PNAS INAUGURAL ARTICLE | ENGINEERING

Data-driven electrolyte design for lithium metal anodes

Sang Cheol Kim[a,1], Solomon T. Oyakhire[b,1], Constantine Athanitis[a], Jingyang Wang[a], Zewen Zhang[a], Wenbo Zhang[a], David T. Boyle[c], Mun Sek Kim[b], Zhiao Yu[c], Xin Gao[a], Tomi Sogade[a], Esther Wu[a], Jian Qin[b], Zhenan Bao[b], Stacey F. Bent[b,d,2], and Yi Cui[a,d,e,2]

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2022.

Contributed by Yi Cui; received September 2, 2022; accepted January 5, 2023; reviewed by Ju Li, Venkat Viswanathan, and Chunsheng Wang

Improving Coulombic efficiency (CE) is key to the adoption of high energy density lithium metal batteries. Liquid electrolyte engineering has emerged as a promising strategy for improving the CE of lithium metal batteries, but its complexity renders the performance prediction and design of electrolytes challenging. Here, we develop machine learning (ML) models that assist and accelerate the design of high-performance electrolytes. Using the elemental composition of electrolytes as the features of our models, we apply linear regression, random forest, and bagging models to identify the critical features for predicting CE. Our models reveal that a reduction in the solvent oxygen content is critical for superior CE. We use the ML models to design electrolyte formulations with fluorine-free solvents that achieve a high CE of 99.70%. This work highlights the promise of data-driven approaches that can accelerate the design of high-performance electrolytes for lithium metal batteries.

battery | energy storage | machine learning | electrolyte

Significance

The liquid electrolyte plays a central role in lithium metal batteries, particularly in dictating the battery cyclability. Electrolyte engineering in recent years has become a promising strategy to improve cyclability in lithium metal batteries. However, owing to the complexity of electrolyte design, prediction of Coulombic efficiency and efficient design of electrolytes remains challenging. In this work, we adopt a data-driven approach to build machine learning (ML) models for electrolyte design. From our models, we extract an important yet previously unidentified insight that low solvent oxygen content can lead to superior cyclability. Leveraging this insight as a strategy, we introduce a series of electrolytes with high stability and cyclability in lithium metal batteries.

Electrification of transportation is crucial for reducing carbon footprint and mitigating climate change (1). For wider deployment of electric vehicles, however, rechargeable batteries with energy densities up to 500 Wh kg^{-1} are in demand (2, 3). The lithium metal anode, regarded as the holy grail of anode technology, is essential for achieving 500 Wh kg^{-1} (2). Despite their high energy densities, lithium metal anodes are not widely deployed in today's smartphones and electric vehicles as they suffer from poor cyclability that stems from excessive side reactions and the formation of inactive lithium (4, 5). Alongside host structures (6–8) and interfacial engineering (9, 10), electrolyte engineering has recently emerged as a promising strategy for improving the cycling efficiency of lithium metal anodes (11–15). Electrolyte engineering has pushed the Coulombic efficiency (CE), a sought-after metric of cycle life defined as the discharge capacity over the charge capacity, to 99.5% (13, 16) and enabled more than 200 cycles in an anode-free configuration (14), propelling the lithium metal anode closer to practical application.

The promise of electrolyte engineering has garnered increased attention and research efforts (11–14, 16–19). The resulting accumulation of data presents an opportunity for a data-driven study that can extract previously overlooked insights and provide new electrolyte engineering solutions. Machine learning (ML) has previously been deployed for battery research, where a major research direction utilizes simulation-derived data to train ML models. Sendek et al. and Ahmad et al. built models to identify potential solid electrolyte materials with high conductivity and predict lithium deposition, respectively (20, 21), while Nakayama et al. developed an ML model to estimate the coordination energy of solvents with Li^+ in liquid electrolytes based on density functional theory (DFT) calculations and physical properties (22). These computation-based studies provide an avenue to effectively screen a large number of materials and molecules. However, first principles-based modeling at the battery cell level is prohibitively expensive (23), and as these studies focus on computed results, many of them stop short of providing experimental validation of the model.

Empirical data-driven approaches can be effective complementary strategies (23). There are numerous studies that utilize ML for battery diagnostics, such as the estimation of the state of charge (SOC) and cycle life (24–27). Severson et al. conducted a series of notable investigations in which they developed an ML model that predicts the cycle life of battery cells (26) and utilized the model in a closed-loop Bayesian optimization algorithm to obtain advanced fast-charging protocols (27). As such, experimental data-driven methods can lead to models that effectively predict battery performance and guide the design of practical solutions. Most of these models, however, are confined to battery diagnostics for a single cell design, and models capable of predicting the performance of various designs and probing molecular intricacies are rare.

In this study, we collect a dataset spanning a large design space to develop supervised ML models that can assist the prediction and optimization of CE of various electrolytes

Author contributions: S.C.K., S.T.O., J.Q., Z.B., S.F.B., and Y.C. designed research; S.C.K., S.T.O., C.A., J.W., Z.Z., D.T.B., M.S.K., Z.Y., X.G., T.S., and E.W. performed research; Z.Z., W.Z., and Z.Y. contributed new reagents/analytic tools; S.C.K., S.T.O., C.A., J.W., Z.Z., W.Z., D.T.B., M.S.K., X.G., and E.W. analyzed data; J.Q. and Z.B. supervised the project; and S.C.K., S.T.O., and Y.C. wrote the paper.

Reviewers: J.L., Massachusetts Institute of Technology; V.V., Carnegie Mellon University; and C.W., University of Maryland at College Park A James Clark School of Engineering.

The authors declare no competing interest.

[1]S.C.K. and S.T.O. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: sbent@stanford.edu or yicui@stanford.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2214357120/-/DCSupplemental.

Published February 27, 2023.

for lithium metal anodes (Fig. 1*A*). Because the electrolyte's effect on CE is highly complex (28) and prohibitively expensive to track using simulations (23), a data-driven approach is particularly fitting for this study. We organized a dataset of 150 entries of Li | Cu cell CE with various liquid electrolytes from the literature and experiments and used the elemental compositions of the liquid electrolytes as features of our model. Using the dataset, we train and test numerous ML models such as linear regression, support vector machine, bagging, and random forest models. We identify that solvent oxygen ratio (sO) is the most important feature, where lower sO correlates with higher CE. Utilizing the model as a design guideline, we introduce novel electrolyte formulations that use fluorine-free solvents that achieve CEs up to 99.70%. We envision that this work can initiate data-driven approaches to battery electrolyte engineering.

Results and Discussion

Data Acquisition and Processing. The availability of data is a key enabler of a data-driven study (23). Considerable attention to electrolyte engineering research has led to a build-up of data on different electrolytes in the literature (11–13, 16, 17), enabling the data collection for our study. We collected 150 data points of Li | Cu cell CE, ranging from 80% to 99.5%, of different electrolytes for our dataset—the complete dataset and the references are available in the supporting files. We used Li|Cu data for our study because such cell configurations have larger dataset representation in the literature, are easy to reproduce across groups, and enable assessments of Li-electrolyte stability without the confounding effects of cathodes. The set of electrolytes includes conventional, high concentration, localized high concentration, fluorinated, dual salt, and additive-enhanced electrolytes, featuring 54 solvents and 17 salts (Fig. 1*B* and *SI Appendix*, Table S1 and Figs. S1–14). As different experimental conditions can affect the measured CE values (29), our Li | Cu cell data were restricted to a narrow current range of 0.4 to 1 mA cm^{-2} and a capacity range of 0.5 to 1 mAh cm^{-2}.

Elemental compositions were selected to be the input features of our model (Table 1 and Fig. 1*C*). The elemental composition of the electrolyte encompasses information on important properties such as solvation and SEI composition. For example, it has been seen that fluorine decoration on solvents can lead to weaker solvation (13, 30, 31), and high-concentration electrolytes that contain large amounts of Li salts—typically composed of inorganic elements—have anion-rich solvation structures (11, 32). Furthermore, because the SEI is formed from the decomposition of the electrolyte, the SEI

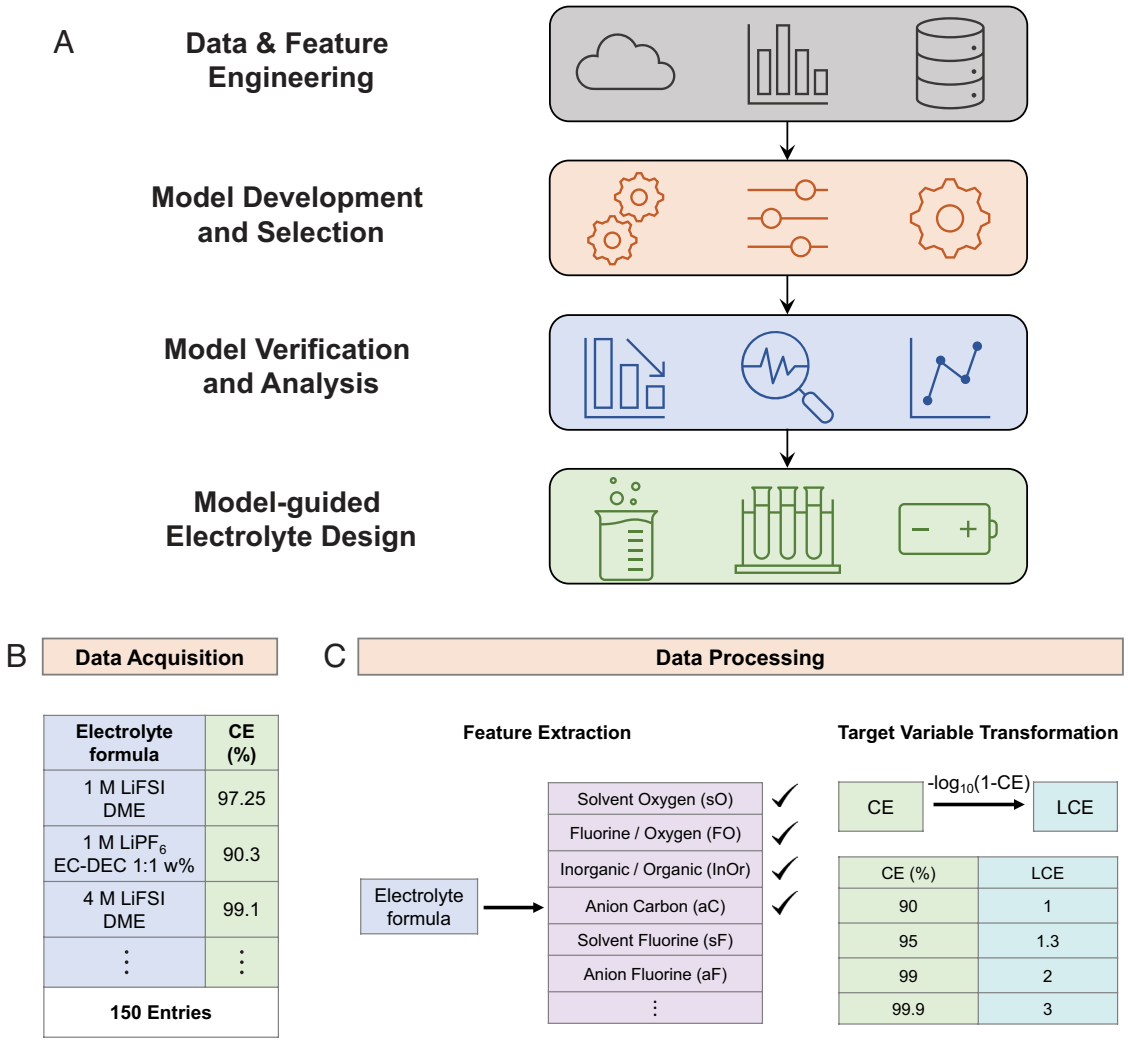

Fig. 1. Data acquisition and processing. (*A*) Workflow from data collection to model development and electrolyte design. (*B*) Li|Cu Coulombic efficiencies (CE) of various electrolytes are collected. (*C*) Elemental composition features are calculated from the electrolyte formula and logarithmic Coulombic efficiency (LCE), obtained through the transformation of CE, is used as the target variable.

Table 1. List of elemental composition features for the ML model

Feature name	Abbreviation
Oxygen ratio	O
Solvent oxygen ratio	sO
Anion oxygen ratio	aO
Carbon ratio	C
Solvent carbon ratio	sC
Anion carbon ratio	aC
Fluorine ratio	F
Solvent fluorine ratio	sF
Anion fluorine ratio	aF
Fluorine/oxygen ratio	FO
Fluorine/carbon ratio	FC
Oxygen/carbon ratio	OC
Inorganic/organic ratio	InOr

composition is intimately related to the electrolyte composition (13, 17). We provide detailed information by specifying the location of the elements and also add elemental ratios as features (Table 1). It is noteworthy that even though elemental composition is beneficial for our ML approach, it does not completely capture molecular nuances like the geometric arrangement of elements in isomers. Nuances of this type have been shown to affect electrolyte properties and battery performance (33, 34), so future studies should encode such molecular details to improve ML robustness.

Although CE is a commonly used metric, it is not necessarily the most relevant and effective metric as a target variable for a ML study. CE has an upper bound of 100%, where many data points are concentrated (*SI Appendix*, Figs. S2–S14). As we approach this limit, significant differences in electrolyte properties will translate to infinitesimally small CE changes. Owing to these reasons, we transformed CE into a more relevant metric named logarithmic Coulombic efficiency (LCE), defined as LCE = $-\log_{10}(1 - \text{CE})$ (Fig. 1*C*). The transformation yields data more suitable for training an ML model (*SI Appendix*, Figs. S2–S14). The LCE metric is highly interpretable as it systematically reflects an increase in CE; for example, an LCE of 1 translates to a 90% CE, an LCE of 2 represents a 99% CE, and an LCE of 3 represents a 99.9% CE (Fig. 1*C*).

Model Development. Like many reports of ML applied to physical sciences (26, 35), our dataset is relatively small, due to the paucity of unique electrolyte mixtures reported in the literature. Owing to the small dataset size and a significant number of input features used in this study, we employ forward stepwise selection (36, 37) to reduce the number of features required for the model. By starting with a model that contains no feature, forward stepwise selection sequentially selects the regression feature that minimizes the residual sum of squares (RSS) the most during each inclusion of a new feature. The dimensionality of this forward stepwise selection is constrained to eight features to avoid overfitting on our training data, indicating that we are selecting the best eight features from the 13 features in our dataset. As shown in Fig. 2*A*, the forward stepwise selection algorithm evolves from a model that contains only solvent oxygen (sO), to a model that includes seven other features. The sole inclusion of sO when the model

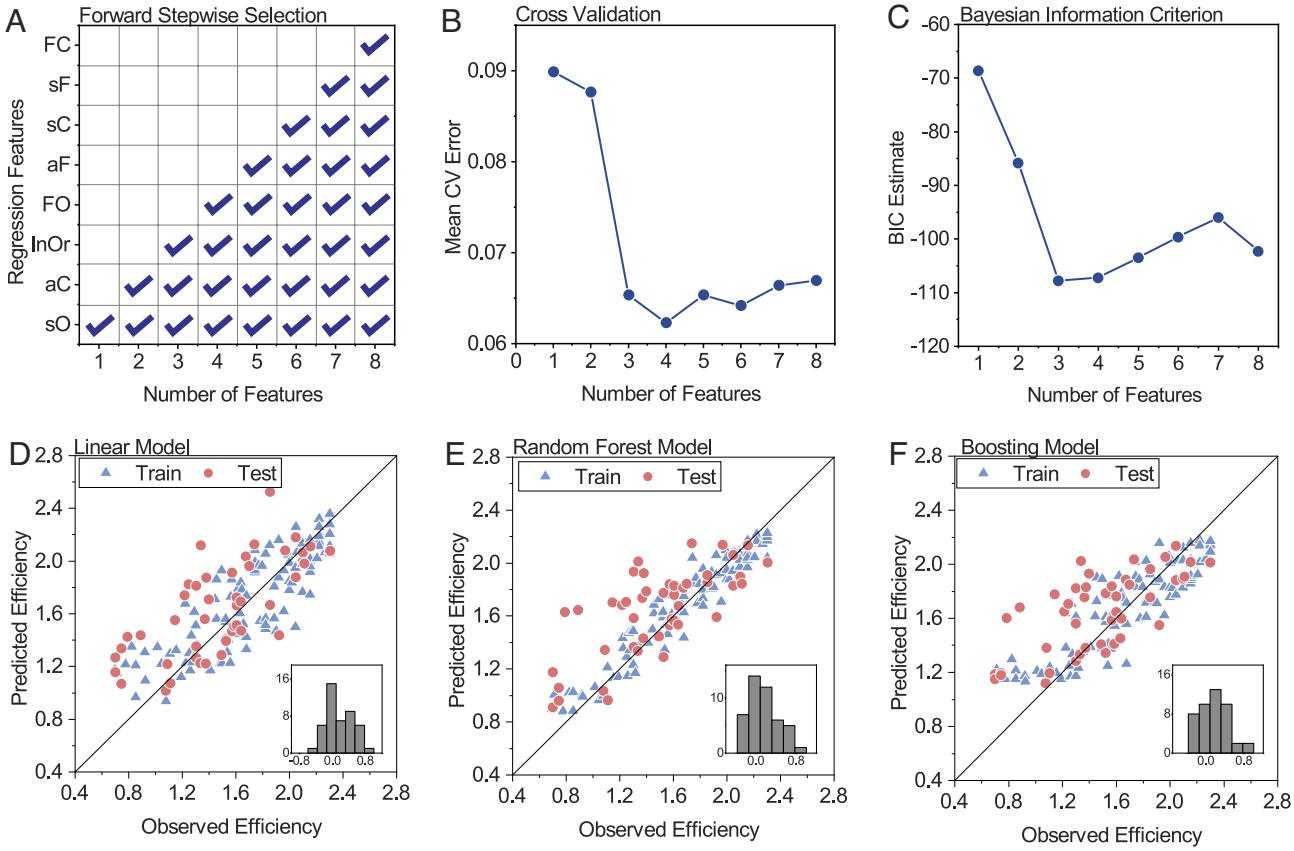

Fig. 2. Model development. (*A*) Reduction of feature space using forward stepwise selection. (*B* and *C*) Cross-validation errors and BIC estimates reveal that the four-feature model shows the smallest errors. (*D*–*F*) Plots of predicted efficiency vs. observed efficiency for the four-feature linear model, random forest model, and boosting model, respectively. Insets show histograms of residuals (predicted–observed) for the test data.

Table 2. Test MSE estimates of various ML algorithms trained using four features (sO, aC, InOr, and FO)

Algorithm	Linear regression	Random forest	Boosting	Bagging
Test MSE	0.343	0.333	0.345	0.341

is reduced to one feature suggests that sO is the most important feature for explaining the variance in LCE. As the number of features increases, aC, InOr, and FO are sequentially included in the model, showing that they are also important for explaining the variance in LCE.

As the number of features increases, the RSS reduces, consistent with our expectations (*SI Appendix*, Fig. S15) (36, 38). However, this apparent improvement in model robustness can be misleading because an increase in the number of features makes models more prone to overfitting. As such, we apply fourfold cross-validation (CV) to identify the number of features that appropriately balances the degree of bias and the extent of variance in our models, as other studies have shown (35, 39). Our CV result shows that the four-feature model has the lowest mean CV error (Fig. 2*B*), which suggests that the model is less susceptible to variation when tested with unseen data. The stability of the four-feature model against unseen data is also verified using multiple partitions of the training dataset (*SI Appendix*, Fig. S16). To further show the robustness of the four-feature model, we calculated the Bayesian information criterion (BIC) for each of the models selected in Fig. 2*A*. The BIC estimate penalizes models with a high number of features because their prediction errors on training sets are usually underestimates of the test error (40). From Fig. 2*C*, the BIC estimate for the four-feature model is also one of the lowest, further suggesting that the four-feature model is robust and useful for test predictions. Using linear regression, we also show comparable test and training mean-squared error (MSE) estimates of 0.34 and 0.23, respectively, for the four-feature model.

We also benchmark our four-feature linear model to other ML models like bagging, random forest, and boosting, showing comparable test performance across the models (Table 2). This shows that the four-feature linear model is robust for predicting LCE. Also, by plotting predicted efficiency against observed efficiency, we observe that most of the test data points for the linear, random forest, and boosting models lie close to the line of perfect prediction and are distributed similarly to the training data points (Fig. 2 *D*–*F*). This suggests that the selected four features are stable for predicting LCE against unseen data. Even though the random forest, bagging, and boosting models provide robust predictions, similar to what other studies have shown (41, 42), we restrict the majority of our model analysis in subsequent segments of the paper to the four-feature linear model because of its high interpretability.

Model Verification and Analysis. To further verify the robustness of our ML models, we prepared a list of candidate electrolytes unseen by the model that represent a large portion of the electrolyte design space to be tested against our model. The list was created by a combination of the following components: three carbonate solvents—ethylene carbonate (EC), dimethyl carbonate (DMC), and fluoroethylene carbonate (FEC); three ether solvents—dimethoxyethane (DME), 1,3-dioxolane (DOL), and tetraethylene glycol dimethyl ether (TEGDME); three salts—lithium bis(fluorosulfonyl)imide (LiFSI), lithium hexafluorophosphate ($LiPF_6$), and lithium bis(trifluoromethanesulfonyl)imide (LiTFSI); four salt concentrations—1 to 4 M; two additives—vinylene carbonate (VC) and lithium nitrate; and the use of a fluorinated diluent, 1,1,2,2-tetrafluoroethyl-2,2,3,3-tetrafluoropropyl ether (TTE). The list includes carbonate-based electrolytes, ether-based electrolytes, high-concentration electrolytes (HCE), localized high-concentration electrolytes (LHCE), and additive-enhanced electrolytes (*SI Appendix*, Table S2). Among the 34 candidate electrolytes, we selected 13 electrolytes across the different types of electrolytes to test our model. Fig. 3*A* shows the predicted efficiencies plotted against measured efficiencies, where the formula for each numbered electrolyte can be found in *SI Appendix*, Table S2. We observe a close match between measured and predicted efficiencies and a rms error of 0.17. This further validates that our model works reliably well across many commonly used electrolytes.

One of the advantages of a linear model is its interpretability (43). Our model has the form

$$LCE = -4.481sO + 0.052FO + 0.188InOr - 13.460aC + 2.112,$$

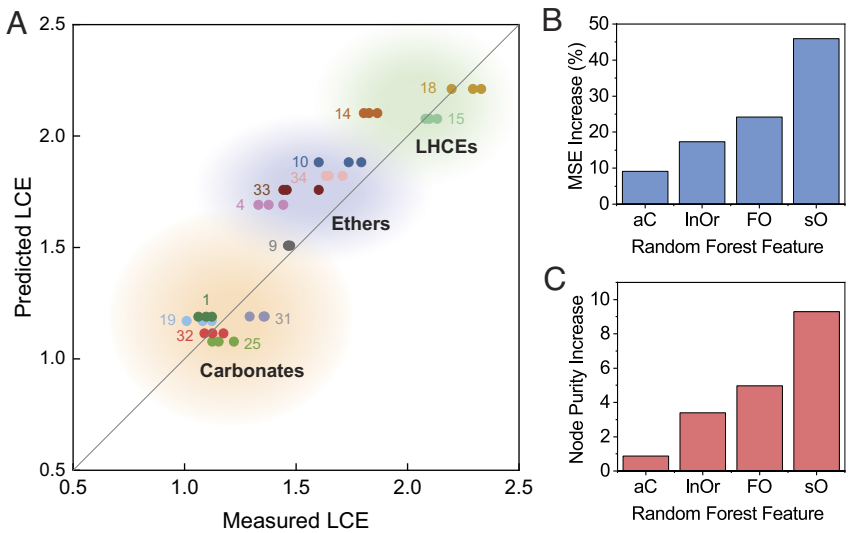

Fig. 3. Model verification and analysis. (*A*) Experimental verification of the model through LCE measurements of 13 electrolytes unseen by the model, which shows agreement between predicted and measured efficiencies. (*B*) Mean-squared error increase and (C) node purity increase show the relative importance of the four features.

consisting of four features—sO, fluorine to oxygen ratio (FO), inorganic to organic elemental ratio (InOr), and anion carbon ratio (aC). sO and aC have negative coefficients, indicating that decreasing solvent oxygen content and anion carbon content leads to superior CE. On the other hand, FO and InOr have positive coefficients, signifying that greater inorganic content (defined here as elements excluding carbon and hydrogen) and fluorine content over oxygen content are favorable for higher CE. To further evaluate the relative importance of the four features, we used the random forest model, averaged over 500 decision trees, to analyze the importance of the four features based on percentage MSE increase and node purity (Fig. 3 *B* and *C*). Our results show that the order of importance of these features is aC, InOr, FO, then sO, with sO being the most important feature. It is noteworthy that these features are not collinear, as evidenced by their low variance inflation factors (*SI Appendix*, Fig. S17).

The most notable yet previously unidentified insight extracted from this analysis is that sO is the most important determinant of CE (Fig. 3 *B* and *C*). An example that illustrates this point is the comparison between carbonates and ether electrolytes; ethers contain fewer oxygen atoms than carbonates and are known to be more stable with Li metal (*SI Appendix*, Fig. S18) (28, 44). Another example is the comparison between DME and diethoxyethane (DEE); DEE, which has a lower sO than DME, has a significantly improved CE (*SI Appendix*, Fig. S18) (19). This suggests that reducing sO may be a promising strategy for improving CE, which we deploy in the following section to design novel high-performance electrolytes. FO comes second in importance; increasing fluorine content while decreasing oxygen content has been shown to correlate with high performance (13, 16, 17, 45). InOr is another significant factor to LCE, where higher inorganic content in the SEI has often been correlated to improved CE (16, 30, 46). Lastly, aC shows a negative impact on CE, which can be exemplified by the comparison between LiFSI and LiTFSI.

The importance of solvent oxygen content over fluorine content is unexpected, as fluorine has been the recent focus of electrolyte design for lithium metal batteries (13, 31, 44, 45, 47, 48). It is thought to be a critical element in electrolytes (44) that leads to the formation of LiF in the SEI, which has been correlated with superior cycling performance (17, 49). In addition, fluorinated solvents have been instrumental in the success of many high-performance electrolytes (12, 13, 16, 28). Indeed, fluorine appears in the FO term and is a major contributor to the InOr term, both of which indicate that fluorine may play an important role. However, our model analysis reveals that sO is the strongest predictor of CE, and scatterplots show that high CEs above 99% can be attained at 0 solvent fluorine content (*SI Appendix*, Fig. S9), highlighting the role of solvent oxygen which had previously received little attention. A plausible hypothesis is that the effects of fluorine on solvation and the SEI may be attainable by tuning solvent oxygen content. Fluorine decoration of solvents is known to weaken the solvation of Li^+ (13, 30), and similar effects may be attainable through reducing the solvent oxygen content. In localized high-concentration electrolytes, the fluorinated co-solvents play a role of diluting the solution while being inactive in solvation and could potentially be substituted with a non-solvating non-fluorinated co-solvent. In the following discussion, we explore the effectiveness of tuning solvent oxygen content as a strategy to achieve high CEs.

Model-guided Electrolyte Design. Our model analysis revealed that solvent oxygen content is the most important feature in our model, and reducing sO may be a strategy to improve CE. We deployed this strategy to design novel high-performance electrolytes. Fig. 4*A* shows the sO of several ether solvents. We sought to find ether solvents that have lower sO values than DME and identified methyl butyl ether (MBE), methyl t-butyl ether (MTBE), and dibutyl ether (DBE) with sO values of 0.056, 0.056, and 0.037, respectively. Using these solvents mixed with non-solvating diluents (50), we designed four electrolytes that our model predicted to have high performance. As shown in Fig. 4*B*, these electrolytes show average CEs as high as 99.70%. Electrolytes using MTBE and DBE mixed with toluene are particularly notable, not only for the high CE of 99.64% and 99.70%, but also for their potential for mass production; MTBE and DBE are common solvents used in the petroleum industry (51, 52), and toluene is an important aromatic hydrocarbon solvent (53). The fluorine-free solvents of this electrolyte are interesting from a fundamental aspect as well, as most high-performance electrolytes use highly fluorinated solvents and co-solvents (12, 13, 16, 17, 54–56) that can alter solvation and the SEI. Fluorinated co-solvents are key to the solvation structure and high performance of localized high-concentration electrolytes (LHCE) (12, 16, 55). In addition, lithium fluoride (LiF) in the SEI that arises from the decomposition of fluorinated species is believed to play a major role in stabilizing the lithium metal cyclability (17, 54, 56). Therefore, understanding the working mechanism of these electrolytes is key to understanding the ML model.

To characterize the solvation properties of these electrolytes, we compared the solvation free energy of Li^+ of 1 M LiFSI MTBE-toluene against a reference electrolyte of 1 M LiFSI DME using a technique recently developed in our group (Fig. 4*C*). This technique probes the relative solvation energy of an electrolyte by measuring the potential of an electrochemical cell that contains symmetric electrodes and asymmetric electrolytes (30). Our measurement shows that 1 M LiFSI MTBE-toluene has much weaker solvation, revealing a difference of 403 meV or 38.9 kJ mol^{-1}. Solvation is intricately related to the SEI, which we characterized using cryogenic electron microscopy. The HR-TEM images, shown in Fig. 4 *D* and *E* and *SI Appendix*, Figs. S19 and S20, reveal that 1 M LiFSI MTBE-toluene shows a thin SEI of ~7 nm, approximately 2 nm thinner than 1 M LiFSI DME. In addition, the stronger contrast of the SEI of 1 M LiFSI MTBE-toluene suggests denser and more inorganic nature of the SEI. Energy-dispersive X-ray spectroscopy (EDX) illustrates that 1 M LiFSI MTBE-toluene has a more oxygen-rich SEI (Fig. 4*F* and *SI Appendix*, Fig. S21). This is particularly interesting as the solvent with less oxygen leads to a more oxygen-rich SEI. We infer that this effect is due to suppressed solvent decomposition that likely resulted from the decrease in sO, which leads to a more anion-derived SEI, a phenomenon seen in other high-performance electrolytes (12, 13, 16). Fig. 4 *G* and *H* and *SI Appendix*, Figs. S22 and S23 show the scanning electron microscopy (SEM) images of the 0.5 mAh cm^{-2} of lithium deposited on copper foil. Insets at high magnification show that the 1 M LiFSI MTBE-toluene has larger lithium deposits with smaller surface area in comparison with 1 M LiFSI DME. A particularly interesting observation is that, whereas 1 M LiFSI DME shows full Li coverage, 1 M LiFSI MTBE-toluene has islands of lithium deposits. This further demonstrates the smaller surface area of the lithium deposits formed in our electrolyte and also hints that during the course of deposition, growth of existing Li nuclei may be favored over nucleation of Li.

We also verify the electrolytes' oxidative stability and the compatibility with commercial cathodes (*SI Appendix*, Figs. S24–S26). The oxidative stability of our MTBE-toluene electrolyte likely originates from an increase in reduction and oxidation voltage conferred by toluene, consistent with previous reports (57, 58).

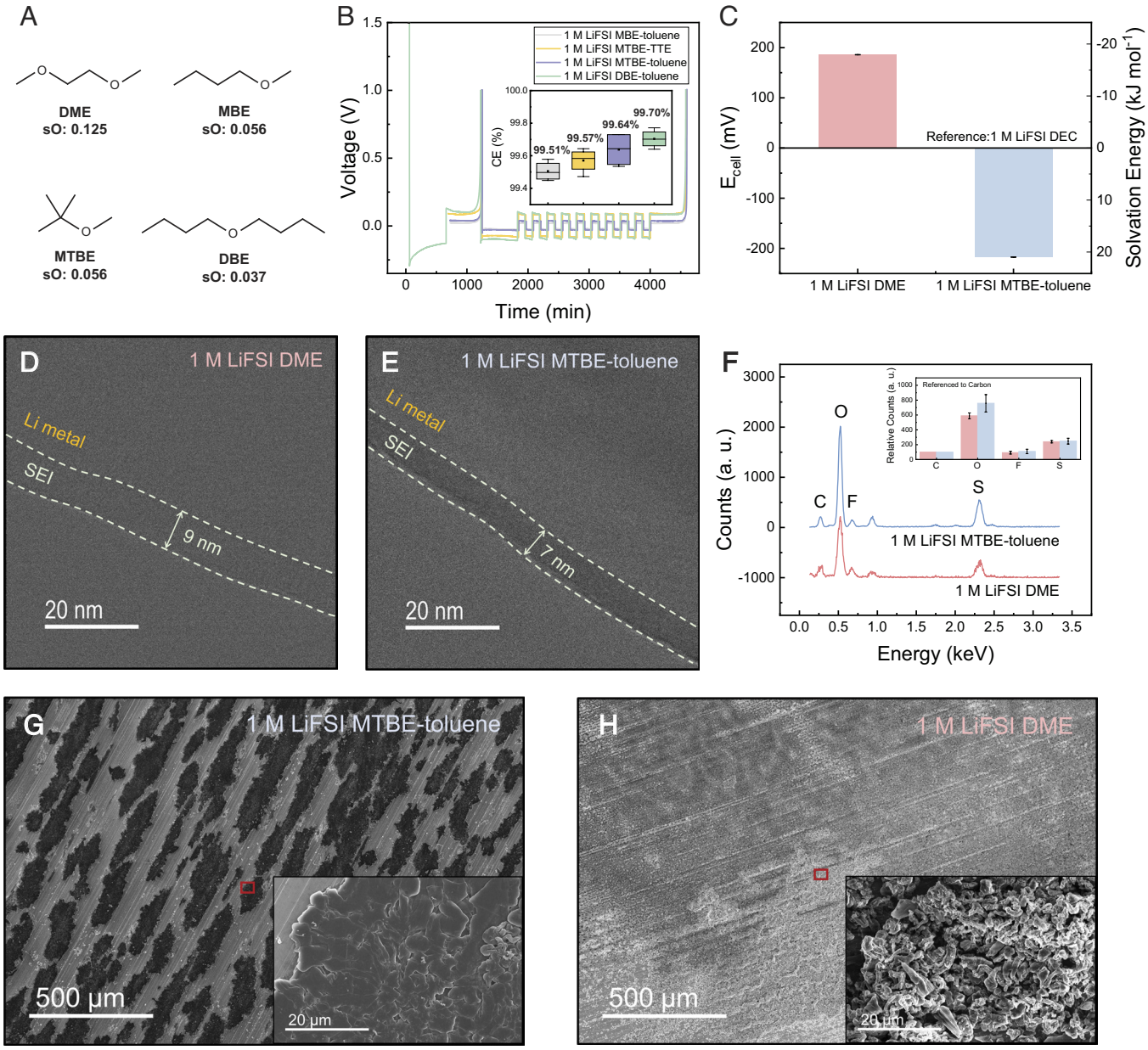

Fig. 4. Model-guided design of electrolytes. (*A*) Molecular structure and sOs of four solvents, showing that our model promotes the use of MBE, MTBE, and DBE, which have lower sO than commonly used DME. (*B*) CE measurements of the four novel electrolytes show high CEs up to 99.70%. (*C*) Solvation energy of 1 M LiFSI DME and 1 M LiFSI MTBE-toluene, which demonstrates that the latter has significantly weaker solvation. (*D* and *E*) Cryo-TEM images of the SEI show that 1 M LiFSI MTBE-toluene has a thinner SEI. (*F*) Cryo-EDX shows that SEI of 1 M LiFSI MTBE-toluene has greater oxygen content. (*G* and *H*) SEM images of Li deposits on copper using the two electrolytes show that 1 M LiFSI MTBE-toluene has larger deposits with a smaller surface area.

Overall, we see that the electrolyte with a superior CE has weaker solvation and a more anion-derived SEI as well as larger Li deposits, which is consistent with our previous proposition (30). We also demonstrate that tuning the solvent oxygen content can achieve high CE for lithium metal anodes, adding support to the hypothesis that the effect of fluorinated solvents in weakening solvation could be emulated by decreasing sO. Taken together, our results indicate that low sO improves CE by reducing the likelihood of solvent decomposition, increasing the formation of favorable anion-rich SEI species, and reducing the contact area of Li deposits.

Our model unveils important insights that can accelerate the design of high-performance electrolytes. Nevertheless, there are ample opportunities for improvements to our model and future developments of complex models with greater accuracy and wider applicability. Our model is trained on a dataset comprised of literature data, which is concentrated toward electrolytes that work reasonably well. Due to the nature of our dataset, insights regarding unexplored classes of electrolytes may be limited. Feeding additional data or interweaving computational methods with our model can be effective approaches to add to the generalizability of the model. Also, greater sensitivity to additives such as FEC and lithium nitrate that are used in small quantities could be improved by distinguishing additives from solvents and salts. Furthermore, capturing the molecular structural information could be possible by using simplified molecular-input line-entry system (SMILES) (59). We envision that our model can serve as a first-generation model that can be modified and improved for a more powerful model with improved predictive capabilities.

Conclusions

This report introduces a data-driven approach to electrolyte optimization. We built ML models upon Li | Cu CE data from literature and experiments, and we found that among the different features that affect CE, solvent oxygen content stands out as the most decisive factor. Leveraging the insight as a design strategy, we introduced a series of electrolytes using fluorine-free solvents that can achieve high CEs up to 99.70%. Our model and electrolytes suggest that careful control of solvent oxygen content can lead to weak solvation and inorganic SEI that form the basis for high Li metal compatibility. We believe that this work highlights the promise of data-driven approaches in building models and identifying overlooked insights, and we hope that our work can provide some directions to reaching CEs of 99.9% and beyond.

Methods

Data Acquisition and Processing. Data containing Coulombic efficiencies of Cu | Li cells with current and capacity close to 0.5 mA cm^{-2} and 1.0 mAh cm^{-2} were collected from the literature. Using the electrolyte formula, the elemental composition data of each electrolyte were extracted. Number of moles per liter of solvent was calculated by using the molar mass and density. Using the chemical formula of the solvents, the elemental composition of the solvents could be calculated. Density of solvent mixtures was assumed to be weighted averages of the components, and the salt content was approximated to have negligible effect on the total volume of the solution. The anion elemental composition could also be calculated by using the molar concentration and the salt chemical formula. Elemental composition could be calculated by taking the ratio between the number of atoms of the element over the total number of atoms in a liter of solution. For example, carbon solvent ratio could be calculated by taking the ratio between the number of carbon atoms in the solvent with the total number of atoms in the electrolyte. Additives were included as either salt or solvent, depending on their chemical forms. An example of the feature data extraction process is shown in Supporting Note. Inorganic content refers to all elements except carbon and hydrogen and lithium. Organic content refers to only carbon. InOr is a ratio between inorganic content and organic content. Hydrogen and lithium were excluded from the two categories, because they are well known to be present in both organic and inorganic materials. CE of the acquired data is converted into LCE through the equation LCE = −log (1 – CE).

ML Model Development. This work involved model fitting (estimating the coefficient values) and model selection (establishing the model structure). To perform both simultaneously, subset selection was used.

A linear model of the form:

$$\hat{y}_i = \hat{w}^T x_i,$$

was initially proposed, where $\hat{y}_i$ represents the predicted LCE of a given electrolyte i, x_i represents a p-dimensional feature vector for a given electrolyte, and $\hat{w}^T$ represents the p-dimensional coefficient vector. To perform subset selection, all possible combinations of all features for each number of desired features (1–7) were examined against the training data, and the best feature combinations were selected using residual sum of squares (RSS) and Bayesian information criterion (BIC).

The formula for RSS is

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2,$$

where y_i is the n-dimensional vector for all experimentally observed LCEs and n represents the number of observations.

The formula for BIC is

$$BIC = \frac{1}{n\hat{\sigma}^2} (RSS + \log(n) d\hat{\sigma}^2),$$

where $\hat{\sigma}^2$ is an estimate of the variance of error associated with each LCE measurement and d is the number of features.

The selected subsets were subsequently tested against the training data using fourfold cross-validation. The best performing models were selected based on CV error estimates ($CV_{(K)}$) which were calculated thus:

$$CV_{(K)} = \sum_{k=1}^{K} \frac{n_k}{n} MSE_k,$$

where K is the number of folds in which the data were divided for CV, n_k is the number of observations in each of the K folds, and MSE_k is the MSE for the observations in each of the K folds, with the MSE calculated thus:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

Subsequently, the robustness of the selected features was measured against test data, using rms error (RMSE). RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}.$$

In summary, we divided our literature data into training and test datasets. We selected the best feature combinations and models using subset selection, then we applied the models to the test datasets. In addition, we generated a new test dataset that contains untested electrolytes, then tested our models against the new data. All the data processing and prediction were performed in R using the MASS, ISLR, leaps, tree, gbm, and Random Forest libraries.

CE Measurement. All electrolytes were prepared and handled in an argon-filled glovebox with a O_2 concentration <0.2 ppm and H_2O concentration <0.01 ppm. All electrolyte materials were used as received. LiFSI (Fluolyte) and $LiPF_6$ (Sigma-Aldrich) were used as salts and EC (Sigma-Aldrich), FEC (BASF), DMC (Sigma-Aldrich), DME (Sigma-Aldrich), TEGDME (Sigma-Aldrich), DOL (Sigma-Aldrich), MTBE (Sigma-Aldrich), THF (Sigma-Aldrich), TTE (SynQuest), and toluene (Sigma-Aldrich) were used as solvents. In an argon-filled glovebox, Li | Cu batteries were assembled into type 2032 coin cells. Each coin cell used 1 cm^2 (750 μm thick) Li foil (Alfa Aesar), Cu foil, a separator (Celgard 2325), and 40 μL of electrolyte. Land Instruments battery tester was used to cycle the coin cells. For each cell, initially 5 mA h cm^{-2} of Li metal was deposited on Cu and stripped (formation cycle). Then 5 mA h cm^{-2} of Li metal was deposited again, to act as a Li reservoir. Then Li was repeatedly stripped and plated 1 mA h cm^{-2} for nine cycles. The remaining Li on Cu was then stripped, and the average CE was calculated by dividing the total stripping capacity by the total plating capacity after the formation cycle. For all steps, the current density was 0.5 mA cm^{-2}. Three cells were tested for each electrolyte.

Cryo-EM Characterization. For cryogenic electron microscopy characterization, lithium was electrodeposited directly onto copper TEM grids by using the grids as electrodes inside type 2032 coin cells. After Li metal was deposited, the cells were disassembled in an argon-filled glovebox and rinsed to remove salts while minimizing artifacts. The sample was then sealed in a container and taken out of the glovebox. While the container was submerged in a bath of liquid nitrogen (LN2), it was crushed to release the sample to the cryogen without any exposure to the atmosphere. Maintaining the sample in LN2, it was loaded onto the cryo-EM holder (Gatan 626). The cryo-EM holder is specially designed to prevent air exposure of the sample, which preserves the specimen in its native state. The holder was inserted into the TEM column. Throughout the characterization process, the sample temperature was maintained at around −178 °C.

The cryo-EM characterizations were conducted using a FEI Titan 80–300 environmental (scanning) TEM. The instrument was operated at an accelerating voltage of 300 kV and was equipped with an aberration corrector, which was tuned before each sample analysis. For all characterizations including HR-TEM and EDS, electron dose rate was carefully controlled to prevent potential damage to the samples. HR-TEM imaging was performed on a Gatan K3 IS direct electron detector camera in low dose mode, with an electron dose of 150 to 300 e$^-$ Å^{-2}.

Electrolyte and Electrode Characterization. Solvation energy measurements were done by using a method developed previously in our group, using an H cell with Li metal as electrodes and asymmetric electrolytes. The electrolyte of interest was placed in a half-cell, while a reference electrolyte of 1 M LiFSI DEC was

placed in the other. An open circuit potential was measured to probe the relative solvation energy. Scanning electron microscopy (SEM) was conducted using an FEI Magellan 400 XHR. 0.5 mAh cm^{-2} of Li was deposited onto Cu current collectors. X-ray photoelectron spectroscopy (XPS) signals were collected on a PHI VersaProbe 1 scanning XPS microprobe with an Al Kα source.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. S.T.O. acknowledges support from the Knight-Hennessy Scholarship. Z.Z. acknowledges the Stanford Interdisciplinary Graduate Fellowship. D.T.B. acknowledges support from the NSF Graduate Research Fellowship Program (NSF GRFP). The cryo-EM part is supported by the Department of Energy, Office of Basic Energy Sciences, Division of Materials Science and Engineering under contract DE-AC02-76SF00515. The battery and electrolyte measurement part was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Vehicle Technologies, of the U.S. Department of Energy under the Battery Materials Research (BMR) Program, and the Battery500 Consortium program. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF). The HR-TEM images were taken with a K3 IS camera and support is courtesy of Gatan.

Author affiliations: [a]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305; [b]Department of Chemical Engineering, Stanford University, Stanford, CA 94305; [c]Department of Chemistry, Stanford University, Stanford, CA 94305; [d]Department of Energy Science and Engineering, Stanford University, Stanford, CA 94305; and [e]Stanford Institute for Materials and Energy Sciences, Stanford Linear Accelerator Center National Accelerator Laboratory, Menlo Park, CA 94025

1. P. Agre, M. Molina, S. Chu, The real climate debate. *Nature.* **550**, S62–S64 (2017).
2. J. Liu *et al.*, Pathways for practical high-energy long-cycling lithium metal batteries. *Nat. Energy.* **4**, 180–186 (2019).
3. H. Liu *et al.*, Ultrahigh coulombic efficiency electrolyte enables Li||SPAN batteries with superior cycling performance. *Mater. Today* **42**, 17–28 (2021).
4. D. Lin, Y. Liu, Y. Cui, Reviving the lithium metal anode for high-energy batteries. *Nat. Nanotechnol.* **12**, 194–206 (2017).
5. M. D. Tikekar, S. Choudhury, Z. Tu, L. A. Archer, Design principles for electrolytes and interfaces for stable lithium-metal batteries. *Nat. Energy.* **1**, 1–7 (2016).
6. H. Wang *et al.*, Wrinkled graphene cages as hosts for high-capacity li metal anodes shown by cryogenic electron microscopy. *Nano Lett.* **19**, 1326–1335 (2019), 10.1021/acs.nanolett.8b04906.
7. D. Lin *et al.*, Layered reduced graphene oxide with nanoscale interlayer gaps as a stable host for lithium metal anodes. *Nat. Nanotechnol.* **11**, 626–632 (2016).
8. J. Zhao *et al.*, Air-stable and freestanding lithium alloy/graphene foil as an alternative to lithium metal anodes. *Nat. Nanotechnol.* **12**, 993–999 (2017).
9. Y. Liu *et al.*, An ultrastrong double-layer nanodiamond interface for stable lithium metal anodes. *Joule* **2**, 1595–1609 (2018).
10. M. S. Kim *et al.*, Langmuir-Blodgett artificial solid-electrolyte interphases for practical lithium metal batteries. *Nat. Energy* **3**, 889–898 (2018).
11. J. Qian *et al.*, High rate and stable cycling of lithium metal anode. *Nat. Commun.* **6**, 6362 (2015), 10.1038/ncomms7362.
12. X. Ren *et al.*, Enabling high-voltage lithium-metal batteries under practical conditions. *Joule* **3**, 1662–1676 (2019).
13. Z. Yu *et al.*, Molecular design for electrolyte solvents enabling energy-dense and long-cycling lithium metal batteries. *Nat. Energy* **5**, **526–533** (2020).
14. A. J. Louli *et al.*, Diagnosing and correcting anode-free cell failure via electrolyte and morphological analysis. *Nat. Energy* **5**, 693–702 (2020).
15. C. S. Rustomji *et al.*, Liquefied gas electrolytes for electrochemical energy storage devices. *Science* **356**, eaal4263 (2017), 10.1126/science.aal4263.
16. X. Cao *et al.*, Monolithic solid–electrolyte interphases formed in fluorinated orthoformate-based electrolytes minimize Li depletion and pulverization. *Nat. Energy* **4**, 796–805 (2019).
17. X. Fan *et al.*, All-temperature batteries enabled by fluorinated electrolytes with non-polar solvents. *Nat. Energy* **4**, 882–890 (2019).
18. H. Wang *et al.*, Correlating li-ion solvation structures and electrode potential temperature coefficients. *J. Am. Chem. Soc.* **143**, 2264–2271 (2021), 10.1021/jacs.0c10587.
19. Y. Chen *et al.*, Steric effect tuned ion solvation enabling stable cycling of high-voltage lithium metal battery. *J. Am. Chem. Soc.* **143**, 18703–18713 (2021), 10.1021/jacs.1c09006.
20. A. D. Sendek *et al.*, Machine learning-assisted discovery of solid li-ion conducting materials. *Chem. Mater.* **31**, 342–352 (2018).
21. M. Uppaluri, A. Subramaniam, L. Mishra, V. Viswanathan, V. R. Subramanian, Can a transport model predict inverse signatures in lithium metal batteries without modifying kinetics? *J. Electrochem. Soc.* **167**, 160547 (2021).
22. T. Nakayama, Y. Igarashi, K. Sodeyama, M. Okada, Material search for Li-ion battery electrolytes through an exhaustive search with a Gaussian process. *Chem. Phys. Lett.* **731**, 136622 (2019).
23. M. Aykol, P. Herring, A. Anapolsky, Machine learning for continuous innovation in battery technologies. *Nat. Rev. Mater.* **5**, 725–727 (2020).
24. C. Vidal, P. Malysz, P. Kollmeyer, A. Emadi, Machine learning applied to electrified vehicle battery state of charge and state of health estimation: State-of-the-art. *IEEE Access.* **8**, 52796–52814 (2020).
25. G. w. You, S. Park, D. Oh, Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach. *Appl. Energy.* **176**, 92–103 (2016).
26. K. A. Severson *et al.*, Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy.* **4**, 383–391 (2019).
27. P. M. Attia *et al.*, Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **578**, 397–402 (2020).
28. J. G. Zhang, W. Xu, J. Xiao, X. Cao, J. Liu, Lithium metal anodes with nonaqueous electrolytes. *Chem. Rev.* **120**, 13312–13348 (2020).
29. B. D. Adams, J. Zheng, X. Ren, W. Xu, J. G. Zhang, Accurate determination of coulombic efficiency for lithium metal anodes and lithium metal batteries. *Adv. Energy Mater.* **8**, 1–11 (2018).
30. S. C. Kim *et al.*, Potentiometric measurement to probe solvation energy and its correlation to lithium battery cyclability. *J. Am. Chem. Soc.* **143**, 10301–10308 (2021).
31. X. Fan *et al.*, All-temperature batteries enabled by fluorinated electrolytes with non-polar solvents. *Nat. Energy* **4**, 882–890 (2019).
32. Y. Yamada, J. Wang, S. Ko, E. Watanabe, A. Yamada, Advances and issues in developing salt-concentrated battery electrolytes. *Nat. Energy* **4**, 269–280 (2019).
33. V. Viswanathan, Y. Zhang, Not all fluorination is the same: Unique effects of fluorine functionalization of ethylene carbonate for tuning solid-electrolyte interphase in li metal batteries. *Langmuir.* **36**, 11450–11466 (2020).
34. Y. Zhang, V. Viswanathan, Design rules for selecting fluorinated linear organic solvents for li metal batteries. *J. Phys. Chem. Lett.* **12**, 5821–5828 (2021).
35. Y. Zhang, C. Ling, A strategy to apply machine learning to small datasets in materials science. *NPJ Comput. Mater.* **4**, 28–33 (2018).
36. G. Casella, S. Fienberg, I. Olkin, *Springer Texts in Statistics* (Springer, 2006), **vol. 102**, http://books.google.com/books?id=9tv0tal8l6YC.
37. M. a. Hall, L. a. Smith, Practical feature subset selection for machine learning. *Comput. Sci.* **98**, 181–191 (1998).
38. M. Kuhn, K. Johnson, *Applied Predictive Modeling* (Springer, New York, NY, 2013).
39. E. D. Cubuk, A. D. Sendek, E. J. Reed, Screening billions of candidates for solid lithium-ion conductors: A transfer learning approach for small data. *J. Chem. Phys.* **150**, 214701 (2019), 10.1063/1.5093220.
40. G. Casella, S. Fienberg, I. Olkin, *Springer Texts in Statistics* (Springer, New York, NY, 2006), **vol. 102**.
41. D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, Predicting reaction performance in C-N cross-coupling using machine learning. *Science* **360**, 186–190, 10.1126/science.aar5169.
42. M. K. Nielsen, D. T. Ahneman, O. Riera, A. G. Doyle, Deoxyfluorination with sulfonyl fluorides: Navigating reaction space with machine learning. *J. Am. Chem. Soc.* **140**, 5004–5008 (2018).
43. Ethem Alpaydin, *Machine Learning New AI* (*MIT Press Essential Knowledge Series*, The MIT Press, Cambridge, MA, 2017), **vol. 91**, https://www.wook.pt/livro/machine-learning-ethem-alpaydin/17932580.
44. G. M. Hobold *et al.*, Moving beyond 99.9% Coulombic efficiency for lithium anodes in liquid electrolytes. *Nat. Energy* **6**, 951–960 (2021).
45. Z. Yu *et al.*, Rational solvent molecule tuning for high-performance lithium metal battery electrolytes. *Nat. Energy* **7**, 94–106 (2022), 10.1038/s41560-021-00962-y.
46. S. Liu *et al.*, An inorganic-rich solid electrolyte interphase for advanced lithium-metal batteries in carbonate electrolytes. *Angew. Chem. Int. Ed. Engl.* **60**, 3661–3671 (2021).
47. X. Fan *et al.*, Highly fluorinated interphases enable high-voltage li-metal batteries. *Chem* **4**, 174–185 (2018).
48. L. Suo *et al.*, Fluorine-donating electrolytes enable highly reversible 5-V-class Li metal batteries. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1156–1161, 10.1073/pnas.1712895115.
49. J. Chen *et al.*, Electrolyte design for LiF-rich solid–electrolyte interfaces to enable high-performance microsized alloy anodes for batteries. *Nat. Energy* **5**, 386–397 (2020).
50. C. V. Amanchukwu, X. Kong, J. Qin, Y. Cui, Z. Bao, Nonpolar alkanes modify lithium-ion solvation for improved lithium deposition and stripping. *Adv. Energy Mater.* **9**, 1–11 (2019).
51. I. Vinoth Kanna, P. Paturu, A study of hydrogen as an alternative fuel. *Int. J. Ambient Energy* **41**, 1433–1436 (2020).
52. G. Rubio-Pérez, N. Muñoz-Rujas, A. Srhiyer, E. A. Montero, F. Aguilar, Isobaric vapor-liquid equilibrium, density and speed of sound of binary mixtures 2,2,4-trimethylpentane + 1-butanol or dibutyl ether (DBE) at 101.3 kPa. *Fluid Phase Equilib.* **475**, 10–17 (2018).
53. A. M. Niziolek, O. Onel, C. A. Floudas, Production of benzene, toluene, and xylenes from natural gas via methanol: Process synthesis and global optimization. *AIChE J.* **62**, 1531–1556 (2016).
54. J. Chen *et al.*, Electrolyte design for Li metal-free Li batteries. *Mater. Today* **39**, 118–126 (2020).
55. S. Chen *et al.*, High-voltage lithium-metal batteries enabled by localized high-concentration electrolytes. *Adv. Mater.* **30**, 1–7 (2018).
56. S. Liu *et al.*, An inorganic-rich solid electrolyte interphase for advanced lithium-metal batteries in carbonate electrolytes. *Angew. Chem. Int. Ed. Engl.* **60**, 3661–3671 (2021).
57. D. J. Yoo *et al.*, Tuning the electron density of aromatic solvent for stable solid-electrolyte-interphase layer in carbonate-based lithium metal batteries. *Adv. Energy Mater* **8**, 1802365 (2018), 10.1002/aenm.201802365.
58. L. Cheng *et al.*, The origin of high electrolyte-electrode interfacial resistances in lithium cells containing garnet type solid electrolytes. *Phys. Chem. Chem. Phys.* **16**, 18294–18300 (2014).
59. G. A. Pinheiro *et al.*, Machine learning prediction of nine molecular properties based on the SMILES representation of the QM9 quantum-chemistry dataset. *J. Phys. Chem. A* **124**, 9854–9866 (2020), 10.1021/acs.jpca.0c05969.