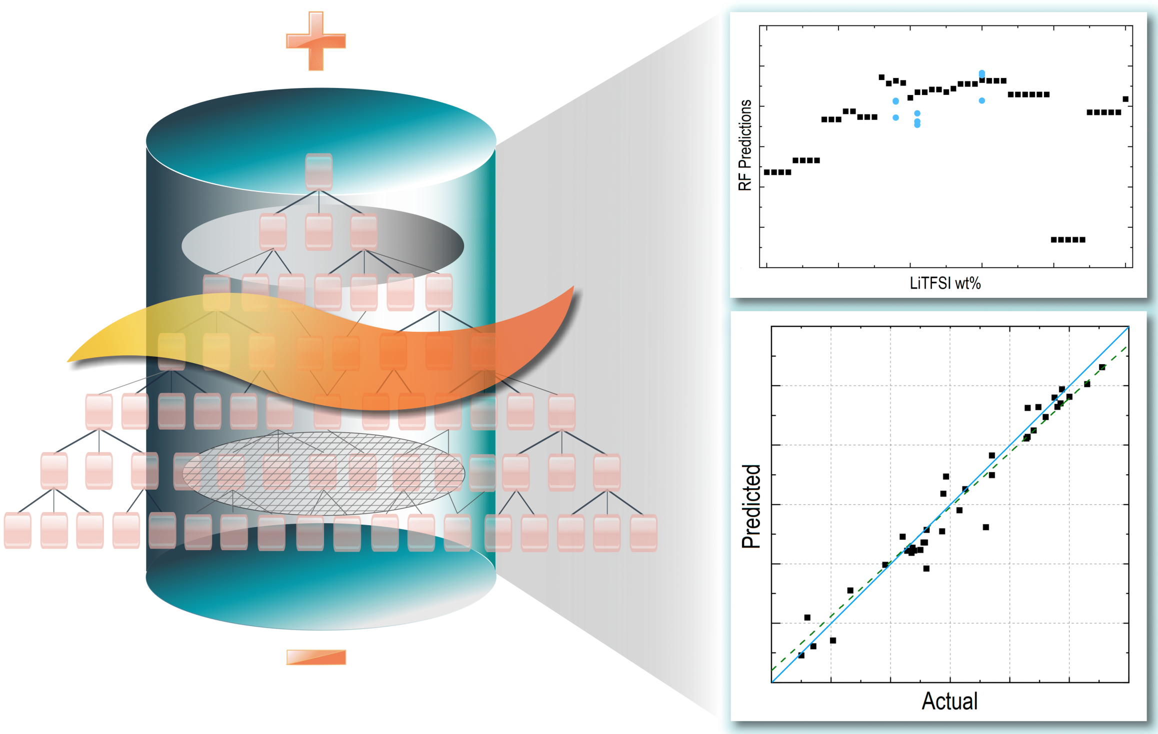
Graphical Abstract

A Data Science Approach for Advanced Solid Polymer Electrolyte Design

Marianne Liu, Conrad Clement, Kathy Liu, Xuming Wang, Taylor D. Sparks



Highlights

A Data Science Approach for Advanced Solid Polymer Electrolyte Design

Marianne Liu, Conrad Clement, Kathy Liu, Xuming Wang, Taylor D. Sparks

- Machine learning models are analyzed for solid polymer electrolyte materials discovery

- High predictive capability was attained using a random forest model

- Data combined from literature and independent experimentation yields efficient model predictivity

A Data Science Approach for Advanced Solid Polymer Electrolyte Design

Marianne Liu[a,b], Conrad Clement[c], Kathy Liu[d], Xuming Wang[b], Taylor D. Sparks[b]

[a] *West High School, Salt Lake City, UT, 84103*
[b] *Materials Science & Engineering Department, University of Utah, Salt Lake City, UT, 84112*
[c] *Mathematics Department, University of Utah, Salt Lake City, UT, 84112*
[d] *Materials Science & Engineering Department, Stanford University, Stanford, CA, 94305*

Abstract

Solid polymer electrolytes (SPEs) stand to revolutionize battery technology innovation by making batteries non-flammable, flexible, and more sustainable. However, SPE breakthroughs are limited by the highly time and resource-intensive nature of battery research. Even when suitable materials are discovered, optimizing the composition and experimental conditions presents another critical barrier to SPE realization. In this work, a data-driven approach to SPE development is presented. First, data is collected and analyzed from published literature, and then supplemented with independent experimentation to complete the SPE dataset. Then, six different models (linear regression, lasso regression, ridge regression, decision tree, random forest, and radial basis function SVM) were tested. The random forest model is identified as the most suitable model with the greatest predictive capability, validated by independent experimentation and by comparing predicted activation energies to those reported in literature using raw predictions from the model. The random forest model is calculated to predict conductivity with a root-mean-square-error of 0.332 log(S/cm). By applying machine learning to incorporate important parameters of SPE synthesis, this study provides a foundation for accelerated SPE innovation.

Keywords: lithium ion battery, electrolyte, materials informatics, solid polymer electrolyte

1. Introduction

Batteries are a crucial component of all portable electronics; however, novel breakthroughs are barred by the highly time and resource-intensive nature of battery research, making it necessary to find a solution for more efficient battery innovation. Current commercial batteries most commonly use flammable liquid electrolytes; however, solid polymer electrolytes (SPEs) have the potential to replace liquid electrolytes to develop higher-performing batteries with higher safety, design flexibility, and application capabilities [1, 2]. With increased emphasis on SPEs in recent years [3, 4, 5], it becomes crucial to develop an effective method for more efficient SPE innovation.

Ionic conductivity is a critical indicator of SPE performance. The composition and microstructure, which are functions of a material's processing history, will drastically impact the ionic conductivity. Optimizing these factors requires lengthy research time and is highly resource intensive. It is possible that machine learning (ML) can be used to develop actionable relationships between composition, processing, microstructure and properties [6]. ML could also quicken the innovation process for SPEs, as the current timeline of commercializing lithium ion batteries is very long [7].

Past research into using ML techniques for battery innovation has been mixed. For example, Kauwe et al. [7] showed that even with extensive experimental data, machine learning models attempting to predict device level performance, such as energy density and capacity retention, are very difficult given the enormous number of variables present in the anode, cathode, and electrolyte components of the device. On the other hand, predicting performance of individual materials within these device components has shown promising results. Ahmad et al. [8] screened over 12,000 inorganic solids and found six solid electrolytes that were predicted to be stable in the presence of dendrite initiation. Ellis et al. [9] developed a ML model capable of determining the concentrations of major components in lithium-ion battery electrolytes with an approximate accuracy of 3-5 weight percentage (wt.%). Jalem et al. worked with multiple researchers to develop a method of predicting the Li-ion migration barrier as a substitute measurement of ionic conductivity and to find potential solid oxide electrolyte materials with low Li-ion hopping energies (E_a) [10, 11, 12, 13]. Sendek et al. used both computational screening [14] and ML methods [15] capable of identifying Li-ion conductors more efficiently and accurately than traditional screening methods to analyze 12,831 materials for promising candidates for solid state electrolytes in lithium-ion batteries. They found 21 candidates with high ionic conductivity, robust stability, and low cost. Fujimura et al. combined DFT calculations with ML techniques and used both theoretical and experimental datasets to predict the conductivity of various compositions of LISICON-type materials at 373 K [16]. The transport properties of garnet-type metal oxides were then evaluated using support vector regression (SVR), revealing the chemical composition–structure–ionic conductivity relationships [17]. Others have investigated ionic transport mechanisms [18, 19], predicted the voltage of electrode materials [20], and explored potential SPE materials [21, 22] using ML as well.

While efforts have been made to approximate and predict properties of electrolytes relating to ionic conductivity, research into predicting the ionic conductivity itself using ML modelling techniques is scarce. In contrast, this study provides an analysis of ML techniques to predict the ionic conductivity of an SPE given its composition and operating temperature by combining information available in the literature with carefully designed experiments. Many ML techniques, and particularly deep learning approaches, require large dataset sizes (greater than 10^3) that are difficult to procure for many materials research projects [23, 24]. Furthermore, the lengthy battery creation and testing procedures also limits the amount of data available for ML algorithms to train from. Thus, an efficient model designed to operate in the limitation of scarce data could be enormously impactful in accelerating materials science innovations.

Polyethylene oxide (PEO) is among the most popularly researched SPE hosts due to its salt-solvation abilities, stability, and commercial availability. PEO in the amorphous phase affords segmental mobility for high ionic conductivity, especially with bulky anion salts such as $Li(CF_3SO_2)_2N$ [lithium bistrifluoromethanesulfonate imide, LiTSFI]. Experimental and analytical methods have been employed to characterize and optimize the kinetics, dynamics, and ionic transport properties of PEO-LiTSFI polymer electrolytes, revealing the critical roles of crystalline domains, solvents, plasticizers, and additives [25]. Enabling PEO-LiTSFI to sufficiently function as an SPE material in solid-state batteries requires higher ionic conductivities equal to those of current commercial batteries containing liquid electrolytes (approximately 10^{-3} S/cm), especially around room temperature, motivating developments in synthesis methods, modifications, and derivatives [26]. With the ever-expanding range of techniques to amplify PEO-LiTSFI performance, data-driven characterization methods are highly desirable to accelerate innovation and optimization for application in batteries. A number of computational approaches have delved into characterizing ionic transport mechanisms, reporting atomic-scale analysis of structure-dynamics properties and ion diffusion models [27, 28, 29]. Modular synthesis, electrochemical characterization, and molecular simulation were used to demonstrate conductivity as a function of available lithium cation solvation sites [27, 28]. Quantum-chemistry-based molecular dynamics simulations have revealed ion transfer occurs through simultaneous intersegmental hopping and PEO chain movement [29].

A Bayesian optimization algorithm integrated with learned coarse-grained molecular dynamics (CGMD) has previously been used to gain a comprehensive description of the relationships between the lithium conductivity and material properties at the molecular level to improve components on current PEO-LiTFSI SPEs [30]. Besides this, few have used ML as a tool in furthering PEO-LiTFSI SPE innovation despite the large amount of research studying the PEO-LiTFSI system, which suggests opportunities to discover new methods of SPE innovation via ML. It's understood that PEO-LiTFSI is not the highest performing electrolyte system, however, it has been extensively researched and therefore is more favorable to use for ML. This study focuses only on linear PEO chains for consistency within the dataset and uses EO/Li ratio, a crucial parameter for ion transport and SPE performance, as a precursor to the study of other parameters (such as additives, processing conditions, PEO structure, and type of lithium salt) that determine the chemical and physical properties of an SPE. It also seeks to investigate relationships between compositional and electrochemical properties from a less computationally demanding machine learning perspective, applying and comparing a number of different machine learning algorithms to predict PEO-LiTFSI conductivity from experimentally reported values. We feed only two parameters into each regression model: temperature and ethylene oxide (EO)/Li ratio. We intentionally chose a simple system to demonstrate the utility of ML for battery research and materials discovery, and to provide a foundation for further research into this field.

2. Methods

2.1. Data Collection and Aggregation

Machine learning requires data that is representative of the problem domain at hand. When extracting data from literature, there may be a lack of parity in the provided information, which means it may be necessary to supplement existing data in literature with those from independent experimentation. This result was witnessed first hand during data collection: the original intention was to use only one dataset with information collected purely from literature, but as more data was added, it was shown that the accuracy of several regression models (namely the linear, lasso, and ridge regression models) began to decrease. To that end, we investigated the influence of using different sets of data for different informational models, and thus proposed three data schemes:

Dataset	Description
1	The first chronologically compiled dataset; consisting of 76 data points collected purely through literature review [31, 32, 33, 34, 35, 36]
2	The second chronologically compiled dataset; a superset of all 76 data points from Dataset 1, with an additional 70 data points collected purely through literature review [31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41]
3	The third chronologically compiled dataset; a superset of all 146 data points from Dataset 2, combined with an additional 47 data points collected from targeted experimentation where attention was paid to class balance and representation [31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41]

Table 1: We make a distinction between Datasets 1 and 2 because it was observed that the addition of more literature data in Dataset 2 resulted in less accurate predictions. This is further explained in the paragraph preceding this table.

Almost 200 literature samples were analyzed by extrapolating data from conductivity vs. temperature plots for different EO/Li ratios. Collected data was standardized to the extent possible across all literature to minimize inconsistencies between articles in measurement and assembly methods. For example, for a fair comparison and to eliminate other potentially contributing factors, it was ensured that all reported electrolytes found in literature contained no additives. Additionally, reported values were converted into standardized units: LiTSFI wt% for EO/Li ratio using PEO molecular weight, Celsius for temperature, and logarithmically scaled conductivity (S/cm). Other variables that are likely present in different reports in literature, including synthesis methods, testing equipment, and testing methods, are not covered in this study but are ripe for future work.

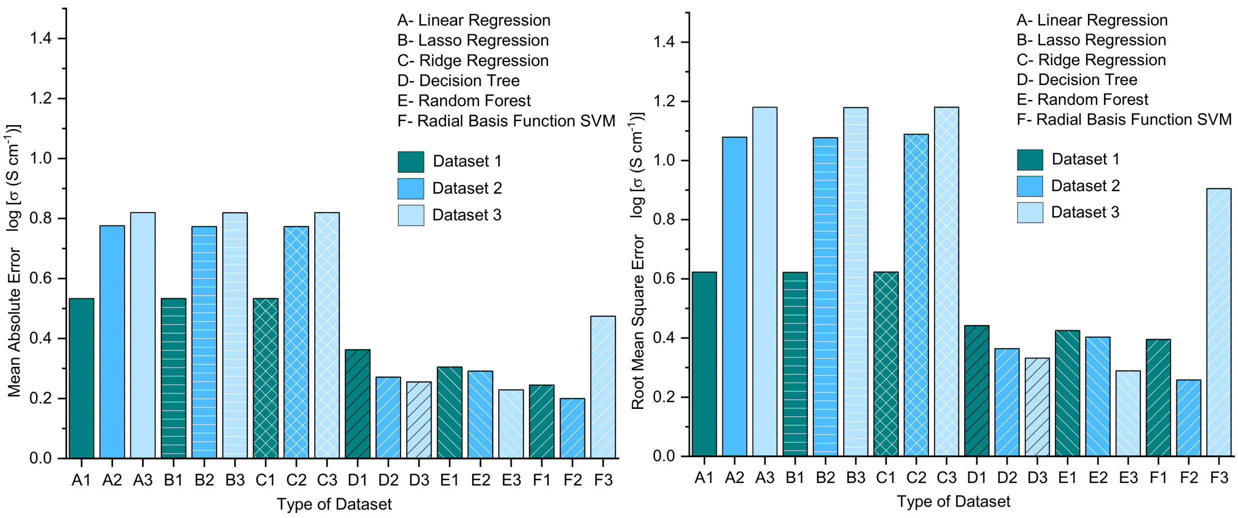
To collect data from independently conducted experiments, SPEs were synthesized and measured. LiTSFI (Lithium bis(trifluoromethanesulfonyl)imide ($LiN(SO_2CF_3)_2$, 99.5%), supplied by MilliporeSigma, was dissolved in acetonitrile (CH_3CN), added to PEO (MilliporeSigma, $M_w \approx 4 \times 10^6$ g/mol) according to the specified LiTFSI wt.%, and stirred for 4 h at room temperature. 1 mL of solution was dripped into a 1.6 cm diameter circular mold for overnight evaporation in an argon-filled glove box. Three SPEs of each LiTFSI wt.% (15, 25, 35, 45, 55, and 65%) were created and tested at 25, 30, 35, 45, 65°C. To test conductivity, SPEs were sandwiched between two stainless steel discs and placed in a cell holder with metal contacts that applied constant pressure. AC impedance spectroscopy was then tested with a Gamry Instruments Potentiostat to measure the resistance of the electrolytes using an AC amplitude of 10 mW between sweep frequencies of 100 kHz to 10 mHz. The ionic conductivity was calculated from the electrolyte resistance (R_b) obtained from the intercept of the AC impedance spectra with the real axis [42], using:

$$\sigma = \frac{l}{S R_b}$$

where l is the electrolyte thickness, S is the surface area, R_b is the bulk resistance (obtained from the x-axis intercept of the corresponding impedance spectra), and σ is the ionic conductivity.

2.2. Model Creation

Six different regression models were created using Python (version 3.7): linear regression, lasso regression, ridge regression, decision tree, random forest [43], and a support vector machine (SVM) [44]. The SVM uses a radial basis function kernel in which the distance between two feature vectors x_1 and x_2, each representing



(a) Mean Absolute Error

(b) Root Mean Square Error

Figure 1: Calculated error rates for each ML model, shown along the y-axis. The datasets are both color-coded and labeled along the x-axis. Both the MAE and RMSE values are equal for linear regression, lasso regression, and ridge regression models due to rounding, and are not because of plotting error.

particular data points, is given by

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2}\right)$$

A standard scaler was used on each dataset before being fitted to each model. Hyperparameter optimization was performed on all models using a grid search with fivefold cross-validation. Finally, each model was evaluated exactly once on the designated test dataset.

Packages such as numpy [45], pandas [46], and sklearn (version 0.21.3) [47] were all imported and used in the program. Each input dataset was split into training and testing sets through the train_test_split function in sklearn with a test size of 20% and random state of 5. Because the random state remained unchanged, the same training and test sets were also used on all of the models. The LiTFSI wt.% predicted to have the highest conductivity was recorded as the ”peak conductivity value” for each model.

2.3. Experimental Model Validation

The best performing ML model's predictions were validated by independently conducted experiments. Three SPE compositions near the wt.% resulting in peak conductivity were created in the lab according to experimental procedures. SPE's of 28, 31, and 40 wt% LiTFSI were created and tested at temperatures of 25, 35, 45, 55, and 65°C. These results were then compared with the predicted values of conductivity from the best performing ML model.

2.4. Theoretical Model Validation

Further validation to the models previously developed was conducted using the Arrhenius equation to calculate and compare predicted activation energies to those in literature. The ML models allow predictions that bypass lengthy temperature testing. Furthermore, comparing activation energies allows comparisons of predictions across multiple temperatures, whereas the experimental validation only compares composition. The Arrhenius equation was used to predict activation energy, E_a (eV), using ML-predicted conductivities according to:

$$E_a = -R\left[\frac{d(log\sigma)}{d(1000/T)}\right]$$

where R is the universal gas constant (8.314 kJ/mol), σ is the conductivity (S/cm), and T is the temperature in Kelvin. Calculated E_a were compared with values reported in literature.

3. Results and Discussion

Predictions from each model were compared using mean absolute error (MAE) and root mean square error (RMSE), as shown in Figure 1a and 1b, respectively. Parity plots were created to visually represent the differences between predicted and actual conductivity values for each model, in which both values were plotted against each other (Figure 2). A model with 100% accuracy would display a perfectly diagonal trendline, represented as a solid blue line in each graph. Predicted vs. actual conductivities are shown in black dots for each graph and the fitted linear trendline for these data points are shown as dotted green lines. Parity plots offer a clear depiction of model accuracy and are commonly used in ML literature.

3.1. Dataset Comparison

By looking at Figure 1a and 1b, we see that the decision tree, random forest, and radial basis function SVM all increase in accuracy with the addition of 70 more data points in Dataset 2 compared with Dataset 1, which was expected. The exact opposite was observed with the linear, lasso, and ridge regression models, in that they consistently predicted more accurately with Dataset 1 than Dataset 2. For example, the RMSE for the linear regression is 0.623 log(S/cm) for Dataset 1 and increases to 1.079 log(S/cm) when more literature data is added with Dataset 2. This result was surprising at first: it is expected that more data will increase model accuracy by providing additional samples to train from.

Further investigation of these 70 literature points that decreased the accuracy of those models revealed that a particularly low conductivity was reported by M. Marzantowicz [33], despite higher conductivities being reported for similar LiTFSI wt% by other authors. This outlier data can be seen by the extremely low ionic conductivity value from 50 to 55 LiTFSI wt% in Figure 4, as well as the scattering of activation energy in Figure 6. The effects of this outlier on the models are further exacerbated because few researchers have synthesized electrolytes containing such high amounts of LiTFSI. This result demonstrates the necessity for researchers to report detailed procedures for accurate literature comparison. For example, literature seldom reports key factors such as: stirring methods, oxygen and humidity levels in the stirring environment, and drying environment.

The comparison of Datasets 1 and 2 also illuminates the value of learning from data that has been carefully curated, which can often only be done through independent experiments. To this end, we now attempt to show ML performance with Dataset 3, which additionally includes experimental data specific to this study. It was hypothesized that the addition of independently conducted experiments would decrease the existing noise in Dataset 2. And, as seen in Figure 1a and 1b, Dataset 3 produced mixed results.

On one hand, the linear, lasso, and ridge regression models all decreased in accuracy once again, which makes it clear that these models are not properly suited for this type of problem. On the other hand, the addition of experimental data in Dataset 3 resulted in more accurate predictions from the decision tree and random forest models once again. For example, the random forest model had an RMSE of 0.403 log(S/cm) using Dataset 2, which decreases to 0.289 log(S/cm) using Dataset 3. This tells us that these two models are capturing important information in the feature space and are more appropriate in this situation. The radial basis function SVM performed strangely using Dataset 3. With Datasets 1 and 2, it performed the most accurate out of all the models, however with Dataset 3, its errors dramatically increase. From this, we conclude that the SVM is not a robust model for this situation because it failed to generalize well to additional data.

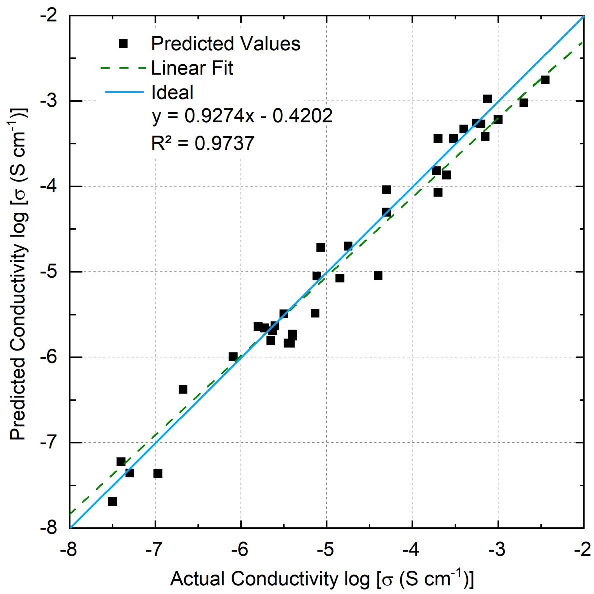


Figure 2: Parity plot of the random forest model fitted to Dataset 3.

3.2. Model Comparison

Linear, lasso, and ridge regression all perform equally with the worst accuracy out of all six models created previously (MAE: 0.820 log(S/cm), RMSE: 1.180 log(S/cm)) when using Dataset 3, as shown in Figure 1a and 1b. The low predictive accuracy is not surprising because they are all simple models; however, simple models can be useful for extrapolation, as demonstrated by Kauwe et al. [48]. In addition, linear regression cannot capture the power dependency of conductivity on the non-linearly related parameters of temperature and composition. The decision tree model has higher accuracy than the linear regression model

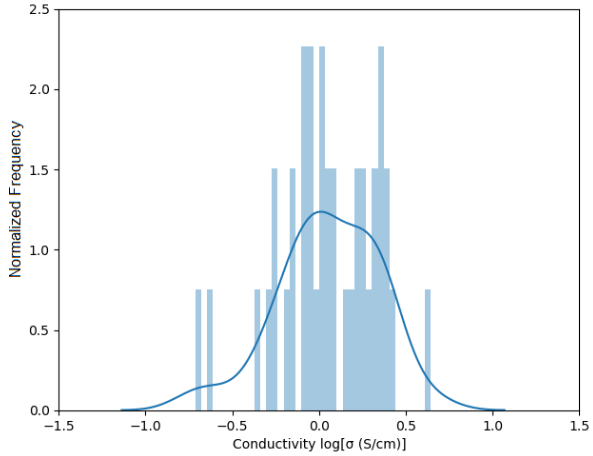


Figure 3: A residual histogram depicting the random forest model's error. The magnitude of error is displayed along the x-axis and the proportion of occurrences of error on the y-axis.

(MAE: 0.255 log(S/cm), RMSE: 0.332 log(S/cm)) using Dataset 3. The random forest model performs with the highest accuracy (MAE: 0.229 log(S/cm), RMSE: 0.289 log(S/cm)), which translates to a percent error of 4.808%. For hyperparameter optimization, we used the following hyperparameters in a grid search for the random forest: max_depth, min_samples_leaf, min_samples_split, and n_estimators. From looking at Figure 2, the random forest model also shows the most similar trendline to the optimal trendline, with a slope of 0.9737. This makes sense because a random forest reduces any outlier decision tree predictions that may exist. In comparison to literature accuracy, Fujimura [16] reports an optimized predicted average error of 0.373, which is within an order of magnitude of the MAE (0.248) reported herein. To our knowledge, there are not many other existing studies for direct comparison of model accuracy in predicting the ionic conductivity of SPEs.

Figure 3 shows a residual plot of the error from the random forest model's predictions. The normalized error has upper and lower bounds of 1 and -1, meaning most of the data has an error near 0, with the greatest error 1 order of magnitude of conductivity away. The trained random forest model performs better than other types of machine learning models created to predict the characteristics of batteries currently reported.

When comparing the RF's predictive performance with that of the radial basis function SVM, figures 1a and 1b show that the SVM performs with a MAE of 0.474 log(S/cm) and RMSE of 0.905 log(S/cm). As mentioned before, even though the SVM performed the best out of all the models using Datasets 1 and 2, it dramatically increased in error using Dataset 3 which tells us it is not a robust model capable of handling new data well. Moreover, because the SVM had poor prediction accuracy, we do not see evidence of a relationship between salt composition, temperature, and conductivity captured by transforming the dimensionality of the model space. Therefore, it is shown that the random forest model is the best predictor of ionic conductivity.

3.3. Model Validation

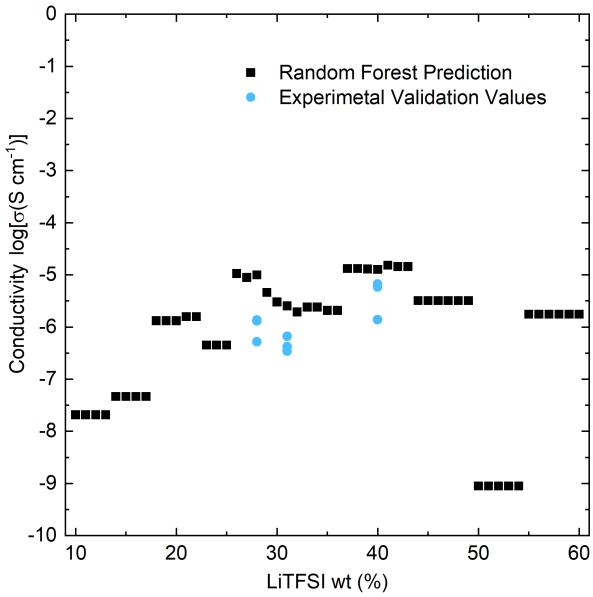


Figure 4: Random forest model's predictions of conductivity for LiTFSI composition from 10% to 60% at room temperature (25°C) in black while the experimental conductivities are shown in blue. The sudden decrease of conductivity seen at around 50 LiTFSI wt% is due to an outlier in the dataset that dramatically skewed the model. This is further discussed in section 3.1 paragraph 2.

After creating a working model with the random forest algorithm, predictions were validated with physical experimentation, as shown in Figure 4. It is also seen that the random forest model predicted a peak conductivity of 41-43 wt% LiTSFI. We believe that the random forest model's predictions show a step-like dependence as a result of only sampling a few concentrations. If the dataset included a greater range of LiTFSI wt% concentrations at smaller intervals, then the graph would gradually smooth out.

These experimental validations found that the random forest model's predictions were fairly accurate with a MAE of 0.253 log(S/cm) and a RMSE of 0.453 log(S/cm). The reason why these calculated MAE and

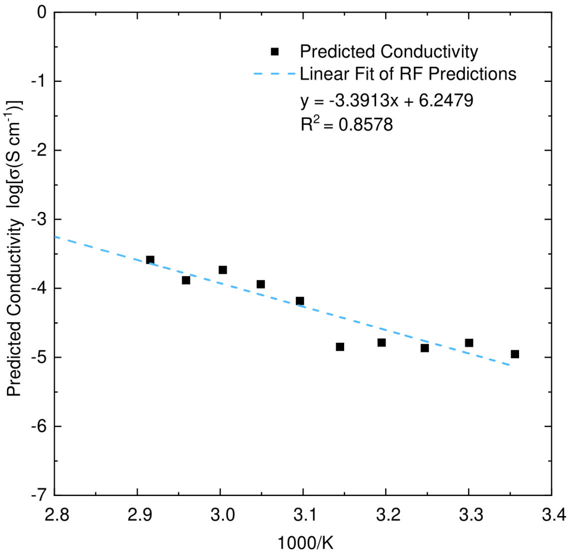


Figure 5: Random forest predictions on temperature dependence of conductivity for a 40 wt% LiTFSI SPE across different operating temperatures

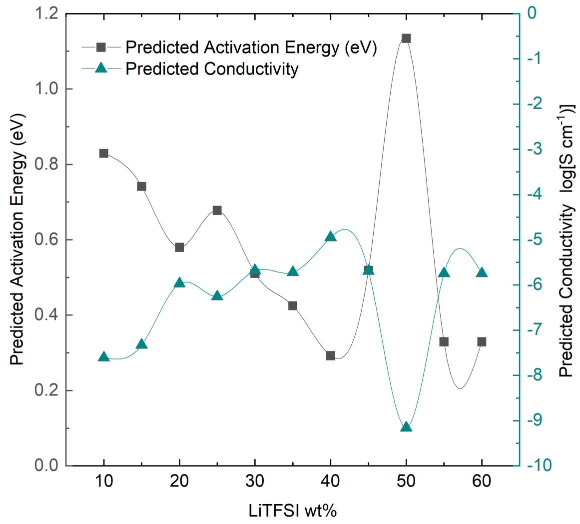


Figure 6: Correlation between the random forest model's predicted values for activation energy, E_a, and conductivity at 25°C. As discussed in section 3.1 paragraph 2, the dramatic decrease in conductivity and consequently, increase in E_a, is due to an outlier found in the dataset.

RMSE are slightly higher than those reported in the model comparisons above is possibly due to the small LiTSFI wt% range tested experimentally. Nonetheless, these reported MAE and RMSE still demonstrate a relatively low amount of error predicted by the random forest model and provides further validation towards its predictions.

Next, we compared E_a calculated from the random forest model predictions with those reported in literature. A lower E_a directly correlates to faster Li^+ ion diffusion. In Figure 9, a linear trend is observed: conductivity increases as temperature increases (the value of 1000/T decreases), which is a relationship that has been previously observed in literature, thus confirming the ability of the random forest model to track theoretical trends. It is assumed herein that the pre-exponential factor remains constant because the electrolyte constituents remain unchanged. The pre-exponential factor does not vary based on the ratio of constituents (LiTSFI wt%).

Lower E_a means easier ionic transportation and higher ionic conductivity. Figure 10 shows that the random forest model does in fact predict this inverse relationship between E_a and conductivity. When comparing the predicted values of E_a from Figure 6 with those that are reported in literature, Chen [49] and Zhao [50] report E_a of 77.56 and 75.5 kJ/mol for a PEO-LiTFSI SPE of 18 wt% LiTFSI, respectively. The random forest model predicts a similar value of 71.53 kJ/mol at 15 wt% LiTFSI, demonstrating high E_a prediction accuracy. This can be harnessed to uncover E_a in an accelerated manner for uncharacterized SPE materials.

4. Conclusion

This study presents an investigation in SPE material development by comparing various ML methods for accuracy and relevance to SPE composition and temperature in order to determine an ideal model.

After analysis, it was found that a dataset combining data from both literature and independently conducted experiments (Dataset 3) results in the most accurate predictions from the decision tree and random forest models. Independently controlled experiments are demonstrated to enable key standardization for accurate data inputs and fill in gaps for conductivity data from compositions that were not accounted for in literature. Overall, using a combination of data from experiments and literature review is shown to be an effective approach for high-accuracy prediction of SPE conductivity by composition.

Furthermore, this study demonstrates that the addition of data without homogeneous parameters can be detrimental for accurate ML model predictions, as varied control parameters muddle trend prediction (an apples-to-apples comparison is necessary). It was seen that including more data from a purely literature (Dataset 2) made predictions less accurate for the decision and random forest models. These results highlight a crucial future direction for battery researchers: experimental parameters and reporting procedures must be standardized to provide both meaningful independent results as well as aggregable information to help stabilize models, enabling accelerated collaboration between materials and computational researchers.

The random forest model was found to be the best predictive model compared with all other tested models: linear regression, lasso regression, ridge regression, decision tree, and the radial basis function SVM. The SVM developed was seen to be less accurate than the random forest model. This shows that while transforming the model through selection of a kernel function can improve its accuracy to some extent, SPE prediction specifically benefited most from a random forest model. The low MAE and RMSE are significantly better than previously reported studies.

The LiTFSI wt% that resulted in the highest conductivity was found to be between 41% and 43%, which was validated by independently conducted experiments. Further validation using the Arrhenius Equation to calculate E_a for different compositions of SPEs found that the random forest model predicted similar E_a as those reported in literature and predicted similar trends seen between conductivity and temperature.

This study was able to achieve high resolution and predictive abilities that normally would not have been possible due to material, time, and human error constraints. These results provide a foundation in a method for researchers to advance battery innovation and have the potential for many applications to related materials research in batteries.

This study builds on a growing body of evidence using materials informatics for materials development to incorporate aspects of solid electrolyte synthesis that researchers have not thoroughly investigated, such as the effects of additives, stirring time, stirring speed, stirring method, drying temperature, drying environment. Optimizing PEO molecular weight is another parameter to consider, as it was not explored in this study.

By studying the relationship of temperature and lithium salt concentration for one type of electrolyte (LiTFSI-PEO), this model can also be extended to include other types of commonly used salts and can perhaps be used to predict conductivities for bisalt and trisalt components as well. The model can also be expanded to evaluate materials for other battery components, including cathodes, anodes, electrode-electrolyte interfaces, and potential additive components.

5. Data Availability

The code and data used in this article can be found at https://github.com/mliu7051/SPE-Design.

6. Acknowledgements

TDS acknowledges support from the National Science Foundation through CAREER Award DMR 1651668. ML acknowledges Professor Dmitry Bedrov and Kaai Kauwe from the Materials Science and Engineering Department at the University of Utah for key discussions. ML would also like to thank teachers Melissa Anderson and Todd Vawdrey at West High School for their continued inspiration and encouragement.

References

[1] RC Agrawal and GP Pandey. Solid polymer electrolytes: materials designing and all-solid-state battery applications: an overview. *Journal of Physics D: Applied Physics*, 41(22):223001, 2008.

[2] La Li, Zheng Lou, Di Chen, Kai Jiang, Wei Han, and Guozhen Shen. Recent advances in flexible/stretchable supercapacitors for wearable electronics. *Small*, 14(43):1702829, 2018.

[3] Mukul D Tikekar, Snehashis Choudhury, Zhengyuan Tu, and Lynden A Archer. Design principles for electrolytes and interfaces for stable lithium-metal batteries. *Nature Energy*, 1(9):1–7, 2016.

[4] Liping Yue, Jun Ma, Jianjun Zhang, Jingwen Zhao, Shanmu Dong, Zhihong Liu, Guanglei Cui, and Liquan Chen. All solid-state polymer electrolytes for high-performance lithium ion batteries. *Energy Storage Materials*, 5:139–164, 2016.

[5] Thomas F Miller III, Zhen-Gang Wang, Geoffrey W Coates, and Nitash P Balsara. Designing polymer electrolytes for safe and high capacity rechargeable lithium batteries. *Accounts of chemical research*, 50(3):590–593, 2017.

[6] Krishna Rajan. Materials informatics. *Materials Today*, 8(10):38–45, 2005.

[7] Steven K Kauwe, Trevor David Rhone, and Taylor D Sparks. Data-driven studies of li-ion-battery materials. *Crystals*, 9(1):54, 2019.

[8] Zeeshan Ahmad, Tian Xie, Chinmay Maheshwari, Jeffrey C Grossman, and Venkatasubramanian Viswanathan. Machine learning enabled computational screening of inorganic solid electrolytes for suppression of dendrite formation in lithium metal anodes. *ACS central science*, 4(8):996–1006, 2018.

[9] LD Ellis, S Buteau, Samuel G Hames, LM Thompson, DS Hall, and JR Dahn. A new method for determining the concentration of electrolyte components in lithium-ion cells, using fourier transform infrared spectroscopy and machine learning. *Journal of The Electrochemical Society*, 165(2):A256, 2018.

[10] Randy Jalem, Takahiro Aoyama, Masanobu Nakayama, and Masayuki Nogami. Multivariate method-assisted ab initio study of olivine-type limxo4 (main group m2+–x5+ and m3+–x4+) compositions as potential solid electrolytes. *Chemistry of Materials*, 24(7):1357–1364, 2012.

[11] Randy Jalem, Masanobu Nakayama, and Toshihiro Kasuga. An efficient rule-based screening approach for discovering fast lithium ion conductors using density functional theory and artificial neural networks. *Journal of Materials Chemistry A*, 2(3):720–734, 2014.

[12] Randy Jalem, Mayumi Kimura, Masanobu Nakayama, and Toshihiro Kasuga. Informatics-aided density functional theory study on the li ion transport of tavorite-type limto4f (m3+–t5+, m2+–t6+). *Journal of chemical information and modeling*, 55(6):1158–1168, 2015.

[13] Randy Jalem, Kenta Kanamori, Ichiro Takeuchi, Masanobu Nakayama, Hisatsugu Yamasaki, and Toshiya Saito. Bayesian-driven first-principles calculations for accelerating exploration of fast ion conductors for rechargeable battery application. *Scientific reports*, 8(1):1–10, 2018.

[14] Austin D Sendek, Qian Yang, Ekin D Cubuk, Karel-Alexander N Duerloo, Yi Cui, and Evan J Reed. Holistic computational structure screening of more than 12000 candidates for solid lithium-ion conductor materials. *Energy & Environmental Science*, 10(1):306–320, 2017.

[15] Austin D Sendek, Ekin D Cubuk, Evan R Antoniuk, Gowoon Cheon, Yi Cui, and Evan J Reed. Machine learning-assisted discovery of solid li-ion conducting materials. *Chemistry of Materials*, 31(2):342–352, 2018.

[16] Koji Fujimura, Atsuto Seko, Yukinori Koyama, Akihide Kuwabara, Ippei Kishida, Kazuki Shitara, Craig AJ Fisher, Hiroki Moriwake, and Isao Tanaka. Accelerated materials design of lithium superionic conductors based on first-principles calculations and machine learning algorithms. *Advanced Energy Materials*, 3(8):980–985, 2013.

[17] Natalia Kireeva and Vladislav S Pervov. Materials space of solid-state electrolytes: Unraveling chemical composition–structure–ionic conductivity relationships in garnet-type metal oxides using cheminformatics virtual screening approaches. *Physical Chemistry Chemical Physics*, 19(31):20904–20918, 2017.

[18] A Maitra and A Heuer. Cation transport in polymer electrolytes: a microscopic approach. *Physical review letters*, 98(22):227802, 2007.

[19] Mahati Chintapalli, Ksenia Timachova, Kevin R Olson, Sue J Mecham, Didier Devaux, Joseph M DeSimone, and Nitash P Balsara. Relationship between conductivity, ion diffusion, and transference number in perfluoropolyether electrolytes. *Macromolecules*, 49(9):3508–3515, 2016.

[20] Rajendra P Joshi, Jesse Eickholt, Liling Li, Marco Fornari, Veronica Barone, and Juan E Peralta. Machine learning the voltage of electrode materials in metal-ion batteries. *ACS applied materials & interfaces*, 11(20):18494–18503, 2019.

[21] Arun Mannodi-Kanakkithodi, Anand Chandrasekaran, Chiho Kim, Tran Doan Huan, Ghanshyam Pilania, Venkatesh Botu, and Rampi Ramprasad. Scoping the polymer genome: A roadmap for rational polymer dielectrics design and beyond. *Materials Today*, 21(7):785–796, 2018.

[22] Kan Hatakeyama-Sato, Toshiki Tezuka, Yoshinori Nishikitani, Hiroyuki Nishide, and Kenichi Oyaizu. Synthesis of lithium-ion conducting polymers designed by machine learning-based prediction and screening. *Chemistry Letters*, 48(2):130–132, 2019.

[23] Ram Seshadri and Taylor D Sparks. Perspective: interactive material property databases through aggregation of literature data. *APL Materials*, 4(5):053206, 2016.

[24] Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.

[25] Ludvig Edman, Marca M Doeff, Anders Ferry, John Kerr, and Lutgard C De Jonghe. Transport properties of the solid polymer electrolyte system p (eo) n litfsi. *The Journal of Physical Chemistry B*, 104(15):3476–3480, 2000.

[26] Zhigang Xue, Dan He, and Xiaolin Xie. Poly (ethylene oxide)-based electrolytes for lithium-ion batteries. *Journal of Materials Chemistry A*, 3(38):19218–19253, 2015.

[27] Michael A Webb, Yukyung Jung, Danielle M Pesko, Brett M Savoie, Umi Yamamoto, Geoffrey W Coates, Nitash P Balsara, Zhen-Gang Wang, and Thomas F Miller III. Systematic computational and experimental investigation of lithium-ion transport mechanisms in polyester-based polymer electrolytes. *ACS central science*, 1(4):198–205, 2015.

[28] Bing Sun, Jonas Mindemark, Evgeny V Morozov, Luciano T Costa, Martin Bergman, Patrik Johansson, Yuan Fang, István Furó, and Daniel Brandell. Ion transport in polycarbonate based solid polymer electrolytes: experimental and computational investigations. *Physical Chemistry Chemical Physics*, 18(14):9504–9513, 2016.

[29] Oleg Borodin and Grant D Smith. Mechanism of ion transport in amorphous poly (ethylene oxide)/litfsi from molecular dynamics simulations. *Macromolecules*, 39(4):1620–1629, 2006.

[30] Yanming Wang, Tian Xie, Arthur France-Lanord, Arthur Berkley, Jeremiah A Johnson, Yang Shao-Horn, and Jeffrey C Grossman. Towards designing highly conductive polymer electrolyte by machine learning assisted coarse-grained molecular dynamics. *Chemistry of Materials*, 2020.

[31] S Das and A Ghosh. Ion conduction and relaxation in peo-litfsi-al2o3 polymer nanocomposite electrolytes. *Journal of Applied Physics*, 117(17):174103, 2015.

[32] Heng Zhang, Chengyong Liu, Liping Zheng, Fei Xu, Wenfang Feng, Hong Li, Xuejie Huang, Michel Armand, Jin Nie, and Zhibin Zhou. Lithium bis (fluorosulfonyl) imide/poly (ethylene oxide) polymer electrolyte. *Electrochimica Acta*, 133:529–538, 2014.

[33] M Marzantowicz, JR Dygas, F Krok, A Łasińska, Z Florjańczyk, E Zygadło-Monikowska, and A Affek. Crystallization and melting of peo: Litfsi polymer electrolytes investigated simultaneously by impedance spectroscopy and polarizing microscopy. *Electrochimica acta*, 50(19):3969–3977, 2005.

[34] Yue Lin, Yun Cheng, Jie Li, Jan D Miller, Jin Liu, and Xuming Wang. Biocompatible and biodegradable solid polymer electrolytes for high voltage and high temperature lithium batteries. *RSC advances*, 7(40):24856–24863, 2017.

[35] Jiayu Wan, Jin Xie, Xian Kong, Zhe Liu, Kai Liu, Feifei Shi, Allen Pei, Hao Chen, Wei Chen, Jun Chen, et al. Ultrathin, flexible, solid polymer composite electrolyte enabled with aligned nanoporous host for lithium batteries. *Nature nanotechnology*, 14(7):705–711, 2019.

[36] Yuewu Zhu, Jie Li, and Jin Liu. A bifunctional ion-electron conducting interlayer for high energy density all-solid-state lithium-sulfur battery. *Journal of Power Sources*, 351:17–25, 2017.

[37] OE Geiculescu, Jin Yang, Shuang Zhou, G Shafer, Yuan Xie, J Albright, SE Creager, WT Pennington, and DD DesMarteau. Solid polymer electrolytes from polyanionic lithium salts based on the litfsi anion structure. *Journal of The Electrochemical Society*, 151(9):A1363–A1368, 2004.

[38] Kathy Liu, Yue Lin, Jan D Miller, Jin Liu, and Xuming Wang. Study of sucrose based room temperature solid polymer electrolyte for lithium sulfur battery. *Journal of The Electrochemical Society*, 164(2):A447–A452, 2017.

[39] Yue Lin, Xuming Wang, Jin Liu, and Jan D Miller. Natural halloysite nano-clay electrolyte for advanced all-solid-state lithium-sulfur batteries. *Nano Energy*, 31:478–485, 2017.

[40] Kai Zhu, Yexiang Liu, and Jin Liu. A fast charging/discharging

all-solid-state lithium ion battery based on peo-mil-53 (al)-litfsi thin film electrolyte. *RSC Advances*, 4(80):42278–42284, 2014.

[41] Qiang Ma, Heng Zhang, Chongwang Zhou, Liping Zheng, Pengfei Cheng, Jin Nie, Wenfang Feng, Yong-Sheng Hu, Hong Li, Xuejie Huang, et al. Single lithium-ion conducting polymer electrolytes based on a super-delocalized polyanion. *Angewandte Chemie International Edition*, 55(7):2521–2525, 2016.

[42] Alex Szendrei, Taylor D Sparks, and Anil V Virkar. Measurement of polarization resistance of lsm+ ysz electrodes on ysz using ac and dc methods. *ECS Transactions*, 91(1):1363–1369, 2019.

[43] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[44] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[45] Stéfan van der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.

[46] Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.

[47] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[48] Steven K Kauwe, Jake Graser, Ryan Murdock, and Taylor D Sparks. Can machine learning find extraordinary materials? *Computational Materials Science*, 174:109498, 2020.

[49] Bo Chen, Zhen Huang, Xiaotian Chen, Yanran Zhao, Qiang Xu, Peng Long, Shaojie Chen, and Xiaoxiong Xu. A new composite solid electrolyte peo/li10gep2s12/sn for all-solid-state lithium battery. *Electrochimica Acta*, 210:905–914, 2016.

[50] Yanran Zhao, Chuan Wu, Gang Peng, Xiaotian Chen, Xiayin Yao, Ying Bai, Feng Wu, Shaojie Chen, and Xiaoxiong Xu. A new solid polymer electrolyte incorporating li10gep2s12 into a polyethylene oxide matrix for all-solid-state lithium batteries. *Journal of Power Sources*, 301:47–53, 2016.