ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Exploring the Possibility of Machine Learning for Predicting Ionic Conductivity of Solid-State Electrolytes

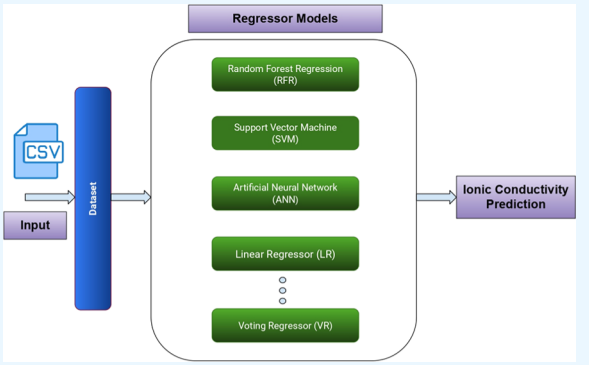
Atul Kumar Mishra,[‖] Snehal Rajput,[‖] Meera Karamta, and Indrajit Mukhopadhyay*

Cite This: *ACS Omega* 2023, 8, 16419−16427

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Unlike conventional liquid electrolytes, solid-state electrolytes (SSEs) have gained increased attention in the domain of all-solid-state lithium-ion batteries (ASSBs) due to their safety features, higher energy/power density, better electrochemical stability, and a broader electrochemical window. SSEs, however, face several difficulties, such as poorer ionic conductivity, complicated interfaces, and unstable physical characteristics. Vast research is still needed to find compatible and appropriate SSEs with improved properties for ASSBs. Traditional trial-and-error procedures to find novel and sophisticated SSEs require vast resources and time. Machine learning (ML), which has emerged as an effective and trustworthy tool for screening new functional materials, was recently used to forecast new SSEs for ASSBs. In this study, we developed an ML-based architecture to predict ionic conductivity by utilizing the characteristics of activation energy, operating temperature, lattice parameters, and unit cell volume of various SSEs. Additionally, the feature set can identify distinct patterns in the data set that can be verified using a correlation map. Because they are more reliable, the ensemble-based predictor models can more precisely forecast ionic conductivity. The prediction can be strengthened even further, and the overfitting issue can be resolved by stacking numerous ensemble models. The data set was split into 70:30 ratios to train and test with eight predictor models. The maximum mean-squared error and mean absolute error in training and testing for the random forest regressor (RFR) model were obtained as 0.001 and 0.003, respectively.



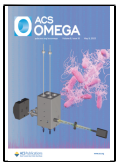
1. INTRODUCTION

After commercialization in the early 1990s, the LIBs have gained tremendous success in the field of portable electronic devices and EVs[1] due to various reasons such as high specific energy (361 Wh kg^{-1}),[2] high operating voltage (∼4 V),[3] and long life cycle. However, with the advent of next-generation smart electronic devices and long-range EVs, a battery system with high energy density, safety, and life span is the need of the hour.[4,5] To manufacture Li-ion batteries with improved safety and wider operating temperature ranges, one alternative is to substitute liquid electrolytes with inorganic SSEs.[6,7] As a result, this field has gained tremendous attention for producing various SSEs, some of which such as sulfide-based thio-LISICON (LIthium Super-Ionic CONductor), etc. have shown ionic conductivities as high as 10^{-2} S cm^{-1} at room temperature, which is close to that of the liquid electrolytes.[8,9] Also, Li-ion batteries with SSEs, i.e., the ASSBs, have the following advantages over conventional LIBs desired for the applications mentioned above, i.e., in the field of portable electronic devices and EVs:

- ASSBs with Li-metal anodes have high specific energy density (>500 Wh/kg), high volumetric energy density (>1500 Wh/L), and potentially lower cost (<$100/kWh).[10−12]
- ASSBs are exceptionally safe since SSEs have better mechanical strength (a shear modulus of 100 GPa) and safety, as these are leakproof and inflammable.[13,14]
- The SSEs used in the ASSBs have a broader electrochemical stability window (∼6 V vs Li^+/Li).[15]
- High mechanical strength of SSEs (a shear modulus of 100 GPa) ensures long life and dendrite suppression in the ASSBs.[13]
- ASSBs are easier to design since they have less packaging and state-of-charge monitoring circuit requirements.
- In ASSBs, SSEs serve as separators and electrolytes at the same time.[14]

Hence, there is a significant need for more effective energy storage and power devices, and an efficient technique of building

Received: March 2, 2023
Accepted: April 13, 2023
Published: April 26, 2023



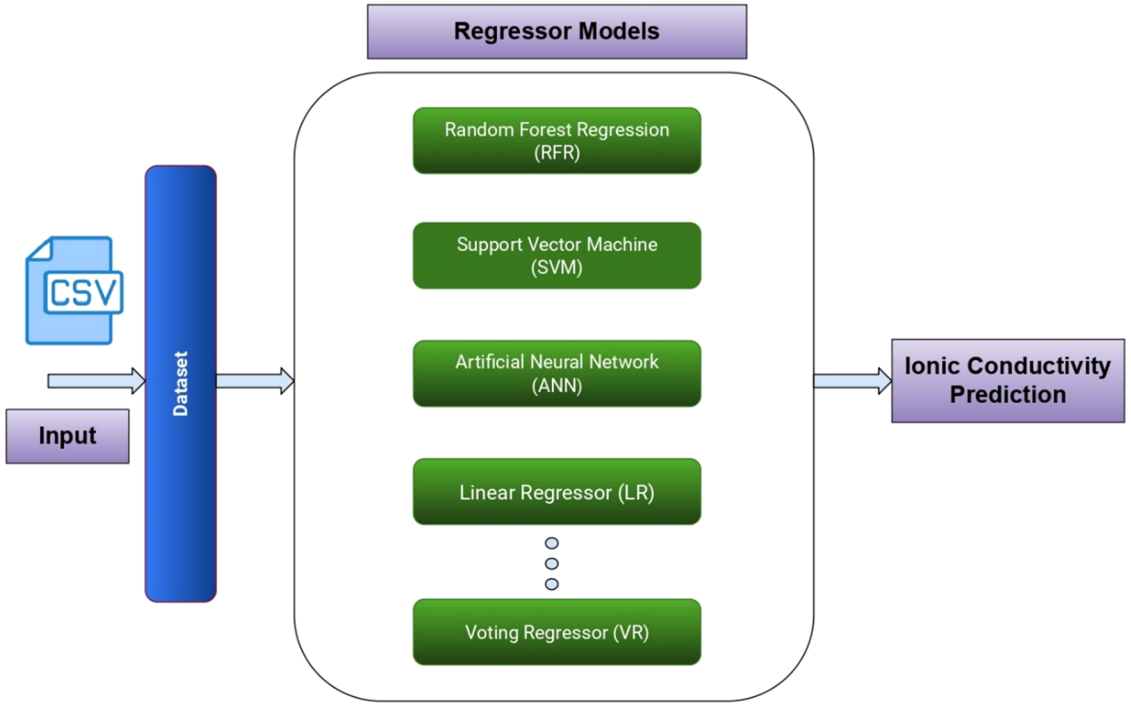


Figure 1. Proposed methodology to predict the ionic conductivity of SSEs.

ionic conductors based on a repeatable and systematic methodology is crucial for long-term success in this domain. Traditional techniques of searching for optimal compositions with superior properties entail trial-and-error procedures that take a long time to produce and verify new SSEs, resulting in poor progress and low efficiency. With the emergence of high-performance computation, first-principles calculation approaches such as DFT are being commonly utilized for the quantitative investigation of ionic conduction in the crystals.[16−20]

Recently, various ML-based data-driven approaches have been proposed for modeling desirable properties of the SSEs. Homma et al. used ML for compositional optimization of ternary Li_3PO_4−Li_3BO_3−Li_2SO_4 and obtained the best lithium-ion conductivity, i.e., 4.9×10^{-4} S/cm (300 °C) for 25:14:61, respectively.[21] The "ML" term was coined by Arthur Samuel, and it is defined as follows: A branch of artificial intelligence that allows systems to learn and develop from data without being explicitly programmed.[22] Machines can be trained to discover patterns in and correlations between incoming data and automate regular activities using huge amounts of computation for a single or numerous specific tasks.[23] Our target problem of predicting the ionic conductivity of SSEs falls under the regression task. The unsupervised learning algorithm is given no labels and is left to find structure in its data on its own. The goal of this algorithm is to find a hidden pattern in the data set.[24,25] The semisupervised learning algorithm is a hybrid of supervised and unsupervised learning. It employs a little quantity of labeled data and a large amount of unlabeled data, allowing it to benefit from both unsupervised and supervised learning while avoiding the difficulties associated with locating huge amounts of labeled data[26,27]

The significant growth of ML during the last few decades has extended the use of this data-driven technique throughout research, commerce, and business.[28] There has recently been a surge in interest in using ML to solve difficulties in materials science. Inorganic materials, in particular, have been represented using ML techniques, which have also been used to forecast fundamental properties, build atomic potential, discover functional candidates, assess complex reaction networks, and guide experimental design.[29−35] The key to these breakthroughs is the ability to reliably forecast behavior in unknown areas by quantitatively learning the pattern from sufficient training samples. Furthermore, ML has been used in the battery industry to estimate the state of charge or cycle life of operating cells[36,37] and to aid in the fabrication of LIB electrodes.[38] However, until Chen et al. looked into an opportunity to employ such skills to improve the performance and overall quality of SSE films used in ASSBs, the application of ML in the fabrication of flexible SSE films remained untapped. Their research looks at the multivariable interdependencies between performance and manufacturing characteristics to predict the quality of SSE films.[39] Similarly, in the domain of SSEs, ML has got a wide application to forecast the best structure, create innovative materials, and predict ionic conductivity.[40−43] The ML algorithm has successfully been used to screen the new SSEs by the Materials Genome Initiative founded in 2011.[44−46] ML is frequently used in attribute forecasting to reflect molecular materials and targeted characteristics.

In summary, from the literature survey, we can summarize ML algorithms as faster and scaler in comparison to the first principles-based DFT, which takes weeks for doing the same calculation. The schematic workflow diagram of our proposed model is shown in Figure 1. In this work, we targeted predicting the ionic conductivity of SSEs, where different ML models were explored to robustly predict ionic conductivity. Also, we tried to get an insight into features' contribution to ionic conductivity prediction using a correlation map. The structure of the remainder of the paper is as follows: The data set is described in Section 2; ionic conductivity prediction methods with eight predictor models are discussed in Section 3; Section 4 contains

https://doi.org/10.1021/acsomega.3c01400
ACS Omega 2023, 8, 16419−16427

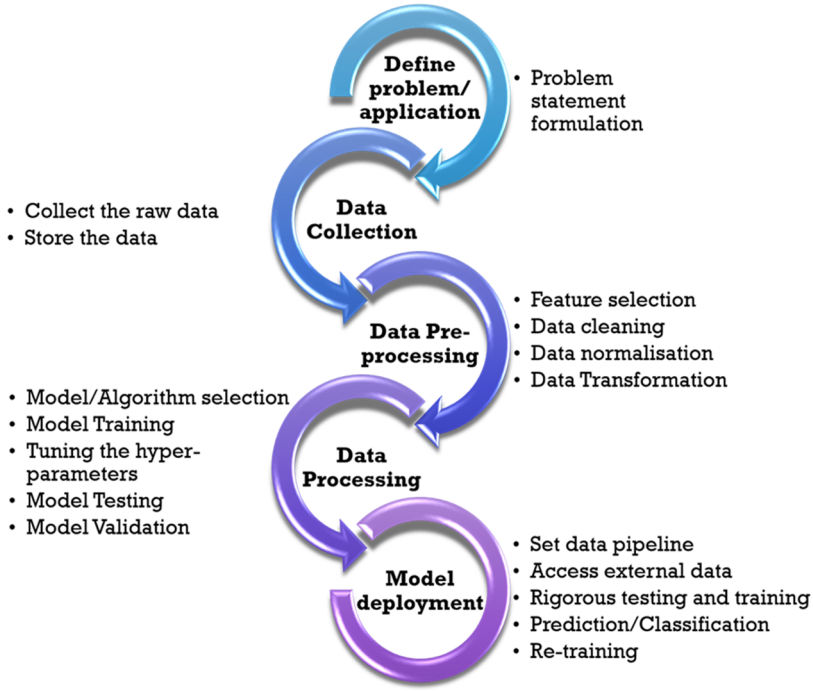


Figure 2. Generic ML application flow diagram.

results and discussions, and finally the conclusion of the paper is included in Section 5.[40]

2. DATA SET

Data sets are the most fundamental resource required for any ML model. Since the ML model learns from the pattern of the given data to either predict or classify the label/labels, the input data set plays a crucial role not only to meet the objective of the problem but also to achieve good performance on the unseen data. The problem to be solved using ML is often represented by data in multiple formats and/or a variety of structures.

Multiple standards and variances in the data set make the ionic conductivity prediction a little bit challenging. However, appropriate data sets can make it much easier to access material information. For the screening of SSE, the material scientists apply the materials databases such as ICSD (fiz-karlsruhe.de/icsd),[47] Material project (materialsproject.org),[48] AFLOW (aflowlib.org),[49] OQMD (oqmd.org),[44] Computational Materials Repository (cmr.fysik.dtu.dk),[50] Crystallography Open Database (crystallography.net),[51] and MATGEN (matgen.nscc-gz.cn).[52] Ionic conductivity is a difficult concept to grasp, and hence most databases lack it. As a result, some material scientists have attempted to automatically collect material synthesis parameters from tens of thousands of academic publications using text mining, i.e., ML and natural language processing techniques, to integrate and compile them into usable datasets for ML and have successfully performed practical applications.

The concentration of charge carriers, the temperature of the crystal, the availability of vacant-accessible sites, which are controlled by the density of defects in the crystal, and the ease with which an ion can jump to another site are the factors that influence the ionic conductivity of the SSEs. Activation energy controls the last parameter mentioned above, namely, the ease with which an ion can migrate to an adjacent site. The phenomenological quantity known as "activation energy" can be thought of as the free energy barrier that an ion must overcome to successfully hop between the two sites. Activation energy is the most important component influencing the ionic conductivity of a crystal since its dependency on other parameters is exponential. Experiments can be used to measure it quite easily. The Arrhenius expression, given in eq 1, is most typically used to calculate activation energies.[53]

$$s = (A/T)\exp(-E_a/k_B T) \quad (1)$$

Here, the ionic conductivity at temperature T (in kelvin) is denoted by "s", the Boltzmann constant is denoted by k_B, the activation energy is denoted by E_a, and the pre-exponential factor is denoted by A. All remaining parameters that determine ionic conductivity, other than the activation energy, are contained in the pre-exponential component, A.

Keeping the above-mentioned dependence of ionic conductivity on the activation energy and temperature in the eye, here, we also have compiled a data set of 120 SSEs from the literature that includes the ionic conductivity, activation energy, lattice parameter, and unit cell volume at a given temperature.[20,54,55] However, unlike ionic conductivity, activation energy, and temperature, there is no clear-cut relationship between the ionic conductivity and the lattice parameter or unit cell volume.

3. PROPOSED METHODOLOGY

Before diving deep into explaining the algorithms used in this work, we present in brief the generic steps involved in ML applications (shown schematically in Figure 2).

3.1. Problem Definition and Data Collection. The first step to ML application is defining the problem. The definition of the problem is important to label the system inputs and outputs. Most physical, engineering or design processes have multiple system variables and hence problem definition involves identifying which control parameter is being used to solve the problem at hand. As mentioned earlier, the feasibility of ML

https://doi.org/10.1021/acsomega.3c01400
ACS Omega 2023, 8, 16419−16427

applications to solve a particular problem lies in the availability of data. Once the problem is defined, it is important to gather the data which are commonly available in tabular form as CSV files or as JSON files. Also, the available data must be readable.

The available data is then organized and stored. The data can be classified as structured, semi-structured, or unstructured based on the extent they are organized. Structured data stored in the form of tables have some relation between the values in the rows and columns. Such datasets are easy to update as they possess concurrency.[56] They are preferred when multiple users are working simultaneously on the dataset for transaction management. Semi-structured and Unstructured datasets have the lowest level of relational attributes within. Thus they do not provide comparable performance. However, these datasets are preferable when we need to possess good flexibility and scalability.

3.2. Data Pre-processing. A huge amount of data is available and stored for the ML model to analyze and interpret. However, it is not necessary that all the variables or features in a dataset may be useful. Thus, feature selection is an important task in pre-processing. The idea is to select the best set of features that provide the most information.[56] Redundancy in the data can lead to instability in the algorithm. The generality of the model application is reduced. For feature selection, techniques such as Correlation coefficient, Chi-square, Dispersion ratio, Exhaustive Feature selection, and LASSO Regression can be employed.[57] Modern tools include optimization-based algorithms, meta-heuristic techniques, and nature-inspired algorithms[58,59] for selecting the variables which lead to an efficient ML model.

It is also important to check whether there exists any correlation, collinearity, or multicollinearity in the dataset. Higher collinearity amongst the features leads to difficulty in the interpretation of the output relation with every single feature. Changes in one feature can also change the correlated feature. The correlation matrix and Variance Inflation Factor[60] are used to detect when multiple features have collinearity.

The most common Data cleaning task includes removing outliers from the datasets. Outliers in the dataset are the values that are far off the allowed or assumed data range. Anomaly detection identifies the error's existence and can be used by the ML model to act on it. Five-Number Summary and Isolation Forest[61] are some common methods used for anomaly detection or outlier detection. In data normalization, the values of numeric columns in the dataset are changed to use a common scale, without compromising the information or distorting the ranges of values. Normalization is needed for the ML model to process data correctly and reduce the mathematical or numerical instability of the algorithm.

3.3. Data Processing and Model Deployment. Once the datasets are processed, we configure the model, by selecting the type of algorithm, and defining its hyperparameters. Most ML models fall into either Classification or Regression models.

Part of the data is used for training the model and part of it is used for testing the model. Usually, an 80%-20% ratio is used in training and testing respectively. Upon training the algorithm, the hyperparameters are tuned. Hyperparameters are specific to the ML model selected. The hyperparameters iterate over multiple combinations. The parameter measures accuracy until it finds the "best" model.

Model validation is required to check whether the model represents the system/process behavior accurately. ML models are validated by comparing the output to an independent field or experimental data sets that align with the simulated scenario.[62] Rigorous training and testing of ML algorithm are followed by model deployment which is done to create a proper user interface for working with the model.

We have used eight different ML algorithms ranging from simple/white-box to complex/black-box models. White-box ML models are simple and easy to interpret but lack performance, whereas Black-box models are highly efficient and are capable enough to capture the non-linear relationship between features but these are complex and lack interpretability. The predictors model used in the proposed work is 1. RFR, 2. SVM, 3. GBR, 4. ANN, 5. DT, 6. LR, 7. XGB, 8. VR (an ensemble of GBR, LR, and RFR models). RFR, GBR, and XGB models are based on ensemble learning. Ensemble learning is a general meta-ML approach that seeks to improve predictive performance by combining predictions from multiple different or same type models which are shown in Figure 3.

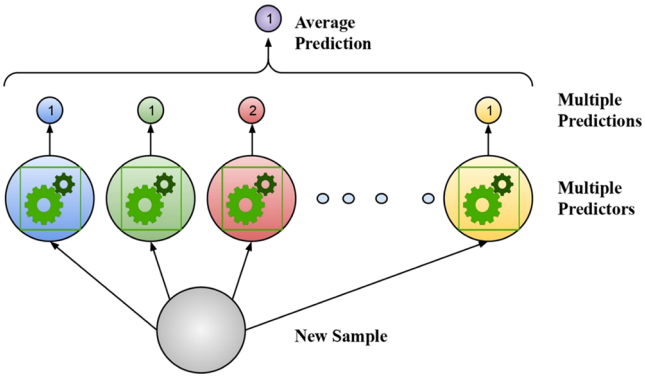


Figure 3. Working of ensemble learning-based model.

It combines multiple weak models into a single strong prediction which results in robust prediction and overcomes overfitting.[63] Again, ensemble learning methods can belong to either: the bagging, boosting, or stacking method. RFR is based on a bagging algorithm, where many DTs were created on different samples of the same dataset and averaged the prediction from different DTs for a sample.[63,64] Whereas, GBR and XGB are based on boosting method where multiple models are created sequentially which corrects the prediction of the predecessor model and outputs the average of the prediction.[65,66] And VR model is based on a stacked algorithm where prediction from multiple predictors is used as input to forecast the final prediction.[67] Further, we have used the LR model which is a linear model where output features are calculated using a linear combination of input features. Further, we have used the SVM model, where the model finds a hyperplane separating different classes having a maximum margin.[68] In this work, we have used SVM with two kernel functions: Radial kernel Function and polynomial function, where kernel functions are used to capture nonlinearity from the dataset. Furthermore, the parameters used in different predictor models are fine-tuned using grid search, which is shown in the following Table 1:

Also, we have used a shallow ANN model with two hidden layers and its structure is shown in Figure 4. ANN model mimics neurons of the human brain. It is the complex (nonlinear) model which can learn the complex relationship between features. The dataset was split into 70-30 ratios to train and test each predictor model.

https://doi.org/10.1021/acsomega.3c01400
ACS Omega 2023, 8, 16419−16427

Table 1. The Parameters Used in Different Predictor Models

model	fine-tuned parameter
RFR	bootstrap, maximum depth of the tree, maximum features for the best split, minimum samples to be required at the child node, minimum samples required for the split at the intermediate node, and numbers of trees used.
SVM forKernel (RBF, Polynomial)	for radial basis function (RBF) kernel: L2 penalty (regularization) parameter.
	for polynomial (poly) kernel: L2 penalty (regularization), degree of the polynomial kernel function, epsilon value - which signifies no penalty if the distance between the actual and predicted value is less than or equal to a specified value.
GBR	learning rate, maximum depth of the tree, maximum features for the best split, minimum samples to be required at the child node, minimum samples required for the split at the intermediate node, and numbers of trees used.
ANN	numbers of epochs, learning rate, and momentum for the Adam optimizer.
DT	maximum depth of the tree, maximum features for the best split, minimum samples to be required at the child node, minimum samples required for the split at the intermediate node, and the technique used to split at each node.
LR	used all default parameters.
XGB	learning rate, maximum depth of the tree, maximum features for the best split, minimum samples to be required at the child node, and minimum samples required for the split at the intermediate node.
VR	used all default parameters.

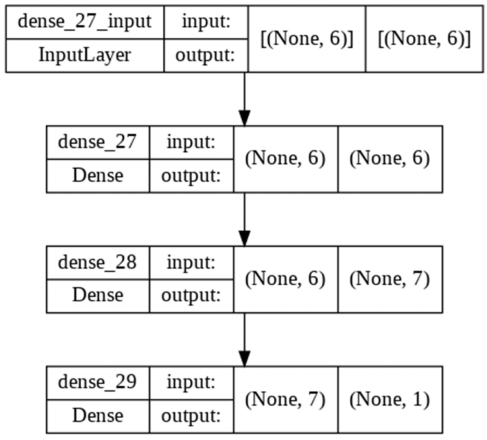


Figure 4. Proposed structure of ANN model.

In summary, we have used eight predictors ranging from linear to nonlinear models that can capture linear and/or non-linear relationships between features, which can help predictor models to generalize well on an unseen dataset.

4. RESULTS AND DISCUSSIONS

We have used MAE and MSE metrics to quantify the performance of different predictor models.[69] MAE is more robust to outliers because it averages the absolute differences between actual and predicted values over all the samples from the dataset[70] which is defined as

$$\mathrm{MAE} = \frac{1}{n}\sum_{k=1}^{n} |Y - \hat{Y}| \tag{2}$$

whereas MSE is less robust to outliers but unlike MAE, it can accurately capture the performance when handling larger error values.[71] MSE measures the squared of the difference between actual and predicted value, which is defined as

$$\mathrm{MSE} = \frac{1}{n}\sum_{k=1}^{n} (Y - \hat{Y}) \tag{3}$$

where n is the number of samples, Y is the actual target value and $\hat{Y}$ is the predicted value.

A graphical representation of the prediction accuracy of the ANN model is presented in Figure 5. For brevity, plots for only the ANN model are shown here. The "Ground Truth" datapoints represent the actual ionic conductivity of the sample dataset. We compare the predicted values of the ionic conductivity obtained by ANN model with the actual values in Figure 5. The loss function from the ANN model on the training and validation dataset is shown in Figure 6. However similar plots can be developed for other ML models as well. The actual ability of prediction is better gauged by using the indicators as MSE and MAE defined earlier. Table 2 shows performance comparisons from different predictor models on the Training and Testing dataset.

It can be observed that predictions from the ensemble models are more robust compared to ANN and LR models. ANN didn't perform well when compared to other models, the probable reasons can be: the complex nature of the model and the smaller size of the input dataset. The MSE from different predictors is approximately zero indicating that the models are not overfitting. Also, the difference in performance measures of the Training and Testing dataset is less which again reflects the robustness of the predictor models used. We have plotted the samples from the Training and Testing set for the Voting Regressor model, which is a stack of three ensemble models i.e., RFR, GBR, and XGR model. Similarly, to gain an insight into the feature set used to train the predictor models, we generated a correlation map which is shown in following Figure 7. (Listing of Feature sets is mentioned in Appendix). The correlation map uses Pearson correlation Index to measure the strength of the linear relationship between features, which ranges from −1 to +1, where 0 represents no correlation, +1 represents strong positive correlation and −1 represents strong negative correlation.[72]

A positive correlation indicates that if the value of parameter A increases, the value of parameter B will also increase, whereas a negative correlation indicates that if A increases, B decreases. We can observe that most of the features are weakly correlated and carry unique information to capture ionic conductivity patterns from the dataset.

In summary, we observe that ensemble-learning-based models are more robust than other predictor models and the feature set carries unique information that contributes significantly to the ionic conductivity prediction.

5. CONCLUSIONS

Traditional trial-and-error procedures to find novel and sophisticated SSEs require huge resources and time. ML has emerged as an effective and trustworthy tool for screening new functional materials, which has been used recently to forecast new SSEs for ASSBs. In this work, we proposed ML-based architecture to predict ionic conductivity using the features

https://doi.org/10.1021/acsomega.3c01400
ACS Omega 2023, 8, 16419−16427

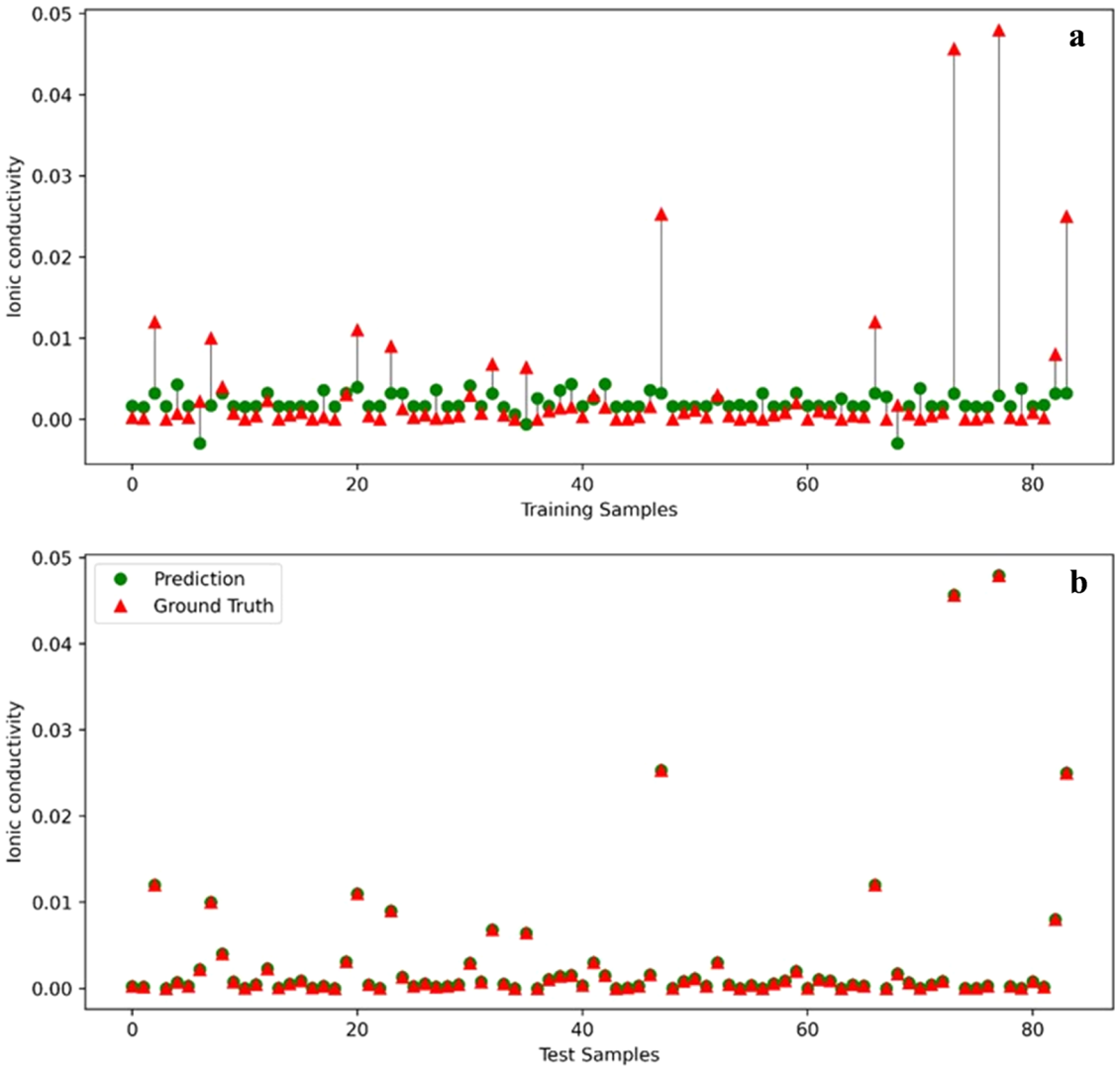


Figure 5. Comparison of predicted v/s actual values of Ionic Conductivity (S cm^{-1}) by ANN: (a) Training Dataset (b) Testing Dataset.

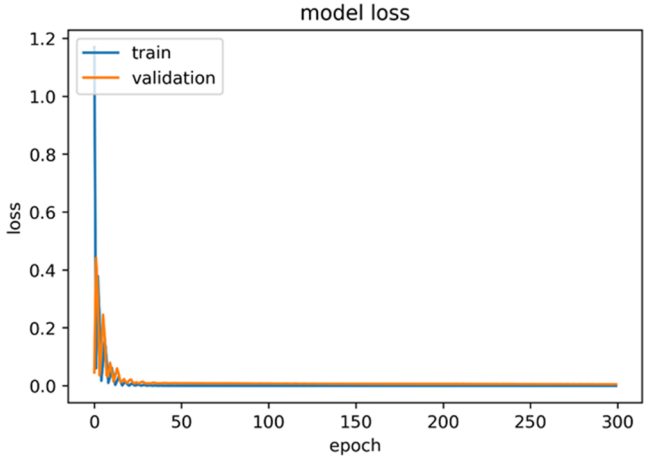


Figure 6. Loss function of Training and Validation dataset using ANN model.

Activation Energy, Temperature, Lattice parameters, and Volume of the Unit Cell. Also, feature sets can capture unique patterns from the dataset, which can be validated from a correlation map. The ensemble-based predictor models are

Table 2. Performance Comparisons from Different Predictor Models on the Training and Testing Dataset

dataset	model	MSE	MAE
training	RFR	0.001	0
	GBR	0	0.001
	LR	0	0.004
	SVM	0	0.022
	ANN	0	0.006
	XGB	0	0.001
	SVM	0	0.022
	DT	0	0.001
	VR	1.08×10^{-05}	0.002
testing	RFR	0.003	0
	GBR	0	0.002
	LR	0	0.003
	SVM	0.001	0.023
	ANN	0	0.002
	XGB	0	0.002
	SVM	0.001	0.023
	DT	0.001	0.007
	VR	2.30×10^{-05}	0.003

https://doi.org/10.1021/acsomega.3c01400
ACS Omega 2023, 8, 16419−16427

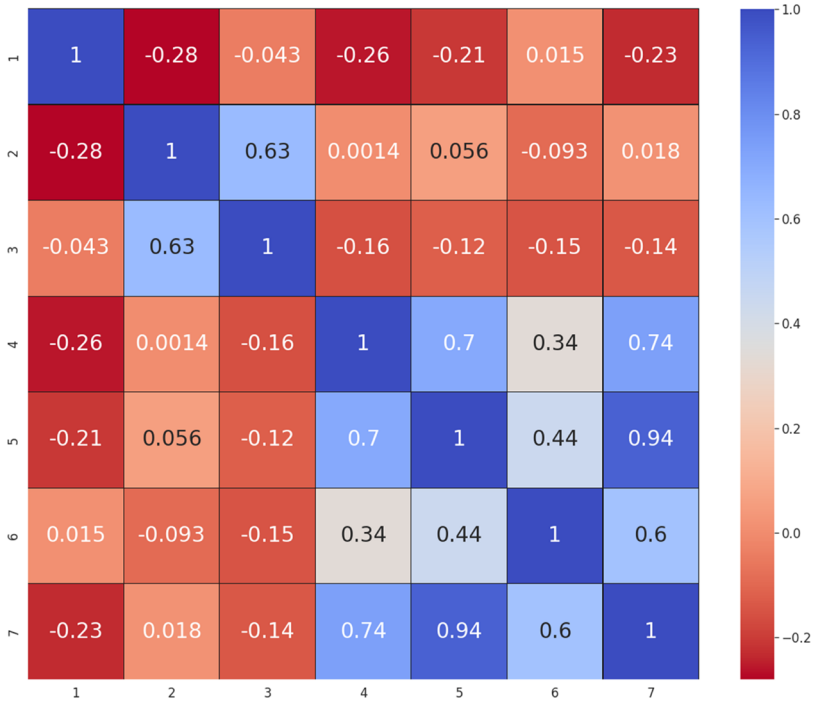


Figure 7. Correlation Map of Input and Output Feature (the numbers from 1 to 7 on the horizontal and vertical axis correspond to Ionic conductivity (S cm^{-1}), Activation Energy (eV), Temperature (°C), Lattice Parameters *a*, *b*, and, *c* (in Å), and the volume of Unit Cell ($Å^3$), respectively).

more robust and hence can predict ionic conductivity more accurately. Also, stacking multiple ensemble models can strengthen the prediction even more and can overcome the overfitting problem. The dataset was split into 70-30 ratios to train and test with 8 predictor models. The maximum MSE and MAE in training and testing for the RFR model were obtained at 0.001 and 0.003 respectively.

5.1. Future Work. We will include more features to predict ionic conductivity and observe their role in the target feature (ionic conductivity) prediction using ensemble models. We may further look through various model agnostic methods such as Shapley Additive exPlanations and Local Interpretable Model-agnostic Explanations, which can depict the global and local impact of input features on target features.

APPENDIX

sr. no.	feature set
1	ionic conductivity (S cm^{-1})
2	activation energy (eV)
3	temperature (°C)
4	lattice parameters (*a* in Å)
5	lattice parameters (*b* in Å)
6	lattice parameters (*c* in Å)
7	volume of the unit cell ($Å^3$)

AUTHOR INFORMATION

Corresponding Author

Indrajit Mukhopadhyay − *Solar Research and Development Center, Department of Solar Energy, Pandit Deendayal Energy University, Gandhinagar 382007 Gujarat, India;* orcid.org/0000-0003-3756-6131; Email: Indrajit.M@sse.pdpu.ac.in

Authors

Atul Kumar Mishra − *Solar Research and Development Center, Department of Solar Energy, Pandit Deendayal Energy University, Gandhinagar 382007 Gujarat, India*

Snehal Rajput − *Department of Computer Science Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar 382007 Gujarat, India*

Meera Karamta − *Department of Electrical Engineering, School of Technology, Pandit Deendayal Energy University, Gandhinagar 382007 Gujarat, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c01400

Author Contributions

‖A.K.M. and S.R. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are grateful to Pandit Deendayal Energy University (PDEU) for providing the necessary facilities to carry out this investigation. Financial support from Department of Science and Technology (DST) Government of India under project no. DST/TMD/MES/2017/32(G) and Solar Research and Development Centre (SRDC), PDEU is deeply acknowledged.

ABBREVIATIONS

LIB, lithium-ion battery; EV, electric vehicle; SSE, solid-state electrolyte; ASSB, all-solid-state Li-ion battery; LISICON, LIthium SuperIonic CONductor; DFT, density functional theory; ML, machine learning; RFR, random forest regressor; SVM, support vector machine; GBR, gradient boosting regressor; ANN, artificial neural network; DT, decision tree; LR, linear regressor; XGB, extreme gradient boosting; VR,

https://doi.org/10.1021/acsomega.3c01400
ACS Omega 2023, 8, 16419−16427

voting regressor; MAE, mean absolute error; MSE, mean squared error

REFERENCES

(1) Mauger, A.; Julien, C. M. Critical Review on Lithium-Ion Batteries: Are They Safe? Sustainable? *Ionics* **2017**, *23*, 1933−1947.

(2) Wu, F.; Kim, G.; Kuenzel, M.; Zhang, H.; Asenbauer, J.; Geiger, D.; Kaiser, U.; Passerini, S. Elucidating the Effect of Iron Doping on the Electrochemical Performance of Cobalt-free Lithium-rich Layered Cathode Materials. *Adv. Energy Mater.* **2019**, *9*, No. 1902445.

(3) Manthiram, A. A Reflection on Lithium-Ion Battery Cathode Chemistry. *Nat. Commun.* **2020**, *11*, No. 1550.

(4) Zhang, X.-Q.; Zhao, C.-Z.; Huang, J.-Q.; Zhang, Q. Recent Advances in Energy Chemical Engineering of Next-Generation Lithium Batteries. *Engineering* **2018**, *4*, 831−847.

(5) Xu, W.; Wang, J.; Ding, F.; Chen, X.; Nasybulin, E.; Zhang, Y.; Zhang, J.-G. Lithium Metal Anodes for Rechargeable Batteries. *Energy Environ. Sci.* **2014**, *7*, 513−537.

(6) Horowitz, Y.; Schmidt, C.; Yoon, D.; Riegger, L. M.; Katzenmeier, L.; Bosch, G. M.; Noked, M.; Ein-Eli, Y.; Janek, J.; Zeier, W. G.; et al. Between Liquid and All Solid: A Prospect on Electrolyte Future in Lithium-Ion Batteries for Electric Vehicles. *Energy Technol.* **2020**, *8*, No. 2000580.

(7) Zhang, Q.; Cao, D.; Ma, Y.; Natan, A.; Aurora, P.; Zhu, H. Sulfide-based Solid-state Electrolytes: Synthesis, Stability, and Potential for All-solid-state Batteries. *Adv. Mater.* **2019**, *31*, No. 1901131.

(8) Seino, Y.; Ota, T.; Takada, K.; Hayashi, A.; Tatsumisago, M. A Sulphide Lithium Super Ion Conductor Is Superior to Liquid Ion Conductors for Use in Rechargeable Batteries. *Energy Environ. Sci.* **2014**, *7*, 627−631.

(9) Fu, Z. H.; Chen, X.; Yao, N.; Shen, X.; Ma, X. X.; Feng, S.; Wang, S.; Zhang, R.; Zhang, L.; Zhang, Q. The Chemical Origin of Temperature-Dependent Lithium-Ion Concerted Diffusion in Sulfide Solid Electrolyte Li10GeP2S12. *J. Energy Chem.* **2022**, *70*, 59.

(10) Albertus, P.; Anandan, V.; Ban, C.; Balsara, N.; Belharouak, I.; Buettner-Garrett, J.; Chen, Z.; Daniel, C.; Doeff, M.; Dudney, N. J.; Dunn, B.; Harris, S. J.; Herle, S.; Herbert, E.; Kalnaus, S.; Libera, J. A.; Lu, D.; Martin, S.; McCloskey, B. D.; McDowell, M. T.; Meng, Y. S.; Nanda, J.; Sakamoto, J.; Self, E. C.; Tepavcevic, S.; Wachsman, E.; Wang, C.; Westover, A. S.; Xiao, J.; Yersak, T. Challenges for and Pathways toward Li-Metal-Based All-Solid-State Batteries. *ACS Energy Lett.* **2021**, *6*, 1399−1404.

(11) Masias, A.; Marcicki, J.; Paxton, W. A. Opportunities and Challenges of Lithium Ion Batteries in Automotive Applications. *ACS Energy Lett.* **2021**, *6*, 621−630.

(12) Sun, Y. K. Promising All-Solid-State Batteries for Future Electric Vehicles. *ACS Energy Lett.* **2020**, *5*, 3221−3223.

(13) Ni, J. E.; Case, E. D.; Sakamoto, J. S.; Rangasamy, E.; Wolfenstine, J. B. Room Temperature Elastic Moduli and Vickers Hardness of Hot-Pressed LLZO Cubic Garnet. *J. Mater. Sci.* **2012**, *47*, 7978−7985.

(14) Mishra, A. K.; Chaliyawala, H. A.; Patel, R.; Paneliya, S.; Vanpariya, A.; Patel, P.; Ray, A.; Pati, R.; Mukhopadhyay, I. Inorganic Solid State Electrolytes: Insights on Current and Future Scope. *J. Electrochem. Soc.* **2021**, *168*, 080536.

(15) Thangadurai, V.; Pinzaru, D.; Narayanan, S.; Baral, A. K. Fast Solid-State Li Ion Conducting Garnet-Type Structure Metal Oxides for Energy Storage. *J. Phys. Chem. Lett.* **2015**, *6*, 292−299.

(16) Mueller, T.; Hautier, G.; Jain, A.; Ceder, G. Evaluation of Tavorite-Structured Cathode Materials for Lithium-Ion Batteries Using High-Throughput Computing. *Chem. Mater.* **2011**, *23*, 3854−3862.

(17) Hautier, G.; Jain, A.; Chen, H.; Moore, C.; Ong, S. P.; Ceder, G. Novel Mixed Polyanions Lithium-Ion Battery Cathode Materials Predicted by High-Throughput Ab Initio Computations. *J. Mater. Chem.* **2011**, *21*, 17147−17153.

(18) Jalem, R.; Aoyama, T.; Nakayama, M.; Nogami, M. Multivariate Method-Assisted Ab Initio Study of Olivine-Type LiMXO4 (Main Group M2+−X5+ and M3+−X4+) Compositions as Potential Solid Electrolytes. *Chem. Mater.* **2012**, *24*, 1357−1364.

(19) Xu, Y.; Zong, Y.; Hippalgaonkar, K. Machine Learning-Assisted Cross-Domain Prediction of Ionic Conductivity in Sodium and Lithium-Based Superionic Conductors Using Facile Descriptors. *J. Phys. Commun.* **2020**, *4*, 055015.

(20) Sendek, A. D.; Yang, Q.; Cubuk, E. D.; Duerloo, K. A. N.; Cui, Y.; Reed, E. J. Holistic Computational Structure Screening of More than 12 000 Candidates for Solid Lithium-Ion Conductor Materials. *Energy Environ. Sci.* **2017**, *10*, 306−320.

(21) Homma, K.; Liu, Y.; Sumita, M.; Tamura, R.; Fushimi, N.; Iwata, J.; Tsuda, K.; Kaneta, C. Optimization of a Heterogeneous Ternary Li3PO4-Li3BO3-Li2SO4Mixture for Li-Ion Conductivity by Machine Learning. *J. Phys. Chem. C* **2020**, *124*, 12865−12870.

(22) Samuel, A. L. Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.* **2000**, *44*, 206−226.

(23) Kleesiek, J.; Murray, J. M.; Strack, C.; Kaissis, G.; Braren, R. A Primer on Machine Learning. *Der Radiologe* **2020**, *60*, 24−31.

(24) Glielmo, A.; Husic, B. E.; Rodriguez, A.; Clementi, C.; Noé, F.; Laio, A. Unsupervised Learning Methods for Molecular Simulation Data. *Chem. Rev.* **2021**, *121*, 9722−9758.

(25) Tao, K.; Wang, Z.; Han, Y.; Li, J. Rapid Discovery of Inorganic-Organic Solid Composite Electrolytes by Unsupervised Learning. *Chem. Eng. J.* **2023**, *454*, No. 140151.

(26) Mehyadin, A. E.; Abdulazeez, A. M. Classification Based on Semi-Supervised Learning: A Review. *Iraqi J. Comput. Inf.* **2021**, *47*, 1−11.

(27) Alloghani, M.; Al-Jumeily, D.; Mustafina, J.; Hussain, A.; Aljaaf, A. J. *A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science*, 2020.

(28) Jordan, M. I.; Mitchell, T. M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255.

(29) Kalidindi, S. R.; de Graef, M. Materials Data Science: Current Status and Future Outlook. *Annu. Rev. Mater. Res.* **2015**, *45*, 171.

(30) Schütt, K. T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K. R.; Gross, E. K. U. How to Represent Crystal Structures for Machine Learning: Towards Fast Prediction of Electronic Properties. *Phys. Rev. B* **2014**, *89*, 205118.

(31) de Jong, M.; Chen, W.; Notestine, R.; Persson, K.; Ceder, G.; Jain, A.; Asta, M.; Gamst, A. A Statistical Learning Framework for Materials Science: Application to Elastic Moduli of k-Nary Inorganic Polycrystalline Compounds. *Sci. Rep.* **2016**, *6*, No. 34256.

(32) Li, Z.; Kermode, J. R.; de Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*, 096405.

(33) Mannodi-Kanakkithodi, A.; Huan, T. D.; Ramprasad, R. Mining Materials Design Rules from Data: The Example of Polymer Dielectrics. *Chem. Mater.* **2017**, *29*, 9001.

(34) Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K. To Address Surface Reaction Network Complexity Using Scaling Relations Machine Learning and DFT Calculations. *Nat. Commun.* **2017**, *8*, No. 14621.

(35) Kim, E.; Huang, K.; Jegelka, S.; Olivetti, E. Virtual Screening of Inorganic Materials Synthesis Parameters with Deep Learning. *NPJ Comput. Mater.* **2017**, *3*, 53.

(36) Severson, K. A.; Attia, P. M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M. H.; Aykol, M.; Herring, P. K.; Fraggedakis, D.; Bazant, M. Z.; Harris, S. J.; Chueh, W. C.; Braatz, R. D. Data-Driven Prediction of Battery Cycle Life before Capacity Degradation. *Nat Energy* **2019**, *4*, 383−391.

(37) Chemali, E.; Kollmeyer, P. J.; Preindl, M.; Emadi, A. State-of-Charge Estimation of Li-Ion Batteries Using Deep Neural Networks: A Machine Learning Approach. *J. Power Sources* **2018**, *400*, 242.

(38) Duquesnoy, M.; Lombardo, T.; Chouchane, M.; Primo, E. N.; Franco, A. A. Data-Driven Assessment of Electrode Calendering Process by Combining Experimental Results, in Silico Mesostructures Generation and Machine Learning. *J. Power Sources* **2020**, *480*, 229103.

(39) Chen, Y. T.; Duquesnoy, M.; Tan, D. H. S.; Doux, J. M.; Yang, H.; Deysher, G.; Ridley, P.; Franco, A. A.; Meng, Y. S.; Chen, Z. Fabrication of High-Quality Thin Solid-State Electrolyte Films Assisted by Machine Learning. *ACS Energy Lett.* **2021**, *6*, 1639.

https://doi.org/10.1021/acsomega.3c01400
ACS Omega 2023, 8, 16419−16427

(40) Guo, H.; Wang, Q.; Stuke, A.; Urban, A.; Artrith, N. Accelerated Atomistic Modeling of Solid-State Battery Materials With Machine Learning. *Front. Energy Res.* **2021**, *9*, 695902.
(41) Adhyatma, A.; Xu, Y.; Hawari, N. H.; Satria Palar, P.; Sumboja, A. Improving Ionic Conductivity of Doped Li7La3Zr2O12 Using Optimized Machine Learning with Simplistic Descriptors. *Mater. Lett.* **2022**, *308*, 131159.
(42) Liu, Y.; Guo, B.; Zou, X.; Li, Y.; Shi, S. Machine Learning Assisted Materials Design and Discovery for Rechargeable Batteries. *Energy Storage Mater.* **2020**, *31*, 434.
(43) Sendek, A. D.; Cubuk, E. D.; Antoniuk, E. R.; Cheon, G.; Cui, Y.; Reed, E. J. Machine Learning-Assisted Discovery of Solid Li-Ion Conducting Materials. *Chem. Mater.* **2019**, *31*, 342.
(44) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501−1509.
(45) Cubuk, E. D.; Sendek, A. D.; Reed, E. J. Screening Billions of Candidates for Solid Lithium-Ion Conductors: A Transfer Learning Approach for Small Data. *J. Chem. Phys.* **2019**, *150*, No. 214701.
(46) US, N. S. and T. C.*Materials Genome Initiative for Global Competitiveness*, Executive Office of the President, National Science and Technology Council, 2011.
(47) Bergerhoff, G.; Hundt, R.; Sievers, R.; Brown, I. D. The Inorganic Crystal Structure Data Base. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 66.
(48) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 011002.
(49) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L. W.; Sanvito, S.; Buongiorno-Nardelli, M.; Mingo, N.; Levy, O. AFLOWLIB.ORG: A Distributed Materials Properties Repository from High-Throughput Ab Initio Calculations. *Comput. Mater. Sci.* **2012**, *58*, 227.
(50) Landis, D. D.; Hummelshøj, J. S.; Nestorov, S.; Greeley, J.; Dułłak, M.; Bligaard, T.; Nørskov, J. K.; Jacobsen, K. W. The Computational Materials Repository. *Comput Sci Eng* **2012**, *14*, 51.
(51) Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; le Bail, A. Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. *Nucleic Acids Res.* **2012**, *40*, D420.
(52) Zhang, L.; He, B.; Zhao, Q.; Zou, Z.; Chi, S.; Mi, P.; Ye, A.; Li, Y.; Wang, D.; Avdeev, M.; Adams, S.; Shi, S. A Database of Ionic Transport Characteristics for Over 29 000 Inorganic Compounds. *Adv. Funct. Mater.* **2020**, 2003087.
(53) Kanai, K.; Ozawa, S.; Kozawa, T.; Naito, M. Low Temperature Synthesis of Ga-Doped Li7 La3 Zr2 O12 Garnet-Type Solid Electrolyte by Mechanical Method. *Adv. Powder Technol.* **2021**, *32*, 3860.
(54) Mahbub, R.; Huang, K.; Jensen, Z.; Hood, Z. D.; Rupp, J. L. M.; Olivetti, E. A. Text Mining for Processing Conditions of Solid-State Battery Electrolyte. *Electrochem. Commun.* **2020**, *121*, 106860.
(55) Jensen, Z.; Kim, E.; Kwon, S.; Gani, T. Z. H.; Román-Leshkov, Y.; Moliner, M.; Corma, A.; Olivetti, E. A Machine Learning Approach to Zeolite Synthesis Enabled by Automatic Literature Data Extraction. *ACS Cent Sci* **2019**, *5*, 892.
(56) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. Statistical Learning. In *An Introduction To Statistical Learning*, Springer: 2021; 15−57.
(57) Molnar, C. *Interpretable Machine Learning*, 2020; Lulu.com.
(58) Abu Khurma, R.; Aljarah, I.; Sharieh, A.; Elaziz, M. A.; Damaševičius, R.; Krilavičius, T. A Review of the Modification Strategies of the Nature Inspired Algorithms for Feature Selection Problem. *Mathematics* **2022**, *10*, 464.
(59) Alsahaf, A.; Petkov, N.; Shenoy, V.; Azzopardi, G. A Framework for Feature Selection through Boosting. *Expert Syst. Appl.* **2022**, *187*, 115895.
(60) Shirali, K.; Amineh, F.; Reza, G. S. H. Post-Traumatic Stress Disorder among Iranian Pre-Hospital Emergency Services Personnel. *Int. J. Adv. Biotechnol. Res.* **2017**, *8*, 1019−1028.
(61) Kolaczyk, E. D.; Csardi, G. *Statistical Analysis of Network Data with R Introduction*, Springer: Berlin, Germany, 2014; Vol. *65*.
(62) Kerr, L. A.; Goethel, D. R. Simulation Modeling as a Tool for Synthesis of Stock Identification Information. *Stock Identification Methods: Applications in Fishery Science*, 2nd ed.; Acdemic Press: 2013.
(63) Zounemat-Kermani, M.; Batelaan, O.; Fadaee, M.; Hinkelmann, R. Ensemble Machine Learning Paradigms in Hydrology: A Review. *J. Hydrol.* **2021**, *598*, 126266.
(64) Ardabili, S.; Mosavi, A.; Várkonyi-Kóczy, A. R. Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods. *Lecture Notes in Networks and Systems*, Springer International Publishing, 2020; p *101*.
(65) Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, 2017.
(66) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189.
(67) Xiong, J.; Zhang, T. Y. Data-Driven Glass-Forming Ability Criterion for Bulk Amorphous Metals with Data Augmentation. *J. Mater. Sci. Technol.* **2022**, *121*, 99.
(68) Noble, W. S. What Is a Support Vector Machine? *Nat. Biotechnol.* **2006**, *24*, 1565.
(69) Bar-Lev, S. K.; Boikai, B.; Enis, P. On the Mean Squared Error, the Mean Absolute Error and the Like. *Commun. Stat. Theory Methods* **1999**, *28*, 1813.
(70) Qi, J.; Du, J.; Siniscalchi, S. M.; Ma, X.; Lee, C. H. On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression. *IEEE Signal Process Lett.* **2020**, *27*, 1485.
(71) Das, K.; Jiang, J.; Rao, J. N. K. Mean Squared Error of Empirical Predictor. *Ann Stat* **2004**, *32*, 818.
(72) Thirumalai, C.; Duba, A.; Reddy, R. *Decision Making System Using Machine Learning and Pearson for Heart Attack*, Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, 2017.

https://doi.org/10.1021/acsomega.3c01400
ACS Omega 2023, 8, 16419−16427