Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts

Baicheng Weng1,2,3,6, Zhilong Song2,6, Rilong Zhu4, Qingyu Yan4, Qingde Sun2, Corey G. Grice1, Yanfa Yan1✉ & Wan-Jian Yin2,5✉

Symbolic regression (SR) is an approach of interpretable machine learning for building mathematical formulas that best fit certain datasets. In this work, SR is used to guide the design of new oxide perovskite catalysts with improved oxygen evolution reaction (OER) activities. A simple descriptor, $\mu/t$, where $\mu$ and $t$ are the octahedral and tolerance factors, respectively, is identified, which accelerates the discovery of a series of new oxide perovskite catalysts with improved OER activity. We successfully synthesise five new oxide perovskites and characterise their OER activities. Remarkably, four of them, $\text{Cs}_{0.4}\text{La}_{0.6}\text{Mn}_{0.25}\text{Co}_{0.75}\text{O}_3$, $\text{Cs}_{0.3}\text{La}_{0.7}\text{NiO}_3$, $\text{SrNi}_{0.75}\text{Co}_{0.25}\text{O}_3$, and $\text{Sr}_{0.25}\text{Ba}_{0.75}\text{NiO}_3$, are among the oxide perovskite catalysts with the highest intrinsic activities. Our results demonstrate the potential of SR for accelerating the data-driven design and discovery of new materials with improved properties.
Machine learning (ML) is increasingly used in the field of materials informatics as an effective tool for discovering quantitative structure— or composition—property relationships that can accelerate materials design. However, the black-box model of ML is often criticized not to provide new “physical laws”, which limits its potential in certain cases. Symbolic regression (SR) is an approach of interpretable machine learning that simultaneously searches for the optimal mathematical formula of a function and set of parameters in the function. Therefore, SR is capable to deliver interpretable mathematical formulas that can accelerate materials design. Despite the great potential, the application in the field of material science is still limited.

In this communication, we demonstrate that SR can construct a simple descriptor that enables the acceleration of the materials discovery for oxide perovskite catalysts. Oxide perovskites (ABO₃) are an important family of catalysts for OER applications, which are in high demand for renewable energy production and storage, such as hydrogen production from water-splitting and rechargeable metal-air batteries, because of their structural flexibility, compositional versatility, and chemical stability. Moreover, oxide perovskites have recently been extended to the bifunctional application of OER and oxygen reduction reaction. The catalysis activities of oxide perovskite catalysts can be described by descriptors, as demonstrated by various studies over the past sixty years. Several descriptors, such as the reaction free energy and c Occupation9,18, have been successfully used to understand the trend of OER activity and achieved great success in this regard. Nevertheless, those descriptors require prior knowledge based on density functional theory (DFT) calculations therefore bear limited applicability to design new materials, where DFT-calculated values are unknown a priori and highly dependent on the used methodologies. Meanwhile, it is difficult for DFT calculation to accurately determine c Occupation where the surface spin state is not well known. A good descriptor should be simple and yet provide physical insight, which will guide and accelerate the discovery of new perovskite oxide OER catalysts. In this work, we propose that SR is perfectly suitable for identifying suitable descriptor to accelerating the discovery of new perovskite catalysts.

Figure 1 shows the workflow diagram of this study. SR analysis may not require massive datasets, if the datasets used are consistent and reliable. Therefore, we firstly synthesise 18 well-studied oxide perovskite catalysts to produce consistent and comparable datasets of OER activity for SR analysis. A descriptor with the balance of simplicity and accuracy is then chosen and help develop strategies to accelerate the discovery of new oxide perovskites. The generality of the descriptor is confirmed by analysing data reported independently by other research groups.

**Results**

**Data acquisition.** Comparable training data used in SR analysis are of crucial importance for SR in order to produce useful mathematical formulas. Since the first discovery of oxide perovskite LaNiO₃ as OER catalyst in 1970, the chemical management of A- or B-site cations has been used to tune the OER activity, permitted by the structural and chemical flexibility of perovskite structures. The results reported by different groups and produced under different experimental conditions over a period of half a century are summarised in a recent review article. However, the comparability of those data is doubtful due to different environments of experiments and measurements. To ensure meaningful and valuable SR analysis, we synthesised eighteen known oxide perovskite catalysts (Supplementary Fig. 1; Supplementary Table 1). Four samples were made for each perovskite and OER measurement was conducted three times under the same conditions with freshly made catalyst inks. Four each measurement, the Vᵢ(RHE) values at five current densities of 50 mA cm⁻², 5 mA cm⁻², 10 mA cm⁻², 15 mA cm⁻² and 20 mA cm⁻² in linear sweep voltammetry (LSV) curve were adopted for SR analysis. Therefore, there are totally 18 perovskites × 4 samples × 3 measurements × 5 current densities = 1080 data points (Fig. 2a). The values were then normalised by the catalyst loading concentration and Brunauer–Emmett–Teller (BET) surface area (Supplementary Table 2) and shown in Fig. 2a and Supplementary Data 1–5. Details of the materials synthesis, along with the structural and OER characterisation, can be found in the
current densities of 50 µA cm⁻², 10 mA cm⁻², 15 mA cm⁻², and 20 mA cm⁻². The exact values of those data points are provided in Supplementary Data 1.

**SR training.** With the available experimental data shown in Fig. 2a, SR was then adopted to construct mathematical formula linking the materials parameters and V_{RHE}. To ensure that the SR analysis determines mathematical formulas that are useful for our purpose, it is critical to select relevant parameters to be included in the mathematical formulas based on prior knowledge. Considering the importance of previous descriptors⁹,10,12,15,25,26, we chose electronic parameters such as the number of d electrons for TM ions (N_d), electronegativity values χ_A and χ_B, and valence states Q_A, as well as structural parameters such as ionic radii R_A, the tolerance factor t, and the octahedral factor μ, where A and B refer to the A- and B-site cations, respectively (Table 1). The tolerance factor t, defined as \( t = \frac{1}{2} \frac{R_A}{R_B} \), and octahedral factor μ, defined as \( \mu = \frac{R_A}{R_B} \), are commonly used features in ML studies of perovskites.²³,²⁷,²⁸

The mathematical formulas were then generated and selected by using SR with genetic programming (GPSR) as implemented in gplearn code²⁹. The flowchart of GPSR process in this work is described in Fig. 2b. In this work, SR initially builds a population of random mathematical formulas with these parameters as variables. Then, these mathematical formulas breed, mutate, and evolve to form new ones via genetic programming. The derived mathematical formulas compete to model experimental data by evaluating the mean absolute errors (MAEs) between the predicted and experimental V_{RHE}. A grid search of hyperparameters resulted in ~8640 mathematical formulas (desciptors), which were characterised by their MAE’s and complexities, as described in Fig. 3a. The hyperparameters setup can be found in Method part and extended information about GPSR can be found in Supporting Information.

**Descriptor generation and analysis.** Of the produced descriptors, only those with low MAE (high accuracy) and low complexity are suitable for guiding the discovery of new oxide perovskite catalysts. The nine mathematical formulas at the Pareto front [marked as A–I in Fig. 3a] that met the criteria of simplicity and accuracy among the 43,200,000 candidates are shown in Table 2. Among them, \( \mu/t \) is the best compromise between complexity and accuracy. To clearly show the correlation, the \( V_{RHE} \) at current densities of 5 mA cm⁻² are shown in terms of \( \mu/t \) in Fig. 3b. For each perovskite, the average values and error bars are the experimental uncertainties from 12 measurement data (4 samples with each 3 measurements). Interestingly, it shows a linear and monotonic behaviour instead of prevalent volcano shape for conventional descriptors. Such linear correlations remain at other current densities, i.e. 50 µA cm⁻², 10 mA cm⁻², 15 mA cm⁻², and 20 mA cm⁻² as shown in Supplementary Fig. 3. To further verify the generality of this descriptor, we used \( \mu/t \) to fit the experimental work⁹ originally reporting the volcano shape for descriptor \( e_p \) (Fig. 3c). As shown in Fig. 3d, \( \mu/t \) provided a clear linear and monotonic correlation with \( V_{RHE} \), with MAE comparable to the volcano shape for descriptor \( e_p \). Apart from the seminal work of ref. ⁹, the generality of \( \mu/t \) can also be confirmed by recent works³⁰–³² as their data reorganized in Supplementary Fig. 4. For experimental data spanning over sixty years from different groups (Table 6 of ref. ¹⁵), their \( V_{RHE} \) values are reorganized according to their \( \mu/t \) values; despite some discrepancies, a roughly linear correlation was observed for the majority of the data points (Supplementary Fig. 5). Such good correlation reveals that the SR-derived descriptors, e.g., \( \mu/t \) indeed provide meaningful insights for OER activity of oxide perovskites.

The descriptor \( \mu/t \) reveals that the OER activity of oxide perovskite catalysts is closely related to the structural factors of the catalysts; i.e. a smaller \( \mu \) and a larger \( t \) should lead to higher OER activity. Such a simple descriptor is superior to conventional descriptors since it does not require additional DFT calculations and can be directly used for materials design. Accordingly, we
Table 1 Key materials parameters of 23 selected oxide perovskites.

| No. | Materials                        | $t$   | $\mu$  | $R_A(\text{Å})$ | $\chi_A$ | $\chi_B$ | $Q_A$ | $N_d$ | $\mu/t$ |
|-----|---------------------------------|-------|--------|-----------------|----------|----------|-------|-------|---------|
| 1   | LaMnO$_3$                       | 0.993 | 0.430  | 1.36            | 1.1      | 1.55     | 3     | 4     | 0.433   |
| 2   | LaMn$_{0.5}$Ni$_{0.5}$O$_3$     | 0.998 | 0.422  | 1.36            | 1.1      | 1.73     | 3     | 5     | 0.423   |
| 3   | LaNiO$_3$                       | 1.003 | 0.415  | 1.36            | 1.1      | 1.91     | 3     | 7     | 0.413   |
| 4   | LaMn$_{0.5}$Cu$_{0.5}$O$_3$     | 0.988 | 0.437  | 1.36            | 1.1      | 1.725    | 3     | 6     | 0.442   |
| 5   | LaNi$_{0.9}$Fe$_{0.1}$O$_3$     | 1.004 | 0.414  | 1.36            | 1.1      | 1.902    | 3     | 6.8   | 0.413   |
| 6   | LaNi$_{0.8}$Fe$_{0.2}$O$_3$     | 1.004 | 0.413  | 1.36            | 1.1      | 1.894    | 3     | 6.6   | 0.412   |
| 7   | LaFeO$_3$                       | 1.009 | 0.407  | 1.36            | 1.1      | 1.83     | 3     | 5     | 0.404   |
| 8   | La$_{0.9}$Pr$_{0.1}$FeO$_3$     | 1.010 | 0.407  | 1.365           | 1.115    | 1.83     | 3     | 5     | 0.403   |
| 9   | PrFeO$_3$                       | 1.012 | 0.407  | 1.365           | 1.115    | 1.83     | 3     | 5     | 0.402   |
| 10  | LaCoO$_3$                       | 1.011 | 0.404  | 1.36            | 1.1      | 1.88     | 3     | 6     | 0.399   |
| 11  | La$_{0.5}$Ca$_{0.5}$CoO$_3$     | 1.011 | 0.398  | 1.35            | 1.05     | 1.88     | 2.5   | 5.5   | 0.394   |
| 12  | La$_{0.8}$Sr$_{0.2}$CoO$_3$     | 1.019 | 0.401  | 1.376           | 1.07     | 1.88     | 2.8   | 5.8   | 0.394   |
| 13  | Sr$_{0.25}$La$_{0.75}$Fe$_{0.5}$Co$_{0.5}$O$_3$ | 1.020 | 0.401  | 1.38            | 1.063    | 1.855    | 2.75  | 5.25  | 0.393   |
| 14  | La$_{0.5}$Sr$_{0.5}$CoO$_3$     | 1.034 | 0.397  | 1.408           | 1.01     | 1.88     | 2.4   | 5.4   | 0.384   |
| 15  | La$_{0.25}$Sr$_{0.75}$CoO$_3$   | 1.042 | 0.395  | 1.424           | 0.98     | 1.88     | 2.2   | 5.2   | 0.379   |
| 16  | SrCoO$_3$                       | 1.049 | 0.393  | 1.44            | 0.95     | 1.88     | 2     | 5     | 0.374   |
| 17  | Ba$_{0.5}$Sr$_{0.5}$Co$_{0.8}$Fe$_{0.2}$O$_3$ | 1.082 | 0.391  | 1.525           | 0.92     | 1.876    | 2     | 4.8   | 0.361   |
| 18  | BaFeO$_3$                       | 1.119 | 0.385  | 1.61            | 0.89     | 1.83     | 2     | 4     | 0.344   |

The key materials parameters include the tolerance factor ($t$), octahedral factor ($\mu$), ionic radii of A-site ($R_A$) and B-site ($R_B$), electronegativity of A-site ($\chi_A$) and B-site ($\chi_B$), valence state of A-site ($Q_A$), and number of d electrons on TM B-site ($N_d$). The materials are ordered by the value of $\mu/t$ in each dataset of conventional and new perovskites. (See Supplementary Table 7 for calculation details.)

Fig. 3 Descriptor generation and performance. a Pareto front of MAE vs. complexity of 8640 mathematical formulas shown via density plot. b $E_{\text{RHE}}$ vs. $\mu/t$ (black diamonds: conventional perovskites; red dots: new perovskites). The current densities are normalised by BET surface areas (Supplementary Table 2) and loading amount. c Figure 2 from the study of Suntivich et al. reproduced with permission from the American Association for the Advancement of Science. d Reformatted plot according to descriptor $\mu/t$ of c. The MAE (Pearson correlation coefficient) for c, d were 20.6 meV (0.923) and 21.0 meV (0.928), respectively. (The error bar in b is produced by the maximum and minimum values in experiments data.)
Cs0.5La0.5MnO3, Cs0.5La0.5Mn0.25Co0.75O3, and Cs0.5La0.5Mn0.5-Co0.25O3 showed an impurity phase of MnO4−δ (main diffraction peaks at 12° and 24°). Ba0.75Pr0.25Ni0.5S2O3 contained Pr2O3 and NiO impurity phases. Five compounds including Cs0.5La0.6Mn0.25Co0.75O3, Cs0.5La0.4NiO3, Cs0.5La0.6Mn0.5Ni0.5-O3, Sr2O2Ba0.75NiO3, and SrNiO2Co0.25O3, formed pure perovskite phases, as by confirmedPXRD (Supplementary Fig. 6).

The OER activities of these five new pure oxide perovskites were then characterised (Fig. 4a–c). Cs0.5La0.6Mn0.25Co0.75O3, Cs0.5La0.7NiO3, SrNiO2Co0.25O3, and Sr2O2Ba0.75NiO3 showed lower V_RHE values (higher OER activity) than BSCF did. The specific activities are also compared with the state-of-the-art perovskite oxide catalysts.20 We found that our materials are among the oxide perovskite catalysts with the highest specific activities (Supplementary Fig. 7).

Remarkably, the experimental V_RHE values of these new oxide perovskite catalysts follow the same trend of SR-derived descriptor, μ/t, as shown in Fig. 3b. To further verify the descriptor, the SR procedure is repeated with the inclusion of five new predicted perovskites. Most of derived mathematical formulas that had been residing near the Pareto front (Fig. 3a), including μ/t, remain (Supplementary Fig. 8 and Supplementary Table 3); this persistence shows that the addition of more training examples does not generate a significant alteration in the model’s response, indicating that the model remained predictive with these new perovskites. It is worth noting that we have selected a very limited number of compositions for experimental synthesis and characterisation because of limited resources. It is highly anticipated that more of these predicted oxide perovskite catalysts with high OER activities can be experimentally synthesised and their OER activities will be verified.

The stability of the four new oxide perovskite catalysts with OER activities higher than previously reported oxide perovskite catalysts were tested galvanostatically at 10 mA cm−2 disk current (Fig. 4d). We selected a higher disk current density for stability testing to verify the activity decay under strong polarisation conditions. Cs0.5La0.2Mn0.25Co0.75O3, Cs0.5La0.4NiO3, SrNiO2Co0.25O3, and Sr2O2Ba0.75NiO3 showed lower activity degradation than BSCF. In particular, the Sr2O2Ba0.75NiO3 electrode maintained a stable V_RHE over 12 h of stability testing without significant decay. Under the same conditions, the BSCF sample showed a much faster degradation rate, with only 90% retention after 9 h. After OER durability tests, the Sr2O2Ba0.75NiO3 electrode maintained its original morphology. Scanning transmission electron microscopy (STEM) and high-resolution transmission electron microscopy images revealed no significant surface amorphization. The surfaces of the Sr2O2Ba0.75NiO3 particles maintained good crystallinity after stability tests, as confirmed by clear observation of the same lattice spacings (Fig. 5) and elemental analysis (Supplementary Table 4). Recent work has shown that increasing the valence states of 3d-TMs such as Ni and Co from 2+/3+ to 3+/4+ can boost the OER activities of LaCoO3 and LaNiO3.33 Interestingly, apart from increasing t, Cs1+ substitution on the A site is a viable route to enhance the valence states of TM B-site ions in oxide perovskites. This correlates with the SR-derived descriptor, μ/t, since increasing the valence states inevitably reduces the ionic radii of TMs, which in turn reduces the μ value, and, therefore, reduces μ/t. Meanwhile, recent theoretical reports predicted that SrNiO3 should have high OER activity.34 Unfortunately, the hexagonal close packing of Sr and O atoms prevents the formation of the perovskite structure. To mitigate this issue, La was proposed to partially substitute Sr. However, partial La substitution leads to the formation of a Ruddlesden–Popper crystal structure instead of perovskite structures.31 Interestingly, the descriptor μ/t suggests that partial substitution of Sr using larger Ba atoms can enhance catalytic

**Table 2 The nine mathematical formulas at the Pareto front in Fig. 3a.**

| Point | Formulas | MAE (eV) | Complexity |
|-------|----------|----------|------------|
| A     | $\mu = \frac{1}{t}$ | 0.0286 | 1 |
| B     | $(\mu = \frac{1}{t})^{0.5}$ | 0.0279 | 2 |
| C     | 1.554 $t + 1.092$ | 0.0253 | 3 |
| D     | $(t + 0.289 + Q_{t}^{0.5})^{0.5}$ | 0.0252 | 5 |
| E     | $(1.282 + Q_{t}^{0.5})^{0.5} + \mu$ | 0.0244 | 6 |
| F     | $(Q_{t}^{0.5} + \mu + \frac{1}{t})^{0.5}$ | 0.0232 | 8 |
| G     | $(\mu + 1.034 + Q_{t}^{0.5})^{0.5}$ | 0.0225 | 10 |
| H     | $(Q_{t}^{0.5} + 1.034 + (Q_{t}^{0.5} + \mu + \frac{1}{t})^{0.5})^{0.5}$ | 0.0224 | 11 |
| I     | $(Q_{t}^{0.5} + 1 + 0.34\mu + (Q_{t}^{0.5} + \frac{1}{t})^{0.5})^{0.5}$ | 0.0220 | 12 |

The OER activities of these five new pure oxide perovskites are: Ba0.75Sr0.25NiO3, Cs0.4La0.6Mn0.25Co0.75O3, and Cs0.5La0.5Mn0.5-Co0.25O3 showed an impurity phase of MnO4−δ (main diffraction peaks at 12° and 24°). Ba0.75Pr0.25Ni0.5S2O3 contained Pr2O3 and NiO impurity phases. Five compounds including Cs0.5La0.6Mn0.25Co0.75O3, Cs0.5La0.4NiO3, Cs0.5La0.6Mn0.5Ni0.5-O3, Sr2O2Ba0.75NiO3, and SrNiO2Co0.25O3, formed pure perovskite phases, as by confirmed PXRD (Supplementary Fig. 6).

The OER activities of these five new pure oxide perovskites were then characterised (Fig. 4a–c). Cs0.5La0.6Mn0.25Co0.75O3, Cs0.5La0.7NiO3, SrNiO2Co0.25O3, and Sr2O2Ba0.75NiO3 showed lower V_RHE values (higher OER activity) than BSCF did. The specific activities are also compared with the state-of-the-art perovskite oxide catalysts.20 We found that our materials are among the oxide perovskite catalysts with the highest specific activities (Supplementary Fig. 7).

Remarkably, the experimental V_RHE values of these new oxide perovskite catalysts follow the same trend of SR-derived descriptor, μ/t, as shown in Fig. 3b. To further verify the descriptor, the SR procedure is repeated with the inclusion of five new predicted perovskites. Most of derived mathematical formulas that had been residing near the Pareto front (Fig. 3a), including μ/t, remain (Supplementary Fig. 8 and Supplementary Table 3); this persistence shows that the addition of more training examples does not generate a significant alteration in the model’s response, indicating that the model remained predictive with these new perovskites. It is worth noting that we have selected a very limited number of compositions for experimental synthesis and characterisation because of limited resources. It is highly anticipated that more of these predicted oxide perovskite catalysts with high OER activities can be experimentally synthesised and their OER activities will be verified.

The stability of the four new oxide perovskite catalysts with OER activities higher than previously reported oxide perovskite catalysts were tested galvanostatically at 10 mA cm−2 disk current (Fig. 4d). We selected a higher disk current density for stability testing to verify the activity decay under strong polarisation conditions. Cs0.5La0.2Mn0.25Co0.75O3, Cs0.5La0.4NiO3, SrNiO2Co0.25O3, and Sr2O2Ba0.75NiO3 showed lower activity degradation than BSCF. In particular, the Sr2O2Ba0.75NiO3 electrode maintained a stable V_RHE over 12 h of stability testing without significant decay. Under the same conditions, the BSCF sample showed a much faster degradation rate, with only 90% retention after 9 h. After OER durability tests, the Sr2O2Ba0.75NiO3 electrode maintained its original morphology. Scanning transmission electron microscopy (STEM) and high-resolution transmission electron microscopy images revealed no significant surface amorphization. The surfaces of the Sr2O2Ba0.75NiO3 particles maintained good crystallinity after stability tests, as confirmed by clear observation of the same lattice spacings (Fig. 5) and elemental analysis (Supplementary Table 4). Recent work has shown that increasing the valence states of 3d-TMs such as Ni and Co from 2+/3+ to 3+/4+ can boost the OER activities of LaCoO3 and LaNiO3.33 Interestingly, apart from increasing t, Cs1+ substitution on the A site is a viable route to enhance the valence states of TM B-site ions in oxide perovskites. This correlates with the SR-derived descriptor, μ/t, since increasing the valence states inevitably reduces the ionic radii of TMs, which in turn reduces the μ value, and, therefore, reduces μ/t. Meanwhile, recent theoretical reports predicted that SrNiO3 should have high OER activity.34 Unfortunately, the hexagonal close packing of Sr and O atoms prevents the formation of the perovskite structure. To mitigate this issue, La was proposed to partially substitute Sr. However, partial La substitution leads to the formation of a Ruddlesden–Popper crystal structure instead of perovskite structures.31 Interestingly, the descriptor μ/t suggests that partial substitution of Sr using larger Ba atoms can enhance catalytic
activity. Our experiments showed that Ba$_{0.75}$Sr$_{0.25}$NiO$_3$ can be synthesised with the perovskite structure and its OER activity is even higher than BSCF (Fig. 5), demonstrating the usefulness of the SR-derived descriptor.

**Discussion**

Those results show that even with a small dataset, the SR analysis could provide simple and meaningful descriptors that enabled us to discover new oxide OER catalysts with improved activities, which is consistent with successful application of small data in materials design by adaptive ML. The descriptor of $\mu/t$ implies that the catalytic activity of oxide perovskites is closely related to their structural stability, i.e. a lower stability leads to a high activity. Feature analysis in SR process shows that $\mu$, $t$, and $Q_A$ correlate with the catalytic activity more than $R_A$, $N_B$, $\chi_A$, and $\chi_B$ (Supplementary Fig. 9). Considering the $t$ and $\mu$ are functions of $r_A$ and $r_B$, we also trained SR model based on the parameters of $r_A$, $r_B$, $N_B$, $\chi_A$, $\chi_B$, $Q_A$ without $t$, $\mu$. The results are shown in Supplementary Fig. 13 and Supplementary Table 5. However, the MAE of descriptors at the same $\mu$ and $t$ are mostly larger than the $\mu$ and $t$ of existing oxide perovskite catalysts reported in the literature. We anticipate that more of the predicted new oxide perovskite catalysts can be synthesised and their OER activities verified. Our results demonstrate that SR is a powerful ML technique to discover physically meaningful descriptors when sufficient comparable data is available. This work suggests a new direction for discovering functional materials with improved activities.

**Methods**

**Symbolic regression.** Symbolic regression analysis using a genetic algorithm was performed using gplearn, a Python library that extends scikit-learn, a machine learning tool, for symbolic regression. The hyper-parameters setup for gplearn is shown in the Supplementary Information.

The meanings of genetic operations of pc, ps, ph, and pp above can be found in the Supplementary Information.

**Experimental synthesis of oxide perovskites.** The oxide perovskites were synthesised using a modified Pechini method following by thermal calcination at 850–1000°C under dry air/oxygen atmospheres. Briefly, the acetate or nitrate precursors of the perovskite oxides (4 mmol) were mixed in methanol/H$_2$O (10 mL, 2:1 v/v), and citric acid (10 mmol) was added to obtain a clear sol. The mixture was dried at 120°C and the remaining solid was calcinated at 500°C for 1 h in air. Then, the obtained powder was ground into fine powder and pressed into pellets with a diameter of 15 mm using a hydraulic press at 20 MPa. Finally,
the pellets were calcined at 850–1000 °C for 6 h under dry air/oxygen atmospheres.

**Crystal structure characterisation.** The structure and phase of the synthesised materials were examined by PXRD (Ultima III, Rigaku, Japan) and Raman spectroscopy (Bruker FT Raman Spectrometer with a laser wavelength of 532 nm). The morphology of the films was characterised using transmission electron microscopy (TEM; JEOL 3011, Japan), scanning transmission electron microscopy (STEM; Hitachi HD-2300A, Japan), and high-resolution TEM (HRTEM; Hitachi HD-3010A, Japan). Elemental compositions were determined using energy-dispersive X-ray spectroscopy (EDS; Oxford Instruments, UK) and inductively coupled plasma mass spectrometry (ICP-MS; Thermo Scientific XSeries 2 ICPMS, USA). The catalyst surface area was determined using Brunauer–Emmett–Teller (BET) analysis, using a BEBELOR-mini II (BEL, Japan Inc.) under a flow of N2 gas.

**OER characterisation.** OER characterisation was performed on a glassy carbon rotating disk electrode. First, 2 mg of catalyst was dissolved in 2 mL ethanol and 100 μL Nafion solution was added. Then, the mixture was sonicated for 30 min to form a homogeneous mixture. Subsequently, 90 μL of the slurry was loaded onto the surface of a glassy carbon electrode (GCE; 0.196 cm²) and the electrode was dried at room temperature. The electrolyte was purified to remove trace Fe using Ni (OH)₂ powder. The OER measurements were performed using a Voltalab PGZ-301 potentiostat/galvanostat (Radiometer Analytical, France), with a Pt foil and a Ag/AgCl electrode used as the counter and reference electrodes, respectively. The loading amount of the catalysts was 0.168 mg cm⁻². All potentials were plotted versus the reversible hydrogen electrode (RHE) as $E_{\text{RHE}} = E_{\text{AgCl}} + 0.197 + 0.0591 \times \text{pH}$. All linear sweep voltammetry measurements were performed at a scan rate of 5 mV s⁻¹. All OER measurements were iR-compensated (98%). Each measurement was conducted three times under the same conditions. The error bars denote variations observed from sample synthesis and OER measurements. The stability test was performed using the controlled current electrolysis method. PXRD measurements verified that all the obtained materials had the perovskite structure.

To evaluate the intrinsic activities, the current densities were normalised by the loading amount and the BET surface areas in order to exclude the increase in current as a result of high loading content and higher surface area. Normalisation was performed according to the expression: $i = \frac{(\text{mA cm}^{-2})}{(\text{oxide current})} = i = \frac{(\text{mA cm}^{-2} \text{ oxide current})}{(\text{loading amount (g cm}^{-2}) \times \text{BET surface area (cm}^{-2} \text{ g}^{-1}))}$. Here, $i$ (mA cm⁻² oxide current) was denoted as the normalised specific activity, while $i$ (mA g⁻¹ oxide current) = $j$ (mA cm⁻² disk current) = (loading amount (g cm⁻²)) refers to the mass activity.

**Data availability**

The data of measured $V_{\text{OER}}$ values for all oxide perovskites and 3545 potential oxide perovskites listed by the amount of $\mu t$ are provided online.

Received: 26 February 2020; Accepted: 16 June 2020; Published online: 14 July 2020

Table 3 The setup of hyper-parameters in gplearn for GPSR.

| Parameter         | Value                  |
|-------------------|------------------------|
| population size   | 5000                   |
| Generations       | 20                     |
| stopping criteria | 0.01 (eV)              |
| pc                | 0.5, 0.95 (step = 0.025) |
| ps                | (1-pp)/3, (0.92-pp)/3 (step = 0.01) |
| ph                | pp                     |
| pp                | 1-pp-ps-ph             |
| function set      | add, sub, mul, div, sqrt |
| parsimony coefficient | 0.0005, 0.0015 (step = 0.0005) |
| tournament size   | 20                     |
| metric            | mean absolute error (MAE) |
| constant range    | (−1)                   |

The explanation of each hyperparameter are as follows: population size the number of mathematical formulas in each generation, generations the max number of generations, stopping criteria MAE value that the program stops, pc crossover probability, ps subtree mutation probability, ph host mutation probability, pp point mutation probability, function set basic building blocks containing mathematical operators, parsimony coefficient a constant that penalizes large individuals by adjusting their MAE to make them less favorable for selection, tournament size the number of individuals in each tournament, metric measures how well an individual fits, constant range the range of constants included in mathematical formula.

References

1. Wang, Y., Wagner, N. & Rondinelli, J. M. Symbolic regression in materials science. *MRS Commun.* 9, 793–805 (2019).
2. Schmidt, M. & Lipson, H. Distilling free-form natural laws from experimental data. *Science* 324, 81–85 (2009).
3. Lu, S. et al. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* 9, 3405 (2018).
4. Xue, D. et al. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* 7, 11241 (2016).
5. Xue, D. et al. Accelerated search for BaTiO₃-based piezoelectrics with vertical morphotropic phase boundary using Bayesian learning. *Proc. Natl Acad. Sci. USA* 113, 13301–13306 (2016).
6. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* 5, 83 (2019).
7. Vellidio, A., Martin-Guerrero, J. D. & Lisbon, P. J. G. Making machine learning models interpretable. In *Proc. ESANN 2012, 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Vol. 12, 163–172 (2012).
8. Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M. & Ghiungherlli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immunity of offered candidates. *Phys. Rev. Mater.* 2, 83802 (2018).
9. Suntivich, J., May, K. J., Gasteiger, H. A., Goodenough, J. B. & Shao-Horn, Y. 
   A perovskite oxide optimized for oxygen evolution catalysis from molecular orbital principles. 
   *Science* **334**, 1383–1385 (2011).

10. Wei, C. et al. Recommended practices and benchmark activity for hydrogen 
   and oxygen electrocatalysis in water splitting and fuel cells. *Adv. Mater.* **31**, 
   1806296 (2019).

11. Vojvodic, A. & Nørskov, J. K. Chemistry: optimizing perovskites for the water- 
   splitting reaction. *Science* **334**, 1355–1356 (2011).

12. She, Z. W. et al. Combining theory and experiment in electrocatalysis: Insights 
   into materials design. *Science* **355**, eaad4998 (2017).

13. Yin, W. J. et al. Oxide perovskites, double perovskites and derivatives for 
   electrocatalysis, photocatalysis, and photovoltaics. *Energy Environ. Sci.* **12**, 
   442–462 (2019).

14. Bradley, K., Giagloglou, K., Hayden, B. E., Jungius, H. & Vian, C. Reversible 
   perovskite electrocatalysts for oxygen reduction/oxygen evolution. *Chem. Sci.* 
   **10**, 4609–4617 (2019).

15. Retuerto, M. et al. La$_{1.5}$Sr$_{0.5}$NiMn$_{0.5}$Ru$_{0.5}$O$_6$ double perovskite with enhanced 
   ORR/OER bifunctional catalytic activity. *ACS Appl. Mater. Interfaces* **11**, 
   23454–23464 (2019).

16. Bockris, J. O. The Electrocatalysis of oxygen evolution on perovskites. *J. 
   Electrochem. Soc.* **131**, 290 (1984).

17. Man, I. C. et al. Universality in oxygen evolution electrocatalysis on oxide 
   surfaces. *ChemCatChem* **3**, 1159–1165 (2011).

18. Hwang, J. et al. Perovskite oxides and oxygen electrocatalysis. *Science* **358**, 
   751–756 (2017).

19. Jacobs, R., Hwang, J., Shao-Horn, Y. & Morgan, D. Assessing correlations of 
   perovskite catalytic performance with electronic structure descriptors. *Chem. 
   Mater.* **31**, 785–797 (2019).

20. Haverkort, M. W. et al. Spin state transition in LaCoO$_3$ studied using soft X- 
   ray absorption spectroscopy and magnetic circular dichroism. *Phys. Rev. Lett.* 
   **97**, 38–41 (2006).

21. Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C. & Scheffler, M. Big 
   data of materials science: critical role of the descriptor. *Phys. Rev. Lett.* **114**, 
   105503 (2015).

22. Vladislavleva, E. Model-based problem solving through symbolic regression 
   via pareto genetic programming. 169–172 (CentER, Center for Economic Research, 2008).

23. Bartel, C. J. et al. New tolerance factor to predict the stability of perovskite 
   oxides and halides. *Sci. Adv.* 5, eaav0693 (2019).

24. Meadowcroft, D. B. Low-cost oxygen electrode material. *Nature* **226**, 847–848 (1970).

25. Hong, W. T., Welsch, R. E. & Shao-Horn, Y. Descriptors of oxygen- 
   evolution activity for oxides: a statistical evaluation. *J. Phys. Chem. C.* **120**, 
   78–86 (2016).

26. Davies, D. W., Butler, K. T. & Walsh, A. Data-driven discovery of photoactive quaternary oxides using first-principles machine learning. *Chem. Mater.* **31**, 
   7221–7230 (2019).

27. Li, Z., Xu, Q., Sun, Q., Hou, Z. & Yin, W. J. Thermodynamic stability 
   landscape of halide double perovskites via high-throughput computing and 
   machine learning. *Adv. Funct. Mater.* **29**, 1807280 (2019).

28. Stephens, T. gplearn. https://gplearn.readthedocs.io/en/latest/intro.html.

29. Petrie, J. R. et al. Enhanced bifunctional oxygen catalysis in strained LaNiO$_3$ 
   perovskites. *J. Am. Chem. Soc.* **138**, 2488–2491 (2016).

30. Forslund, R. P. et al. Exceptional electrocatalytic oxygen evolution via tunable 
   charge transfer interactions in La$_2$Sr$_{1-x}$Ni$_x$Fe$_{1-y}$O$_{4+x}$ Ruddlesden-Popper 
   oxides. *Nat. Commun.* **9**, 3150 (2018).

31. Honz, R. K. & Ramezanpour, F. Remarkable oxygen-evolution activity of a 
   perovskite oxide from the Ca$_2$–Sr$_x$Fe$_{1-y}$O$_{4+x}$ series. *Angew. Chem.* **131**, 
   2082–2085 (2019).

32. Weng, B. et al. A layered Na$_2$–NiFe$_2$O$_4$ double oxide oxygen evolution 
   reaction electrocatalyst for highly efficient water-splitting. *Energy Environ. Sci.* 
   **10**, 121–128 (2017).

33. Rong, X., Parolin, J. & Kolpak, A. M. A fundamental relationship between 
   reaction mechanism and stability in metal oxide catalysts for oxygen 
   evolution. *ACS Catal.* **6**, 1153–1158 (2016).

34. Feng, L. M. et al. Formability of ABO$_3$ cubic perovskites. *J. Phys. Chem. Solids* 
   **69**, 967–974 (2008).

Acknowledgements

W.Y. acknowledges funding support from the National Key Research and Development 
Program of China (grant No. 2016YFB0700700); National Natural Science Foundation of 
China (grant No. 11674237, 11974257) and the Priority Academic Program Development 
of Jiangsu Higher Education Institutions (PAPD). The theoretical work was carried 
out at the National Supercomputer Center in Tianjin and the calculations were per- 
formed on TianHe-1(A). This paper presents results from an NSF project (grant No. 
CBET–1433401) under the ‘NSF 14–15: NSF/DOE Partnership on Advanced Frontiers in 
Renewable Hydrogen Fuel Production via Solar Water Splitting Technologies’ project, 
which was co-sponsored by the National Science Foundation, Division of Chemical, 
Bioengineering, Environmental, and Transport Systems (CBET), and the U.S. Depart- 
ment of Energy, Office of Energy Efficiency and Renewable Energy, Fuel Cell Technol- 
ogies Office.

Author contributions

W.Y. and Y.Y. conceived the idea and supervised the project. W.Y. supervised the 
theoretical study. Y.Y. supervised the experimental study. W.Y., Y.Y., and Q.S. performed 
data collection and analysis. Z.S. and W.Y. carried out SR calculations and analysis. B.W., 
C.G., Z.Z., and Q.Y. performed materials synthesis and OER characterisation. W.Y., Y.Y., 
and B.W. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467- 
020-17263-9.

Correspondence and requests for materials should be addressed to Y.Y. or W.-J.Y.

Peer review information *Nature Communications* thanks Marcel Risch, Christopher 
Bartel and the other, anonymous, reviewer for their contribution to the peer review of 
this work. Peer reviewer reports are available.

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.

© The Author(s) 2020

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, 
adaptation, distribution and reproduction in any medium or format, as long as you give 
appropriate credit to the original author(s) and the source, provide a link to the Creative 
Commons license, and indicate if changes were made. The images or other third party 
material in this article are included in the article’s Creative Commons license, unless 
indicated otherwise in a credit line to the material. If material is not included in the 
article’s Creative Commons license and your intended use is not permitted by statutory 
regulation or exceeds the permitted use, you will need to obtain permission directly from 
the copyright holder. To view a copy of this license, visit http://creativecommons.org/ 
licenses/by/4.0/.