Accelerated design and discovery of perovskites with high conductivity for energy applications through machine learning

Pikee Priya\textsuperscript{1,2} and N. R. Aluru\textsuperscript{1,2}

We use machine learning tools for the design and discovery of ABO\textsubscript{3}-type perovskite oxides for various energy applications, using over 7000 data points from the literature. We demonstrate a robust learning framework for efficient and accurate prediction of total conductivity of perovskites and their classification based on the type of charge carrier at different conditions of temperature and environment. After evaluating a set of >100 features, we identify average ionic radius, minimum electronegativity, minimum atomic mass, minimum formation energy of oxides for all B-site, and B-site dopant ions of the perovskite as the crucial and relevant predictors for determining conductivity and the type of charge carriers. The models are validated by predicting the conductivity of compounds absent in the training set. We screen 1793 undoped and 95,832 A-site and B-site doped perovskites to report the perovskites with high conductivities, which can be used for different energy applications, depending on the type of the charge carriers.

npj Computational Materials (2021) 7:90; https://doi.org/10.1038/s41524-021-00551-3

INTRODUCTION

Advances in technology and science rely on the screening and discovery of high-performance and novel functional materials. Discovery of materials for alternative green energy technologies have become a necessity to reduce the reliance on fossil fuels and decrease the carbon footprint\textsuperscript{1,2}. Green energy technologies like the fuel and electrolyzer cells, alkali-ion batteries, gas (e.g., carbon dioxide and oxygen) sensors, gas separation membranes, membrane reactors, and solar cells, all rely on total conductivity be it of the type ionic, electronic, or mixed\textsuperscript{3-6}. Solid oxide fuel/electrolyzer cells (SOFCs/SOECs) use materials of different charge carrier types as electrodes (mixed conductors) or electrolyte (ionic conductors) for conversion of chemical energy to electrical energy and vice versa\textsuperscript{7-11}. Oxygen separation membranes mostly make use of mixed ionic/electronic conduction to compensate for charge differences due to ionic transfer from higher partial pressures to low partial pressures of oxygen, through electronic transport\textsuperscript{4}. Dense proton-conducting membranes are used in membrane reactors for dehydrogenation of gaseous hydrocarbons to produce clean hydrogen fuel\textsuperscript{12,13}. However, most of these applications currently are limited to high temperatures (800–1000 °C), where the kinetics of various chemical reactions and transport of charge carriers are relatively fast. However, high temperatures may sometimes lead to high polarization resistances, as shown by Barfod et al.\textsuperscript{14}, and reducing the polarization resistance is also an active area of research. Research efforts over the years have been directed toward extending these technologies for low temperature applications for wider applicability and portability\textsuperscript{15-22}, as seen in Fig. 1a, which shows the state of the art conductivities of different solid oxides. Over the years, the general trend for progress has been diagonally upward toward regions of higher conductivity and lower temperature. This has been made possible through extensive experiments by researchers. The use of trial-and-error experimentation for discovery and characterization of advanced materials can be very time-consuming, expensive, and inefficient. Statistical data-driven methods can accelerate this process through high-throughput screening of materials.

In recent work, high-throughput computational and experimental techniques were employed for discovery and design of novel materials for energy applications\textsuperscript{23-25}. For example, high-throughput density functional theory (DFT) methods were used to screen: perovskite oxides for cathode materials\textsuperscript{26} or materials for anodes of SOFCs\textsuperscript{27}, perovskite oxides for thermochemical splitting of water\textsuperscript{28}, materials for electrocatalytic hydrogen evolution\textsuperscript{29,30} and oxygen reduction\textsuperscript{31}, and fast Li-ion conductors\textsuperscript{32} and electrodes\textsuperscript{33} for Li-ion batteries for energy storage. Such studies also led to the significant data on computed and experimental properties of solid oxides. An active area of current research is to develop new methods for data extraction and establishing trends from existing datasets in scientific literature\textsuperscript{34-36}. For instance, data-intensive machine learning (ML) techniques are currently widely used in materials science\textsuperscript{37}. Data science-based methods have already been used in cheminformatics\textsuperscript{38-40}, pattern recognition\textsuperscript{41-43}, event forecasting\textsuperscript{44-46}, decision making\textsuperscript{47-50}, etc. The availability of data on properties of interest for candidates within a well-defined chemical space can assist statistical training and prediction approaches leading to design of new compounds. Some recent examples of materials’ property predictions using this approach include predictions of molecular\textsuperscript{51,52} and periodic systems’ properties\textsuperscript{53-56}, transition states\textsuperscript{57}, structure classifications\textsuperscript{58-60}, dielectric properties\textsuperscript{61}, and predictions of bandgaps\textsuperscript{52}. These existing material properties in open databases along with published experimental data provide an opportunity for efficient design and discovery of advanced materials, with enhanced functionalities.

The ability to identify solid oxide materials rapidly and accurately, with high conductivities is of much interest for a range of applications that might either require oxide materials for metallic, semiconductor, or resistive properties to name a few. It is very difficult to estimate conductivities without direct experiments using multi-probe equipment in a furnace for direct current
conductivities or alternating current impedance spectroscopy. Estimating the bandgap through computational approaches might be useful. However, more advanced quantum calculations, such as the GW method or those using hybrid functionals, are computationally expensive, and are thus inefficient for screening of materials based on bandgaps. Here, we develop a ML model, where one can use available attributes (also referred to as features) in open literature, of a material to directly predict conductivities in an efficient and accurate manner.

In this paper, we aim to build a validated statistical learning model for solid oxide perovskites using data from experimental observations of total conductivity, with the type of the majority charge carrier and material properties from open materials databases. The goal is to create a framework for fast exploration of the multidimensional chemical space of perovskites for applications based on total conductivity and carrier type. The perovskite structure is represented by the chemical formula ABO3, where the A cations are generally of larger radii and have a 12-fold

Fig. 1 Motivation and methodology. a The progress in solid oxide conductivities over decades moving diagonally up toward higher conductivities at lower temperatures. The data has been taken from references. We aim to accelerate this process through machine learning (ML). b The flow chart for the methodology adopted in this work. Total conductivity and charge carrier data from literature and 111 features from open databases are used to train a regressor ML model for total conductivity, to screen perovskites for high conductivities. These are then tested for stability based on energy above hull and tolerance factor. The stable perovskites are then classified for majority charge carrier using the charge carrier data from literature and 112 features (+ total conductivity), resulting in the prediction of new perovskite chemistries for different charge carriers. c Periodic table (adapted) showing the complexity of screened perovskite (ABO3) chemistries for conductivity. All the A, B, and M (which is A- or B-site dopant) elements considered are shown. The elements have been chosen depending on the stability for AO and A2O3 type oxides and BO2 and B2O3 type oxides for type I (AO + BO2) and type II (A2O3 + B2O3) perovskites with M2O3 or MO type dopants, respectively.
coordination, while the relatively smaller B metal ions occupy sixfold coordinated positions in the oxygen octahedra. The A-site ions typically have +2 or +3 nominal charge states, while the charge state of the B-site cations is +4 or +3 for charge neutrality.

To scan the entire composition space of perovskites, we considered two classes depending on the charge states of the A- and B-sites: type I, which is defined as AO + BO2 perovskite and type II, which is defined as A2O3 + B2O3 perovskite. All the thermodynamically stable AO/MO, BO2, and A2O3/B2O3/M2O3 (M being the dopant for the A- or B-site) type oxides with energy above hull of 0 in the Materials Project database were used for the predictions. The elements in the periodic table showing the complexity of the perovskite chemistry, that were screened for A, B, and M are illustrated in Fig. 1c. The methodology of screening undoped and doped perovskites with high conductivity through ML regression and classifying them based on charge carriers through ML classification is illustrated in Fig. 1b, and discussed in detail in “Methods”.

RESULTS AND DISCUSSION

The important chemical features from the ML regressor model turned out to be the minimum electronegativity and the average ionic radius of B-site ions, while for the classifier model they were minimum atomic mass of B-site ions and minimum formation energy of B-site oxides. The variation of the conductivity and ionic radius of B-site ions, while for the classiﬁer model they were the minimum electronegativity of B-site ions and lower average B-site ion size. All the proton-conducting perovskites (which are of interest for low temperature conductivities) have relatively high B-site atomic mass or lower formation energy of B-site oxide. The range of conductivities for each charge carrier is shown in the inset.

The conductivity of the 7230 different perovskites considered from literature with respect to the most important features from the XG-Boost regression model showing higher conductivities for higher minimum electronegativity of B-site ions and lower average B-site ion size. The type of charge carrier (proton, oxygen, mixed protonic/electronic, mixed oxide/electronic, mixed protonic/oxygen, and electronic conductors) with respect to the most important features from the XG-Boost classifier model, showing that the lower B size atomic mass and higher minimum formation energy of B-site oxides regions are mostly electronic or mixed oxide/electronic conductors. All the proton-conducting perovskites (which are of interest for low temperature conductivities) have relatively high B-site atomic mass or lower formation energy of B-site oxide. The range of conductivities for each charge carrier is shown in the inset.

Fig. 2 Important features for the regression and classiﬁcation ML models. a The conductivity of the 7230 different perovskites considered from literature with respect to the most important features from the XG-Boost regression model showing higher conductivities for higher minimum electronegativity of B-site ions and lower average B-site ion size. b The type of charge carrier (proton, oxygen, mixed protonic/electronic, mixed oxide/electronic, mixed protonic/oxygen, and electronic conductors) with respect to the most important features from the XG-Boost classifier model, showing that the lower B size atomic mass and higher minimum formation energy of B-site oxides regions are mostly electronic or mixed oxide/electronic conductors. All the proton-conducting perovskites (which are of interest for low temperature conductivities) have relatively high B-site atomic mass or lower formation energy of B-site oxide. The range of conductivities for each charge carrier is shown in the inset.

We observe both semiconductor and metallic electrical behavior among the screened perovskites from the ML model as seen in Fig. 3a, b, showing total conductivities at 400 and 600 °C for type I, and corresponding Fig. 3c, d for type II perovskites. As seen from the classiﬁcation maps in Fig. 4, at three different atmospheres of wet (3% water vapor) H2, wet (3% water vapor) air, and dry O2, we observe atmosphere dependent change in charge carriers—for example, from ionic conduction (mostly protonic in wet atmospheres due to the reaction H2O + VO2+ + O2 → 2HO) toward more mixed electronic or purely electronic type behavior with increase in O2 partial pressures as O2 fills in the oxide vacancies creating holes as per the reaction: 1/2O2 → O2+ + 2e, leading to increased electronic conductivity. Interestingly, many lanthanide group perovskites of type II become oxide and proton conducting at high oxygen atmospheres as seen in Fig. 4f, which have not been investigated because of the rarity of these elements. For the electronic conductivities. The list of features with their feature importance from the XG-Boost models are listed in Supplementary Notes of the Supplementary Information.

Screening of stable pure perovskites

The conductivity of perovskites depends on the chemistry, off-stoichiometry, as well as the effect of environment, such as temperature and atmosphere. In perovskites, the conductivity is the sum of electronic (negatively charged electrons and positively charged holes) and ionic conductivities, which may be due to the movement of protons interstitially attaching to O atoms or oxide ions via substitutions of the O sites. The total conductivity is \( \sigma_{\text{total}} = \sigma_{\text{e}} + \sigma_{\text{h}} + \sigma_{\text{o}} + \sigma_{\text{h,ox}} \), where each term represents conductivity due to protons, oxide ions, electrons, and holes, respectively. One conductivity term may prevail over the other depending on the majority charge carrier. Also, depending on the effect of temperature, the electrical behavior of materials can either be metallic if the conductivity decreases with temperature due to the decrease in the mean free path of electrons enhanced by increased defects at high temperatures or semiconductor type if the conductivity increases with temperature, where the majority charge carrier can be ions, which diffuse faster at higher temperatures.

We observe both semiconductor and metallic electrical behavior among the screened perovskites from the ML model as seen in Fig. 3a, b, showing total conductivities at 400 and 600 °C for type I, and corresponding Fig. 3c, d for type II perovskites. As seen from the classification maps in Fig. 4, at three different atmospheres of wet (3% water vapor) H2, wet (3% water vapor) air, and dry O2, we observe atmosphere dependent change in charge carriers—for example, from ionic conduction (mostly protonic in wet atmospheres due to the reaction H2O + VO2+ + O2 → 2HO) toward more mixed electronic or purely electronic type behavior with increase in O2 partial pressures as O2 fills in the oxide vacancies creating holes as per the reaction: 1/2O2 → O2+ + 2e, leading to increased electronic conductivity. Interestingly, many lanthanide group perovskites of type II become oxide and proton conducting at high oxygen atmospheres as seen in Fig. 4f, which have not been investigated because of the rarity of these elements. For the electronic conductivities. The list of features with their feature importance from the XG-Boost models are listed in Supplementary Notes of the Supplementary Information.
conduction, we do not distinguish between p-type or n-type electronic conductivity, which is not the primary aim of this work. The trends of effect of temperature and atmosphere are predicted by the XG-Boost models reasonably well.

One of the reasons we developed these ML models extracting the data from literature was to see if we could design materials for low temperature applications. From Fig. 3a, which are conductivity maps for screened type I perovskites at 400 °C, we see that the conductivities of some perovskites like manganates, chromates, cobaltates, and germanates (columns 4–7 in Fig. 3a) are high, which are mostly classified as mixed oxide/electronic and can be used as cathodes/anodes for SOFCs, oxygen separation membranes or other energy applications. Also, there are other perovskites like osmates, iridates, platinates, plumbates, and bismuthates (columns 23–27 in Fig. 3a), which have high conductivities at 400 °C and are classified mostly as proton or mixed protonic/electronic conducting oxides in wet H2/air conditions (corresponding Fig. 4a, b), which can be used for hydrogen separation membranes or electrodes for fuel cells again keeping in mind the stability and electrocatalytic behavior (water splitting) for fuel cell anodes. However, many of these oxides become electronic conductors at very high oxygen partial pressures, as seen in corresponding Fig. 4c. Figure 4a–c have the majority charge carrier types plotted for the same position of the elements as in Fig. 3a, b.

As thermally activated proton diffusion is much easier than oxide ion diffusion due to the small size of the protons, which diffuse interstitially, proton conductors are more likely to provide a solution to low temperature applications. We find many viable substitutes to BaZrO3 and BaCeO3, which are some of the best-known proton-conducting perovskite systems till date.69,70 Niobates, stannates, hafnates, and thorates (columns 10, 16, 21, and 29 in Fig. 3a) are some systems, which are already known to be proton conductors, and have been predicted to have higher conductivities and can perform better than zirconates (column 9 in Fig. 3a) at low temperatures. We also predict palladates (column 15 in Figs. 3a and 4a) as proton conductors, which have not been studied much in literature.

Among the type II perovskites, from Fig. 3c, we find that antimonates (column 20), manganates (column 27), chromates (column 19), vanadates (column 24), titanates (column 29), ferrates (column 22), bismuthates (column 21), and gallates (column 23) in the decreasing order are among the high conductivity perovskites at 400 °C. While gallates are known to be good oxide conductors, manganates and chromates71,72 are reported as mixed oxide/electronic conductors. Some of the ferrates (column 22), titanates

![Image](https://example.com/image.png)
antimonates (column 20), and bismuthates (column 21) are oxide ion conductors, while others along with vanadates \(^73,74\) (column 24) are electronic conductors as also predicted by our classification model (see Fig. 4d, e). The conductivity of type II (Fig. 3c, d) perovskites in general is found to be lower than that of type I (Fig. 3a, b).

Stability of perovskites is of major importance when considering their feasibility for energy applications through high-throughput screening. The thermodynamic stability has been theoretically dealt with by many researchers in the past. The stability of perovskites under oxygen reduction reaction (ORR) environments has been reported by Jacobs et al.\(^26\) for use as cathodes in SOFCs. Similarly, high-throughput screening of stability of perovskites for electrocatalytic water splitting has been performed in a previous work by Emery et al.\(^28\). Emery et al.\(^28\) suggests two criteria for the stability of perovskites: (a) the energy above hull calculated through DFT, which determines whether the composition is energetically at its minimum and possible, and (b) the tolerance factor of a perovskite, which also determines its crystal structure. Based on this, we have chosen the stable perovskites using the energy above hull <0.05 eV (The estimated error of high-throughput DFT calculations of heat of formation for OQMD is 0.096 eV atom\(^{-1}\)) compared to experiments, while the stability criteria chosen by their group is <0.025 eV atom\(^{-1}\) at 0 K and tolerance factors between 0.7 and 1.1 at 300 K. Energy above hull seems to be a more conservative determination of stability than tolerance factors. The charts shown in Supplementary Fig. 3 in the Supplementary Information give us an idea of the stable perovskites based on the two factors. These stability criteria exclude the environmental effects in the operating environments of different energy devices, which might be oxidizing or reducing. The reducing atmosphere as on the fuel side of a SOFC, might make the perovskites prone to reduction and at extreme circumstances even formation of hydrides\(^75\). Some perovskites might also split the water electrochemically\(^28\). In the oxidizing atmosphere as on the air side, the A-site cations in the perovskites may oxidize and segregate on to the surface\(^76,77\).

Apart from screening the perovskites based on temperature and atmosphere, we also screened the best performing proton, oxide, mixed oxide/electronic, and mixed protonic/electronic perovskites of type I and type II based on the lower activation energy for conductivity and higher conductivities at 400 °C in wet air conditions, as shown in Fig. 5. Figure 5 shows the results for all the stable perovskites. While screening them for various applications, we make sure that they have a lower activation energy preferably positive for ionic conduction. To make sure our predictions are accurate, we compared the activation energies from literature to our predicted values. The segmented linear sections for conductivities vs 1/temperature plots were used for the calculations. The comparison of the predicted and experimental values is shown in Fig. 5a and the root mean-squared error (RMSE) for calculation of the activation energies was found to be 0.047 eV.

Among the proton conductors shown in Fig. 5b, BaBiO\(_3\), BaTbO\(_3\), PmEuO\(_3\), CeErO\(_3\), EuPbO\(_3\), and SbTmO\(_3\) have conductivities three to four orders of magnitude above the well-known zirconates (YbZrO\(_3\) is predicted to be the one with highest conductivity among the zirconates). However, these have an activation energy which is negative indicating co-existence of some electronic conduction, also predicted by classification in dry O\(_2\) environment in Fig. 4c making them useful for specific applications needing mixed protonic/electronic conductivities. Also, AgNbO\(_3\) is among one of the best predicted proton conductors and is also known as a photocatalyst for water...
splitting\textsuperscript{78}. We, however, have some not so well-known candidates with good performance like EuNbO\textsubscript{3} and EuSnO\textsubscript{3} with conductivities two orders of magnitude higher than the zirconates. While SrNbO\textsubscript{3} and BaSnO\textsubscript{3} have been investigated for proton diffusion in the past\textsuperscript{79–81}, EuNbO\textsubscript{3} and EuSnO\textsubscript{3} have not yet been studied. Also, studies for lanthanide elements as B-sites in perovskites are scarce. Among these, InLaO\textsubscript{3}, GdCeO\textsubscript{3}, and EuNdO\textsubscript{3} are predicted to be proton conducting, as seen in Fig. 5b.

Among the oxide conductors, we have predicted a jump of two orders of magnitude higher than the gallates, as seen in Fig. 5c. SrTiO\textsubscript{3} have oxide ions as the prominent charge carrier in wet air conditions, which has been extensively studied in the past\textsuperscript{82,83}. SrTiO\textsubscript{3} is a potential candidate for an intermediate temperature (~400 °C) oxide conductor with predicted total conductivity an order of magnitude higher than the gallates. Acceptor doped SrTiO\textsubscript{3} has been reported to show a semiconductor behavior with transference number >0.6 at 400 °C (ref. 82). The conductivity of oxide ions depends not only on the vacancy formation, but also on the vacancy migration energies. SrTiO\textsubscript{3} has a lower oxygen vacancy migration energy of 0.5 eV (ref. 84) than the well-known LaGaO\textsubscript{3} (0.6 eV)\textsuperscript{85}. Although, SrTiO\textsubscript{3} may have a higher conductivity, the oxygen ion transference number might not be optimum. Also, the ML model is unable to predict structural or phase changes in SrTiO\textsubscript{3} (ref. 86) or any other perovskite. Among the gallates, the ScGaO\textsubscript{3} ($\sigma$ = 0.0011 Scm\textsuperscript{-1}) has a total conductivity that is two orders of magnitude higher compared to the more commonly known LSGM ($\sigma$ = 0.00002 Scm\textsuperscript{-1})\textsuperscript{87,88}, which again has not yet been studied, to our knowledge.

**Fig. 5** Prediction of stable perovskites of different charge carrier types. a) Comparison of the predicted activation energies with the experimental values. The RMSE for the prediction is 0.047 eV; screening of stable perovskites with energy above hull <0.05 eV and 0.7 < tolerance factor < 1.1. (type I: darker symbols with black labels for highest conductivities at 400 °C for a given B-site and type II: lighter with dark blue labels for highest conductivities at 400 °C for a given B-site), b) proton-conducting, c) oxide-conducting, d) protonic/electronic, and e) mixed oxide/electronic perovskites based on conductivities and activation energies. The predictions are made in wet air.
Among the mixed protonic oxides, several new candidates, such as EuGeO$_3$, CaPbO$_3$, YbIrO$_3$, YbOsO$_3$, and EuMoO$_3$ are predicted to have high conductivities, along with the well-known stannates (BaSnO$_3$)$^{90}$, as seen in Fig. 5d. Here, mixed conductors refer to mixed electronic-ionic conductors, not different types of ions. MnSnO$_3$ predicted as a potential mixed protonic/electronic perovskite has previously been used for catalytic reduction of CO$_2$, which involves transfer of both protons and electrons$^{89}$. Among the mixed oxide/electronic perovskites (Fig. 5e), we find some perovskites with lanthanide groups as B-sites, such as SbHoO$_3$ and TbDyO$_3$ perform better or as good as the well-known cobaltates and manganates (SrCoO$_3$, SrMnO$_3$)$^{90,91}$. A complete list of the screened perovskites with predicted conductivities and classification is provided in the Supplementary Datasets.

The studied perovskites have already been reported by many researchers for various energy applications in the past. For example, BaMoO$_3$ has been studied for hydrogen evolution reaction$^{92}$, and SrTiO$_3$ has been reported for low temperature fuel cell electrolyte$^{93}$. Addition of transition metal, Ru to La$_{0.7}$Sr$_{0.3}$CoO enhances the electrocatalytic behavior$^{94}$, indicating ruthenates to be possibly used as oxide or mixed ionic–electronic conductors electrodes for fuel cells$^{95}$. Stannates like BaSnO$_3$ has already been used for SO$_2$ gas$^{96}$, formaldehyde$^{97}$, and liquefied petroleum gas sensor$^{98}$.

**Screening of doped perovskites**

We next use the ML model to screen the A- and B-site doped chemistries, which are stable by considering cases, where the tolerance factors were between 0.7 and 1.1, as well as energy above hull of the undoped perovskite <0.05 eV. We considered a nominal 5% doping of the A- and B-sites to avoid clustering. Many researchers have found no phase segregation at low dopant concentrations (<5%) in perovskites and oxides in general$^{99–102}$. The dopant oxides of type M$_2$O$_3$ for type I and MO for type II are used. In general, type I perovskites have higher conductivities compared to the type II perovskites for all classes of charge carriers as seen in Fig. 6, which is a summary of all the stable perovskites screened. While A-site doping does not increase the maximum total conductivity above the pure perovskites, B-site doping increases it for each of the charge carrier type. We find over 4000 (2926 B-site doped type I and 1151 B-site doped type II perovskites) proton conductors with conductivities above the well-known BZY. Also, there are over 300 candidates (5 undoped and 318 B-site doped) for oxide ion conduction above the well-known ceria nanocomposites for intermediate to low temperature applications. There are also several not so well-known candidates for mixed protonic/oxide electronic conduction like GdAlO$_3$ and doped (Ni$_{0.5}$Sc)CoO$_3$, (Fe$_{0.5}$Sr)BiO$_3$ and Y(Cr$_{0.5}$Ag)$_3$O$_3$ for relevant applications, as seen in Fig. 6. The mixed ionic/electronic conductors, which are often used as cathodes in fuel cells have been screened by Jacobs et al.$^{26}$ based on ORR catalytic activity and stability. We find that the predicted ferrates, cobaltates, and aluminates (BaFe$_{0.875}$Al$_{0.125}$O$_3$, SrCo$_{0.75}$Fe$_{0.25}$O$_3$) for better catalytic activity by Jacobs et al.$^{26}$, might also have high total conductivities contributed by oxide and electronic charge carriers.

To test the validity of the ML model without using the conductivities for all the carrier types, we used the XG-Boost regressor model to predict conductivities of a class of charge carrier, which was excluded from the training set. We then tried to predict the conductivities of the class which was excluded for validation. The comparison of the testing, validation, and cross-validation performance of the model with each excluded charge carrier type is shown in Fig. 7. The corresponding mean absolute errors for the different ML models excluding a class of charge carrier type from testing or validation set, and the mean absolute error for the predicted log(conductivities) of the excluded charge carrier class is also reported in Table 1. The predicted conductivities are a maximum one order of magnitude off for the ionic and ionic–electronic conductors and about two orders of magnitude off for the electronic conductors. Training data from all classes of charge carriers is therefore necessary for making valid predictions for all conductor classes of perovskites.

We have developed a ML model by data mining conductivities for 7230 different perovskite chemistries from published literature under different conditions of temperature and atmosphere. We chose 111 different features related to chemistry, physical, electrical, and mechanical properties of A-, B-, and M- (A/B-site dopant) sites in perovskites and correlated them to total conductivities. We also classified them based on the type of charge carriers, which may either be protonic, oxide, mixed protonic/electronic, mixed oxide/electronic, mixed protonic/oxide, or electronic. The regression and classification were made possible through a XG-Boost regressor and classifier model, respectively. We identify average ionic radius, minimum electronegativity, minimum atomic mass, minimum formation energy of oxides for B-site ions and its dopants, as relevant predictors for determining conductivity and the type of charge carriers. We screened 1793 undoped and 95,832 A and B-site doped AO + BO$_3$ (5% M$_2$O$_3$) and A$_2$O$_3$ + B$_2$O$_3$ (5% MO) perovskites for high conductivities, and...
classified them for majority charge carrier type. Among the proton conductors, the predictions of some potential candidates were four orders of magnitude higher than the well-known BaZr(Y)O₃ system. Eliminating perovskites with negative activation energies, which is indicative of electronic contribution, EuNbO₃ and EuSnO₃ are predicted to be good candidates for low temperature proton-conducting electrolytes. ScGaO₃ has a total conductivity two orders of magnitude higher than LSGM, a well-known oxide conductor. The models are validated by predicting the conductivity of compounds absent in the training set. The study shows that ML models using literature data are computationally inexpensive and effective means of obtaining useful guidance for material design and experimentation.

**METHODS**

**Machine learning regression model for total conductivity**

The crucial elements for any ML problem are the target and the features or descriptors that the target is a function of. For the problem considered here, the total conductivity is the target and we chose 111 different features that adequately characterize the difference in electrical properties, without recourse to a specific hypothesis. These features are related to chemistry, structure, electronic, mechanical, and physical properties calculated from DFT from open databases like the Materials Project or those present in the literature. For reasons of internal consistency and descriptor availability, some calculated properties were also used. A detailed description of the features is provided in Supplementary Notes.

The total conductivity of different perovskites was extracted from literature for 7230 perovskite cases (585 compositions). A wide range of perovskite...
chemistries with varying total conductivities depending on the type of charge carrier at different temperatures and atmospheres were scanned from 118 different journal publications. Data extraction was performed by hand using WebPlot Digitizer. A detailed breakup of the different perovskite cases and compositions for each of the charge carrier is provided in Supplementary Fig. 1a, b in Supplementary Information.

Microstructure, such as the grain size distribution, plays a very important role in determining the conductivity of perovskites. The grain size information was available for only 1124 (out of 7230) perovskite cases and compositions for each of the charge carrier type as reported in the literature. The errors would have been lower if the grain sizes would have been included in the analysis. Although found important, we could not include it due to the lack of enough data in journal papers, which reported the conductivities with the grain size information. The grain size distribution with a mean and standard deviation was even more scarce, so we collected just the average grain size from journal papers which reported them.

With these data in hand, we evaluated the accuracy of several ML methods using 85% and 10% of the data as the training and testing set, respectively, and the remaining 5% for cross-validation. For all the models, we also used five k-fold cross-validation to optimize the hyperparameters. This procedure has a single parameter, denoted as k, that refers to the number of groups that a given data sample is to be split into so that every split dataset serves as a test set for that iteration (carries for k iterations). ML package, scikit-learn was used for the purpose. For the linear regression models, we evaluated various regularization methods (such as LASSO, least absolute shrinkage and selection operator), ridge regression, and elastic net, but regularization worsened the predictive performance. We also found that the support vector machines model provided no improvement over the linear regression models. However, a seven-layered neural network with 100 nodes in each layer and k-nearest neighbor regression with five nearest neighbors, each delivered improvement over these methods. Moreover, we found that the RF algorithm with XG-Boost provided even better predictive performance. RF algorithms operate by randomly selecting a subset of features and constructing decision trees based on the limited data, which are then averaged out for the final prediction. XG-Boost uses a sequence of RF models learning sequentially from the previous model. By combining several of these decision tree models on the data subsets, an accurate prediction of the conductivities could be made without overfitting. A comparison of these methods is provided in the Supplementary Methods (Supplementary Fig. 2). The comparison of the XG-Boost model predictions with the experimental results is shown in Fig. 8 with the x-axis being the experimental observation. The test set RMSE for the XG-Boost model is 0.24 with a coefficient of determination $R^2$ value of 0.987, and the RMSE value is 0.25 with $R^2$ of 0.986 for cross-validation set. The coefficient of determination is given by $R^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$ and RMSE = $\sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$, where $y_i$ is the ground truth, $\hat{y}_i$ is the predicted value and $\bar{y}$ is the average of the sample.

| Excluded class | Mean absolute error test set | Mean absolute error validation set | Mean absolute error excluded class |
|----------------|-----------------------------|-----------------------------------|-----------------------------------|
| H              | 0.097                       | 0.086                             | 0.904                             |
| O              | 0.099                       | 0.110                             | 1.275                             |
| H + E          | 0.124                       | 0.104                             | 1.381                             |
| O + E          | 0.134                       | 0.107                             | 0.648                             |
| O + H          | 0.106                       | 0.105                             | 0.396                             |
| E              | 0.109                       | 0.103                             | 2.535                             |

Table 1. Mean absolute error for the different ML models excluding a class of charge carrier type from training, testing, or validation set. The mean absolute error for the predicted log(conductivities) of the excluded charge carrier class is also shown. The predicted conductivities are maximum one order of magnitude off for the ionic and ionic-electronic conductors and about two orders of magnitude off for the electronic conductors.

![Fig. 8 Performance of the machine learning algorithm.](image)

The comparison of ground truth total conductivities (along the x-axis) with predicted total conductivities for test set (10%) and cross-validation set (5%) with error bars. The root mean-squared error (RMSE) in prediction of the test set is 0.24, while that of the cross-validation set is 0.25.

Table 2. Confusion matrix for the cross-validation set (5% of the total data) for charge carrier prediction.

| Predicted | Actual |      |      |      |      |
|-----------|--------|------|------|------|------|
|           | H      | O    | H + E| O + E| O + H|
| H         | 111    | 0    | 0    | 0    | 0    |
| O         | 0      | 36   | 0    | 1    | 0    |
| H + E     | 0      | 0    | 17   | 0    | 0    |
| O + E     | 0      | 0    | 0    | 88   | 1    |
| O + H     | 0      | 0    | 1    | 0    | 27   |
| E         | 0      | 0    | 0    | 0    | 79   |

Machine learning classification model for charge carrier type

While the total conductivities of all the perovskites could be predicted through regression using the data from all different classes of conductors, it might be useful if we could also predict the type of charge carrier. The charge carriers in perovskites can be electrons, positively charged holes, oxygen vacancies (oxide ions) or protons, or a combination of them. Relying on the chemistry, temperature, and atmosphere (partial pressures of hydrogen, oxygen gases, and water vapor) based classification in literature, we develop a classifier model to group solid oxide perovskites into proton, oxide, mixed protonic/electronic, mixed oxide/electronic, mixed protonic/oxygen, or electronic conductors referred to as H, O, H, O + E, O + H, and E, respectively, in Figs. 2b, 4a–f and 6. Although the physics of conduction is different in these charge carrier types, we assume that the features of the ML model, are able to capture the conductivities and differences in the charge carrier types.

The classification XG-Boost model used the same feature list as conductivity along with total conductivity as an additional feature. Less than 30% of the data reported transference numbers for different perovskites making the exact demarcation of the transference numbers for determining a charge carrier type difficult. We have assigned the majority charge carrier as reported in the literature or the charge carriers whose transference numbers are >0.6. For the XG-Boost classification model, the error for the test set was 1.46% and the error for the cross-validation set was 0.83% for the classification of charge carriers. Also, the XG-Boost classification model was tested for five and ten different random test sets splitting the data into 5/10 parts. The average error based on this method was 1.4% and 1.5% for classification based on five splits and ten splits, respectively, which was higher than the errors using a random set.
for cross-validation. Such discrepancies have also been reported by other researchers\(^{14}\). The comparison of the tested and predicted number of cases for the cross-validation set is provided in Table 2.

The tested and cross-validated ML model was then used to screen all different permutations and combinations of stable oxides for A, B and 5% of A/B-site dopant M sites for the following classes of perovskites: (a) pure perovskites of type I (AO + BO\(_3\)); (b) pure perovskites of type II (A\(_2\)O\(_3\) + B\(_2\)O\(_3\)); (c) A-site doped type I (0.95 AO + 0.025 M\(_2\)O\(_3\) + BO\(_3\)); (d) A-site doped type II (0.475 A\(_2\)O\(_3\) + 0.05 MO + B\(_2\)O\(_3\)); (e) B-site doped type I (AO + 0.95 BO\(_3\) + 0.025 M\(_2\)O\(_3\)); and (f) B-site doped type II (A\(_2\)O\(_3\) + 0.475 B\(_2\)O\(_3\) + 0.05 MO).

**DATA AVAILABILITY**

The data of the training and test sets from the literature are available along with the manuscript. The dataset used for training, testing, and validation, and the corresponding python script to train and validate the XG-Boost model is also provided. The predicted conductivity and charge carrier data for the doped and undoped perovskites is provided in the tables in the dataset accompanying the manuscript. All the datasets are provided at https://figshare.com/s/10b18051e266a4df4f18c.

**CODE AVAILABILITY**

The data generation scripts will be provided on request.

Received: 23 March 2020; Accepted: 10 May 2021; Published online: 10 June 2021

**REFERENCES**

1. Hussein, A. K. Applications of nanotechnology in renewable energies - a comprehensive overview and understanding. Renew. Sustain. Energy Rev. 42, 460–476 (2015).
2. Valero, A., Valero, A., Calvo, G. & Ortego, A. Material bottlenecks in the future development of green technologies. Renew. Sustain. Energy Rev. 93, 178–200 (2018).
3. Mohamad, A. A. Absorbency and conductivity of quasi-solid-state polymer electrolytes for dye-sensitized solar cells: a characterization review. J. Power Sources 329, 57–71 (2016).
4. Sunarso, J., Hashim, S. S., Zhu, N. & Zhou, W. Perovskite oxides applications in high temperature oxygen separation, solid oxide fuel cell and membrane reactor: a review. Prog. Energy Combust. Sci. 61, 57–77 (2017).
5. Jiang, C., Li, H. & Wang, C. Recent progress in solid-state electrolytes for alkalii-ion batteries. Sci. Bull. 62, 1473–1490 (2017).
6. Shao, T. et al. Recent research on strategies to improve ion conduction in alkalii metal-ion batteries. Batter. Supercaps 2, 403–427 (2019).
7. Jiang, S. P. & Chan, S. H. A review of anode materials development in solid oxide fuel cells. J. Mater. Sci. 39, 4405–4439 (2004).
8. Fabbri, E., Pergolesi, D. & Traversa, E. Materials challenges toward proton-conducting oxide fuel cells: a critical review. Chem. Soc. Rev. 39, 4355–4369 (2010).
9. Tespin, E. V. & Kharton, V. V. Electrode materials and reaction mechanisms in solid oxide fuel cells: a brief review: I Electrochemical behavior vs. materials science aspects. J. Solid State Electrochem. 12, 1367–1391 (2008).
10. Ma, L., Priya, P. & Aluru, N. R. A multiscale model for electrochemical reactions in LSCF based solid oxide cells. J. Electrochem. Soc. 165, F1223–F1241 (2018).
11. Priya, P. & Aluru, N. R. A multiscale framework to predict electrochemical characteristics of yttrium doped Barium Zirconate based solid oxide cells. J. Power Sources 481, 229690 (2021).
12. Dong, X., Xin, W., Xu, N. & Li, K. Dense ceramic catalytic membranes and membrane reactors for energy and environmental applications. Chem. Commun. 47, 10886–10902 (2011).
13. Sundsmacher, K., Rikhio-Struckmann, L. K. & Galvita, V. Solid electrolyte membrane reactors: status and trends. Catal. Today 104, 185–199 (2005).
14. Lu, H. J. et al. Error assessment and optimal cross-validation approaches in machine learning applied to impurity diffusion. Comput. Mater. Sci. 169, 109075 (2019).
15. Xia, C. et al. Shaping triple-conducting semiconductor BaCo\(_{0.95}\)Fe\(_{0.05}\)Zr\(_{0.00}\)Y\(_{0.00}\)O\(_{3}\)) into an electrolyte for low-temperature solid oxide fuel cells. Nat. Commun. 10, 1–9 (2019).

16. Lim, Y., Lee, H., Hong, S. & Kim, Y. B. Co-sputtered nanocomposite nickel cermet anode for high-performance low-temperature solid oxide fuel cells. J. Power Sources 412, 160–169 (2019).
17. Miao, L., Hou, J., Gong, Z., Jin, Z. & Liu, W. A high-performance cobalt-free Ruddlesden-Popper phase cathode La\(_{2}\)Sr\(_{0.5}\)Ni\(_{0.5}\)Fe\(_{0.5}\)O\(_{4}\) for low temperature proton-conducting solid oxide fuel cells. Int. J. Hydrog. Energy 44, 7531–7537 (2019).
18. Meng, Y. et al. Novel high ionic conductivity electrolyte membrane based on semiconductor La\(_{0.25}\)Sr\(_{0.75}\)Co\(_{0.5}\)Fe\(_{0.5}\)O\(_{3}\) for low-temperature solid oxide fuel cells. J. Power Sources 421, 33–40 (2019).
19. Batool, R. et al. Structural and electrochemical study of Ba\(_{0.75}\)Cu\(_{0.15}\)Ni\(_{0.10}\)Zn\(_{0.60}\)BO\(_{2}\); (c) A-site doped type I (0.95 AO + 0.025 M\(_2\)O\(_3\) + BO\(_3\)); (d) A-site doped type II (0.475 A\(_2\)O\(_3\) + 0.05 MO + B\(_2\)O\(_3\)); (e) B-site doped type I (AO + 0.95 BO\(_3\) + 0.025 M\(_2\)O\(_3\)); and (f) B-site doped type II (A\(_2\)O\(_3\) + 0.475 B\(_2\)O\(_3\) + 0.05 MO).

**CODE AVAILABILITY**

The data generation scripts will be provided on request.

npj Computational Materials (2021) 90 Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences
ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (USA) under grant 
1545907. This research is part of the Blue Waters sustained-petascale computing 
project, which is supported by the National Science Foundation (awards OCI-0725070 
and ACI-1238993) and the state of Illinois. Assistance from Dr. Debashish Das in 
collecting the data is gratefully acknowledged.

AUTHOR CONTRIBUTIONS

P.P. optimized the model for calculating conductivities and classification of charge 
carriers, did the analysis, and had primary writing responsibilities. N.R.A. guided and 
supervised all aspects, edited the manuscript, and led the project.

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/s41524-021-00551-3.

Correspondence and requests for materials should be addressed to P.P.

Reprints and permission information is available at http://www.nature.com/ 
reprints

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