ARTICLE   OPEN

Check for updates

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

Pikee Priya [ID][1,2 ✉] and N. R. Aluru [ID][1,2]

We use machine learning tools for the design and discovery of ABO_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.

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[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[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[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[4]. Dense proton-conducting membranes are used in membrane reactors for dehydrogenation of gaseous hydrocarbons to produce clean hydrogen fuel[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.[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[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[23–25]. For example, high-throughput density functional theory (DFT) methods were used to screen: perovskite oxides for cathode materials[26] or materials for anodes of SOFCs[27], perovskite oxides for thermochemical splitting of water[28], materials for electrocatalytic hydrogen evolution[29,30] and oxygen reduction[31], and fast Li-ion conductors[32] and electrodes[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[34–36]. For instance, data-intensive machine learning (ML) techniques are currently widely used in materials science[37]. Data science-based methods have already been used in cheminformatics[38–40], pattern recognition[41–43], event forecasting[44–46], decision making[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[51,52] and periodic systems' properties[53–56], transition states[57], structure classifications[58–60], dielectric properties[61], and predictions of bandgaps[62]. 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

[1]Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [2]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ✉email: pikeepriya@gmail.com
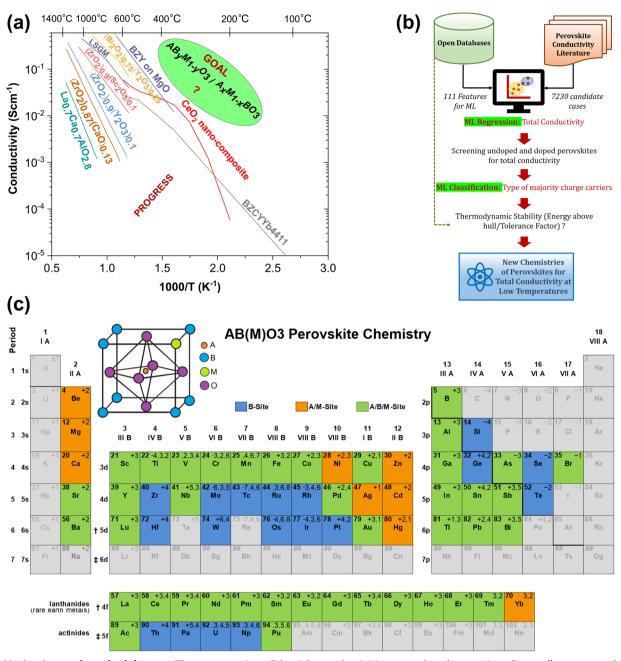
Fig. 1 Motivation and methodology. a The progress in solid oxide conductivities over decades moving diagonally up toward higher conductivities at lower temperatures. The data has been taken from references[69,70,116,117]. We aim to accelerate this process through machine learning (ML). **b** The flow 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[118] (adapted) showing the complexity of screened perovskite (ABO_3) 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 A_2O_3 type oxides and BO_2 and B_2O_3 type oxides for type I (AO + BO_2) and type II (A_2O_3 + B_2O_3) perovskites with M_2O_3 or MO type dopants, respectively.

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[63] or those using hybrid functionals[64] 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 ABO_3, where the A cations are generally of larger radii and have a 12-fold
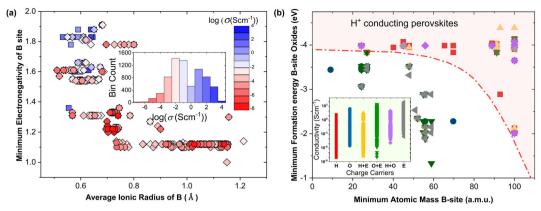
Fig. 2 Important features for the regression and classification ML models. a The conductivity of the 7230 different perovskites considered from literature with respect to the most important features from the XG-Boost regression model showing higher conductivities for higher minimum electronegativity 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/oxide, 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.

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 $AO + BO_2$ perovskite and type II, which is defined as $A_2O_3 + B_2O_3$ perovskite. All the thermodynamically stable AO/MO, BO_2, and $A_2O_3/B_2O_3/M_2O_3$ (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 charge carriers with these chemical features for the perovskites in the data collected from the literature are shown in Fig. 2a, b. Figure 2a shows that, higher electronegativity and smaller B-site radii lead to higher conductivities, which is reasonable as the smaller atoms in the periodic table are more electronegative and electronegativity of the B atoms closer to that of oxygen (electronegativity of oxygen is 3.44) lead to higher metallic characteristics enabling higher conductivities. On the other hand, in Fig. 2b, we see that most of the perovskites with higher atomic mass of B-site and lower formation energy of B-site constituting oxides are proton or mixed proton conducting. We hypothesize that the lower formation energies lead to stronger B–O bonds in the perovskites, which makes oxide ion conduction difficult, making proton conduction preferable. Also, heavier B-site ions have larger radii that may lead to adjustment of positions of oxygen atoms in the BO_6 octahedra in the perovskite structures farther from the central B-site ion creating more free-cell volume, which makes proton diffusion easier and hence the higher proton

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[65] or oxide ions via substitutions of the O sites[66]. The total conductivity is, $\sigma_{total} = \sigma_{H^+} + \sigma_{O^{2-}} + \sigma_{e^-} + \sigma_{h^+}$, 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[67,68] or semiconductor type if the conductivity increases with temperature, where the majority charge carrier can be ions, which diffuse faster at higher temperatures. We screened type I and type II perovskites for higher conductivities at temperatures of 600 and 400 °C and classified them with respect to their majority charge carriers as shown in Figs. 3 and 4.

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) H_2, wet (3% water vapor) air, and dry O_2, we observe atmosphere dependent change in charge carriers—for example, from ionic conduction (mostly protonic in wet atmospheres due to the reaction $H_2O + V_O^{\cdot\cdot} + O_O^x \rightarrow 2OH_O^{\cdot}$) toward more mixed electronic or purely electronic type behavior with increase in O_2 partial pressures as O_2 fills in the oxide vacancies creating holes as per the reaction: $1/2O_2 + V_O^{\cdot\cdot} \rightarrow O_O^x + 2h^{\cdot}$, 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
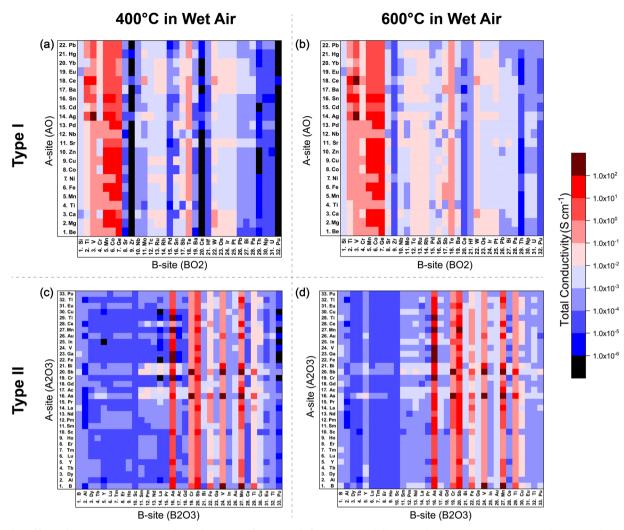
Fig. 3 The effect of temperature on total conductivity of screened (from ML model) perovskites. Type I ($AO + BO_2$), shown in plots **a** and **b**, and type II ($A_2O_3 + B_2O_3$), shown in plots **c** and **d**, perovskites. The conductivities are shown at a temperature of 400 °C (**a**, **c**) and 600°C (**b**, **d**). The increase in conductivity (e.g., indicated by change of color as observed in column 9 for zirconates in **a** and **b**) with temperature indicates ionic conduction, while a decrease in conductivity (e.g., indicated by change in color in the lanthanide-based perovskites in rows 11–13 and columns 6–9), with temperature mostly implies electronic contribution to conductivity.

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 H_2/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 $BaZrO_3$ and $BaCeO_3$, 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 chromates[71,72] are reported as mixed oxide/electronic conductors. Some of the ferrates (column 22), titanates
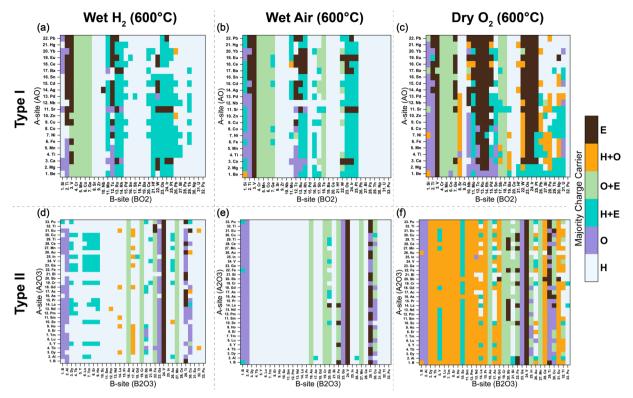
Fig. 4 The effect of atmosphere on total conductivity of screened (from ML model) perovskites. Type I (AO + BO$_2$), shown in plots **a–c**, and type II (A$_2$O$_3$ + A$_2$O$_3$), shown in plots **d–f**, perovskites. The conductivities are shown at atmospheres of wet (3% H$_2$O) H$_2$ (**a, d**); wet air (**b, e**); and pure O$_2$ (**c, f**). The white region refers to proton conductors, purple to oxide ion conductors, dark green to mixed protonic/electronic, light green to mixed oxide/electronic, orange to mixed protonic/oxide, and brown to purely electronic conductors. There is a transition from protonic to electronic conductivity indicated by increase in electronic (brown) and mixed protonic and mixed oxide (green) colors due to inclusion of holes in presence of O$_2$.

(column 29), 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}[53] compared to experiments, while the stability criteria chosen by their group is <0.025 eV atom^{-1}[28].) 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 (YbZrO3 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
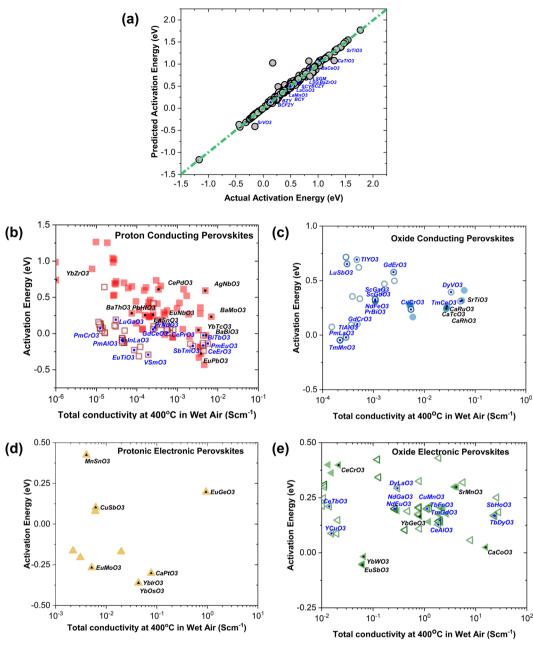
Fig. 5 Prediction of stable perovskites of different charge carrier types. a Comparison of the predicted activation energies with the experimental values. The RMSE for the prediction is 0.047 eV; screening of stable perovskites with energy above hull < 0.05 eV and 0.7 < tolerance factor < 1.1. (type 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.

splitting[78]. We, however, have some not so well-known candidates with good performance like $EuNbO_3$ and $EuSnO_3$ with conductivities two orders of magnitude higher than the zirconates. While $SrNbO_3$ and $BaSnO_3$ have been investigated for proton diffusion in the past[79–81], $EuNbO_3$ and $EuSnO_3$ have not yet been studied. Also, studies for lanthanide elements as B-sites in perovskites are scarce. Among these, $InLaO_3$, $GdCeO_3$, and $EuNdO_3$ 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 in conductivity over the well-known gallates, as seen in Fig. 5c. $SrTiO_3$ have oxide ions as the prominent charge carrier in wet air conditions, which has been extensively studied in the past[82,83]. $SrTiO_3$ 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 $SrTiO_3$ 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. $SrTiO_3$ has a lower oxygen vacancy migration energy of 0.5 eV (ref. [84]) than the well-known $LaGaO_3$ (0.6 eV)[85]. Although, $SrTiO_3$ 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 $SrTiO_3$ (ref. [86]) or any other perovskite. Among the gallates, the $ScGaO_3$ ($\sigma = 0.0011$ Scm^{-1}) has a total conductivity that is two orders of magnitude higher compared to the more commonly known LSGM ($\sigma = 0.00002$ Scm^{-1})[87,88], which again has not yet been studied, to our knowledge.
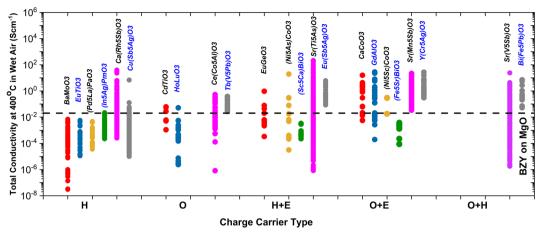
Fig. 6 Summary of all the screened stable perovskites of different charge carriers. Different colored symbols indicate different classes of perovskites: red solid symbols are pure type I (AO + BO$_2$) perovskites, blue are type II (A$_2$O$_3$ + B$_2$O$_3$) perovskites, yellow are the A-site doped type I perovskites, green are A-site doped type II, the purple are the B-site doped type I, and gray are the B-site doped type II perovskites. Type I are doped with M$_2$O$_3$ and type II are doped with MO type oxides. The top candidate for each class has been labeled. The conductivity of BZY at 400 °C is shown by a black line for reference.

Among the mixed protonic oxides, several new candidates, such as EuGeO$_3$, CaPtO$_3$, YbIrO$_3$, YbOsO$_3$, and EuMoO$_3$ are predicted to have high conductivities, along with the well-known stannates (BaSnO$_3$)[80], 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 liquified 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$_5$Sc)CoO$_3$, (Fe$_5$Sr)BiO$_3$ and Y(Cr$_5$Ag)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$_2$ (5% M$_2$O$_3$) and A$_2$O$_3$ + B$_2$O$_3$ (5% MO) perovskites for high conductivities, and
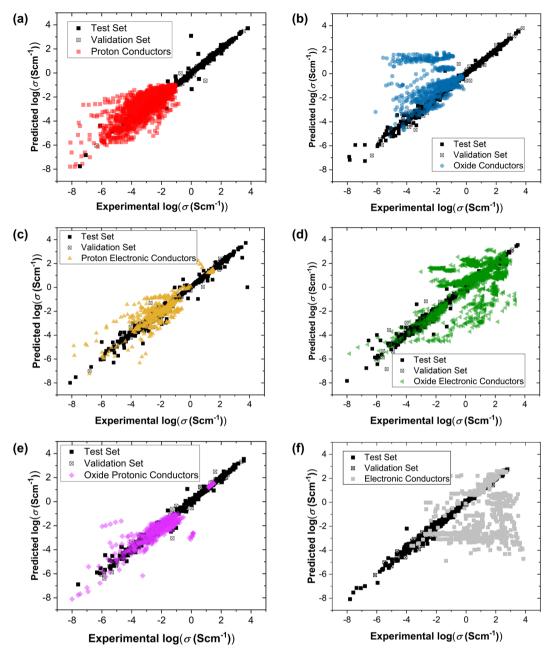
Fig. 7 The comparison of the predicted conductivities from the ML model with the experimental values for different set of data excluding a given class of material. Excluding **a** proton (H), **b** oxide (O), **c** protonic electronic (H + E), **d** oxide electronic (O + E), **e** oxide protonic (O+H), and **f** electronic (E) perovskite conductors. The performance of the model in predicting the excluded class is also shown.

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)O$_3$ system. Eliminating perovskites with negative activation energies, which is indicative of electronic contribution, EuNbO$_3$ and EuSnO$_3$ are predicted to be good candidates for low temperature proton-conducting electrolytes. ScGaO$_3$ 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

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.

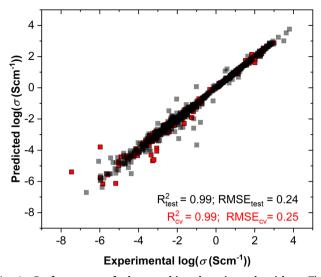| Excluded class	Mean absolute error test set	Mean absolute error validation set	Mean absolute error excluded class
H	0.097	0.086	0.904
O	0.099	0.110	1.275
H + E	0.124	0.104	1.381
O + E	0.134	0.107	0.648
O + H	0.106	0.105	0.396
E	0.109	0.103	2.535

Fig. 8 Performance of the machine learning algorithm. 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.

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

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[103]. 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[104–106], plays a very important role in determining the conductivity of perovskites. The grain size information was available for only 1124 (out of 7230) perovskite candidates in 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 gran 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[107] was used for the purpose. For the linear regression models, we evaluated various regularization methods (such as LASSO[108] (least absolute shrinkage and selection operator), ridge regression[109], and elastic net[110]), but regularization worsened the predictive performance. We also found that the support vector machines[111] model provided no improvement over the linear regression models. However, a seven-layered neural network[112] with 100 nodes in each layer and k-nearest neighbor[113] regression with five nearest neighbors, each delivered improvement over these methods. Moreover, we found that the RF[114] algorithm with XG-Boosting[115] 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-Boosting 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, $R^2 = \frac{\sum_{i=1}^{N}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2}$ and

$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2}$, where y_i is the ground truth, $\widehat{y_i}$ is the predicted value and $\overline{y}$ is the average of the sample.

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/oxide, or electronic conductors referred to as H, O, H + E, 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 have 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 ($AO+BO_2$); (b) pure perovskites of type II ($A_2O_3 + B_2O_3$); (c) A-site doped type I ($0.95\ AO + 0.025\ M_2O_3 + BO_2$); (d) A-site doped type II ($0.475\ A_2O_3 + 0.05\ MO+B_2O_3$); (e) B-site doped type I ($AO + 0.95\ BO_2 + 0.025\ M_2O_3$); and (f) B-site doped type II ($A_2O_3 + 0.475\ B_2O_3 + 0.05\ MO$).

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/10b18051e26fa4d4f18c.

CODE AVAILABILITY

The data generation scripts will be provided on request.

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 alkali-ion batteries. *Sci. Bull.* **62**, 1473–1490 (2017).
6. Shao, T. et al. Recent research on strategies to improve ion conduction in alkali 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. Tsipis, 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**, F1232–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**, 228969 (2021).
12. Dong, X., Jin, W., Xu, N. & Li, K. Dense ceramic catalytic membranes and membrane reactors for energy and environmental applications. *Chem. Commun.* **47**, 10886–10902 (2011).
13. Sundmacher, K., Rihko-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 $BaCo_{0.4}Fe_{0.4}Zr_{0.1}Y_{0.1}O_{3-δ}$ 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 $La_{1.2}Sr_{0.8}Ni_{0.6}Fe_{0.4}O_{4+Δ}$ 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 $La_{0.65}Sr_{0.3}Ce_{0.05}Cr_{0.5}Fe_{0.5}O_{3-Δ}$ for low-temperature solid oxide fuel cells. *J. Power Sources* **421**, 33–40 (2019).
19. Batool, R. et al. Structural and electrochemical study of $Ba_{0.15}Cu_{0.15}Ni_{0.10}Zn_{0.60}$ oxide anode for low temperature solid oxide fuel cell. *J. Alloy. Compd.* **780**, 653–659 (2019).
20. Fan, L., Zhu, B., Su, P. C. & He, C. Nanomaterials and technologies for low temperature solid oxide fuel cells: recent advances, challenges and opportunities. *Nano Energy* **45**, 148–176 (2018).
21. Kim, C. et al. Facile fabrication strategy of highly dense gadolinium-doped ceria/yttria-stabilized zirconia bilayer electrolyte via cold isostatic pressing for low temperature solid oxide fuel cells. *J. Power Sources* **415**, 112–118 (2019).
22. Chen, G. et al. Investigation of layered $Ni_{0.8}Co_{0.15}Al_{0.05}LiO_2$ in electrode for low-temperature solid oxide fuel cells. *Int. J. Hydrog. Energy* **43**, 417–425 (2018).
23. Jain, A., Shin, Y. & Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev.* **1**, 1–13 (2016).
24. Cheng, L. et al. Accelerating electrolyte discovery for energy storage with high-throughput screening. *J. Phys. Chem. Lett.* **6**, 283–291 (2015).
25. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
26. Jacobs, R., Mayeshiba, T., Booske, J. & Morgan, D. Material discovery and design principles for stable, high activity perovskite cathodes for solid oxide fuel cells. *Adv. Energy Mater.* **8**, 1–12 (2018).
27. Strasser, P. et al. High throughput experimental and theoretical predictive screening of materials - a comparative study of search strategies for new fuel cell anode catalysts. *J. Phys. Chem. B* **107**, 11013–11021 (2003).
28. Emery, A. A., Saal, J. E., Kirklin, S., Hegde, V. I. & Wolverton, C. High-throughput computational screening of perovskites for thermochemical water splitting applications. *Chem. Mater.* **28**, 5621–5634 (2016).
29. Bjorketum, M. et al. Screening of electrocatalytic materials for hydrogen evolution. *Phys. Chem. Chem. Phys.* **12**, 10536–10541 (2010).
30. Greeley, J. & Nørskov, J. K. Large-scale, density functional theory-based screening of alloys for hydrogen evolution. *Surf. Sci.* **601**, 1590–1598 (2007).
31. Greeley, J. & Nørskov, J. K. Combinatorial density functional theory-based screening of surface alloys for the oxygen reduction reaction. *J. Phys. Chem. C* **113**, 4932–4939 (2009).
32. 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. *Nat. Publ. Gr.* **5**, 1–11 (2015).
33. Kirklin, S., Chan, M. K. Y., Trahey, L., Thackeray, M. M. & Wolverton, C. High-throughput screening of high-capacity electrodes for hybrid Li-ion – Li – O 2 cells. *Phys. Chem. Chem. Phys.* **16**, 22073–22082 (2014).
34. Baro, E., Degoul, S., Beuscart, R. & Chazard, E. Toward a literature-driven definition of big data in healthcare. *Biomed Res. Int.* **2015**, 1–9 (2015).
35. Manaseer, S., Alawneh, A. R. & Asoudi, D. Big data investment and knowledge integration in academic libraries. *J. Inf. Stud. Technol.* **2019**, 1–8 (2019).
36. Hirschman, L., Park, J. C., Tsujii, J., Wong, L. & Wu, C. H. Accomplishments and challenges in literature data mining for biology. *Bioinformatics* **18**, 1553–1561 (2002).
37. Rajan, K. *Informatics for Materials Science and Engineering: Data-driven Discovery for Accelerated Experimentation and Application*. (Elsevier Science, 2013).
38. Chepelev, L. L. & Dumontier, M. Chemical entity semantic specification: knowledge representation for efficient semantic cheminformatics and facile data integration. *J. Cheminform.* **3**, 1–19 (2011).
39. Walker, T., Grulke, C. M., Pozefsky, D. & Tropsha, A. Chembench: a cheminformatics workbench. *Bioinformatics* **26**, 3000–3001 (2010).
40. Firdaus Begam, B. & Satheesh Kumar, J. A study on cheminformatics and its applications on modern drug discovery. *Procedia Eng.* **38**, 1264–1275 (2012).
41. Kwon, J. S., Lee, J. M. & Kim,. W. Y. Real-time detection of foreign objects using x-ray imaging for dry food manufacturing line. *Proc. Int. Symp. Consum. Electron. ISCE*, 1–4 (2008).
42. Teow, L. N. & Loe, K. F. Robust vision-based features and classification schemes for off-line handwritten digit recognition. *Pattern Recognit.* **35**, 2355–2364 (2002).
43. Lee, H. & Choi, S. PCA+HMM+SVM for EEG pattern classification. In *Proc. of the IEEE seventh international symposiumon signal processing and its applications*, 541–544 (IEEE, 2003).

44. Li, X. et al. Traffic management and forecasting system based on 3D GIS. In *Proc. 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 991–998 (2015).

45. Brémaud, P. J. & Pointin, Y. B. Forecasting heavy rainfall from rain cell motion using radar data. *J. Hydrol.* **142**, 373–389 (1993).

46. Compton, R., Lee, C., Lu, T. C., De Silva, L. & Macy, M. Detecting future social unrest in unprocessed Twitter data: 'Emerging phenomena and big data'. *In Proc. 2013 IEEE International Conference on Intelligence and Security Informatics*, 56–60 (IEEE ISI, 2013).

47. McAfee, A. & Brynjolfsson, E. Big data: the management revolution. *Harv. Bus. Rev.* **90**, 60–66 (2012).

48. Provost, F. & Fawcett, T. Data science and its relationship to big data and data-driven decision making. *Big Data* **1**, 51–59 (2013).

49. Marsh, J. A., Pane, J. F. & Hamilton, L. S. Making sense of data-driven decision making in education: Evidence from recent RAND research. Santa Monica, CA: RAND Corporation, https://www.rand.org/pubs/occasional_papers/OP170.html (2006).

50. Waller, M. A. & Fawcett, S. E. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *J. Bus. Logist.* **34**, 77–84 (2013).

51. Jain, A. et al. The materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002-1–011002-11 (2013).

52. Kirklin, S. et al. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **1**, 15010 (2015).

53. Canepa, P. et al. High-throughput screening of small-molecule adsorption in MOF. *J. Mater. Chem. A* **1**, 13597–13604 (2013).

54. Wu, Y., Lazic, P., Hautier, G., Persson, K. & Ceder, G. First principles high throughput screening of oxynitrides for water-splitting photocatalysts. *Energy Environ. Sci.* **6**, 157–168 (2013).

55. Castelli, I. E. et al. Computational screening of perovskite metal oxides for optimal solar light capture. *Energy Environ. Sci.* **5**, 5814–5819 (2012).

56. Priya, P., Yan, X. & Chaudhuri, S. Study of intermetallics for corrosion and creep resistant microstructure in Mg-RE and Mg-Al-RE alloys through a data-centric high-throughput DFT framework. *Comput. Mater. Sci.* **175**, 109541 (2020).

57. Li, P. et al. High-throughput theoretical optimization of the hydrogen evolution reaction on MXenes by transition metal modification. *J. Mater. Chem. A* **6**, 4271–4278 (2018).

58. Rasmussen, F. A. & Thygesen, K. S. Computational 2D materials database: electronic structure of transition-metal dichalcogenides and oxides. *J. Phys. Chem. C.* **119**, 13169–13183 (2015).

59. Balluff, J., Diekmann, K., Reiss, G. & Meinert, M. High-throughput screening for antiferromagnetic Heusler compounds using density functional theory. *Phys. Rev. Mat.* **034404**, 1–8 (2017).

60. Nazarian, D., Camp, J. S., Chung, Y. G., Snurr, R. Q. & Sholl, D. S. Large-scale refinement of metal-organic framework structures using density functional theory. *Chem. Mater.* **29**, 2521–2528 (2017).

61. Setyawan, W., Gaume, R. M., Lam, S., Feigelson, R. S. & Curtarolo, S. High-throughput combinatorial database of electronic band structures for inorganic scintillator materials. *ACS Comb. Sci.* **13**, 382–390 (2011).

62. Li, W., Walther, C. F. J., Kuc, A. & Heine, T. Density functional theory and beyond for band-gap screening: Performance for transition-metal oxides and dichalcogenides. *J. Chem. Theory Comput.* **9**, 2950–2958 (2013).

63. Hedin, L. New method for calculating the Onee-particle Green's Function with application to the Electron-Gas problem. *Phys. Rev.* **139**, 796 (1965).

64. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).

65. Jing, Y. & Aluru, N. R. The role of A-site ion on proton diffusion in perovskite oxides (ABO3). *J. Power Sources* **445**, 227327 (2020).

66. Muñoz-García, A. B., Ritzmann, A. M., Pavone, M., Keith, J. A. & Carter, E. A. Oxygen transport in perovskite-type solid oxide fuel cell materials: Insights from quantum mechanics. *Acc. Chem. Res.* **47**, 3340–3348 (2014).

67. Broom, T. Lattice defects and the electrical resistivity of metals. *Adv. Phys.* **3**, 26–83 (1954).

68. Milchberg, H. M., Freeman, R. R., Davey, S. C. & More, R. M. Resistivity of a simple metal from room temperature to 106 K. *Phys. Rev. Lett.* **61**, 2364–2367 (1988).

69. Pergolesi, D. et al. High proton conduction in grain-boundary-free yttrium-doped barium zirconate films grown by pulsed laser deposition. *Nat. Mater.* **9**, 846–852 (2010).

70. Choi, S. et al. Exceptional power density and stability at intermediate temperatures in protonic ceramic fuel cells. *Nat. Energy* **3**, 202–210 (2018).

71. Hancock, C. A. & Slater, P. R. Synthesis of silicon doped SrMO₃ (M = Mn, Co): stabilization of the cubic perovskite and enhancement in conductivity. *Dalt. Trans.* **40**, 5599–5603 (2011).

72. Cascos, V., Alonso, J. A. & Fernández-Díaz, M. T. Nb5+-doped SrCoO₃-δ perovskites as potential cathodes for solid-oxide fuel cells. *Materials* **9**, 1–12 (2016).

73. Zeng, Q., Liu, X., Xie, D., Ni, J. & Ni, C. A strontium-free and iron-based oxygen electrode for solid-oxide electrochemical cells (SOCs). *Int. J. Hydrog. Energy* **44**, 15387–15399 (2019).

74. Moyer, J. A., Eaton, C. & Engel-Herbert, R. Highly conductive SrVO₃ as a bottom electrode for functional perovskite oxides. *Adv. Mater.* **25**, 3578–3582 (2013).

75. Vargaftik, M. N. et al. Formation of palladium hydride complexes upon reduction pf P(II) by hydrogen. *Br. Commun.* **34**, 2381–2384 (1986).

76. Onuma, S. et al. Phase stability of La₁₋ₓCaxCrO₃-δ in oxidizing atmosphere. *J. Solid State Chem.* **170**, 68–74 (2003).

77. Koo, B. et al. Sr segregation in perovskite oxides: why it happens and how it exists. *Joule* **2**, 1476–1499 (2018).

78. Lu, Y., Yu, Q., Zhang, F., Li, G. & Zhang, W. Converting Ag nanowire into one-dimensional silver niobate and their enhanced photocatalytic activity. *Appl. Phys. A Mater. Sci. Process* **122**, 1–5 (2016).

79. Bork, N., Bonanos, N., Rossmeisl, J. & Vegge, T. Simple descriptors for proton-conducting perovskites from density functional theory. *Phys. Rev. B Condens. Matter Mater. Phys.* **82**, 1–6 (2010).

80. Bévillon, É., Geneste, G., Chesnaud, A., Wang, Y. & Dezanneau, G. Ab initio study of La-doped BaSnO₃ proton conductor. *Ionics* **14**, 293–301 (2008).

81. Wang, Y., Chesnaud, A., Bevillon, E. & Dezanneau, G. Properties of Y-doped BaSnO₃ proton conductors. *Solid State Ion.* **214**, 45–55 (2012).

82. Noll, F., Münch, W., Denk, I. & Maier, J. SrTiO3 as a prototype of a mixed conductor conductivities, oxygen diffusion and boundary effects. *Solid State Ion.* **86–88**, 711–717 (1996).

83. Hui, S. & Petric, A. Electrical conductivity of yttrium-doped SrTiO₃: influence of transition metal additives. *Mater. Res. Bull.* **37**, 1215–1231 (2002).

84. Zhang, L. et al. Oxygen vacancy diffusion in bulk SrTiO₃ from density functional theory calculations. *Comput. Mater. Sci.* **118**, 309–315 (2016).

85. Lee, J., Ohba, N. & Asahi, R. First-principles prediction of high oxygen-ion conductivity in trilanthanide gallates Ln₃GaO₆. *Sci. Technol. Adv. Mater.* **20**, 144–159 (2019).

86. El-Mellouhi, F., Brothers, E. N., Lucero, M. J., Bulik, I. W. & Scuseria, G. E. Structural phase transitions of the metal oxide perovskites SrTiO₃, LaAlO₃, and LaTiO₃ studied with a screened hybrid functional. *Phys. Rev. B Condens. Matter Mater. Phys.* **87**, 1–11 (2013).

87. Huang, K., Tichy, R. S. & Goodenough, J. B. Superior perovskite oxide-ion conductor; strontium- and magnesium-doped LaGaO₃: I, phase relationships and electrical properties. *J. Am. Ceram. Soc.* **75**, 2565 (1998).

88. Ishihara, T., Matsuda, H. & Takita, Y. Doped LaGaO₃ perovskite type oxide as a new oxide ionic conductor. *J. Am. Ceram. Soc.* **116**, 3801–3803 (1994).

89. Zhang, X., Li, F., Zhang, Y., Bond, A. M. & Zhang, J. Stannate derived bimetallic nanoparticles for electrocatalytic CO₂ reduction. *J. Mater. Chem. A* **6**, 7851–7858 (2018).

90. Demirel, S. et al. Structural, magnetic, electrical and electrochemical properties of SrCoO₂.₅, Sr₉Co₂Mn₅O₂₁ and SrMnO₃ compounds. *Ceram. Int.* **43**, 14818–14826 (2017).

91. Wei, Q. T., Guo, R. S., Wang, F. H. & Li, H. L. Structure and electrical properties of SrCoO₃-δ doped by CeO₂. *J. Mater. Sci.* **40**, 1317–1319 (2005).

92. Xu, X. et al. From scheelite BaMoO₄ to perovskite BaMoO₃: enhanced electrocatalysis toward the hydrogen evolution in alkaline media. *Compos. Part B Eng.* **198**, 108214 (2020).

93. Chen, G. et al. Electrochemical mechanisms of an advanced low-temperature fuel cell with a SrTiO₃ electrolyte. *J. Mater. Chem. A* **7**, 9638–9645 (2019).

94. Georges, S., Parrour, G., Henault, M. & Fouletier, J. Gradual internal reforming of methane: a demonstration. *Solid State Ion.* **177**, 2109–2112 (2006).

95. Takeda, T., Kanno, R., Kawamoto, Y., Takeda, Y. & Yamamoto, O. New cathode materials for solid oxide fuel cells ruthenium pyrochlores and perovskites. *J. Electrochem. Soc.* **147**, 1730 (2000).

96. Marikutsa, A., Rumyantseva, M., Baranchikov, A. & Gaskov, A. Nanocrystalline BaSnO₃ as an alternative gas sensor material: surface reactivity and high sensitivity to SO₂. *Materials* **8**, 6437–6447 (2015).

97. Bhattacharya, A. et al. Highly responsive and selective formaldehyde sensor based on La3+-doped barium stannate microtubes prepared by electrospinning. *J. Mater. Res.* **34**, 2067–2077 (2019).

98. Reddy, C. V. G., Manorama, S. V. & Rao, V. J. Preparation and characterization of barium stannate: application as a liquefied petroleum gas sensor. *J. Mater. Sci. Mater. Electron* **12**, 137–142 (2001).

99. Blasco, J., Stankiewicz, J. & García, J. Phase segregation in the Gd₁₋ₓSrxFeO₃-δ series. *J. Solid State Chem.* **179**, 898–908 (2006).

100. Shirpour, M. et al. Dopant segregation and space charge effects in proton-conducting BaZrO₃ perovskites. *J. Phys. Chem. C.* **116**, 2453–2461 (2012).

101. Phung, N. et al. The doping mechanism of halide perovskite unveiled by alkaline earth metals. *J. Am. Chem. Soc.* **142**, 2364–2374 (2020).

102. Kumar, P. et al. Investigation of phase segregation in Zn₁₋ₓMg xO systems. *Curr. Appl. Phys.* **12**, 1166–1172 (2012).

103. Rohatdi, A. WebPlotDigitizer, 011002-1–011002-11 https://automeris.io/WebPlotDigitizer (2019).

104. Watanabe, K. et al. Oxygen permeation properties of Co-free perovskite-type oxide membranes based on BaFe[sub 1−y]Zr[sub y]O[sub 3−δ]. *J. Electrochem. Soc.* **156**, E81 (2009).

105. Hueso, L. E. et al. Influence of the grain-size and oxygen stoichiometry on magnetic and transport properties of polycrystalline $La_{0.67}Ca_{0.33}MnO_{3 \pm \delta}$ perovskites. *J. Magn. Magn. Mater.* **189**, 321–328 (1998).

106. Haile, S. M., Staneff, G. & Ryu, K. H. Non-stoichiometry, grain boundary transport and chemical stability of proton conducting perovskites. *J. Mater. Sci.* **36**, 1149–1160 (2001).

107. Varoquaux, G. et al. Scikit-learn. *GetMobile Mob. Comput. Commun.* **19**, 29–33 (2015).

108. Tishbirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **58**, 267–288 (1996).

109. Morozov, V. A. Regularization of incorrectly posed problems and chois of regularization parameter. *USSR Comput. Math. Math. Phys.* **6**, 242–251 (1966).

110. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net (Journal of the Royal Statistical Society. Series B: Statistical Methodology (2005) 67 (301-320)). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 768 (2005).

111. Boser, B. E., Guyon, I. M. & Vapnik, V. N. SVM-A training algorithm for optimal margin classifiers.pdf. In *Proc of the 5th Annual ACM Workshop on Computational Learning Theory* (ACM, 1992).

112. Rumelhart, D. E., Widrow, B. & Lehr, M. A. *The Basic Ideas in Neural Networks,* 37 (Communications of the ACM, 1994).

113. Fix, E. & Hodges, J. L. *Discriminatory analysis :Nonparametric Discrimination: Consistency Properties*, 21–49 (Randolph Field, 1951).

114. Breiman, L. Random forests. *Machine Learning* **45**, 5–32 (2001).

115. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 42, 785–794 (ACM, 2016).

116. L, F. Solid Oxide Fuel Cells. in *Advances in Electrochemical Science and Engineering* **2**, 35–78 (2008).

117. Wang, X. et al. Ceria-based nanocomposite with simultaneous proton and oxygen ion conductivity for low-temperature solid oxide fuel cells. *J. Power Sources* **196**, 2754–2758 (2011).

118. Bigler, J. Periodic-Table.xls http://www.mrbigler.com/documents/Periodic-Table.xls (2007).

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.