Sun, A method of increasing the energy density of layered Ni-rich Li Ni1-2xCoxMnx O-2 cathodes (x=0.05, 0.1, 0.2), J. Mater. Chem. A 6 (2019) 2694–2701.
[18] F. German, A. Hintennach, A. LaCroix, D. Thiemig, S. Oswald, F. Scheiba, M. J. Hoffmann, H. Ehrenberg, Influence of temperature and upper cut-off voltage on the formation of lithium-ion cells, J. Power Sources 264 (2014) 100–107.
[19] L. Beretta, A. Santaniello, Nearest neighbor imputation algorithms: a critical evaluation, BMC Med. Inform. Decis. 3 (2016) 74.
[20] F. Tang, H. Ishwaran, Random forest missing data algorithms, Stat. Anal. Data Min. 6 (2017) 363–377.
[21] M. Kokla, J. Virtanen, M. Kolehmainen, J. Paananen, K. Hanhineva, Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: a comparative study, BMC Bioinforma. 1 (2019) 492.
[22] S. van Buuren, K. Groothuis-Oudshoorn, Mice: multivariate imputation by chained equations in R, J. Stat. Softw. 3 (2011) 1–67.
[23] M.J. Azur, E.A. Stuart, C. Frangakis, P.J. Leaf, Multiple imputation by chained equations: what is it and how does it work? Int. J. Methods Psychiatr. Res. 1 (2011) 40–49.
[24] Y. Deng, C. Chang, M.S. Ido, Q. Long, Multiple imputation for general missing data patterns in the presence of high-dimensional data, Sci. Rep. 6 (2016) 21689.
[25] B.H. Menze, B.M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F. A. Hamprecht, A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data, BMC Bioinform. 1 (2009) 213.

[26] N. Nitta, F.X. Wu, J.T. Lee, G. Yushin, Li-ion battery materials: present and future, Mater. Today (2015).
[27] J.N. Kumar, Q. Li, K.Y.T. Tang, T. Buonassisi, A.L. Gonzalez-Oyarce, J. Ye, Machine learning enables polymer cloud-point engineering via inverse design, npj Comput. Mater. 1 (2019) 73.
[28] Y.C. Xie, C. Zhang, X.Q. Hu, C. Zhang, S.P. Kelley, J.L. Atwood, J. Lin, Machine learning assisted synthesis of metal-organic nanocapsules, J. Am. Chem. Soc. 3 (2020) 1475–1481.
[29] X. Chen, D. Chen, M.Y. Weng, Y. Jiang, G.W. Wei, F. Pan, Topology-based machine learning strategy for cluster structure prediction, J. Phys. Chem. Lett. 11 (2020) 4392–4401.
[30] Y. Xia, J. Zheng, C. Wang, M. Gu, Designing principle for Ni-rich cathode materials with high energy density for practical applications, Nano Energy 49 (2018) 434–452.
[31] S.H. Lee, S. Lee, B.S. Jin, H.S. Kim, Optimized electrochemical performance of Ni rich LiNi0.91Co0.06Mn0.03O2 cathodes for high-energy lithium ion batteries, Sci. Rep. 9 (2019) 8901.
[32] X.K. Lu, A. Bertei, D.P. Finegan, C. Tan, S.R. Daemi, J.S. Weaving, K.B. O'Regan, T. M.M. Heenan, G. Hinds, E. Kendrick, D.J.L. Brett, P.R. Shearing, 3D microstructure design of lithium-ion battery electrodes assisted by X-ray nano-computed tomography and modelling, Nat. Commun. 1 (2020) 2079.
[33] R. Jung, M. Metzger, F. Maglia, C. Stinner, H.A. Gasteiger, Oxygen release and its effect on the cycling stability of LiNixMnyCozO2 (NMC) cathode materials for Li-ion batteries, J. Electrochem. Soc. 7 (2017) A1361–A1377.
[34] J. Ahn, D. Susanto, J.K. Noh, G. Ali, B.W. Cho, K.Y. Chung, J.H. Kim, S.H. Oh, Achieving high capacity and rate capability in layered lithium transition metal oxide cathodes for lithium-ion batteries, J. Power Sources 360 (2017) 575–584.
[35] T. Li, X.-Z. Yuan, L. Zhang, D. Song, K. Shi, C. Bock, Degradation mechanisms and mitigation strategies of nickel-rich NMC-based lithium-ion batteries, Electrochem. Energy Rev. 1 (2020) 43–80.
[36] X.K. Yang, R.Z. Yu, L. Ge, D. Wang, Q.L. Zhao, X.Y. Wang, Y.S. Bai, H. Yuan, H. B. Shu, Facile synthesis and performances of nanosized Li2TiO3-based shell encapsulated LiMn1/3Ni1/3Co1/3O2 microspheres, J. Mater. Chem. A 22 (2014) 8362–8368.
[37] X. Zhang, W.J. Jiang, A. Mauger, Qilu, F. Gendron, C.M. Julien, Minimization of the cation mixing in Li1+x(NMC)1−xO2 as cathode material, J. Power Sources 5 (2010) 1292–1301.
[38] Q.K. Guo, J.L. Huang, Z. Liang, H. Potapenko, M.M. Zhou, X.D. Tang, S.W. Zhong, The use of a single-crystal nickel-rich layered NCM cathode for excellent cycle performance of lithium-ion batteries, New J. Chem. 7 (2021) 3652–3659.
[39] S. Hong, et al., Reducing time to discovery: materials and molecular modeling, imaging, informatics, and integration, ACS Nano 15 (3) (2021) 3971–3995, https://doi.org/10.1021/acsnano.1c00211.