



PAPER

OPEN ACCESS

RECEIVED
25 December 2019REVISED
29 April 2020ACCEPTED FOR PUBLICATION
13 May 2020PUBLISHED
25 May 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Machine learning-assisted cross-domain prediction of ionic conductivity in sodium and lithium-based superionic conductors using facile descriptors

Yijie Xu^{1,2}, Yun Zong^{2,4} and Kedar Hippalgaonkar^{2,3,4} ¹ Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, United Kingdom² Institute of Materials Research and Engineering (IMRE), A*STAR Agency for Science, Technology and Research, 2 Fusionopolis Way, Innovis #08-03, 138634, Singapore³ School of Material Science and Engineering, Nanyang Technological University, 639798, Singapore⁴ Authors to whom any correspondence should be addressed.E-mail: y-zong@imre.a-star.edu.sg and kedar@ntu.edu.sg**Keywords:** machine learning, batteries, NASICON, classification, descriptors, ionic conductivitySupplementary material for this article is available [online](#)

Abstract

Solid state lithium- and sodium-ion batteries utilize solid ionically conducting compounds as electrolytes. However, the ionic conductivity of such materials tends to be lower than their liquid counterparts, necessitating research efforts into finding suitable alternatives. The process of electrolyte screening is often based on a mixture of domain expertise and trial-and-error, both of which are time and resource-intensive. In this work, we present a novel machine-learning based approach to predict the ionic conductivity of sodium and lithium-based SICON compounds. Using primarily theoretical elemental feature descriptors derivable from tabulated information on the unit cell and the atomic properties of the components of a target compound on a limited dataset of 70 NASICON-examples, we have designed a logistic regression-based model capable of distinguishing between poor and good superionic conductors with a validation accuracy of over 84%. Moreover, we demonstrate how such a system is capable of cross-domain classification on lithium-based examples with the same accuracy, despite being introduced to zero lithium-based compounds during training. Through a systematic permutation-based evaluation process, we reduced the number of considered features from 47 to 7, reduction of over 83%, while simultaneously improving model performance. The contributions of different electronic and structural features to overall ionic conductivity is also discussed, and contrasted with accepted theories in literature. Our results demonstrate the utility of such a facile tool in providing opportunities for initial screening of potential candidates as solid-state electrolytes through the use of existing data examples and simple tabulated or calculated features, reducing the time-to-market of such materials by helping to focus efforts on promising candidates. Given enough data utilizing suitable descriptors, high accurate cross-domain classifiers could be created for experimentalists, improving laboratory and computational efficiency.

Introduction

Lithium ion batteries (LIBs) have become the forefront of energy storage technology for a wide range of applications, including mobile devices and electric vehicles. Similarly, Na-ion batteries have shown great promise as being a safer, more cost-effective alternative for larger grid-storage applications. However, both systems suffer from well publicized stability and safety issues [1, 2]. One of these originates in their use of liquid electrolytes consisting of an ionic salt dissolved in an organic solvent, such as LiPF_6 or NaPF_6 in ethylene and dimethyl carbonate. While such electrolytes possess high ionic conductivity while remaining affordable,

Table 1. Categorization of descriptors used in study.

| Category | Examples | Count |
|------------------------------|---|-------|
| Simple molecular descriptors | Formulaic species count, numeric species occupancy | 23 |
| Structural descriptors | Unit cell parameters, unit cell volume, ionic radii | 18 |
| Electronic descriptors | Electronegativity, ionic conductivity | 8 |

their inherent flammability increase the risk of catastrophic failure scenarios through mechanical damage, side reactions, and pressure build-up originating from detrimental thermal runaway scenarios [3].

In contrast, solid ionic electrolytes offer higher energy densities while improving operational safety by reducing the overall flammability and limiting the prevalence of side reactions and consequently allowing for the use of higher voltage cathodes. One of the primary issues in their development has been their lower ionic conductivities when compared to their liquid counterparts, often in the scale of several orders of magnitudes [4].

NASICON compounds refer to materials of the general simplified formula $\text{NaMP}_3\text{O}_{12}$, consisting of a skeleton made up of corner-sharing MO_6 octahedra and PO_4 tetrahedra, with channels accommodating the mobile Na cations [5]. First demonstrated as being suitable sodium-based solid electrolyte materials by Lalere *et al* in 2014, they have been well characterized and studied since the 1970s, resulting in a wealth of ionic conductivity data being available for analysis [6].

However, the development of suitable candidates has traditionally required either trial-and-error or systematic approaches. While the former approach is possible [7], both approaches would be prohibitively resource and labour-intensive [8]. To accelerate this search process, data-driven approaches utilizing machine learning have rapidly become more popular in modelling desirable properties in chemistry, covering a variety of applications including the prediction of ternary oxide structures [9], identification of new high temperature piezoelectrics [10], prediction of compound thermodynamic stability [11], screening of new catalyst materials [12], optimization of hydrothermal synthesis methods [13], the design of new electrolyte materials [14], and the prediction of reaction productions in organic synthesis [15].

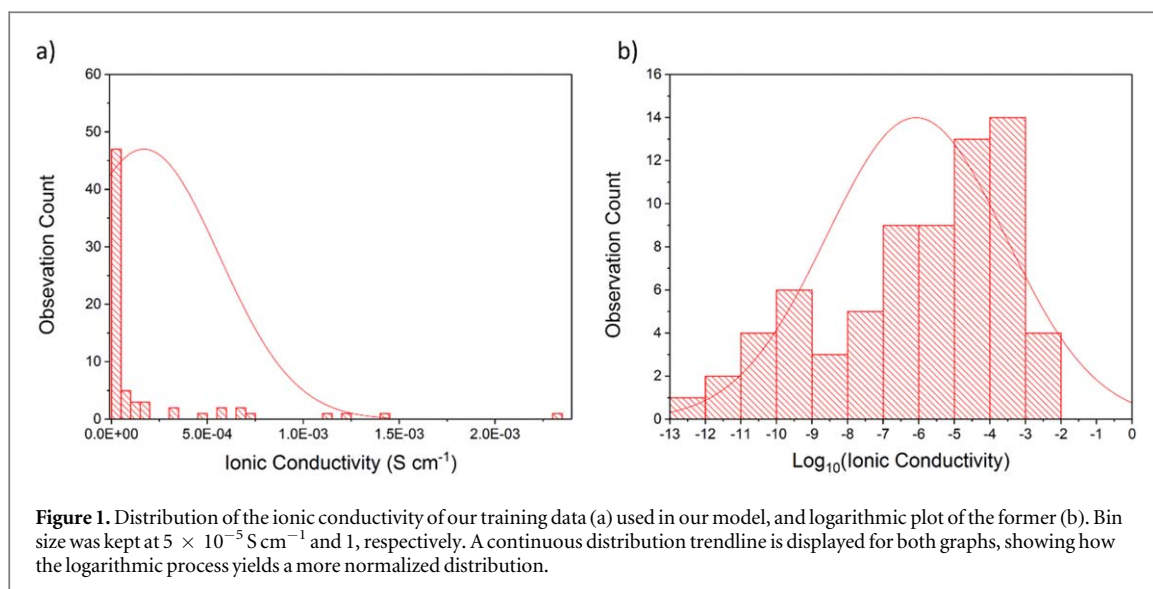
These machine learning approaches share the commonality in their attempt to capture some universal and possibly non-obvious information using certain feature-based representation that can be used to correlate to a particular desired material property for a targeted application. In their seminal work, Raccuglia *et al* converted existing knowledge about failed and successful hydrothermal experiments into calculated cheminformatical features together and tabulated atomic properties in order to predict successful conditions for vanadium selenite crystallization [13]. Most importantly, they possess a capability to extrapolate beyond their training domains in order to predict properties of foreign examples not found within their training dataset, which is achieved by feeding the model a diverse set of training examples [15]. In the case of ionic conductivity of solid-state electrolytes, a suitable set of easily computed or measured material descriptors or features may be obtained that allow for the generation of a predictive model capable of guiding future research direction with only limited available data at a rate consistently shown to be better than random trial-and-error.

In addition to allowing for data-based evaluation of literature hypotheses on the strongest factors related to ionic conduction, a machine-learning approach is significantly more robust against human bias when evaluating the potential of different compounds. In this study, we've chosen to utilize a data-driven approach relying on a predictor model trained on primarily theoretically-derived elemental and structural features, as the electronic structure-based calculations of ionic conductivity often require computationally expensive and complex calculations, which are undesirable for simple screening initiatives.

Training set

A training dataset consisting of 70 experimentally synthesized NASICON materials of R3c symmetry, along with a test dataset consisting of 57 experimentally synthesized NASICON and LISICON materials, was collected from literature to generate a feature set containing experimental ionic conductivity values measured at 25 °C, along with simple molecular descriptors, structural descriptors, and electronic descriptors [5, 16–47]. The ionic conductivity values were obtained either from tabulated sources or from Arrhenius plots, with the approach detailed in figure S2. The separation of these descriptors was done by category, and is shown below in table 1.

The formulas of our NASICON materials and accompanying ionic conductivities are listed in tables S1(a) and (b) is available online at stacks.iop.org/JPCO/4/055015/mmedia, while the methods of determination for each descriptor (henceforth feature) are listed in table S1c. A distribution of ionic conductivities along with the logarithmically transformed values is shown in figure 1. The measured ionic conductivity values of our dataset



ranged from 10^{-3} to $10^{-13} \text{ S cm}^{-1}$ in magnitude, although a strong over-representation is observed for lower ionically conducting examples, particularly of those conducting less than $10^{-8} \text{ S cm}^{-1}$.

The decision to include poorly conductive labelled examples was primarily due to both necessity to obtain data, and supported by the work by Raccuglia *et al* who demonstrated the importance of exposing any machine learning model to negative examples in order to contribute to more accurate predictions [13]. However, the poor representation of highly conducting examples in the dataset may influence the performance of our model when classifying such examples.

Besides the experimentally reported ionic conductivities, the dataset was augmented with a subset of features that may or may not be correlated to the ionic conductivity. A full list of these features together with their descriptions is exhibited in table S1(b), along with their respective Pearson correlations with experimentally measured ionic conductivity. As Pearson correlation coefficients measure the degree of linear correlation between two values, a high correlation with our desired target feature of ionic conductivity could artificially increase our classification accuracy. These features were derived from a combination of unit cell parameters and elemental descriptors derived from tabulated information, chosen based on domain expertise as well as literature suggestions, and targeted compositional as well as physical and electronic information of our compounds [5]. While the Inorganic Crystal Structure Database was consulted for detailed crystal structure information on the compounds as a possible model feature, this was both sparse and incomplete for our dataset, and were hence not considered in our study [48].

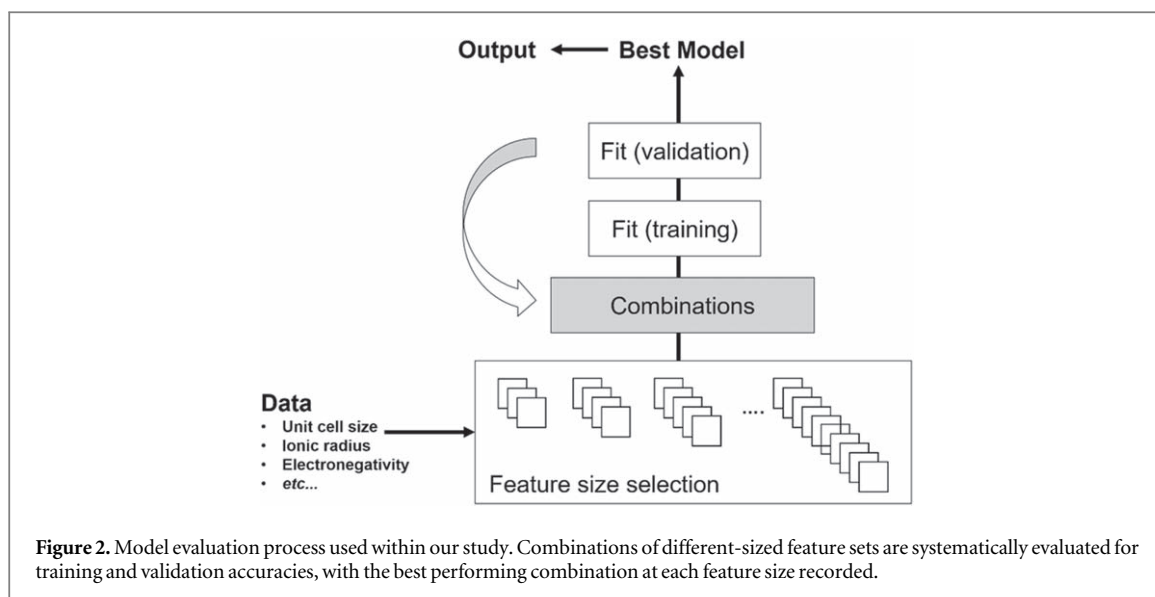
Model selection

To build an ionic conductivity predictor using our dataset, a Logistic Regression classification model was implemented. The motivation for the use of logistic regression is given in supplementary information S3. Briefly, logistic regression relies upon the use of a sigmoid function of the form shown below in equation (1):

$$P_{\text{LogReg}}(\tilde{\sigma}^i = 1) = \left[1 + \exp\left(-\sum_{j=0}^n \theta_j x_j^i\right) \right]^{-1} \quad (1)$$

Where θ_j refers to the regression coefficients of each feature x_j , and n is the total number of features. The sigmoid function component of the logistic regression equation converts the raw scalar value into one between 0 and 1, representing that probability that the particular material belonging to class 1, or excellent ionic conductor. The final predicted output of the logistic regression model is dependent on the class with greater likelihood, i.e. $\tilde{\sigma}^{(i)} = 1$ if the material i is predicted to be an excellent ionic conductor ($P > = 50\%$), and $\tilde{\sigma}^{(i)} = 0$ if the material i if the material is predicted to be a poor ionic conductor ($P < 50\%$).

We established a decision boundary of $1 \times 10^{-6} \text{ S cm}^{-1}$ by examining our dataset, to ensure an approximate equal numerical representation across the two classes. In support of this, ionic conductivity values above the boundary are classified as good ionic conductors (with a label of 1), and those below classified as poor ionic conductors (with a label of 0). Using our criteria, approximately 42.9% of our training dataset were considered class 0 and 57.1% considered class 1. Alternative decision boundaries were explored, but found to yield significantly more unbalanced output classes. Imbalanced class distributions in datasets have been well



studied for classification applications and have been shown to be linked to poorer classification performance [49].

Feature selection

47 different computed or experimentally measured features were collected for both the training and test datasets of NASICON structures. A complete list of considered features can be found in table S1(b). These carefully selected features encompass both structural and electronic attributes of their compounds, and were chosen for their computational simplicity, wherein every attribute could be derived from tabulated information associated with its unit formula.

To build a primarily theoretically-based predictive model free from the need of experimental inputs, experimentally-measured features such as activation energies were excluded from the feature list of the final model. Features related to specific elements in the test set, such as the number of a constituent atom not found within the training dataset, were excluded in order to build predictive robustness, prevent overfitting, and to evaluate classification accuracy on test examples featuring foreign elements.

The final model features used in the modelling process, together with their individual Pearson correlations with ionic conductivity for the training dataset, are given in table S1.

Evaluation

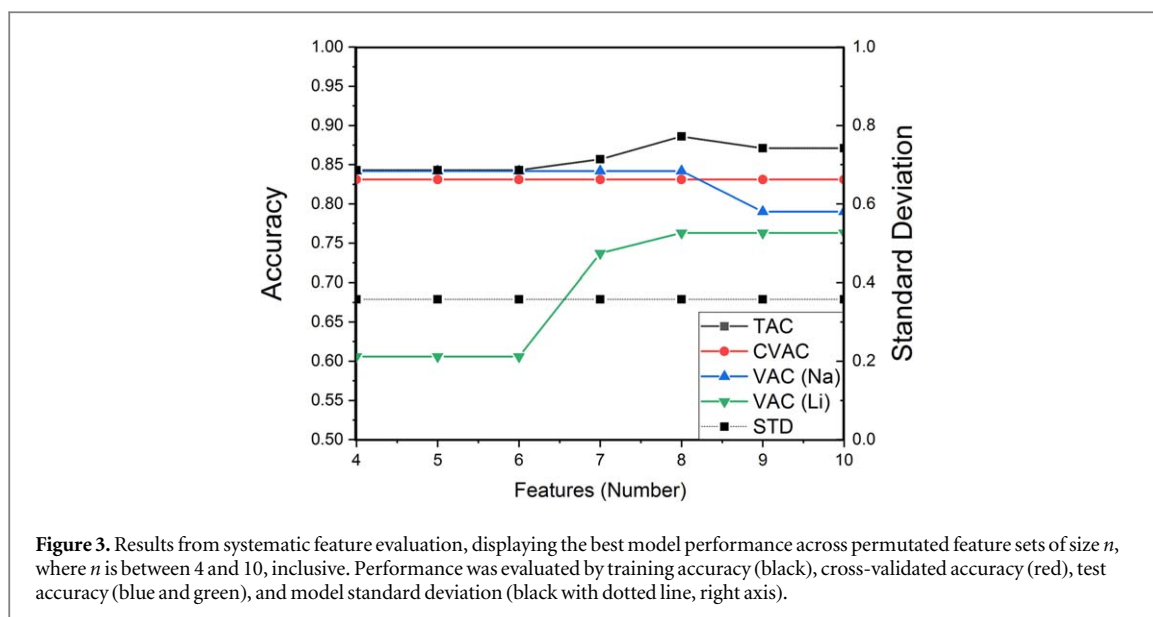
The pipeline for our feature evaluation process is shown below in figure 2, with pseudo-code detailing the process in section S1d in the supplementary.

To identify a model with the best-performing feature combination a systematic evaluation of the features was performed as for each n -sized feature set (where n is between 4 and 10, inclusive): a logistic regression model was constructed using all possible permuted combinations of features for that fixed feature set size. The number of features used to fit the data was chosen in accordance to literature evidence of the use of keeping the minimum 5:1 ratio between data points and the number of features to avoid overfitting and promote good generalization capability [50]. Although it is considered standard practice to split a dataset into training and validation subsets at a ratio of roughly 80:20, the small size of our dataset necessitated the use of fold-based cross-validation approaches.

For a specific feature set size, the training dataset was fitted to their respective ionic conductivity class labels utilizing the permuted subset of features and the resulting training accuracy (TAC), or the difference between the predicted and true outputs, was calculated. In the context of machine learning, a model possessing an accuracy of 0.7 would mean that given 100 examples, an average of 70 would be correctly classified. This is given by equation (2):

$$TAC = \frac{1}{M} \sum_{i=1}^M (\tilde{\sigma}^i = \hat{\sigma}^i) \quad (2)$$

Where $\tilde{\sigma}^i$ and $\hat{\sigma}^i$ refer to the predicted and true class of a training example from a M -sized dataset. Values of 1 and 0 were added into the sum for each correct and incorrect prediction, respectively, before the average was taken to obtain the accuracy value. From the TAC results of all models at the specified feature set size, models



with the TAC of more than 0.8 were selected for cross-validation evaluation. Due to the small size of our dataset, the cross-validated accuracy (CVAC) was computed through K-fold stratified cross-validation (SCV), with a K value of 7. Briefly, 10 stratified datapoints were systematically removed as a validation set, with the model retrained on the remaining N-10 data points, following which the accuracy was measured on the validation subset. This was then repeated for the length of the entire dataset, with an overall mean accuracy and standard deviations recorded. This can be described by equation (3):

$$CVAC = \frac{1}{M} \sum_{i=1}^M (\tilde{\sigma}^i = \hat{\sigma}_{cv}^i) \quad (3)$$

Where $\tilde{\sigma}^i$ and $\hat{\sigma}_{cv}^i$ refer to the predicted and true class of a cross-validation example from a M-sized dataset. Values of 1 and 0 were added into the sum for each correct and incorrect prediction, respectively, before the average was taken to obtain the accuracy value. The CVAC was used as a measure of predictive capacity of each model at the specified feature set size, giving a good summary of the level of under- or overfitting observed with each feature combination. Generally, models that exhibit high training accuracy together with lower cross-validation accuracy suggest that overfitting has taken place, as the model fails to generalize or extrapolate outside its training dataset. In contrast, models with low training accuracy together with low cross-validation accuracy suggest the presence of underfitting, or that the model lacks sufficient data or suitable features for prediction.

Following validation analysis, models of each feature set size with high CVAC scores and low standard deviations were evaluated using an unseen test set, consisting of 19 NASICON compounds from literature with documented room-temperature ionic conductivities and an additional 38 LISICON compounds added as an evaluation of the model's extrapolation capabilities [51–63]. These lithium compounds were added due to the similar behavior of the two systems in electrochemical storage applications and as a test of the generalization capabilities of our model beyond its original training domain. The test accuracy or validation misclassification rate (VAC) was evaluated in the same manner as the TAC mentioned above. Since each set of cross-validation predictions was made on an experimental sample that was removed from the training set, the CVAC together with the VAC captures the potential predictive power of the model on unseen data and provides a way to evaluate its usefulness to researchers in the field.

Results and discussion

Model performance

The results of our evaluation process are displayed in figure 3 and table 2, respectively. Figure 3 displays the accuracies and standard deviation metrics of the best feature combination at a specified feature set size n , while table 2 lists the exact feature combination that gives rise to the performance metrics specified in the figure 2.

Overall, a maximum training accuracy of 0.886 was achieved with a feature size of 8. The highest CVAC accuracy and VAC accuracy values of 0.831, 0.842 (VAC-Na), and 0.763 (VAC-Li) was also achieved by the model with a feature size of 8. Past this feature number, test accuracy (VAC-Na) drops to 0.790, which was attributed to the noise introduced by the additional features. This high similarity in accuracy levels between the majority of feature sizes can be explained by the small size of the validation and test sets, and it is expected that

Table 2. Output accuracies of models trained with permuted feature combinations across different feature set sizes. Training and cross-validation accuracies are based on the training dataset, and validation accuracy is based on the test dataset. The bolded entry represents the final selected optimal model. Here, the symbols D1, D2, and D3 refer to the elements occupying the Na, M1, and M2 positions in NaM₁M₂P₃O₁₂. Detailed descriptions of the individual feature labels can be found in table S1(b).

| Feature set size | Feature combination | TAC | CVAC | STD | VAC (Na) | VAC (Li) |
|------------------|---|--------------|--------------|--------------|--------------|--------------|
| 10 | 'a', 'c', 'h', 'd', 'D1ionic', 'D2ionic', 'PO4', 'D1eneg', 'D2eneg', 'D3eneg' | 0.871 | 0.831 | 0.358 | 0.790 | 0.763 |
| 9 | 'a', 'c', 'h', 'd', 'D2ionic', 'PO4', 'D1eneg', 'D2eneg', 'D3eneg' | 0.871 | 0.831 | 0.358 | 0.790 | 0.763 |
| 8 | 'c', 'h', 'd', 'D1ionic', 'D2ionic', 'PO4', 'D2eneg', 'D3eneg' | 0.886 | 0.831 | 0.358 | 0.842 | 0.763 |
| 7 | 'c', 'h', 'd', 'D2ionic', 'PO4', 'D2eneg', 'D3eneg' | 0.857 | 0.831 | 0.358 | 0.842 | 0.737 |
| 6 | 'a', 'a/c', 'D1ionic', 'PO4', 'D2eneg', 'D3eneg' | 0.843 | 0.831 | 0.358 | 0.842 | 0.606 |
| 5 | 'a', 'D1ionic', 'PO4', 'D2eneg', 'D3eneg' | 0.843 | 0.831 | 0.358 | 0.842 | 0.606 |
| 4 | 'a', 'PO4', 'D2eneg', 'D3eneg' | 0.843 | 0.831 | 0.358 | 0.842 | 0.606 |

large sets would show greater discrepancies as a greater number of examples featuring a wider possible range of features would be evaluated [64].

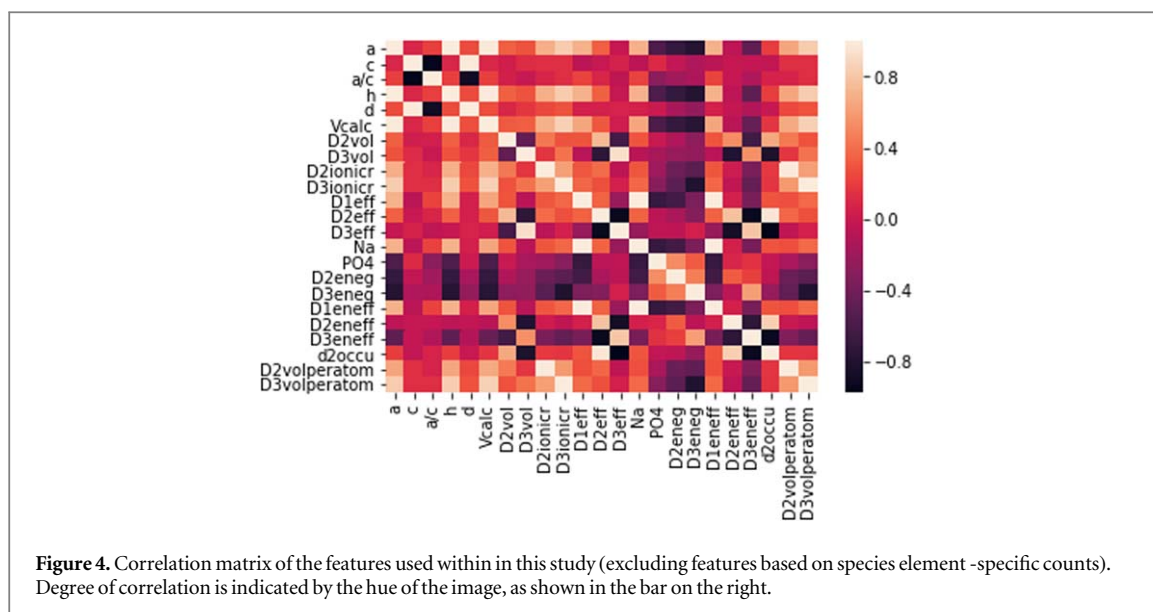
The positive correlation between training accuracy and feature size can be explained by the incorporation of more features such as species electronegativity, ionic radii and structural features such as lattice parameters, unit cell hypotenuses, which are able to capture information universally shared across all compounds. When the feature size was reduced to below 7, the need to maximize training accuracy resulted in strong overfitting to the training dataset being observed using possibly underlying overrepresented features, as evidenced by the poor performance of the model on the lithium test set. Interestingly, if the number of features used exceeded 8 ($n > 8$), the training accuracy of the models decreased by 1.7%, which may be due to the incorporation of the sodium-specific features (namely the ionic radius and electronegativity) that are in conflict with existing feature sets, as the immutable nature of these features could cause them to behave as a constant bias value, shifting the existing fit in the y -axis.

Following the cross-validation process, the extrapolative performance of our models was evaluated using the test set mentioned in the previous section. From our results, we notice that the same validation accuracy is shared among most of our models, which can be explained by the relatively small test dataset. The test accuracies of our sodium and lithium test datasets is similar, at 84.2 % and 76.3 % each, demonstrating the extrapolation capabilities of our model when presented with previously unseen data of a similar domain. The similarity of lithium and sodium-ion conduction has been investigated extensively in literature for energy storage purposes, supporting the outputs of our model [65]. In literature, logistic regression models utilizing more complex computationally-derived features for lithium-ion based solid electrolyte screening have been investigated by Sendek *et al* While the study achieved a higher cross-validation accuracy 90%, the feature complexity of the model demanded extensive computational effort, reducing its effectiveness in rapid screening. The capability of our model to capture much of the same performance using easily derivable elemental information with minimal data pre-processing validates its utility as a preliminary screening tool.

Discussion

As many of our models give the similar levels of validation and sodium test accuracies past $n = 4$, the selection criteria for the final optimal model feature were high accuracy, minimal overfitting on existing training data, and maximal generalization capability. The relationship between the former two factors can be controlled by comparing CVAC with TAC, where a difference between training and validation accuracies would suggest a high level of overfitting. In our study, this difference was arbitrarily set at below 5%. Similarly, the generalization capability was judged by selecting models featuring features with a minimal amount of intra-category cross-correlation. Cross-correlation was measured through a cross-correlation matrix plotted on the features identified in table S1 using the Seaborn statistical package available in Python.

The results from figure 4 demonstrate some explicit intra-category correlations between our features. For example, structural parameters tend to be positively correlated with each other and the overall volume of the unit cell. In particular, features 'a', 'h', and 'Vcalc' were highly cross-correlated, which is understandable given their interdependence during calculations. Similarly, features specific to the cationic structural of the compound, such as the specific electronegative features 'D3eneg' and 'D2eneg', also exhibit strong correlation due to a proportion of compounds exhibiting the same elements for M2 and M3 positions in NaM₁M₂P₃O₁₂. Furthermore, some of the observed correlations are significant in implication to warrant further study. For example, a high unit cell volume is shown to be correlated with a high number of sodium ions, lower count of



phosphate groups and fewer electronegative structural cationic species, suggesting that bond ionicity is partially related to the volume of a unit cell.

As such, the selection of the final optimal model (bolded under $n = 8$ in table 1) was also based on a small VAC-TAC accuracy difference of 4.4%, high test accuracy, a minimal the number of highly cross-correlated features, and a minimal number (1) of highly Pearson-correlated features. The coefficients for the selected features were then extracted for analysis.

Thus, the final Logistic Regression model can thus be described with equation (4),

$$\sum_{i=0}^n \theta_i X_i = 0.017c + 0.519h + 0.195d + 0.076I_{\text{onic radius}}_{D1} - 0.072I_{\text{onic radius}}_{D2} - 0.928N_{\text{PO}_4} - 0.726E_{\text{lectronegativity}}_{D2} - 0.057E_{\text{lectronegativity}}_{D3} \quad (4)$$

The nature of the coefficients in our equation gives some intuition behind the underlying elemental attributes that affect ionic conductivity and allows us to examine literature theories. More importantly, it facilitates the quantification of the relative importance of each attribute with regards to ionic conductivity through analysis of the magnitude of the coefficients, opening the way for future optimization studies. Equation (4) suggests an increase in ionic conductivity with increasing number of structural parameters—in fact, an increase in the number of structural unit cell features was observed to increase the training accuracy of the model when $n > 7$. As all our examples share the R3c rhombohedral spacegroup, it can be argued these lattice parameters act as a representation of the cell volume as suggested by our cross-correlation analysis. It must be mentioned that the lack of the unit cell volume feature itself in the final models suggests that structural parameters are capturing some information beyond the spacegroup’s calculated volume.

In literature, the role of free unit cell volume in gauging ionic conductivity was first proposed by Sammels *et al* who theorized that a large difference between unit cell volume and the ionic volume of its constituents would lead to enhanced ionic conductivity, and concluded that a cubic or pseudo-cubic unit cell structure was preferable for such applications [66]. Experimentally, a positive correlation between unit cell volume and ionic conductivity has been observed for superionic selenide, sulfide, and oxide compounds by Ong *et al*. Interestingly, such a correlation was observed until an optimum size of an ionic channel is reached, after which further volume increases have negligible effect, although this is not seen in our analysis as feature information on the size of ionic channels was not incorporated [67]. These conclusions are supported by the positive coefficients of our structural parameters in equation (4), which given the positive correlation of structural parameters with unit cell volume, support the view of high unit cell volume being linked to higher ionic conductivity. It must be noted however, that the unit cell volume descriptor is not part of the optimal model’s feature sets, suggesting that the underlying relationships may be more complex.

Linked to the unit cell volume and lattice parameters is the ionic radius of any introduced dopant. In equation (4), both the ionic radii of the mobile ion and structural metal components are present and shown to be positively and negatively correlated to the output ionic conductivity, respectively. Furthermore, our cross-

correlational analysis results suggest that both component ionic radii are negatively correlated to higher overall unit cell volume. Recent experimental evidence in spinel ionic compounds has suggested that the need to maximize unit cell free volume must be balanced by a minimization in strain in the host lattice [68]. Hayashi *et al* observed that the ionic radius of any such dopant must be similar to that of the host lattice to minimize structural strain, where the similarity can be judged with a computational tolerance factor [69]. The work of Kim *et al* and Kilner on oxygen ionic conductivity in oxide-fluorites further suggested that a critical ionic radius exists for ionic conductors that results in a lack of both expansion or contraction movements in the host lattice [70, 71]. Experimental evidence has concluded that increased dopant cation size exhibits a negative correlation with ionic conductivity, which has been explained by the steric and electronic interactions of the larger dopant cations and anionic vacancies hindering movements of the latter within the lattice [41, 72]. These results are in agreement with our resulting equation, where increasing ionic radius of the structural metal cations is negatively correlated with predicted ionic conductivity.

The number of phosphate groups has the highest negative effect on the overall ionic conductivity, which is understandable given its high Pearson correlation factor (-0.772). This correlation has been previously described in literature, and was thought to originate from the possibility of negatively charged anionic groups disturbing the pathways of conducting cations and hence reducing the rate of ionic conduction [73]. Concomitant studies by Sendek *et al* arrived at a similar conclusion when modelling ionic electrolyte conductivities from computed electronic features, describing the ideal conduction pathway to be mostly covalent in character, where a cation can rapidly form and break bonds along the axis of travel with minimal energy penalty [74]. Recent studies by Ong *et al* and Reed *et al* have further linked the high electronegativity of the anionic components in solid ionic conductors with poor ionic conductivity, which was attributed to the preference for strong ionic bonding [75, 76]. Interestingly, it has been demonstrated that the strong ionic bonding observed by the anionic component leads to increased electrochemical stability, with the discrepancy best exemplified by the difference between the ionic conductivities of sulfide-based and oxide-based Li-ion conductors [77, 78].

Finally, our model also suggests that compounds featuring structural cations with higher base electronegativity will tend to reduce the overall ionic conductivity. While a high electronegativity in the mobile cation has been shown to reduce ionic conductivity in SICON-type solids, the effect of immobile structural cation electronegativity has not been systematically explored [79]. As mentioned above, ideal conduction pathways should be covalent in character, which would naturally filter out the use of highly electronegative dopants to improve conductivity. A large difference between the electronegativities between different components will lead to a preference for ionic bonding, reducing ionic conductivity. Interestingly, a study on LISICON-type lithium titanium phosphate (LTP) by Wang *et al* suggested that the substitution of the less electronegative Ti(III) with the more electronegative Ti(IV) would increase ionic conductivity, which was theorized to originate from the increased covalent character in the formed Li–O–Ti bond. However, as the study focused on altering the amount of lithium within the structure, and featured no electronic characterization, it remains speculative in nature [80].

While our classification model, by definition, is not capable of inverse design, analysis of our training examples together with model coefficients allows for some predictive capability for likely superionic candidates. We can summarize from the analysis above that a NASICON-type compound of the R3c space group is likely to exhibit an ionic conductivity value of greater than $1 \times 10^{-6} \text{ S cm}^{-1}$ it meets the following qualitative conditions:

- The number of phosphate groups is smaller than 3.
- The electronegativity of the M1 and M2 cations is low.
- The structural parameters of the compound are large.
- The ionic radius of the M2 cation is low.

Analysis of the dataset material candidates suggests that NASICON or LISICON-type structures of the general formula $\text{Na}_{1+x}\text{M1M2Si}_x(\text{P}_{1-x/3}\text{O}_4)_3$ and $\text{Li}_{1+x}\text{M1M2Si}_x(\text{P}_{1-x/3}\text{O}_4)_3$, where M1 and M2 represent transition metal species with suitably low III/IV oxidation state ionic radii and electronegativity (such as chromium, iron, aluminum, or titanium) meet the conditions specified above and would make of interesting candidates to pursue for further research as highly conductive solid-state electrolyte candidates.

Limitations

Despite high training, validation, and test set accuracies, our model is not free from capability limitations. Firstly, the training dataset is small, and the low number of highly conductive examples may influence the classification accuracy of compounds of that class, although this cannot be conclusively determined without comparing the performance of our model to one trained with additional data on highly conductive NASICON compounds.

Secondly, our model's features do not capture any bonding-related or defect-related interactions or effects, relying instead on counterparts derived from isolated constituent atoms. While the structural parameters are related to the number and type of species occupying a unit cell, the features used by the best performing models with high feature count ($n > 6$) do not imply any direct relationship between the formulaic number of a species in a unit cell with ionic conductivity. A positive relationship between having multiple structural cationic species and ionic conductivity has been previously documented in solid oxide electrolytes, which our current model fails to capture [81]. As a result, drastic changes in ionic conductivity as a result of increasing levels of dopants may fail to be correctly modelled. It is important to note that the underlying mechanics and electronic interactions governing ionic conduction are complex and a regression of features obtained from chemical intuition and domain expertise may prove to be unsatisfactory to capture this complexity. These can be captured by descriptors based on DFT-calculations or lattice dynamics such as the average vibrational frequency of a mobile cationic sublattice [74, 82]. However, we aimed to develop a model that would capture some information about the relative importance of different elemental and structural features derivable from structural information on the unit cell and compound constituents to the overall conductivity, while demonstrating the building of screening algorithms using constrained datasets.

In addition, while the formulaic definitions of our materials may be known, the experimental details behind each conductivity measurement are known to be inconsistent. Although ionic conductivity of solid compounds is usually measured through pelletized materials, studies have shown that the density of the generated pellets possess an effect on measured conductivities, with highly-dense sintered pellets exhibiting significantly higher conductivities than their unsintered equivalents [41]. Any systematic data collection initiative must hence maintain a consistent measurement pipeline for maximum accuracy.

Moreover, the lack of more complex features stemming from a variety of materials synthesis pathways that can affect ionic conductivity, such as the polarizability, grain size distribution, grain defect distributions, local mobile ion distributions, lattice stress, and structural restrictions, are not modelled, and it is believed that the adoption of such features would further increase the precision of our model, at the cost of more complex feature calculations [83, 84]. The correlations behind the features identified by our model may also not be causal, and may instead be capturing the importance of some other feature aspect not considered, as supported by the poor individual Pearson correlations to the ionic conductivity of individual features. Guin *et al*'s previous work on statistically modelling NASICON ionic conductivity suggested that ionic radius of the dopants play a critical role; such a relationship was not directly observed here, but may have been captured by the prominence of the dopant electronegativity, as the two has been shown to be linked through the effective ionic potential, which is defined by both the ionic radius and the ionization energies of an element [5, 85]. As previously mentioned, a study Sendek *et al* on lithium-based superionic conductors demonstrated a higher cross validation accuracy when utilizing more complex structural and electronic features, however the requirement of such features would decrease the value of such a tool for experimentalists [74]

Conclusions

We have developed a simple, cost-effective screening algorithm for ionic conductivity for NASICON compounds using widely available, computationally simple elemental features. By using the limited data available in literature, we systematically evaluated feature combinations to identify the model with the highest confidence in predicting ionic conductivity in both Li- and Na- based SICON-type compounds, with a validation accuracy of over 84 and 76%, respectively, using only features derived from the formulaic unit in a facile process. Through this process, we reduced the number of features considered from 47 to 8, a reduction of approximately 83%, while simultaneously increasing the performance of our models. Our models also provide intuition behind the chemistry behind fast ionic conductors. With more available training data, increased confidence and more intuitive conclusions could be achieved in the future, while the introduction of more specialized features such as anionic electronegativity could lead improved cross-domain classification beyond R3c-type conductors.

With the increase in interest in solid ionic electrolytes, the use of efficient high-performance algorithms can drastically cut down on the amount of wasted labour and reduce the time-to-market of research outputs. Our

model allows a researcher to evaluate candidates obtained from elemental permutations or chemical intuition for likely superionic performance, without the need for expensive laboratory or computational experimental work. To better investigate the effects of structural composition and atomic arrangements on ionic conductivity, the development of classification models utilizing DFT-based features and structural features based on simulated XRD data should be pursued, for more accurate electronic and geometric representations, respectively. In the absence of the most suitable features, the introduction of a diverse and larger set of training examples would undoubtedly improve the robustness of our model. We recommend the experimental community to aggressively increase the size of the training data available in literature, with focus on the inclusion of more highly conductive examples in primary or supplementary data, as these will serve to improve classifier robustness, as well as to provide further data to evaluate the performance of our model feature selections.

Acknowledgments

The Engineering and Physical Sciences Research Council are thanked for funding the Centre for Doctoral Training in Molecular Modelling & Materials Science (UCL, UK; EPSRC reference EP/L015862/1) and A*Star (Singapore) are thanked for supporting a studentship for YX. KH acknowledges funding from the Accelerated Materials Development for Manufacturing Program at A*STAR via the AME Programmatic Fund by the Agency for Science, Technology and Research under Grant No. A1898b004.

Code availability

The code and datasets used in this study can be found at https://github.com/EXJUSTICE/LISICON_NASICON_IonicConductivity_Classification

ORCID iDs

Kedar Hippalgaonkar  <https://orcid.org/0000-0002-1270-9047>

References

- [1] Goodenough J B and Park K-S 2013 The Li-ion rechargeable battery: a perspective *JACS* **135** 1167–76
- [2] Kim S W, Seo D H, Ma X, Ceder G and Kang K 2012 Electrode materials for rechargeable sodium-ion batteries: potential alternatives to current lithium-ion batteries *Adv. Energy Mater.* **2** 710–21
- [3] Arbizzani C, Gabrielli G and Mastragostino M 2011 Thermal stability and flammability of electrolytes for lithium-ion batteries *J. Power Sources* **196** 4801–5
- [4] Aravindan V, Gnanaraj J, Madhavi S and Liu H K 2011 Lithium-ion conducting electrolyte salts for lithium batteries *Chemistry–A European Journal* **17** 14326–46
- [5] Guin M and Tietz F 2015 Survey of the transport properties of sodium superionic conductor materials for use in sodium batteries *J. Power Sources* **273** 1056–64
- [6] Lalère F, Leriche J-B, Courty M, Boulineau S, Viallet V, Masquelier C and Seznec V 2014 An all-solid state NASICON sodium battery operating at 200 C *J. Power Sources* **247** 975–80
- [7] Bodner G M and Herron J D 2002 Problem-solving in chemistry *Chemical Education: Towards Research-based Practice* (Berlin: Springer) pp 235–66
- [8] Xiang X-D, Sun X, Briceno G, Lou Y, Wang K-A, Chang H, Wallace-Freedman W G, Chen S-W and Schultz P G 1995 A combinatorial approach to materials discovery *Science* **268** 1738–40
- [9] Hautier G, Fischer C C, Jain A, Mueller T and Ceder G 2010 Finding nature's missing ternary oxide compounds using machine learning and density functional theory *Chem. Mater.* **22** 3762–7
- [10] Balachandran P V, Broderick S R and Rajan K 2011 Identifying the 'inorganic gene' for high-temperature piezoelectric perovskites through statistical learning *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **467** 2271–90
- [11] Meredig B, Agrawal A, Kirklin S, Saal J E, Doak J, Thompson A, Zhang K, Choudhary A and Wolverton C 2014 Combinatorial screening for new materials in unconstrained composition space with machine learning *Phys. Rev. B* **89** 094104
- [12] Ma X, Li Z, Achenie L E and Xin H 2015 Machine-learning-augmented chemisorption model for CO₂ electroreduction catalyst screening *The Journal of Physical Chemistry Letters* **6** 3528–33
- [13] Raccuglia P, Elbert K C, Adler P D, Falk C, Wenny M B, Mollo A, Zeller M, Friedler S A, Schrier J and Norquist A J 2016 Machine-learning-assisted materials discovery using failed experiments *Nature* **533** 73
- [14] Schütter C, Husch T, Viswanathan V, Passerini S, Balducci A and Korth M 2016 Rational design of new electrolyte materials for electrochemical double layer capacitors *J. Power Sources* **326** 541–8
- [15] Wei J N, Duvenaud D and Aspuru-Guzik A 2016 Neural networks for the prediction of organic chemistry reactions *ACS Central Science* **2** 725–32
- [16] Hong H-P 1976 Crystal structures and crystal chemistry in the system Na_{1+x}Zr₂Si_xP_{3-x}O₁₂ *Mater. Res. Bull.* **11** 173–82
- [17] Winand J-M, Rulmont A and Tarte P 1991 Nouvelles solutions solides Li(M^{IV})_{2-x}(N^{IV})_x(PO₄)₃ (L = Li, Na, M, N = Ge, Sn, Ti, Zr, Hf) synthèse et étude par diffraction x et conductivité ionique *J. Solid State Chem.* **93** 341–9
- [18] Aono H and Sugimoto E 1996 Ionic conductivity and sinterability of NASICON-type ceramics: the systems NaM₂(PO₄)₃ + yNa₂O (M = Ge, Ti, Hf, and Zr) *J. Am. Ceram. Soc.* **79** 2786–8

- [19] Bennouna L, Arsalane S, Brochu R, Lee M, Chassaing J and Quarton M 1995 Spécificités des ions Nb^{IV} et Mo^{IV} dans les monophosphates de type Nasicon *J. Solid State Chem.* **114** 224–9
- [20] Goodenough J B, Hong H-P and Kafalas J 1976 Fast Na^+ -ion transport in skeleton structures *Mater. Res. Bull.* **11** 203–20
- [21] Vogel E, Cava R J and Rietman E 1984 Na^+ ion conductivity and crystallographic cell characterization in the Hf -nasicon system $\text{Na}_{1+x}\text{Hf}_2\text{Si}_x\text{P}_{3-x}\text{O}_{12}$ *Solid State Ionics* **14** 1–6
- [22] Ivanov D, Currie J, Bouchard H, Lecours A, Andrian J, Yelon A and Poulin S 1994 Sputtered silicate-limit NASICON thin films for electrochemical sensors *Solid State Ionics* **67** 295–9
- [23] Wang W, Zhang Z, Xiangyang O and Zhao J 1988 Properties and phase relationship of the $\text{Na}_{1+x}\text{Hf}_{2-y}\text{Ti}_y\text{Si}_x\text{P}_{3-x}\text{O}_{12}$ system *Solid State Ionics* **28** 442–5
- [24] Losilla E R, Aranda M A, Bruque S, Paris M A, Sanz J and West A R 1998 Understanding Na mobility in NASICON materials: a rietveld, ^{23}Na and ^{31}P MAS NMR, and impedance study *Chem. Mater.* **10** 665–73
- [25] Losilla E R, Aranda M A, Bruque S, Sanz J, Paris M A, Campo J and West A R 2000 Sodium mobility in the NASICON series $\text{Na}_{1+x}\text{Zr}_{2-x}\text{In}_x(\text{PO}_4)_3$ *Chem. Mater.* **12** 2134–42
- [26] Winand J, Rulmont A and Tarte P 1990 Ionic conductivity of the $\text{Na}_{1+x}\text{M}_x^{\text{III}}\text{Zr}_{2-x}(\text{PO}_4)_3$ systems ($\text{M} = \text{Al, Ga, Cr, Fe, Sc, In, Y, Yb}$) *J. Mater. Sci.* **25** 4008–13
- [27] Miyajima Y, Saito Y, Matsuoka M and Yamamoto Y 1996 Ionic conductivity of NASICON-type $\text{Na}_{1+x}\text{M}_x\text{Zr}_{2-x}\text{P}_3\text{O}_{12}$ ($\text{M}: \text{Yb, Er, Dy}$) *Solid State Ionics* **84** 61–4
- [28] Miyajima Y, Miyoshi T, Tamaki J, Matsuoka M, Yamamoto Y, Masquelier C, Tabuchi M, Saito Y and Kageyama H 1999 Solubility range and ionic conductivity of large trivalent ion doped $\text{Na}_{1+x}\text{M}_x\text{Zr}_{2-x}\text{P}_3\text{O}_{12}$ ($\text{M}: \text{In, Yb, Er, Y, Dy, Tb, Gd}$) solid electrolytes *Solid State Ionics* **124** 201–11
- [29] d'Yvoire F, Pintard-Scrépel M, Bretey E and De la Rochere M 1983 Phase transitions and ionic conduction in 3D skeleton phosphates $\text{A}_3\text{M}_2(\text{PO}_4)_3$: $\text{A} = \text{Li, Na, Ag, K}$; $\text{M} = \text{Cr, Fe}$ *Solid State Ionics* **9** 851–7
- [30] Maldonado-Manso P, Aranda M A, Bruque S, Sanz J and Losilla E R 2005 Nominal versus actual stoichiometries in Al-doped NASICONs: a study of the $\text{Na}_{1.4}\text{Al}_{0.4}\text{M}_{1.6}(\text{PO}_4)_3$ ($\text{M} = \text{Ge, Sn, Ti, Hf, Zr}$) family *Solid State Ionics* **176** 1613–25
- [31] Mouahid F, Bettach M, Zahir M, Maldonado-Manso P, Bruque S, Losilla E and Aranda M 2000 Crystal chemistry and ion conductivity of the $\text{Na}_{1+x}\text{Ti}_{2-x}\text{Al}_x(\text{PO}_4)_3$ ($0 \leq x \leq 0.9$) NASICON series *J. Mater. Chem.* **10** 2748–57
- [32] Takahashi T, Kuwabara K and Shibata M 1980 Solid-state ionic-conductivities of Na^+ ion conductors based on NASICON *Solid State Ionics* **1** 163–75
- [33] Gandi S, Chinta S R, Ojha P K, Surendra Babu M S and Ravuri B R 2018 High Na-ion conducting $\text{Na}_{1+x}[\text{Sn}_x\text{Ge}_{2-x}(\text{PO}_4)_3]$ glass-ceramic electrolytes: structural and electrochemical impedance studies *J. Am. Ceram. Soc.* **101** 167–77
- [34] Kazakevičius E, Kežionis A, Žukauskaitė L, Barré M, Šalkus T, Žalga A, Selskis A and Orliukas A 2015 Characterization of $\text{Na}_{1.3}\text{Al}_{0.3}\text{Zr}_{1.7}(\text{PO}_4)_3$ solid electrolyte ceramics by impedance spectroscopy *Solid State Ionics* **271** 128–33
- [35] Verissimo C, Garrido F M, Alves O L, Calle P, Martínez-Juárez A, Iglesias J E and Rojo J M 1997 Ionic conductivity and structural characterization of $\text{Na}_{1.5}\text{Nb}_{0.3}\text{Zr}_{1.5}(\text{PO}_4)_3$ with NASICON-type structure *Solid State Ionics* **100** 127–34
- [36] Cherkaoui F, Viala J, Delmas C and Hagenmuller P 1986 Crystal chemistry and ionic conductivity of a new Nasicon-related solid solution $\text{Na}_{1+x}\text{Zr}_{2-x}\text{Mg}_x(\text{PO}_4)_3$ *Mater. Res. Bull.* **21** 333–7
- [37] Delmas C, Cherkaoui F and Hagenmuller P 1986 Ionic conductivity in a new NASICON related solid solution: $\text{Na}_{3+y}\text{Cr}_{2-y}\text{Mg}_y(\text{PO}_4)_3$. An optical characterization of the skeleton covalency *Mater. Res. Bull.* **21** 469–77
- [38] Delmas C, Viala J-C, Olazcuaga R, Le Flem G, Hagenmuller P, Cherkaoui F and Brochu R 1981 Ionic conductivity in Nasicon-type phases $\text{Na}_{1+x}\text{Zr}_{2-x}\text{L}_x(\text{PO}_4)_3$ ($\text{L} = \text{Cr, In, Yb}$) *Solid State Ionics* **3** 209–14
- [39] Aono H, bin Idris M A and Sadaoka Y 2004 Ionic conductivity and crystal structure for the $\text{Li}_{3-2x}\text{Cr}_{2-x}\text{Ta}_x(\text{PO}_4)_3$ system *Solid State Ionics* **166** 53–9
- [40] Aono H, Sugimoto E, Sadaoka Y, Imanaka N and Adachi G-Y 1993 Electrical properties and crystal structure of solid electrolyte based on lithium hafnium phosphate $\text{LiHf}_2(\text{PO}_4)_3$ *Solid State Ionics* **62** 309–16
- [41] Aono H, Sugimoto E, Sadaoka Y, Imanaka N and Adachi G-Y 1990 Ionic conductivity and sinterability of lithium titanium phosphate system *Solid State Ionics* **40** 38–42
- [42] Safanama D and Adams S 2017 High efficiency aqueous and hybrid lithium-air batteries enabled by $\text{Li}_{1.5}\text{Al}_{0.5}\text{Ge}_{1.5}(\text{PO}_4)_3$ ceramic anode-protecting membranes *J. Power Sources* **340** 294–301
- [43] Weiss M, Weber D A, Senyshyn A, Janek J R and Zeier W G 2018 Correlating transport and structural properties in $\text{Li}_{1+x}\text{Al}_x\text{Ge}_{2-x}(\text{PO}_4)_3$ (LAGP) prepared from aqueous solution *ACS Applied Materials Interfaces* **10** 10935–44
- [44] Rettenwander D, Welzl A, Pristat S, Tietz F, Taibl S, Redhammer G and Fleig J 2016 A microcontact impedance study on NASICON-type $\text{Li}_{1+x}\text{Al}_x\text{Ti}_{2-x}(\text{PO}_4)_3$ ($0 \leq x \leq 0.5$) single crystals *J. Mater. Chem. A* **4** 1506–13
- [45] Svitan'Ko A, Novikova S, Safronov D and Yaroslavtsev A 2011 Cation mobility in $\text{Li}_{1+x}\text{Ti}_{2-x}\text{Cr}_x(\text{PO}_4)_3$ NASICON-type phosphates *Inorg. Mater.* **47** 1391–5
- [46] Li Y, Liu M, Liu K and Wang C-A 2013 High Li^+ conduction in NASICON-type $\text{Li}_{1+x}\text{Y}_x\text{Zr}_{2-x}(\text{PO}_4)_3$ at room temperature *J. Power Sources* **240** 50–3
- [47] Yue Y, Zhou F and Pang W J M C 1993 Hydrothermal crystallization and structural investigation of $\text{Na}_{1+2x}\text{Zr}_{2-x}\text{Mg}_x(\text{PO}_4)_3$ systems ($0 < x < 1$) *Physics (N.Y.)* **35** 28–30
- [48] Hellenbrandt M 2004 The inorganic crystal structure database (ICSD)—present and future *Crystallogr. Rev.* **10** 17–22
- [49] Mazurowski M A, Habas P A, Zurada J M, Lo J Y, Baker J A and Tourassi G D 2008 Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance *Neural Netw.* **21** 427–36
- [50] Roy K, Kar S and Das R N 2015 *A Primer on QSAR/QSPR Modeling: Fundamental Concepts* (Berlin: Springer)
- [51] Subramanian M, Rudolf P and Clearfield A 1985 The preparation, structure, and conductivity of scandium-substituted NASICONs *J. Solid State Chem.* **60** 172–81
- [52] Feltz A and Barth S 1983 Preparation and conductivity behaviour of $\text{Na}_3\text{M}^{\text{II}}\text{Zr}(\text{PO}_4)_3$, ($\text{M}: \text{Mn, Mg, Zn}$) *Solid State Ionics* **9** 817–21
- [53] Novikova S, Larkovich R, Chekannikov A, Kulova T, Skundin A and Yaroslavtsev A 2018 Electrical conductivity and electrochemical characteristics of $\text{Na}_3\text{V}_2(\text{PO}_4)_3$ -based NASICON-type materials *Inorg. Mater.* **54** 794–804
- [54] Shrivastava O, Kumar N and Chourasia R 2007 Synthesis, crystallographic characterization and ionic conductivity of iron substituted sodium zirconium phosphate $\text{Na}_{1.2}\text{Zr}_{1.8}\text{Fe}_{0.2}(\text{PO}_4)_3$ *J. Mater. Sci.* **42** 2551–6
- [55] Kaus M, Guin M, Yavuz M, Knapp M, Tietz F, Guillon O, Ehrenberg H and Indris S 2017 Fast Na^+ ion conduction in NASICON-type $\text{Na}_{3.4}\text{Sc}_2(\text{SiO}_4)_0.4(\text{PO}_4)_2.6$ Observed by ^{23}Na NMR relaxometry *The Journal of Physical Chemistry C* **121** 1449–54
- [56] Brunet F, Bagdassarov N and Miletich R 2003 $\text{Na}_3\text{Al}_2(\text{PO}_4)_3$, a fast sodium conductor at high pressure: *in situ* impedance spectroscopy characterisation and phase diagram up to 8 GPa *Solid State Ionics* **159** 35–47

- [57] Kosova N, Devyatkina E, Stepanov A and Buzlukov A 2008 Lithium conductivity and lithium diffusion in NASICON-type $\text{Li}_{1+x}\text{Ti}_{2-x}\text{Al}_x(\text{PO}_4)_3$ ($x = 0; 0.3$) prepared by mechanical activation *Ionics* **14** 303–11
- [58] Xie H, Li Y and Goodenough J B 2011 NASICON-type $\text{Li}_{1+2x}\text{Zr}_{2-x}\text{Ca}_x(\text{PO}_4)_3$ with high ionic conductivity at room temperature *RSC Adv.* **1** 1728–31
- [59] Enrique L, Losillaa R and Bruquea S N 1998 NASICON to scandium wolframate transition in $\text{Li}_M\text{Hf}(\text{PO})_3$ ($M = \text{Cr, Fe}$): structure and ionic conductivity *Solid State Ionics* **112** 53–62
- [60] Kumar S and Balaya P 2016 Improved ionic conductivity in NASICON-type Sr^{2+} doped $\text{LiZr}_2(\text{PO}_4)_3$ *Solid State Ionics* **296** 1–6
- [61] Ruffo R, Mari C and Catti M 2001 Structural and electrical characterization of the NASICON-type $\text{Li}_2\text{FeZr}(\text{PO}_4)_3$ and $\text{Li}_2\text{FeTi}(\text{PO}_4)_3$ compounds *Ionics* **7** 105–8
- [62] Safronov D, Stenina I, Maksimychiev A, Shestakov S and Yaroslavtsev A 2009 Phase transitions and ion transport in NASICON materials of composition $\text{Li}_{1+x}\text{Zr}_{2-x}\text{In}_x(\text{PO}_4)_3$ ($x = 0-1$) *Russ. J. Inorg. Chem.* **54** 1697
- [63] Anantharamulu N, Velchuri R, Sarojini T, Madhavi K, Prasad G and Vithal M 2009 Preparation, characterization and conductivity studies of $\text{NaAlSb}(\text{PO}_4)_3$ and $\text{HAlSb}(\text{PO}_4)_3$ *Indian Journal of Engineering and Materials Sciences* **16** (5) 347–54
- [64] Gales M 2008 Introduction to machine learning <https://www.cl.cam.ac.uk/teaching/1011/L101/ml4lp-lect1.pdf>
- [65] Masse R C, Uchaker E and Cao G 2015 Beyond Li-ion: electrode materials for sodium- and magnesium-ion batteries *Sci. China. Mater.* **58** 715–66
- [66] Sammells A F, Cook R L, White J H, Osborne J J and MacDuff R C 1992 Rational selection of advanced solid electrolytes for intermediate temperature fuel cells *Solid State Ionics* **52** 111–23
- [67] Ong S P, Mo Y, Richards W D, Miara L, Lee H S and Ceder G 2013 Phase stability, electrochemical stability and ionic conductivity of the $\text{Li}_{10\pm 1}\text{MP}_2\text{X}_{12}$ ($M = \text{Ge, Si, Sn, Al or P}$, and $X = \text{O, S or Se}$) family of superionic conductors *Energy & Environmental Science* **6** 148–56
- [68] Lybye D, Poulsen F W and Mogensen M 2000 Conductivity of A- and B-site doped LaAlO_3 , LaGaO_3 , LaScO_3 and LaInO_3 perovskites *Solid State Ionics* **128** 91–103
- [69] Hayashi H, Inaba H, Matsuyama M, Lan N, Dokiya M and Tagawa H 1999 Structural consideration on the ionic conductivity of perovskite-type oxides *Solid State Ionics* **122** 1–15
- [70] Kim D J 1989 Lattice parameters, ionic conductivities, and solubility limits in fluorite-structure MO_2 oxide [$M = \text{Hf}^{4+}, \text{Zr}^{4+}, \text{Ce}^{4+}, \text{Th}^{4+}, \text{U}^{4+}$] solid solutions *J. Am. Ceram. Soc.* **72** 1415–21
- [71] Kilner J and Brook R 1982 A study of oxygen ion conductivity in doped non-stoichiometric oxides *Solid State Ionics* **6** 237–52
- [72] Badwal S 1992 Zirconia-based solid electrolytes: microstructure, stability and ionic conductivity *Solid State Ionics* **52** 23–32
- [73] Matsumoto H, Kawasaki Y, Ito N, Enoki M and Ishihara T 2007 Relation between electrical conductivity and chemical stability of BaCeO_3 -based proton conductors with different trivalent dopants *Electrochem. Solid-State Lett.* **10** B77–80
- [74] Sendek A D, Yang Q, Cubuk E D, Duerloo K-A N, Cui Y and Reed E J 2017 Holistic computational structure screening of more than 12000 candidates for solid lithium-ion conductor materials *Energy & Environmental Science* **10** 306–20
- [75] Sendek A D, Cheon G, Pasta M and Reed E J 2019 Quantifying the search for solid Li-ion electrolyte materials by anion: a data-driven perspective *J. Phys. Chem. C* **124** (15) 8067–79
- [76] Deng Z, Wang Z, Chu I-H, Luo J and Ong S P 2016 Elastic properties of alkali superionic conductor electrolytes from first principles calculations *J. Electrochem. Soc.* **163** A67–74
- [77] Jimenez R, Rivera A, Varez A and Sanz J 2009 Li mobility in $\text{Li}_{0.5-x}\text{Na}_x\text{La}_{0.5}\text{TiO}_3$ perovskites ($0 \leq x \leq 0.5$): influence of structural and compositional parameters *Solid State Ionics* **180** 1362–71
- [78] Manthiram A, Yu X and Wang S 2017 Lithium battery chemistries enabled by solid-state electrolytes *Nature Reviews Materials* **2** 16103
- [79] Tamura S, Highchi T and Imanaka N 2010 High Ag^+ ion conduction in NASICON-type solids *Electrochem. Solid-State Lett.* **13** F13–5
- [80] Wang B, Greenblatt M, Wang S and Hwu S 1993 Ionic conductivity of lithium titanium phosphate $\text{Li}_{1+x}\text{Ti}_2(\text{PO}_4)_3$ ($0.2 < x < 1.72$) with NASICON-related structures *Chem. Mater.* **5** 23–6
- [81] Hui S R, Roller J, Yick S, Zhang X, Decès-Petit C, Xie Y, Maric R and Ghosh D 2007 A brief review of the ionic conductivity enhancement for selected oxide electrolytes *J. Power Sources* **172** 493–502
- [82] Muy S, Voss J, Schlem R, Koerver R, Sedlmaier S J, Maglia F, Lamp P, Zeier W G and Shao-Horn Y 2019 High-throughput screening of solid-state Li-Ion conductors using lattice-dynamics descriptors *iScience* **16** 270–82
- [83] Zhang L, Yang K, Mi J, Lu L, Zhao L, Wang L, Li Y and Zeng H 2015 Na_3PSe_4 : a novel chalcogenide solid electrolyte with high ionic conductivity *Adv. Energy Mater.* **5** 1501294
- [84] Omar S, Wachsman E D and Nino J C 2006 A co-doping approach towards enhanced ionic conductivity in fluorite-based electrolytes *Solid State Ionics* **177** 3199–203
- [85] Li K and Xue D 2006 Estimation of electronegativity values of elements in different valence states *J. Phys. Chem. A* **110** 11332–7