scientific reports

OPEN

Discovery of novel Li SSE and anode coatings using interpretable machine learning and high-throughput multi-property screening

Shreyas J. Honrao[1], Xin Yang[2], Balachandran Radhakrishnan[1✉], Shigemasa Kuwata[2], Hideyuki Komatsu[3], Atsushi Ohma[3], Maarten Sierhuis[2] & John W. Lawson[4✉]

All-solid-state batteries with Li metal anode can address the safety issues surrounding traditional Li-ion batteries as well as the demand for higher energy densities. However, the development of solid electrolytes and protective anode coatings possessing high ionic conductivity and good stability with Li metal has proven to be a challenge. Here, we present our informatics approach to explore the Li compound space for promising electrolytes and anode coatings using high-throughput multi-property screening and interpretable machine learning. To do this, we generate a database of battery-related materials properties by computing Li^+ migration barriers and stability windows for over 15,000 Li-containing compounds from Materials Project. We screen through the database for candidates with good thermodynamic and electrochemical stabilities, and low Li^+ migration barriers, identifying promising new candidates such as Li_9S_3N, $LiAlB_2O_5$, $LiYO_2$, $LiSbF_4$, and $Sr_4Li(BN_2)_3$, among others. We train machine learning models, using ensemble methods, to predict migration barriers and oxidation and reduction potentials of these compounds by engineering input features that ensure accuracy and interpretability. Using only a small number of features, our gradient boosting regression models achieve R^2 values of 0.95 and 0.92 on the oxidation and reduction potential prediction tasks, respectively, and 0.86 on the migration barrier prediction task. Finally, we use Shapley additive explanations and permutation feature importance analyses to interpret our machine learning predictions and identify materials properties with the largest impact on predictions in our models. We show that our approach has the potential to enable rapid discovery and design of novel solid electrolytes and anode coatings.

High energy density batteries that are, both, safe from volatile reactions with air and water, and capable of quick charging, are the holy grail of the electric vehicles industry today[1–5]. While advances in controls and engineering of battery modules are paving the way for highly optimized batteries, theoretical limits on battery performance have been largely determined by the underlying choice of materials and electrochemistry[6,7]. Previous developments in battery materials were driven by long cycles of experiments involving several expensive and time-consuming synthesis-and-characterization loops. Recently, high-throughput computations of materials properties, made possible by the exponential growth in computational capabilities of supercomputers and cluster-computing facilities, have resulted in the development of large open-source databases like Materials Project[8], OQMD[9], AFLOWLIB[10], etc. Such databases provide valuable opportunities for many data mining and machine learning (ML) approaches[11–18]. Over the past few years, materials science researchers have implemented ML models to predict stoichiometries and geometries[19–22], atomization energies[23–26], formation energies[27–29], diffusion barriers[30], band gaps[31,32], and many other materials properties. When used in conjunction with experiments

[1]KBR Wyle, Intelligent Systems Division, NASA Ames Research Center, Moffett Field, CA 94035, USA. [2]Research Division, Nissan North America, Santa Clara, CA 95051, USA. [3]Research Division, Nissan Motor Company, Yokosuka, Kanagawa 237-8523, Japan. [4]Intelligent Systems Division, NASA Ames Research Center, Moffett Field, CA 94035, USA. ✉email: grbala@umich.edu; john.w.lawson@nasa.gov

and computations, ML approaches allow for the rapid exploration and discovery of new materials at much lower cost, significantly speeding up the materials design process.

Traditional Li-ion batteries use flammable organic liquid electrolytes that are prone to fire-hazards. Furthermore, the conventional choice of carbon-based anodes limits the specific energies of these batteries to less than 300 Wh/kg[33]. Ionic liquid electrolytes have been widely studied as a safer alternative due to their non-volatility and thermal stability[34], however their high viscosity, and thus relatively low conductivity, have limited their appeal[35]. An all-solid-state battery (ASSB) with a Li metal anode can address the safety issues as well as the demand for higher energy densities. The major obstacle to creating such a battery involves the development of solid electrolytes and protective anode coatings possessing high ionic conductivity, wide electrochemical window, good stability against Li metal, and inertness to environmental elements like water and air.

Computational approaches to understand ionic migration in solids include the nudged elastic band method (NEB)[36] and molecular dynamics (MD) methods based on density functional theory (DFT)[37,38], space topology analysis based on the Voronoi-Dirichlet partitioning of crystal space[39,40], and bond valence (BV) estimation techniques[38,41–44]. Although DFT methods provide the most accurate measurement of migration barriers, they are extremely expensive, thus limiting their applicability for high-throughput screening approaches. The BV method, on the other hand, can be used to quickly identify low mismatch pathways between Li sites, representing probable Li^+ transport paths in these structures. By linking the BV mismatch to the absolute energy scale using a Morse potential, percolating migration barriers can be extracted from the energy landscape in a quick and accurate manner[45]. While the accuracy of the BV approach is contingent upon the development of accurate empirical potentials, results show a consistent trend against DFT values[45,46], making it a quick screening filter for identifying fast ion-conducting solid electrolytes and coatings.

A lot of recent work has focused on screening for fast Li-conductors. Recently, Xiao et al.[38] used BV calculations to identify oxide candidates with low migration barriers from a set of 1000 Li-containing compounds in the inorganic crystal structure database (ICSD)[47]. In two other publications, He[48,49], Zhang[50], and others developed a platform for performing high-throughput screening for over 29,000 inorganic compounds from the ICSD using a combination of topology features, bond valence energies, and NEB calculations. Another study involved the use of topological analysis and ab-initio MD simulations to quantify key structural features of fast ion conductors[51]. First-principles calculations have also been performed to assess electrochemical stability for a wide range of electrolyte/electrode combinations[52,53], where exceptional agreement with experimental results was found. Several studies have focused on screening for coatings for Li and Li-ion batteries[54–56]. In one particular study, it was found that nitrides have a significantly lower reduction potential than sulfides, oxides, and fluorides, making them more suitable for anode coatings[57].

ML approaches to identify promising solid electrolytes are also getting more attention. Recently, Tian et al.[58] performed clustering based on modified XRD patterns of the anion sub-lattice to identify 16 promising Li-conductors. Sendek et al.[59] screened for Li superionic conductors using a logistic regression ML model trained on a relatively small set of 40 experimentally measured ionic conductivity values. The model identified 21 compounds as being possible superionic conductors, 8 of which were actually predicted to have an ionic conductivity $> 10^{-4}$ S/cm based on high temperature DFT-MD simulations[60].

In the following sections, we describe our approach to quickly explore the Li compound space for promising solid electrolyte and anode coating candidates. First, we create an extensive database of battery-related materials properties of electrolyte candidates based on over 15,000 Li compounds catalogued by Materials Project[8]. For each compound, we compute both transport and stability based properties. Our database can be used for multi-property screening and is distinguished from previous work which has been primarily based on single property descriptors and smaller datasets. Second, we screen our database for new compounds and identify over 250 promising candidates spanning different chemistries and structures. We find many compounds with low 3D barriers and wide stability. These candidates include some well-known names as well as several novel and unexpected compounds, which have not been previously reported. Third, we train accurate and interpretable multi-property ML models to predict 3D migration barriers as well as oxidation and reduction potentials. The ML models will permit broader searches for materials outside the original database. Previously, only one group has reported a supervised ML model for solid electrolytes, based on single property predictions of ionic conductivity trained on a small experimental dataset[59]. Arguably, our ML model is an improvement over these previous results and is also the first to predict electrochemical stability. In addition, our models are completely interpretable, and allow us to explain individual machine learning predictions in a way no other models do. We believe our multi-property ML model will be a valuable tool for future electrolyte design.

Database of Li compounds

To identify novel candidates for Li ASSB, we first create a database of 15,446 Li-containing compounds with their computed materials properties. Most well-known solid electrolytes have, in common, the same few characteristics - high ionic conductivity, wide electrochemical stability window, low electronic conductivity, good chemical, thermal, and mechanical stability etc. In order to identify good solid electrolytes and anode coatings, we focus on three important characteristics in this work: i. fast ionic migration; ii. wide electrochemical stability window; and iii. stability against Li metal.

Ionic migration. Ionic conductivity is the property most often used to study ionic migration in solids. The ionic conductivity of a solid measures how easily an ion can move from one site to another through defects in the crystal lattice. While the ionic conductivity clearly depends on the underlying crystal structure, experimentally measured values are also greatly influenced by the synthesis method, sample preparation, and measurement technique[61]. Additionally, the availability of, and access to, experimental ionic conductivity data is mostly lim-

ited; the only way to obtain such data is by text mining through existing publications or having access to internal databases maintained by individual research groups. Even then, the total accessible data is quite small. At the same time, theoretical ionic conductivity measurements using DFT-MD simulations are also quite expensive and not ideal for a high-throughput based approach. Instead, we compute Li^+ migration barriers using the softBV tool[45] developed by Chen *et al.* to measure ionic migration in Li compounds.

The softBV tool is based on the BV method, more generally used to examine the stability of chemical structures or estimate the oxidation state of atoms[42,62,63]. At the heart of the BV method is the valence sum rule which states that the sum of bond valences (S_{ij}) around any atom, i, with neighboring atoms j, should be equal to the atomic valence, (V_i) *i.e.*, the oxidation state.

$$V_i = \sum_j S_{ij} \tag{1}$$

S_{ij}, considered a measure of the electrostatic strength of a bond, can be written as a simple two-parameter algebraic equation:

$$S_{ij} = \left(\frac{R_0 - R_{ij}}{b} \right), \tag{2}$$

where R_{ij} is the observed bond length between atoms i and j, and R_0 and b are bond valence parameters that depend entirely on the nature of the bond. Every pair of elements can be fitted with unique R_0 and b values associated with their interaction. The BV mismatch is defined as

$$|\Delta V_i| = V_i - V_i^{id}, \tag{3}$$

where V_i^{id} is the ideal valence state of atom i. Sites in a crystal structure where $|\Delta V_i| \approx 0$ are considered accessible sites for atom i. The BV method postulates that paths between accessible sites along which $|\Delta V_i|$ remains sufficiently low represent probable ion transport pathways. Chen *et al.* linked the BV mismatch, $|\Delta V_i|$, to the energy scale, in order to compute activation energies, by developing a BV-based force field method using a general Morse-type interaction potential[45]. The isosurfaces of fixed $|\Delta V_i|$ then represent regions that atom i can reach with a certain activation energy.

According to the original definition of the valence sum rule, only interactions with atoms, j, in the first coordination shell are considered in Equation (1). However, it was later demonstrated that by incorporating information from higher coordination shells, R_0 and b can, in most cases, be refined further. The new parameters thus obtained by incorporating higher shell information are referred to as the bond softness parameters, and the resulting tool based on this approach is called softBV.

In addition to describing migration pathways between lattice sites, softBV can also find the lowest energy isosurfaces that percolate through the X, Y, and Z directions of the unit cell. The 1D barrier represents the lowest energy required by a diffusing species to hop across between two opposite faces of a unit cell, in any one of the three directions. The 2D and 3D barriers, similarly, represent the lowest energies required to hop between opposite faces in any two or all three directions, respectively. It should be evident that the 1D barrier $\leq$ 2D barrier $\leq$ 3D barrier for all solids. For the purpose of this study, we only use the 3D barriers. Supplementary Fig. S1 shows the 3D migration pathway for Li^+ in Li_3PS_4, evaluated using the softBV tool. The lowest activation energy required to connect every point on the shaded pathway is the 3D migration barrier, and it provides a quantitative measure of the achievable Li-ion conductivity in Li_3PS_4.

Using the softBV tool, we calculate 3D barriers for Li^+ diffusion in all 15,446 Li compounds in a high-throughput manner. While the accuracy of softBV depends on the accuracy of underlying bond softness parameters and empirical potential energy functions, a consistent trend between Li^+ barriers computed using softBV and DFT has been shown to exist[45,46]. It is important to remember that materials with potential for fast migration may not necessarily show high ionic conductivity unless there is a high concentration of mobile ions to carry the charge[64]. It is, thus, possible that Li compounds with extremely low 3D barriers have large associated defect formation energies that prevent them from being superionic conductors. The ionic conductivity of such materials can generally be enhanced by the introduction of additional charge carriers. Furthermore, since the BV approach does not allow for the structural relaxation of neighboring atoms that accompanies migration, softBV barriers tend to be overestimated in some cases[65]. Hence, softBV barriers should not be used to directly estimate the ionic conductivity.

Electrochemical stability window. The electrochemical stability window of a material represents the electrode potential range in which it is neither oxidized nor reduced. We use the grand potential phase diagram approach[66,67] to calculate the electrochemical stability window of materials. The grand potential phase diagrams represent phase equilibria that are open to Li. Here, the chemical potential of Li (μ_{Li}) is the external input variable that can be controlled. The relevant thermodynamic potential to study phase equilibria is the Li grand potential, defined as

$$e\phi = E - \mu_{Li}N_{Li} \tag{4}$$

where, e is the elementary charge, ϕ is the electrode potential, E is the total energy computed using DFT, and N_{Li} is the number of Li atoms in the system. The resulting phase diagram provides information about the equilibrium phases at different values of μ_{Li}.

As an example, we use the grand potential phase diagram of the ternary Li–P–S system to calculate voltage profiles depicting the lithiation and delithiation of Li_3PS_4 (Supplementary Fig. S2). At low electrode potentials, we see that Li_3PS_4 undergoes reduction and uptakes Li to form Li_2S and P, while at higher potentials, Li_3PS_4 is oxidized and loses Li forming LiS_4 and P_2S_7. The electrochemical stability window of Li_3PS_4 (1.7 − 2.2 V) is the range in which no lithiation or delithiation occurs *i.e.* where Li uptake is zero.

The grand potential analyses performed in this study include only the lowest energy phase at each composition. Thus, only 8,924 Li compounds in the database have reported electrochemical stability windows. In the case of meta-stable compounds that do not lie on the convex hull, the stability window is represented by the oxidation and reduction potentials of their decomposition components in the grand potential phase diagram. We use the DFT energies obtained from Materials Project for our calculations.

Stability against Li metal. Stability against Li metal represents a material's inertness to lithium. A material is considered stable if it does not undergo spontaneous reaction with Li at 0 V. Thus, materials that are stable against Li have a reduction potential (*vs.* Li/Li+) of 0 V. According to Supplementary Fig. S2, Li_3P and Li_2S are the only two phases in the Li–P–S system that are stable against Li metal.

Besides the three characteristics described above, we include in our database DFT computed values of energy above the convex hull (E_{hull}) and band gap (E_g) for all 15,446 Li compounds, also obtained from Materials Project. E_{hull}, which is a measure of thermodynamic stability at 0 K, is defined as the vertical distance of a phase from the convex hull in terms of energy per atom. A stable compound has $E_{hull} = 0$ eV. Compounds with non-zero E_{hull} values are thermodynamically unstable at 0 K, although they may be stabilized at higher temperatures through entropic contributions. We use E_g as a measure of the electronic conductivity in a solid. Materials with $E_g > 0$ eV are known to be bad electronic conductors. In the future, we plan to extend the database to include other properties like elastic moduli, stability in water and air, etc.

Screening

We use the database generated in section "Database of Li compounds" to screen for promising solid electrolyte and anode coating candidates. To identify fast Li-conducting solid electrolytes, we focus on materials with low migration barriers. We select an arbitrary cutoff of 0.5 eV, and screen for all Li compounds with 3D barriers ≤ 0.5 eV.

For protective coatings on the anode, stability against Li metal on one side, and the electrolyte on the other, is more important. Hence, we identify compounds with reduction potentials (*vs.* Li/Li+) of 0 V and an electrochemical stability window wider than 1 V. We restrict the 3D barriers of coatings to 1 eV, to achieve a good balance between electrochemical stability and ionic conduction. Because coatings are generally applied as thin-films, they allow for a comparatively low ionic conductivity.

To ensure that our screened solid electrolytes and coatings are also thermodynamically stable and possess low electronic conductivities, we add a second filtering step to only include compounds with $E_{hull} \leq 30$ meV and $E_g \geq 1$ eV. We use a relatively safe lower bound of 1 eV to ensure minimal electronic conduction across the solid electrolyte or coating layer, accounting for the fact that local and semi-local DFT functionals tend to underestimate band gaps[68]. Finally, compounds with transition metal atoms are generally known to be susceptible to reactions with Li, as transition metals have many stable oxidation states. We, therefore, filter out most compounds with d- and f-block elements.

Although we calculate electrochemical stability windows only for the lowest energy phase at every composition, limiting our search to candidates with $E_{hull} \leq 30$ meV ensures that the stabilities of all identified compounds at a given composition are fairly similar. It should be noted that since the grand potential phase diagram depends on all existing phases in a given chemical subspace, it is prone to change when new phases are discovered or get added to Materials Project. This would affect the electrochemical stability windows of all compounds within the subspace. However, many of these systems are fairly well-studied and this should not be a major issue.

We identify over 250 promising solid electrolyte candidates using the above criteria. Few of them are highlighted in Fig. 1. The screened compounds are classified based on their chemistries into sulfides, halides, oxides, and other compounds. Supplementary Figs. S3–S6 show expanded results for each class. Among the sulfide compounds, well-known superionic conductors like $Li_7P_3S_{11}$ and Li_3PS_4 are identified[69,70]. While $Li_{10}GeP_2S_{12}$ (LGPS) does not show up due to its marginally higher E_{hull} (32 meV), its Si counterpart ($Li_{10}SiP_2S_{12}$) gets picked up by our screening algorithm[71]. In addition, we identify several new sulfides, including Li_9S_3N, as promising electrolyte candidates based on their low 3D barriers. Among halides, the argyrodite Li_6PS_5I is identified[72], but Li_6PS_5Br and Li_6PS_5Cl are excluded due to their higher E_{hull} values. Other known fast Li-conductors like Li_3YBr_6 (0.55 eV), Li_3InCl_6 (0.59 eV), Li_3ScCl_6 (0.56 eV) fall just above the 3D barrier cutoff of 0.5 eV and are also excluded. However, both $LiBF_4$ and $LiSbF_4$ are identified as having extremely low 3D barriers. $LiAlB_2O_5$ is the best performing oxide identified through screening, having a low 3D barrier of 0.175 eV. Other screened oxides include NASICON-type solid electrolytes like $LiTi_2(PO_4)_3$ and $LiZr_2(PO_4)_3$, as well as garnets like $Li_7La_3Zr_2O_{12}$ (LLZO) and $Li_7La_3Hf_2O_{12}$ (LLHfO), all of which have measured ionic conductivities ranging between 10^{-5} to 10^{-3} S/cm[2]. Li_2CO_3 is also identified as a fast Li-conductor with a low 3D migration barrier. DFT studies have confirmed the low migration barrier for Li_2CO_3[73], however it has been shown that an increase in its charge carrier concentration through the introduction of intrinsic defects or dopants is required to overcome the relatively high defect formation energies in this compound[74].

For anode coatings, the stringent requirement of stability against Li metal leaves us with only 26 compounds, including some binaries like Li_2O, Li_2S, Li_2Se, and LiX ($X =$ halide) which have already been tested as thin-film coatings[75,76] (see Supplementary Table S1). We identify $Sr_4Li(BN_2)_3$ as an extremely promising anode coating candidate with a low 3D barrier and wide electrochemical stability window. Similarly, $LiYO_2$ and $Li_7La_3Hf_2O_{12}$,
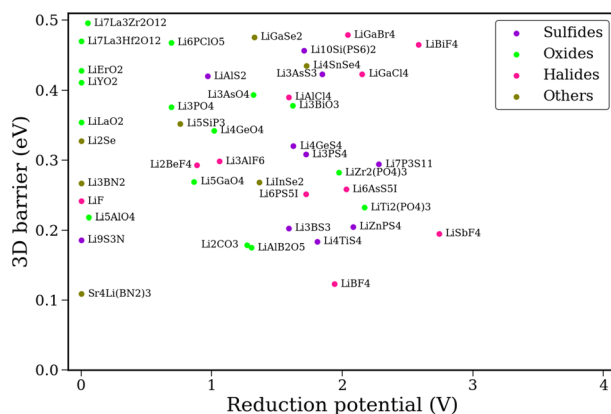
Figure 1. Promising electrolyte and anode coating candidates identified through the screening approach. Complete screening results can be found in Supplementary Figs. S3–S6 and Supplementary Table S1. [3D barrier ≤ 0.5 eV, $E_{hull} \leq 30$ meV, $E_g \geq 1$ eV].

both of which have been previously studied as potential solid electrolyte candidates[77,78], also get picked up by the screening algorithm. The success of such thin-film coatings largely depends on striking the right balance between thickness of the coating, achieved by deposition or chemical reaction methods, and complete coverage of the Li metal anode surface.

Li$^+$ migration pathways for a few chosen electrolyte and anode coating candidates are shown in Supplementary Fig. S7.

Machine learning

Screening, as described above, helps us explore the database of Li compounds for promising solid electrolyte and anode coating candidates. ML, on the other hand, provides us a way to not only extend this search to materials outside the database, but to also explain individual predictions and generate model-level insights that can be used for designing new compounds. Training ML models on properties within the database gives us the ability to quickly test new Li compounds in the future without the need for explicit computations. Additionally, ML models can solve the problem of estimating 3D barriers for compounds for which softBV parameters have not yet been calculated.

We train independent ML models using Li compounds in the database to predict 3D barriers and electrochemical stability windows. ML regression models take a vector $x \in \mathbb{R}^n$ as input and return a value y. To utilize these models for predictions, we therefore construct a vector based data representation that encodes relevant physical information about the crystal structure and/or composition of the compounds. We compare the performance of two ensemble learning algorithms on our regression tasks – random forests (RF)[79] and gradient boosting decision trees (GB)[80,81]. Ensemble methods combine predictions from several base estimators to reduce bias and variance, thus improving performance. They work extremely well with small and medium sized datasets that have a mix of categorical and continuous features spanning various scales. As an additional bonus, ensemble algorithms do not require feature scaling of components of input vectors. We use the implementations of RF and GB regression available from the sci-kit learn python package[82]. We perform hyperparamter selection using a grid-based search of the hyperparameter space and ten-fold cross-validation. R^2 values are obtained by averaging over 20 randomized 90%-10% training-test set splits and are reported for the held-out test set only.

3D barriers. For the task of predicting 3D barriers, we use the crystal structures of Li compounds as input. We identify features that are both physically motivated and easy to compute in order to represent the crystal structures. Based on previous research[2], we know that structures with a large number of connected sites available for mobile ions to occupy, and small migration barriers between these sites, are ideal for fast ionic migration. To leverage this information, we include features describing the Li concentration, local geometry, sublattice chemistry, and topology into our ML models. We use several existing features, like those developed by Sendek et al.[59] and Ward et al.[83], as well as introduce some new ones. We perform feature selection to identify a final set of 22 features that provide a good balance between accuracy and simplicity in our model. The complete list is shown in Table 1 and described in Supplementary section S1.

We use features like *Li atomic fraction* (Li), *mean Li neighbor count* (LNC), and *mean Li-Li bonds per Li atom* (LLB) to feed information about the concentration and arrangement of Li atoms to our models. Other features like *mean sublattice neighbor count* (SNC), *mean sublattice bond ionicity* (SBI), and *mean electronegativity of the sublattice* (ENS) provide similar details about the sublattice elements *i.e.* non-Li constituents in the compounds. Local geometry is also captured in our models through features like *mean Li-Li, Li-Anion*, and *Anion-Anion separation distances* (LLSD, LASD, AASD), *mean neighbor distance variation* (NDV), and *mean ordering parameters* (OP_1, OP_2, OP_3).

Topological features provide details about the channels available for Li$^+$ diffusion within the crystal structure. Such information is incorporated into the model through features including *diameter of the largest free sphere*

No.	Feature	Acronym
1.	Li atomic fraction	Li
2.	Mean Li neighbor count[59]	LNC
3.	Mean Li–Li bonds per Li[59]	LLB
4.	Mean sublattice neighbor count[59]	SNC
5.	Mean sublattice bond ionicity[59]	SBI
6.	Mean electronegativity of sublattice[59]	ENS
7.	Mean Li–Li separation distance[59]	LLSD
8.	Mean Li-anion separation distance[59]	LASD
9.	Mean anion-anion separation distance[59]	AASD
10.	Mean neighbor distance variation[83]	NDV
11.	Mean ordering parameter shell 1[83]	OP_1
12.	Mean ordering parameter shell 2[83]	OP_2
13.	Mean ordering parameter shell 3[83]	OP_3
14.	Diameter of largest free sphere[84]	DLFS
15.	Mean Straight Line Path Width[59]	SLPW
16.	Sublattice packing fraction[59]	SPF
17.	Max packing efficiency[83]	MPE
18.	XRD principal component 1	XRD_1
19.	XRD principal component 2	XRD_2
20.	XRD principal component 3	XRD_3
21.	XRD principal component 4	XRD_4
22.	XRD principal component 5	XRD_5

Table 1. List of 22 structure-based features used to predict 3D barriers for Li^+ migration in Li compounds. A complete description of the individual features can be found in Supplementary section S1.

(DLFS), calculated using the Zeo++ tool[84], *mean straight line path width* (SLPW), *sublattice packing fraction* (SPF), and *maximum packing efficiency* (MPE).

Additionally, in order to also include a single representative fingerprint of the entire crystal structure, we use a reduced version of the powder X-ray diffraction (XRD) pattern, consisting of the first five eigenvectors (XRD_1 to XRD_5) calculated using principal component analysis, as additional inputs to our models. The fraction of variance explained by the first five principal components together is 0.65. Our final choice of 22 assorted features are meant to provide high accuracy in our models while still maintaining interpretability.

Oxidation and reduction potentials. Since electrochemical stability windows are calculated for a single (lowest energy) phase at each composition, we decide to use only the composition of Li compounds as input for training the corresponding ML models. We break the stability window prediction problem into two tasks: predicting oxidation potentials (*vs.* Li/Li+) and reduction potentials (*vs.* Li/Li+), using the same set of input features for both tasks.

As before, we test an assortment of different features to represent the composition of Li compounds in the database. These are based on physical and chemical properties of individual elements present in the compound, weighted by composition. They include eight simple element properties, each measured through three statistics – mean, maximum, and range – resulting in a total of 24 distinct features. In addition, we introduce 4 new features based on oxidation states. The first two measure the total difference between the oxidation states of elements present in the compound and their lowest, or highest, possible oxidation states. For compound X, the two features are given by

$$oxid.state - min.oxid.state(X) = \sum_{i \in \{X\}} |V_i^X - min\{V_i\}|$$

$$max.oxid.state - oxid.state(X) = \sum_{i \in \{X\}} |max\{V_i\} - V_i^X|$$

where, $\{X\}$ is the set of elements present in X, V_i^X is the oxidation state of element i in compound X, and $\{V_i\}$ is the set of possible oxidation states of i. The remaining two features measure the same quantity, but are weighted by the Pauling electronegativities of the elements.

Composition-weighted element properties	Statistics
Standard features	
Atomic number	Mean
Electronegativity	Maximum
Valence d-electrons	Range
Unfilled d-electrons	
Unfilled p-electrons	
Band gap	
Magnetic moment	
Melting temperature	
New oxidation state features	
oxid. state − min. oxid. state	
max. oxid. state − oxid. state	
(oxid. state − min. oxid. state) * E_neg	
(max. oxid. state − oxid. state) * E_neg	

Table 2. List of 28 element property features used for predicting oxidation and reduction potentials (*vs.* Li/Li+). Each of eight standard element properties are weighted by composition and measured through three different statistics to obtain the first 24 features. The other 4 are new oxidation state features introduced in "Oxidation and reduction potentials".

$$(oxid.state − min.oxid.state) * E\_neg\ (X) = \sum_{i \in \{X\}} |V_i^X − min\ \{V_i\}| * EN_i$$

$$(max.oxid.state − oxid.state) * E\_neg\ (X) = \sum_{i \in \{X\}} |max\ \{V_i\} − V_i^X| * EN_i$$

The final list of 28 features is given in Table 2.

Shapley explanations and permutation feature importances. To interpret a ML model, it is useful to know which features have the highest impact on predictions. SHapley Additive exPlanations (SHAP)[85] and Permutation Feature Importance analysis (PFI)[79] are two model inspection techniques that we use to explain the output from our ML models. Shapley explanations use a game theoretic approach to break down individual predictions. Specifically, SHAP measures the impact of having a certain value for a given feature in comparison to the feature taking some baseline value.

On the other hand, PFI provides an overall metric of the most important features for a given model. It measures the change in model score when a single feature vector is randomly permuted; a higher score reflects a higher dependence of the model on the corresponding feature. The shap and rfpimp[86] python packages are used to calculate SHAP and PFI values in this work. We report SHAP values for both training and test data, whereas PFI is measured, using R^2 values, on the test set only.

Although tree-based methods also provide an inbuilt measure of feature importance, based on the decrease in impurity, they are always computed on the training set and, therefore, do not reflect the predictive ability of the feature on unseen data. Furthermore, impurity-based feature importance schemes are strongly biased towards high cardinality features. PFI does not exhibit any such biases. It must be noted that PFI, just like impurity-based metrics, can still be affected by the presence of highly correlated features. When a correlated feature is permuted, the model still has access to the feature through its correlated counterpart, thereby lowering the apparent importance value for both features. A common solution, and one we use for the oxidation and reduction potentials models, is to cluster the correlated features prior to performing feature analysis.

Results and discussion

3D barriers.
The structure-based features introduced above are used to train RF and GB regression models to predict 3D barriers. To avoid outliers, we restrict our models to compounds with 3D barriers ≤ 5 eV. Table 3 shows the cross-validated results on the barrier prediction task. We see that the GB model performs slightly better than the RF model, providing R^2 values of 0.86 on average. To study the importance of feature selection, we also compare the performance of our input features against other structure-based representation schemes from literature. We find that our set of 22 input features performs better than all others at the 3D barrier prediction task. Supplementary Fig. S8 shows parity plots comparing the 3D barriers predicted using the GB model against barriers computed using softBV.

Although our database contains 1D, 2D, and 3D migration barriers calculated by softBV, we only use the 3D barriers for screening and ML. In isotropic materials, where ionic conduction is equally fast in all three dimensions, the three barriers are equal. Other compounds like $LiFePO_4$, $LiCoO_2$ etc. have dominant 1D or 2D conduction pathways[87,88], and thus have 1D and 2D barriers that are significantly lower than their 3D barrier. However, crystal defects and imperfections can potentially obstruct these low barrier conduction pathways, and

Features + model	R^2	Vector length
Current work + GB	0.86	22
Current work + RF	0.84	22
Sendek conductivity features[59] + GB	0.76	20
Coulomb matrix[23] + GB	0.72	200
Powder X-ray diffraction pattern + GB	0.70	128

Table 3. The cross-validated performance of our RF and GB models trained using 22 input features on the 3D barrier prediction task. Also compared are three other structure-based representations from literature (with the same GB model in each case).

| No.	mp_id	Formula	3D barrier (eV)	DFT-MD predicted ionic conductivity $> 10^{-4}$ S/cm
1.	mp-558219	$SrLi(BS_2)_3$	0.340	No
2.	mp-532413	$Li_5B_7S_{13}$	0.487	Yes
3.	mp-643069	Li_2HIO	0.519	Yes
4.	mp-7744	$LiSO_3F$	0.529	Yes
5.	mp-569782	$Sr_2LiCBr_3N_2$	0.548	Marginal
6.	mp-676109	Li_3InCl_6	0.595	Yes
7.	mp-676361	$LiErCl_6$	0.664	Yes
8.	mp-22905	$LiCl$	0.735	No
9.	mp-559238	$CsLi_2BS_3$	0.780	Yes
10.	mp-29410	$Li_2B_2S_5$	0.972^d	Yes
11.	mp-34477	$LiSmS_2$	0.977	No
12.	mp-8430	$KLiS$	1.720^d	No
13.	mp-8751	$RbLiS$	1.727^d	No
14.	mp-554076	$BaLiBS_3$	2.035	No
15.	mp-866665	$LiMgB_3(H_9N)_2$	2.330^d	Yes
16.	mp-19896	Li_2GePbS_4	2.835	No
17.	mp-561095	$LiHo_3Ge_2(O_4F)_2$	3.259	No
18.	mp-15791	$LiErS_2$	4.698^d	No
19.	mp-15790	$LiHoS_2$	4.749^d	No
20.	mp-15789	$LiDyS_2$	4.872	No
21.	mp-15797	$LiErSe_2$	4.884	Marginal

Table 4. Li compounds identified as possible superionic conductors by the Sendek model[59] are arranged in increasing order of their softBV 3D barriers. Where calculated barriers are missing, ML predictions are used instead. The rightmost column indicates whether the compound was predicted to have an ionic conductivity $> 10^{-4}$ S/cm based on high temperature DFT-MD simulations[60]. ^{d} ML prediction

special material preparation and processing is needed to take advantage of them. The 3D barrier provides a much better estimate of the overall ionic migration in these compounds.

To test the validity of this idea, we consider the 21 compounds identified as being possible superionic conductors by the Sendek model[59]. Table 4 ranks the compounds in increasing order of their softBV 3D barriers. Where the calculated barriers are missing, we use predictions from our GB model instead. Among the 21 compounds, 8 are predicted to have an ionic conductivity $> 10^{-4}$ S/cm based on high temperature DFT-MD simulations, while 2 more are identified as marginal cases[60]. 8 of these 10 total compounds also have a low calculated 3D barrier and occupy the top half of the table. However, our models estimate much larger barriers for $LiMgB_3(H_9N)_2$ as well as for $LiErSe_2$, which is structurally similar to $LiErS_2$. On the other hand, our models also perform well on the remaining 11 compounds that the Sendek model identifies as being superionic, but for which DFT-MD simulations predict ionic conductivities $< 10^{-4}$ S/cm. We estimate high 3D barriers for all but one of these compounds. The exception, $SrLi(BS_2)_3$, has the lowest calculated 3D barrier among the 21 compounds and is explored further below. LiCl, which appears to have a moderately low 3D barrier, is already identified as a promising anode coating candidate in "Screening" section. It must be noted that the LiCl structure identified by us is the DFT ground state structure, and slightly differs from the one identified by the Sendek model. Overall, our calculated and predicted 3D barriers show good agreement with high temperature DFT-MD simulation results.

In addition to being good at predicting Li^+ migration barriers, our models are also interpretable. Fig. 2 shows Shapley explanations[89] for individual predictions made by our GB model. Red arrows represent feature effects that drive the predicted 3D barrier higher, while blue arrows represent those that drive the prediction lower.
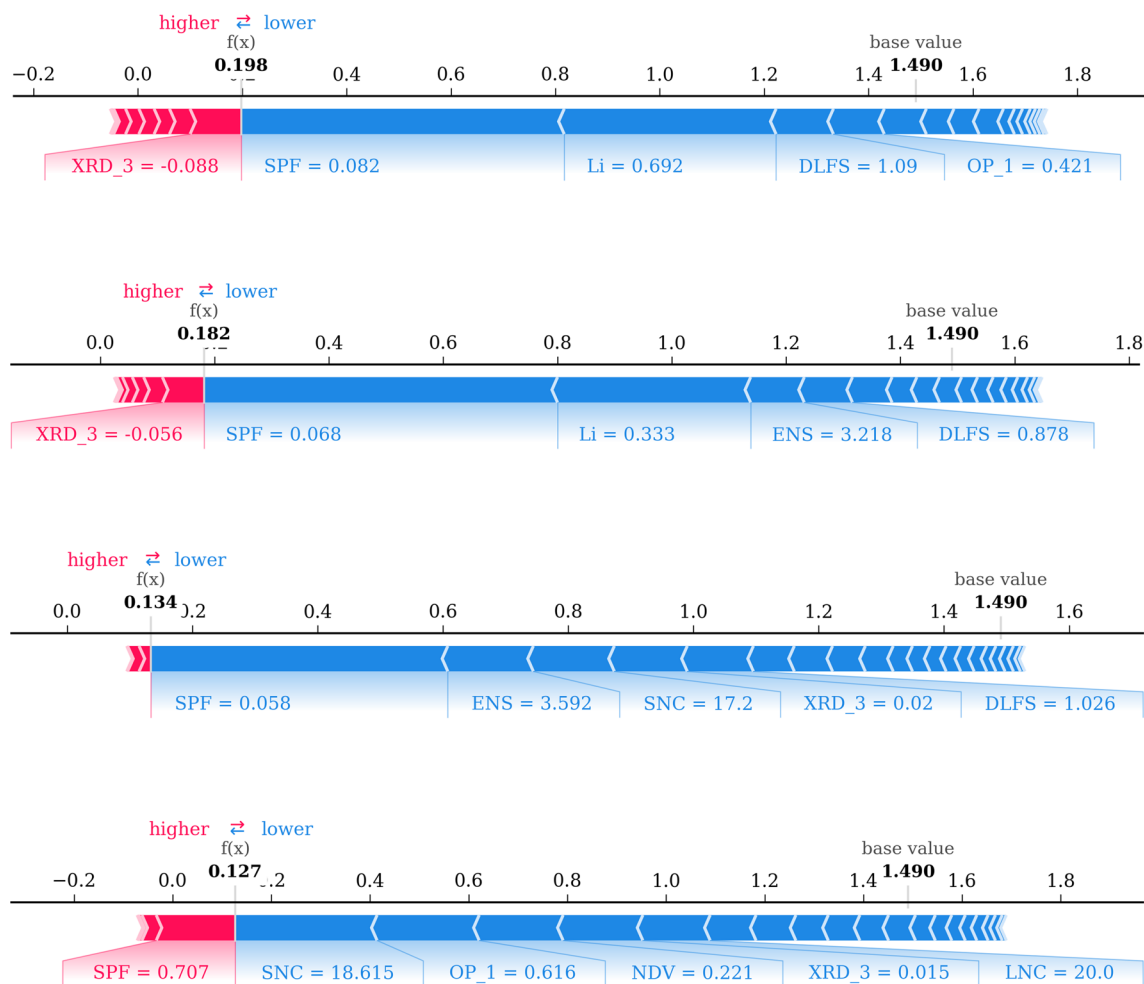
Figure 2. Shapley explanations for individual predictions made by the GB model trained on 3D barriers. Red arrows represent feature effects that drive the predicted 3D barrier higher, while blue arrows represent those that drive the prediction lower. The lengths of the arrows indicate the magnitudes of the effects. From top to bottom: Li_9S_3N, Li_2CO_3, $LiBF_4$, and $Sr_4Li(BN_2)_3$.

The lengths of the arrows indicate the magnitude of the effects, referred to as SHAP values. We first consider 3D barrier predictions of compounds identified by the screening model as potential solid electrolyte candidates in "Screening" section. These are training set predictions that give us an insight into how the GB model uses input features to make decisions. We see that Li_9S_3N, the most promising sulfide compound, has an extremely small *sublattice packing fraction* (SPF) = 0.082 and extremely large *Li fraction* (Li) = 0.692, driving its 3D barrier prediction from the base value of 1.490 eV, which is the GB model's average prediction over the entire training set, down to 0.198 eV. Similarly, Li_2CO_3 also has a very small SPF = 0.068, moderate *Li fraction* = 0.333, as well as high *mean electronegativity of sublattice* (ENS) = 3.218, resulting in a similarly low 3D barrier prediction of 0.182 eV. $LiBF_4$, on the other hand, does not have a large *Li fraction*. However, a small SPF = 0.058, high ENS = 3.592, and lower-than-average *sublattice neighbor count* (SNC) = 17.20, all drive its 3D barrier prediction down to 0.134 eV. All three compounds have reasonably large *diameters of largest free spheres* (DLFS). $Sr_4Li(BN_2)_3$, with one of the lowest computed 3D barriers, is a peculiar case. A small *Li fraction* = 0.071 and extremely large SPF = 0.707, which in fact drives the prediction higher for $Sr_4Li(BN_2)_3$, are counterbalanced by a low SNC = 18.615 and several other negative contributions from local geometry based features.

We use similar plots to explain GB predictions for compounds from Table 4 with contradicting softBV and high temperature DFT-MD results. We see that $LiErSe_2$ has an extremely large predicted 3D barrier of 4.889 eV, even though its ionic conductivity is predicted to be marginal by DFT-MD. Supplementary Fig. S9 reveals that the the major contributors driving the 3D barrier prediction of $LiErSe_2$ higher are a large SPF = 0.384, high *maximum packing efficiency* MPE = 0.522, and low ENS = 2.113, all indicators of sluggish Li^+ migration. Similarly, $LiMgB_3(H_9N)_2$ is also predicted to have an ionic conductivity > 10^{-4} S/cm by DFT-MD, but its small SPF = 0.126 is more than made up for by a very high SNC = 34.5, extremely small *Li fraction* = 0.04 , and low *mean sublattice bond ionicity* (SBI) = 0.284 and ENS = 2.213, resulting in a high predicted 3D barrier of 2.330 eV. Finally, although $SrLi(BS_2)_3$ has a small *Li fraction* = 0.09 and above-average SPF = 0.279, atypical of fast Li^+ conductors, it also has extremely low SNC = 10.30 and LNC = 11.0, and large DLFS = 1.111, explaining its low barrier prediction.
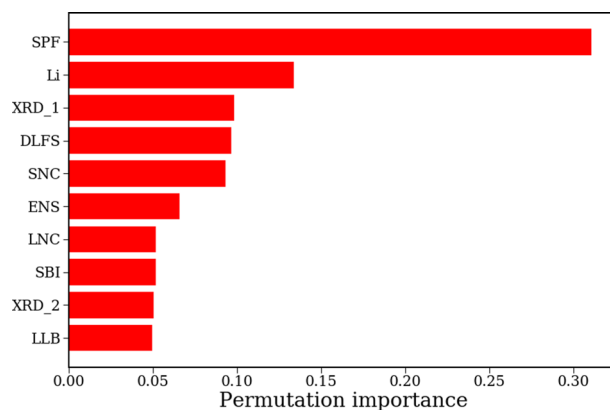
Figure 3. Permutation feature importance plot for the 3D barrier GB model showing the decrease in R^2 values upon random permutation of individual feature vectors. Higher values indicate a larger impact on predictions. Only features with the top 10 PFI scores are shown.

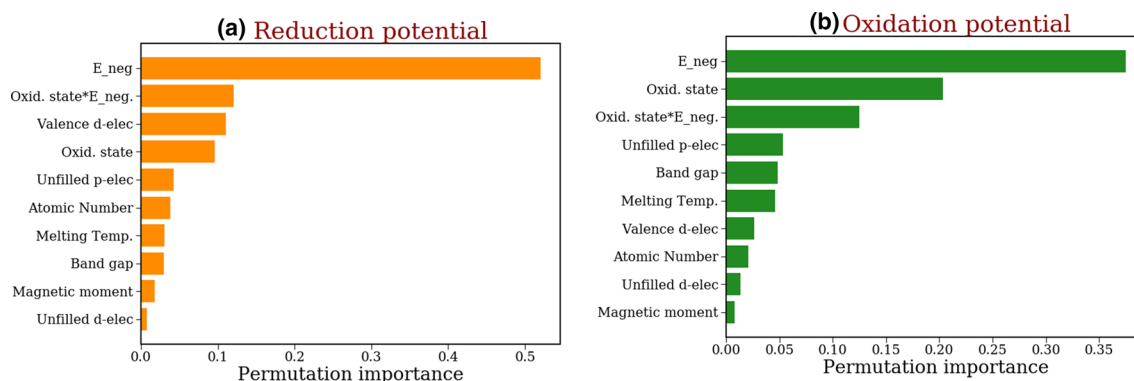| Features + model	Reduction potential	Oxidation potential	Vector length
Current work + GB	0.95	0.92	28
Current work + RF	0.93	0.91	28
Roost[90] + neural networks	0.92	0.92	64
Magpie element features[12] + GB	0.89	0.84	132
Element fractions + GB	0.87	0.87	103

Table 5. The cross-validated performance of our RF and GB models trained using 28 input features on the oxidation and reduction potential prediction tasks. Also compared are three other composition-based representations from literature (with the same GB model in each case).

In addition to providing explanations for individual predictions, our ML models can also be used to study the impact any given feature has on overall predictions through PFI analysis. The PFI values measured on the test set of the GB model are shown in Fig. 3. It is evident from looking at the figure that SPF, by far, has the largest impact on overall predictions of 3D barriers. The feature importance plot indicates that if SPF values were to be randomly permuted for the test set, the resulting R^2 of the GB model would drop by ~ 0.3. In addition to SPF, we see other features discussed above like Li, SNC, and DLFS also have high impact on overall predictions. Both, SPF and DLFS, measure the void space inside the crystal structure available for Li^+ ions to diffuse through, making them excellent predictors of ionic migration. A low SPF, or wide DLFS, implies a low 3D barrier and, thus, fast ionic migration. Similarly, a high *Li fraction* or low SNC indicates a higher proportion of candidate Li sites for Li^+ to hop onto, also favoring fast migration. Although XRD_1, as a feature, is hard to interpret, its overall importance highlights the fact that crystal geometry largely determines the Li^+ migration barrier in these compounds.

Supplementary Fig. S10 shows a heatmap of the Pearson correlation coefficients between each pair of input features and between individual features and the 3D barrier. These coefficients measure the strength of the linear association between two variables, taking values between -1 and 1. Positive values imply a positive correlation whereas negative values imply an inverse correlation. We observe the same positive and negative correlations discussed in the paragraphs above, between the top five features and the 3D barrier.

Oxidation and reduction potentials. Just like the 3D barriers, we also train ML models to predict oxidation and reduction potentials. Table 5 provides a comparison between R^2 values for RF and GB models against other composition-based representation schemes from literature. We again see the GB model marginally outperform the RF model at both tasks. Comparison against other representations confirms the superior predictive performance of our selected features. Supplementary Fig. S11 shows parity plots comparing the oxidation and reduction potentials predicted using GB models against those calculated directly from DFT energies. We see excellent agreement between the two sets of values with no clear outliers.

Shapley explanations for oxidation and reduction potential predictions are less intuitive because of the correlated nature of input features, which measure the same property through different statistics. Hence, we only perform PFI analysis to study overall feature importance for these GB models. We still need to cluster input features describing the same property into individual groups to understand the impact of each element property on the oxidation and reduction potential predictions. Additionally, we also combine the oxidation state features into two groups as shown in Fig. 4.

Figure 4. Permutation feature importance plots for the (**a**) reduction potential and (**b**) oxidation potential GB models. Features describing the same element property, measured through different statistics, are clustered together to avoid highly correlated features.

We see that *electronegativity* features have the largest impact on predictions of oxidation and reduction potentials. The electronegativity of an element represents its tendency to attract electrons, and largely determines whether it gets oxidized or reduced. Hence, the major influence on calculated oxidation and reduction potentials is expected. Additionally, the new oxidation state features introduced in this work also seem to be quite effective at predicting electrochemical windows. These features account for the variations in oxidation states of multivalent elements, making them excellent predictors of stability. Other input features that have a noticeable impact include the number of *valence d-electrons* and *unfilled p-electrons*.

It is important to point out that PFI and SHAP values are not intrinsic predictors of feature importance. They are model dependent, and only represent the importance a given value or feature has on predictions made using that particular model. However, for accurate models with physically motivated and interpretable features, these results can still be useful. For example, $Li_{10}SiP_2O_{12}$ (LSiPO) has a 3D barrier of 0.305 eV. We would expect a compound with the same crystal structure as LSiPO, but with larger sublattice ions, to exhibit a higher 3D barrier due to its large SPF, all other features being equal. We can test this by substituting Sn for Si in the structure. Sn belongs to the same group (IVA) and has a similar electronegativity as Si (1.96 *vs.* 1.90), but a bigger ionic radius. By simply changing the SPF, we find that the 3D barrier of LSnPO increases to 0.518 eV. Similarly, if we substitute the O in LSiPO with S instead, we would expect the new compound to have a lower ENS due to the lower electronegativity of S compared to O (2.58 *vs.* 3.44), as well as a higher SPF due to the larger S ions. This would also drive the barrier higher, which is exactly what we see for LSiPS, whose 3D barrier is 0.457 eV.

The above example shows that, even though PFI analysis and Shapley explanations do not necessarily provide causality, they can still serve as a useful guide for designing new compounds with tailored 3D barriers. Besides substitutions, the introduction of intrinsic defects and dopants can also be used to alter the geometry, electronegativity, and the Li fraction in the compounds. Such an approach has the potential to enable the discovery of promising new solid electrolytes and will be the focus of our future work.

Summary

In summary, we developed a materials informatics approach to explore and identify promising candidates for solid electrolytes and protective anode coatings in all-solid-state Li batteries. By combining high-fidelity DFT calculations with quick BV-based computations, we generated a database of battery-related materials properties for all Li compounds on Materials Project. We screened through the database to identify over 250 electrolyte and 26 anode coating candidates that span a wide range of structures and compositions. We found new compounds like Li_9S_3N, $LiAlB_2O_5$, $LiYO_2$, $LiSbF_4$, and $Sr_4Li(BN_2)_3$, that have extremely low 3D barriers, and provide a good balance of electrochemical stability and fast ionic migration.

We also trained machine learning models, using ensemble methods, to predict 3D migration barriers and oxidation and reduction potentials for Li compounds, to identify promising compounds outside our database. By using an assorted set of physically motivated features, we showed that our models can achieve high accuracy combined with interpretability. We compared our carefully selected features against common descriptors from literature to show that our features achieved better performance. Additionally, we explained individual predictions and provided model-level insights, useful for designing new electrolyte and coating materials in the future.

Our database and the subsequent screening, machine learning, and model interpretation techniques, together, provide a comprehensive strategy to efficiently explore the Li compound space for promising solid electrolyte candidates. Our approach has the potential to accelerate the discovery and design of novel electrolyte and coating materials for Li ASSB. Future work will involve designing and testing new Li compounds using the trained models, as well as accurate DFT simulations and experiments to perform rigorous validation of the screening and machine learning results. We also plan to extend our database to include other properties such as elastic moduli, which are predictors of dendrite formation, and stability in air and water in the future.

Data and code availability

Data not found in the main text or SI is available from the corresponding authors upon request. The developed ML model was implemented using the scikit-learn library in python. All the scripts for generating input features for the ML models are also available upon request from the corresponding authors.

References

1. Pasta, M. *et al.* 2020 roadmap on solid-state batteries. *J. Phys.* **2**, 032008. https://doi.org/10.1088/2515-7655/ab95f4 (2020).
2. Manthiram, A., Yu, X. & Wang, S. Lithium battery chemistries enabled by solid-state electrolytes. *Nat. Rev. Mater.* **2**, 1 (2017).
3. Hwang, J.-Y., Park, S.-J., Yoon, C. S. & Sun, Y.-K. Customizing a Li-metal battery that survives practical operating conditions for electric vehicle applications. *Energy Environ. Sci.* **12**, 2174 (2019).
4. Du, Z., Wood, D. L. III. & Belharouak, I. Enabling fast charging of high energy density Li-ion cells with high lithium ion transport electrolytes. *Electrochem. Commun.* **103**, 109 (2019).
5. Wang, X. *et al.* Toward high-energy-density lithium metal batteries: Opportunities and challenges for solid organic electrolytes. *Adv. Mater.* **32**, 1905219 (2020).
6. Radin, M. D. *et al.* Narrowing the gap between theoretical and practical capacities in Li-ion layered oxide cathode materials. *Adv. Energy Mater.* **7**, 1602888 (2017).
7. Folkson, R. *Alternative Fuels and Advanced Vehicle Technologies for Improved Environmental Performance: Towards Zero Carbon Transportation* (Elsevier, 2014).
8. Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *Appl. Mater.* **1**, 011002 (2013).
9. Kirklin, S. *et al.* The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies. *NPJ Comput. Mater.* **1**, 1 (2015).
10. Curtarolo, S. *et al.* AFLOWLIB. ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227 (2012).
11. Ward, L. *et al.* Matminer: An open source toolkit for materials data mining. *Comput. Mater. Sci.* **152**, 60 (2018).
12. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *NPJ Comput. Mater.* **2**, 1 (2016).
13. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: Recent applications and prospects. *NPJ Comput. Mater.* **3**, 1 (2017).
14. Jain, A., Hautier, G., Ong, S. P. & Persson, K. New opportunities for materials informatics: Resources and data mining techniques for uncovering hidden relationships. *J. Mater. Res.* **31**, 977 (2016).
15. Zhou, Q. *et al.* Learning atoms for materials discovery. *Proc. Natl. Acad. Sci. USA* **115**, E6411 (2018).
16. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
17. Isayev, O. *et al.* Materials cartography: Representing and mining materials space using structural and electronic fingerprints. *Chem. Mater.* **27**, 735 (2015).
18. Ye, W. *et al.* Harnessing the Materials Project for machine-learning and accelerated discovery. *MRS Bull.* **43**, 664 (2018).
19. Meredig, B. *et al.* Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014).
20. Javed, S. G., Khan, A., Majid, A., Mirza, A. M. & Bashir, J. Lattice constant prediction of orthorhombic ABO 3 perovskites using support vector machines. *Comput. Mater. Sci* **39**, 627 (2007).
21. Pilania, G., Wang, C., Jiang, X., Rajasekaran, S. & Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **3**, 2810 (2013).
22. Liu, Y., Zhao, T., Ju, W. & Shi, S. Materials discovery and design using machine learning. *J. Materiom.* **3**, 159 (2017).
23. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
24. Huo, H., & Rupp, M. Unified representation for machine learning of molecules and crystals (2017). arXiv preprint arXiv:1704.06439.
25. von Lilienfeld, O. A., Ramakrishnan, R., Rupp, M. & Knoll, A. Fourier series of atomic radial distribution functions: A molecular fingerprint for machine learning models of quantum chemical properties. *Int. J. Quantum Chem.* **115**, 1084 (2015).
26. Montavon, G., Hansen, K., Fazli, S., Rupp, M., Biegler, F., Ziehe, A., Tkatchenko, A., Lilienfeld, A. V., & Müller, K.- R. Learning invariant representations of molecules for atomization energy prediction. In *Advances in Neural Information Processing Systems* 440–448 (2012)
27. Honrao, S., Anthonio, B. E., Ramanathan, R., Gabriel, J. J. & Hennig, R. G. Machine learning of ab-initio energy landscapes for crystal structure predictions. *Comput. Mater. Sci.* **158**, 414 (2019).
28. Honrao, S. J., Xie, S. R. & Hennig, R. G. Augmenting machine learning of energy landscapes with local structural information. *J. Appl. Phys.* **128**, 085101 (2020).
29. Faber, F., Lindmaa, A., von Lilienfeld, O. A. & Armiento, R. Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem.* **115**, 1094 (2015).
30. Wu, H. *et al.* Robust FCC solute diffusion predictions from ab-initio machine learning methods. *Comput. Mater. Sci.* **134**, 160 (2017).
31. Lee, J., Seko, A., Shitara, K., Nakayama, K. & Tanaka, I. Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques. *Phys. Rev. B* **93**, 115104 (2016).
32. Pilania, G. *et al.* Machine learning bandgaps of double perovskites. *Sci. Rep.* **6**, 19375 (2016).
33. Abraham, K. Prospects and limits of energy storage in batteries. *J. Phys. Chem. Lett.* **6**, 830 (2015).
34. Armand, M., Endres, F., MacFarlane, D. R., Ohno, H., & Scrosati, B. Ionic-liquid materials for the electrochemical challenges of the future. Materials for sustainable energy: a collection of peer-reviewed research and review articles from Nature Publishing Group 129–137 (2011).
35. Giffin, G. A. Ionic liquid-based electrolytes for beyond lithium battery technologies. *J. Mater. Chem. A* **4**, 13378 (2016).
36. Jónsson, H., Mills, G., & Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions (1998).
37. Meng, Y. S. & Arroyo-de Dompablo, M. E. First principles computational materials design for energy storage materials in lithium ion batteries. *Energy Environ. Sci.* **2**, 589 (2009).
38. Xiao, R., Li, H. & Chen, L. High-throughput design and optimization of fast lithium ion conductors by the combination of bond-valence method and density functional theory. *Sci. Rep.* **5**, 1 (2015a).
39. Anurova, N. *et al.* Migration maps of Li+ cations in oxygen-containing compounds. *Solid State Ionics* **179**, 2248 (2008).

40. Polyakov, V. Visualization of conduction channels and the dynamics of ion transport in superionic conductors. *Phys. Solid State* **43**, 655 (2001).
41. Adams, S. & Swenson, J. Pathway models for fast ion conductors by combination of bond valence and reverse Monte Carlo methods. *Solid State Ionics* **154**, 151 (2002).
42. Adams, S. Bond valence analysis of structure-property relationships in solid electrolytes. *J. Power Sources* **159**, 200 (2006).
43. Gao, J. *et al.* Screening possible solid electrolytes by calculating the conduction pathways using Bond Valence method. *Sci. China Phys. Mech. Astron.* **57**, 1526 (2014).
44. Xiao, R., Li, H. & Chen, L. Candidate structures for inorganic lithium solid-state electrolytes identified by high-throughput bond-valence calculations. *J. Materiom.* **1**, 325 (2015b).
45. Chen, H., Wong, L. L. & Adams, S. SoftBV: A software tool for screening the materials genome of inorganic fast ion conductors. *Acta Crystallogr. B* **75**, 18–33 (2019).
46. Nestler, T. *et al.* Combined theoretical approach for identifying battery materials: Al3+ mobility in oxides. *Chem. Mater.* **31**, 737 (2019).
47. Belsky, A., Hellenbrandt, M., Karen, V. L. & Luksch, P. New developments in the Inorganic Crystal Structure Database (ICSD): Accessibility in support of materials research and design. *Acta Crystallogr. B* **58**, 364 (2002).
48. He, B. *et al.* CAVD, towards better characterization of void space for ionic transport analysis. *Sci. Data* **7**, 1–13 (2020a).
49. He, B. *et al.* High-throughput screening platform for solid electrolytes combining hierarchical ion-transport prediction algorithms. *Sci. Data* **7**, 1–14 (2020b).
50. Zhang, L. *et al.* A database of ionic transport characteristics for over 29,000 inorganic compounds. *Adv. Funct. Mater.* **30**, 2003087 (2020).
51. He, X. *et al.* Crystal structural framework of lithium super-ionic conductors. *Adv. Energy Mater.* **9**, 1902078 (2019).
52. Richards, W. D., Miara, L. J., Wang, Y., Kim, J. C. & Ceder, G. Interface stability in solid-state batteries. *Chem. Mater.* **28**, 266 (2016).
53. Zhu, Y., He, X. & Mo, Y. Origin of outstanding stability in the lithium solid electrolyte materials: Insights from thermodynamic analyses based on first-principles calculations. *ACS Appl. Mater. Interfaces* **7**, 23685 (2015).
54. Xiao, Y., Miara, L. J., Wang, Y. & Ceder, G. Computational screening of cathode coatings for solid-state batteries. *Joule* **3**, 1252 (2019).
55. Aykol, M. *et al.* High-throughput computational design of cathode coatings for Li-ion batteries. *Nat. Commun.* **7**, 1 (2016).
56. Nolan, A. M., Liu, Y. & Mo, Y. Solid-state chemistries stable with high-energy cathodes for lithium-ion batteries. *ACS Energy Lett.* **4**, 2444 (2019).
57. Zhu, Y., He, X. & Mo, Y. Strategies based on nitride materials chemistry to stabilize Li metal anode. *Adv. Sci.* **4**, 1600517 (2017).
58. Tian, Y. *et al.* Promises and challenges of next-generation Beyond Li-ion batteries for electric vehicles and grid decarbonization. *Chem. Rev.* **12**, 1623–1669 (2020).
59. Sendek, A. D. *et al.* Holistic computational structure screening of more than 12000 candidates for solid lithium-ion conductor materials. *Energy Environ. Sci.* **10**, 306 (2017).
60. Sendek, A. D. *et al.* Machine learning-assisted discovery of solid Li-ion conducting materials. *Chem. Mater.* **31**, 342 (2018).
61. Thangadurai, V., Narayanan, S. & Pinzaru, D. Garnet-type solid-state fast Li ion conductors for Li batteries: Critical review. *Chem. Soc. Rev.* **43**, 4714 (2014).
62. Adams, S. Relationship between bond valence and bond softness of alkali halides and chalcogenides. *Acta Crystallogr. B* **57**, 278 (2001).
63. Chen, H. & Adams, S. Bond softness sensitive bond-valence parameters for crystal structure plausibility tests. *IUCrJ* **4**, 614 (2017).
64. Ceder, G., Ong, S. P. & Wang, Y. Predictive modeling and design rules for solid electrolytes. *Mrs Bull.* **43**, 746 (2018).
65. Nishitani, Y., Adams, S., Ichikawa, K. & Tsujita, T. Evaluation of magnesium ion migration in inorganic oxides by the bond valence site energy method. *Solid State Ionics* **315**, 111 (2018).
66. Ong, S. P., Wang, L., Kang, B. & Ceder, G. Li–Fe–P–O2 phase diagram from first principles calculations. *Chem. Mater.* **20**, 1798 (2008).
67. Aydinol, M., Kohan, A., Ceder, G., Cho, K. & Joannopoulos, J. Ab initio study of lithium intercalation in metal oxides and metal dichalcogenides. *Phys. Rev. B* **56**, 1354 (1997).
68. Chan, M. & Ceder, G. Efficient band gap prediction for solids. *Phys. Rev. Lett.* **105**, 196403 (2010).
69. Liu, Z. *et al.* Anomalous high ionic conductivity of nanoporous β-Li3PS4. *J. Am. Chem. Soc.* **135**, 975 (2013).
70. Seino, Y., Ota, T., Takada, K., Hayashi, A. & Tatsumisago, M. A sulphide lithium super ion conductor is superior to liquid ion conductors for use in rechargeable batteries. *Energy Environ. Sci.* **7**, 627 (2014).
71. Whiteley, J. M., Woo, J. H., Hu, E., Nam, K.-W. & Lee, S.-H. Empowering the lithium metal battery through a silicon-based superionic conductor. *J. Electrochem. Soc.* **161**, A1812 (2014).
72. Boulineau, S., Courty, M., Tarascon, J.-M. & Viallet, V. Mechanochemical synthesis of Li-argyrodite Li6PS5X (X= Cl, Br, I) as sulfur-based solid electrolytes for all solid state batteries application. *Solid State Ionics* **221**, 1 (2012).
73. Iddir, H. & Curtiss, L. A. Li ion diffusion mechanisms in bulk monoclinic Li2CO3 crystals from density functional studies. *J. Phys. Chem. C* **114**, 20903 (2010).
74. Shi, S., Qi, Y., Li, H. & Hector, L. G. Jr. Defect thermodynamics and diffusion mechanisms in Li2CO3 and implications for the solid electrolyte interphase in Li-ion batteries. *J. Phys. Chem. C* **117**, 8579 (2013).
75. Guo, R. & Gallant, B. M. Li2O solid electrolyte interphase: Probing transport properties at the chemical potential of lithium. *Chem. Mater.* **32**, 5525 (2020).
76. Liu, F. *et al.* A mixed lithium-ion conductive Li2S/Li2Se protection layer for stable lithium metal anode. *Adv. Funct. Mater.* **30**, 2001607 (2020).
77. Zou, Y. & Petric, A. Structure and conductivity of zirconium-doped polycrystalline lithium yttrium oxide. *Mater. Res. Bull.* **28**, 1169 (1993).
78. Zaiß, T., Ortner, M., Murugan, R. & Weppner, W. Fast ionic conduction in cubic hafnium garnet Li 7 La 3 Hf 2 O 12. *Ionics* **16**, 855 (2010).
79. Breiman, L. Random forests. *Mach. Learn.* **45**, 5 (2001).
80. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
81. Friedman, J. H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **38**, 367 (2002).
82. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825 (2011).
83. Ward, L. *et al.* Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B* **96**, 024104 (2017).
84. Willems, T. F., Rycroft, C. H., Kazi, M., Meza, J. C. & Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **149**, 134 (2012).
85. Lundberg, S., Lee, S. -I. A unified approach to interpreting model predictions (2017). arXiv preprint arXiv:1705.07874.
86. Parr, T. P., & Turgutlu, K. Feature importances for scikit-learn machine learning models, https://github.com/parrt/random-forest-importances.
87. Morgan, D., Van der Ven, A. & Ceder, G. Li conductivity in Li x MPO 4 (M= Mn, Fe Co, Ni) olivine materials. *Electrochem. Solid State Let.* **7**, A30 (2003).

88. Ouyang, C., Shi, S., Wang, Z., Huang, X. & Chen, L. First-principles study of Li ion diffusion in LiFePO 4. *Phys. Rev. B* **69**, 104303 (2004).
89. Lundberg, S. M. *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749 (2018).
90. Goodall, R. E., & Lee, A. A. Predicting materials properties without crystal structure: Deep representation learning from stoichiometry (2019). arXiv preprint arXiv:1910.00617.

Author contributions

S.H., B.R., S.K., A.O., and J.L. conceived the study. S.H. and B.R. computed the migration barriers and electrochemical windows, respectively. S.H., X.Y., and B.R. generated features and trained ML models. All authors contributed to the discussion and interpretation of results. J.L., A.O., and M.S. supervised the work and provided valuable feedback. S.H. and B.R. wrote the main manuscript. All authors reviewed the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-94275-5.

Correspondence and requests for materials should be addressed to B.R. or J.W.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.