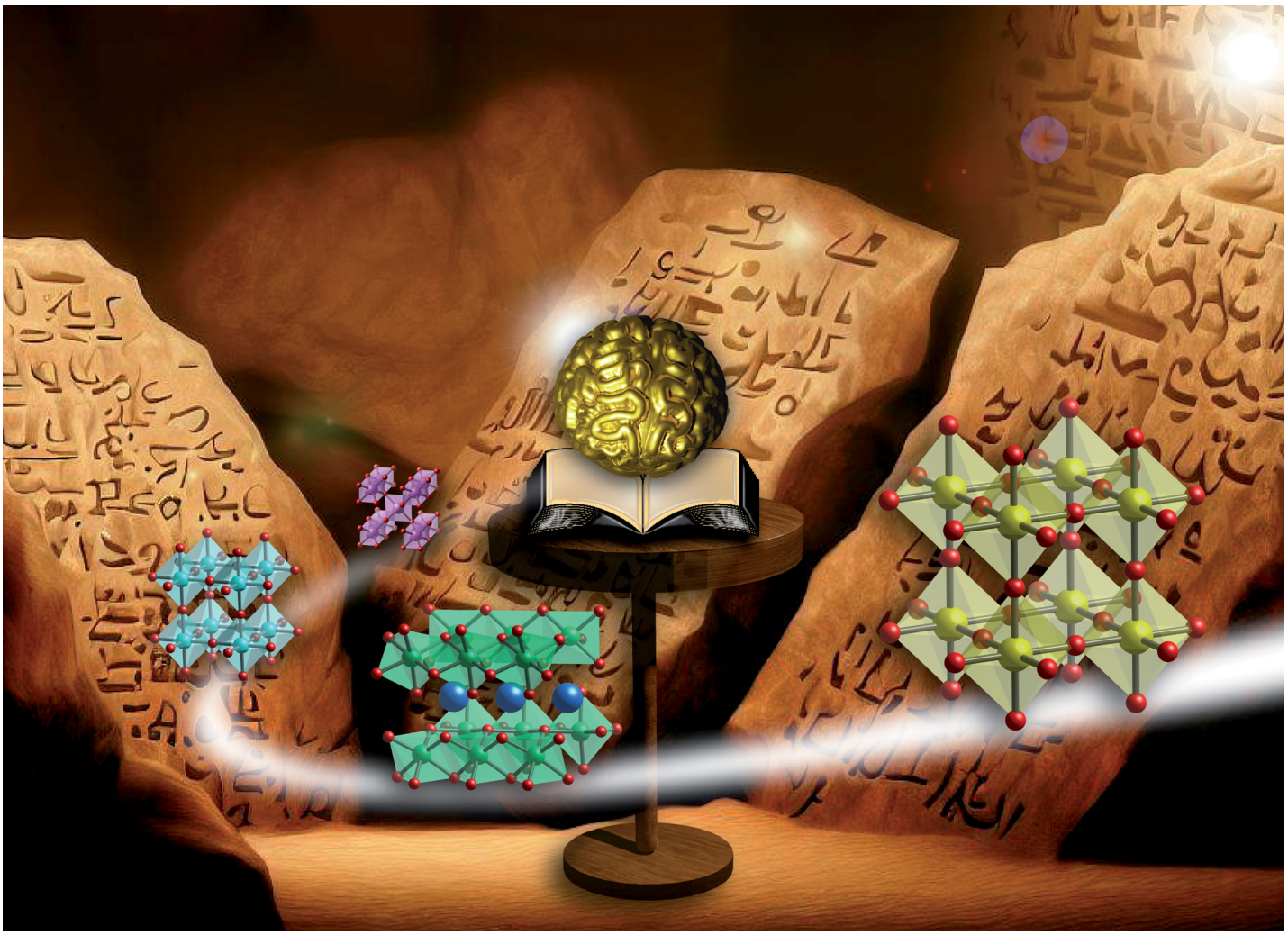


Showcasing research from Prof. Yamaguchi's lab at Tokyo Institute of Technology, performed in educational collaboration with graduate students in Tokyo Tech Academy for Convergence of Materials and Informatics (TAC-MI) program.

Machine learning-aided unraveling of the importance of structural features for the electrocatalytic oxygen evolution reaction on multimetal oxides based on their A-site metal configurations

Artificial intelligence technology, such as machine learning (ML), recently propels social and technological innovation to unprecedented scales. The state-of-the-art ML analysis uncovers crystal structure-based comprehensive descriptors for the performance of multimetal oxide electrocatalysts for oxygen evolution reaction (OER) in hydrogen production *via* water splitting. This finding is useful for rational and rapid design of promising OER electrocatalysts.

As featured in:

See Yuuki Sugawara, Takeo Yamaguchi *et al.*, *Energy Adv.*, 2023, **2**, 1351.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/energy-advances

Registered charity number: 207890

Energy Advances

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Energy Adv.*, 2023, 2, 1351

Received 31st May 2023,
Accepted 22nd July 2023

DOI: 10.1039/d3ya00238a

rsc.li/energy-advances

Open Access Article. Published on 24 mis Gortheren 2023. Downloaded on 10/12/2025 01:57:44.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Machine learning-aided unraveling of the importance of structural features for the electrocatalytic oxygen evolution reaction on multimetal oxides based on their A-site metal configurations†

Yuuki Sugawara, *[a] Xiao Chen,[b] Ryusei Higuchi[c] and Takeo Yamaguchi*[a]

There is a need for comprehensive descriptors to develop prominent electrocatalysts for use in the oxygen evolution reaction (OER) for water splitting. Through machine learning analysis of the data obtained from multimetal oxides that contain A-site alkaline-/rare-earth and B-site transition metals, this study revealed that the OER activities depend on the A-site-related structures.

Efficient and clean hydrogen production technology is crucial for achieving a decarbonized society, making the development of superior electrocatalysts for electrochemical water-splitting anodes *via* the oxygen evolution reaction (OER) an important research area.[1,2] Among the studied compounds that serve as OER electrocatalysts, multimetal oxides with the general formula $A_xB_yO_z$, wherein the A and B sites are occupied by alkaline-/rare-earth and transition metals, respectively, have attracted considerable attention (Fig. 1a).[3–6] The conventional approach to catalyst development through trial and error is time-consuming for identifying highly active materials. Therefore, there is a need to develop comprehensive descriptors for catalytic activity that can facilitate the rapid development of promising materials.

Various electronic[7] and structural factors[8–11] have been proposed as descriptors of OER activity. These factors include the number of d electrons in the e_g orbital of transition metals in perovskite-type multimetal oxides,[12] and the charge-transfer energy,[13] which represents the energy level gap between the occupied p orbitals of oxygen and the unoccupied d orbital of the transition metal. All of these descriptors are associated with the B-site metals, which are considered to be active sites of the OER, whereas A-site metal-based descriptors for the OER remain relatively unknown. Conventionally, the B-site metals are believed to have a major influence on the activity,[14] with the A-site metals playing a minor role in slightly regulating the activity, as shown in Fig. 1a. However, the A-site metals are typically close, within a range of 3–4 Å, to the B-site metals. In addition, the electron orbitals of the A-site metal overlap with those of oxygen, and the strength of the A–O bonds varies considerably and is dependent on the materials.[15] Therefore, the A-site metal also contributes sterically and electronically to the OER process. Therefore, focusing solely on the B-site metals does not provide sufficient insight into OER electrocatalysis, necessitating a more comprehensive understanding of the electrocatalytic behavior of multimetal oxides for the rapidly development of highly active OER electrocatalysts.

Previous studies have focused on the A-site metals of multimetal oxides,[16–21] enabling the prediction of OER activity for multimetal oxides bearing identical B-site metals that were synthesized using specific methods. However, a comprehensive A-site-based descriptor for OER activity has yet to be uncovered. Electronic parameters, such as the energy levels of the electron orbitals, require first-principles calculations using high-performance computers or extensive measurements at synchrotron radiation facilities. Conversely, bulk crystal structure parameters can be obtained through general X-ray diffraction measurements. Moreover, these bulk structural parameters can be easily collected from databases such as the Inorganic Crystal Structure Database (ICSD) without the need for experiments. Therefore, structural parameters can serve as helpful descriptors of catalytic activity. Informatics technologies, such as machine learning (ML), have the capability to analyse vast amounts of data that encompass hundreds of thousands of materials within a short timeframe, making them powerful tools for elucidating such data. ML techniques have been increasingly employed to develop promising catalysts[22–27] and enhance researchers' understanding of catalytic science.[14,28,29]

[a] *Laboratory for Chemistry and Life Science, Institute of Innovative Research, Tokyo Institute of Technology, R1-17, 4259 Nagatsuta, Midori, Yokohama, 226-8503, Japan. E-mail: sugawara.y.aa@m.titech.ac.jp, yamag@res.titech.ac.jp*
[b] *School of Engineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori, Yokohama, Kanagawa 226-8503, Japan*
[c] *Laboratory for Materials and Structures, Institute of Innovative Research, Tokyo Institute of Technology, 4259 Nagatsuta, Midori, Yokohama 226-8503, Japan*

† Electronic supplementary information (ESI) available: Methods, additional tables and figure. See DOI: https://doi.org/10.1039/d3ya00238a

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 24 mis Gortheren 2023. Downloaded on 10/12/2025 01:57:44.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

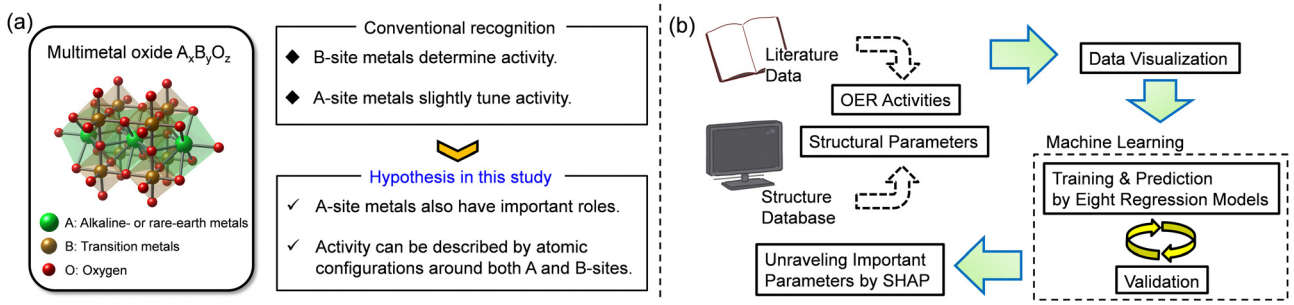


Fig. 1 (a) General formula and typical crystal structure of multimetal oxides, with conventional recognition for OER electrocatalysis in this field and hypothesis in this study. (b) ML approach used in this study to uncover descriptors in terms of the A-site metals.

Recently, we identified a clear correlation between the OER activity and the Fe–O bond length in iron-based multimetal oxides, and used ML to demonstrate that the Fe–O bond length played a more important role for the OER activity of these oxides than the other structural parameters.[30]

In this study, the role of A-site metals in multimetal oxides with respect to their OER activity was investigated by conducting ML analysis using data collected from previously published literature, as shown in Fig. 1b. Our aim is to clarify the importance of conventionally neglected parameters of A-site

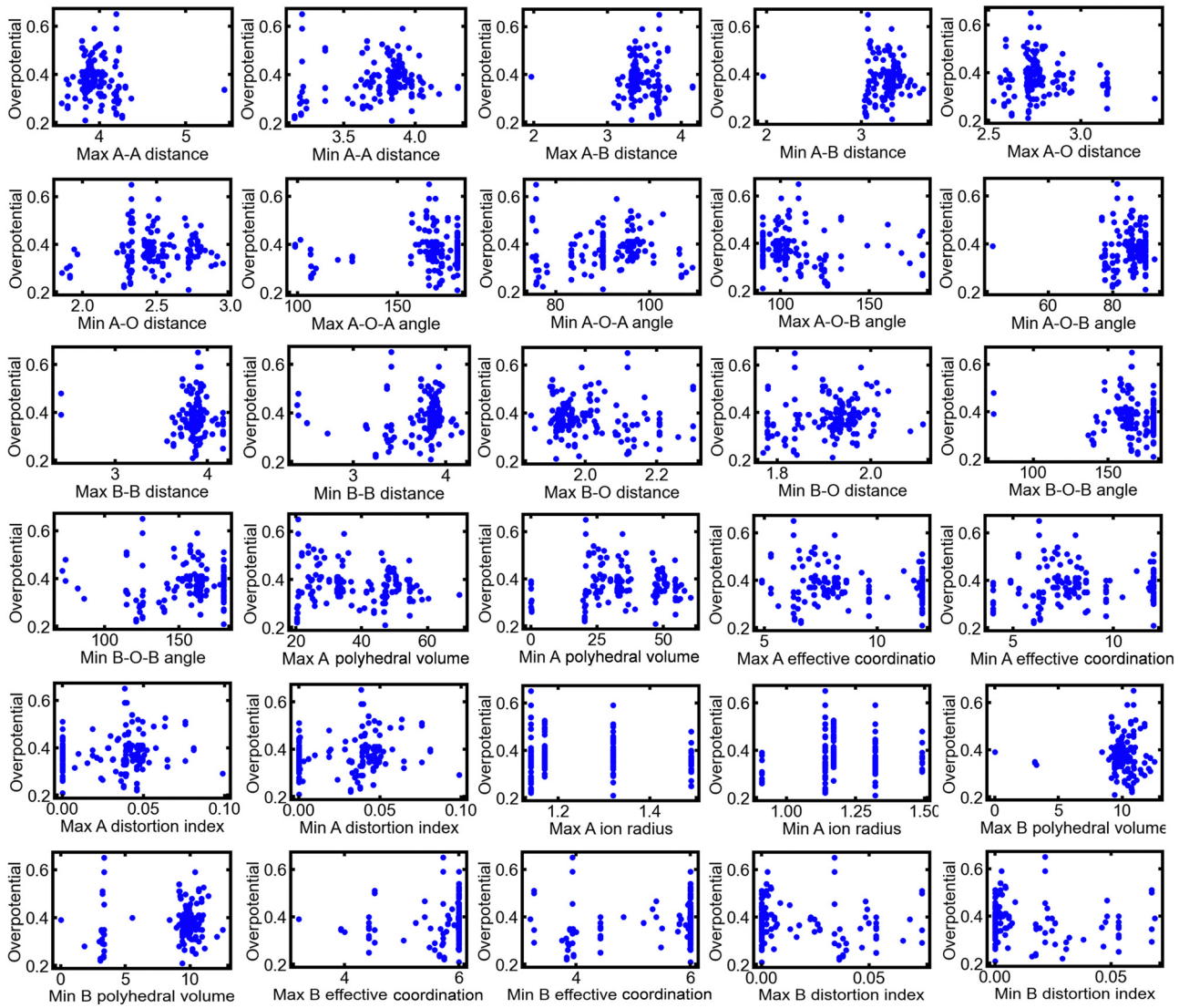


Fig. 2 Plots of the collected dataset from the previously published literature, including each structural parameter and overpotential.

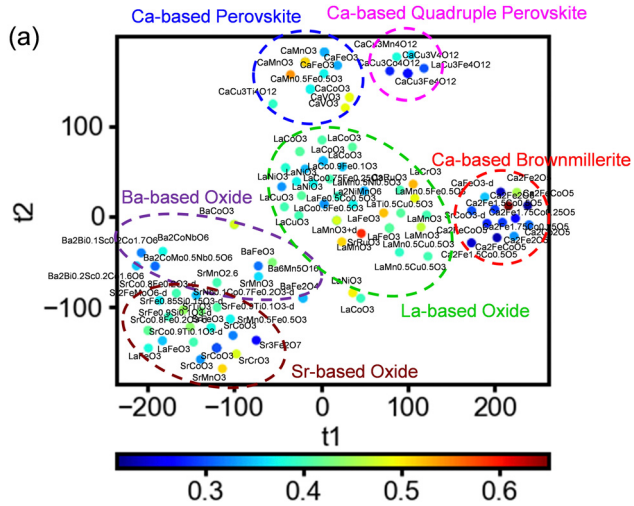
© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 24 mis Gortheren 2023. Downloaded on 10/12/2025 01:57:44.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

metal configurations and propose comprehensive descriptors for OER activity that can be applied to a wide range of multimetal oxides, regardless of the elemental compositions and structural categories.

A total of 154 multimetal oxide OER electrocatalysts reported in 47 published literature sources were collected for analysis in this study. The measured OER overpotentials were extracted from the literature and were used as the objective variables for ML. The details of the literature sources and the chemical formulas of the reported catalysts are presented in Table S1 of the ESI.† In addition, 30 crystal structure parameters for the catalysts were extracted as explanatory variables for ML using VESTA software[31] from CIF files obtained from databases, such as the ICSD and Joint Committee for Powder Diffraction Standards (JCPDS) cards. The collected structural parameters and their definitions are described in the ESI.† Previous studies,[5,32] have achieved a high prediction accuracy through ML using the maximum and minimum values of the parameters related to atomic configurations as explanatory variables rather than their average values. Therefore, in this study, the maximum and minimum values of each parameter in the unit cell of the materials were used for ML. The collected OER overpotentials were plotted against each structural parameter, as shown in Fig. 2. There was no clear correlation between each structural parameter and the OER activity, implying that the OER activity reflected the effect of several structural parameters.

Owing to the use of 30 different structural parameters as explanatory variables, interpreting the relations of the structural parameters to the OER activity becomes challenging due to the large dimension of the dataset. To address this, the analysis was carried out using t-distributed stochastic neighbor embedding (t-SNE).[33] t-SNE is a nonlinear visualization method to simplify high-dimensional data into a low-dimensional space, which preserves the relationships between nearby plots. Sample plots are distributed in a colored map, where similar data points are grouped together. More details of the t-SNE are described in the ESI.† To visualize the data using simpler structural factors, t-SNE was conducted using a reduced dataset comprising 118 materials that contain a single A-site metal component. Fig. 3a shows the output distributions from the t-SNE, where close proximity between the plots indicates a similar relationship between the explanatory and objective variables. In Fig. 3a, the data are clustered by Ca-based brownmillerites, Ca-based perovskites, Ca-based quadruple perovskites, La-based oxides, Sr-based oxides and Ba-based oxides, *i.e.*, the data were distributed based on the A-site metallic element in the materials. By contrast, the B-site metallic elements were distributed in a disorderly fashion, with the position of each B-site element scattered. Moreover, with respect to the ionic radius values of the A-site cations, the ionic radii differed considerably depending on the element, as shown in Fig. 3b and a, where most plots are distributed according to the size of their ionic radii. Oxides containing Ca^{2+} and La^{3+}, which have close ionic radii, were located close to each other on top-right side of the map, whereas oxides containing Sr^{2+} and



(b)

A-site	Crystal ionic radii [A]
Ca^{2+}	1.14
La^{3+}	1.17
Sr^{2+}	1.32
Ba^{2+}	1.49

(c)

B-site	Crystal ionic radii [A]
Ni^{4+}	0.62
Co^{4+}	0.67
Mn^{4+}	0.67
Fe^{3+}	0.69
Co^{3+}	0.69
Ni^{3+}	0.70
Mn^{3+}	0.72

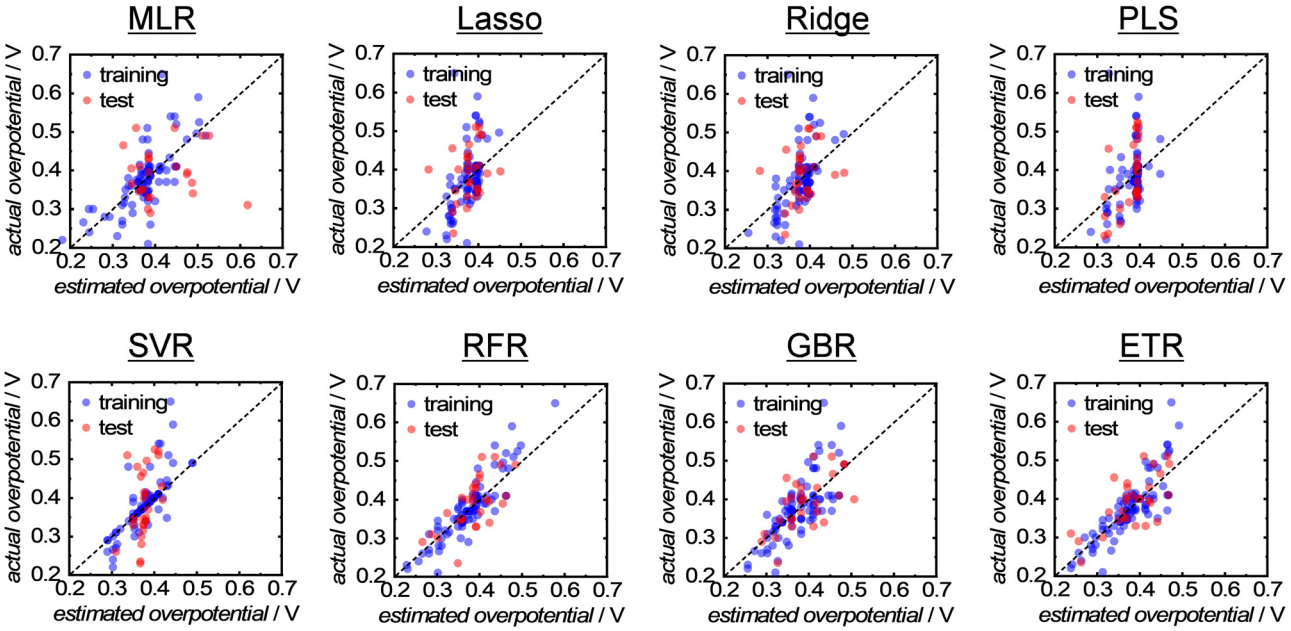
Fig. 3 Data distribution obtained using t-SNE analysis including single A-site-containing multimetal oxides. The plot colors indicate the overpotential of each catalyst. (b) and (c) Ionic radii for the typical A-site and B-site metals, respectively.[34]

Ba^{2+}, with larger ionic radii, were located farther apart on bottom-left side. By contrast, the ionic radii of B-site metals do not change significantly even for different elements, as shown in Fig. 3c. This can result in disordered distributions of the B-site metals observed in Fig. 3a. These results indicated that the relationships between the structural parameters and the OER activity depend on the A-site metallic elements due to the large difference in ionic radii and steric effects in the lattices. Conversely, the ionic radius of the B-site metal changed only slightly for different elements, indicating that the steric effect near the B-sites has a negligible impact on the OER activity. These results indicate the importance of the structure near the A-site metal in determining the OER activity.

Next, to analyze the importance of the structural parameters more precisely, the collected data were used to train eight different ML algorithms and construct predictive models. Detailed information on the ML algorithms can be found in the ESI.† The hyperparameters for each algorithm were optimized using a grid-search method from the range of values provided in Table S2 (ESI†). The prediction accuracy of the constructed models was evaluated by cross-validation using 75% and 25% of samples as the training and test data, respectively, and the errors between the predicted and actual overpotentials were expressed as the root-mean-squared error (RMSE). The results of the ML analysis demonstrated that the decision-tree-based algorithms of random forest regression (RFR), gradient boosted regression (GBR) and extra trees

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 24 mis Gortheren 2023. Downloaded on 10/12/2025 01:57:44.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.



	MLR	Lasso	Ridge	PLS	SVR	RFR	GBR	ETR
RMSE (training)	0.054	0.067	0.064	0.065	0.045	0.035	0.052	0.042
RMSE (test)	0.232	0.067	0.067	0.073	0.079	0.051	0.053	0.051

Fig. 4 Results of ML analysis with eight types of (non)linear regression models for single-A-site-containing multimetal oxides.

regression (ETR) exhibited a higher performance with smaller errors and overfitting, as shown in Fig. 4. The superiority of the decision-tree-based algorithm is consistent with a previous study that analyzed the catalytic activity of methane-oxidation catalysts using ML.[26] Thus, the analysis clarified that the OER activity of multimetal oxides can be accurately predicted using parameters derived from the bulk crystal structure.

In addition, the importance of each explanatory variable in relation to the OER activity was quantitatively estimated using Shapley additive explanations (SHAP) values.[35] This analysis was performed using the analytical results obtained from the ETR algorithm, which exhibited the highest accuracy for the test data among the eight kinds of algorithm and a smaller overfitting than the RFR algorithm. The SHAP method can quantitatively compare the contribution of each explanatory variable for the prediction of an objective variable from ML. The magnitude of the SHAP values also indicates the relative importance of each explanatory variable.[32,36,37] Fig. 5a shows the top six rankings of the SHAP values obtained from the abovementioned ML analysis. Interestingly, all six of the most critical parameters are associated with the structures of A-site metals, which align well with the earlier discussion on the t-SNE result. However, although the t-SNE result in Fig. 3a depicts a distribution based on the ionic radius of A-site metals, the ionic radius of the A-site metals was not ranked in Fig. 5a. This result can be explained by Fig. S1 (ESI†), where it is presumed that the ionic radius alone is not affect the OER activity, but rather the subsequent changes in bond length and angle and interatomic distance, which are influenced by the ionic radius, affect the activity. For example, a change in the ionic radius of the A-site cations affects the tilting and

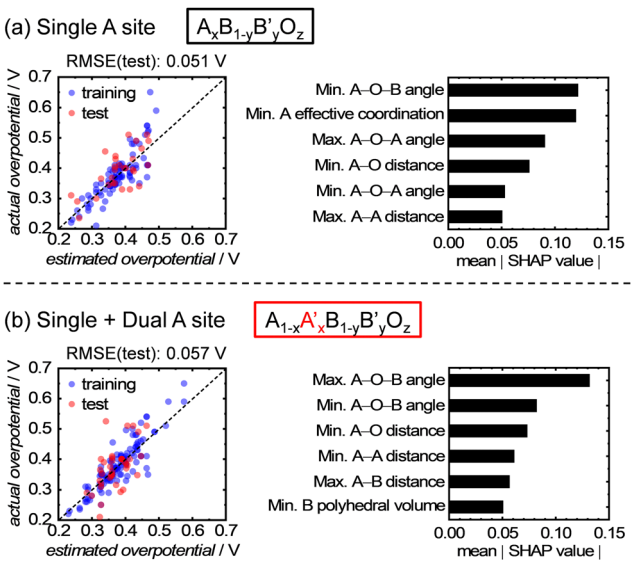


Fig. 5 Comparison of the ML analyses for datasets with (a) only single and (b) single and dual A-site-containing oxides, and the output SHAP values from the ML analysis using the ETR algorithm.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 24 mis Gortheren 2023. Downloaded on 10/12/2025 01:57:44.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

symmetry around the B-site cations,[38] which can influence the OER activity and shows good consistency with the high ranking of the A–O–B angles in Fig. 5. Conversely, the SHAP values of structural parameters involving only B-site metals, such as the B–O distance, are not included in the top ranking in Fig. 5a. This indicates that the atomic configurations of the A-site metals have a strong influence on the OER activity, compared with those of the B-site metals. Remarkably, this contradicts the conventionally believed ideas, where B-site metals as active sites predominantly influence the OER activity.

Finally, the ML analysis was expanded to encompass dual A-site metal components, and ML analysis was performed on 154 materials with single and dual A-site-containing multimetal oxides. The ETR algorithm was used for this analysis because it exhibited the highest accuracy for data with single A-site metal components. As a result, the prediction accuracy decreased slightly compared with the single A-site series, as shown in Fig. 5a and b. This decrease can be attributed to the random occupancy of different A-site cations in the dual A-site series crystals, making the atomic configurations more complex. Nevertheless, the increase in RMSE was approximately 0.006 V, indicating a minimum decline in the prediction accuracy. In addition, the important structural parameters that contribute significantly to the OER activity are almost the same as those in the single A-site series. Therefore, the methodology of prediction using A-site metal configurations is applicable to data from the dual A-site series. Note that A-site alkaline-earth metals in multimetal oxides could be leached out during electrochemical potential cycles, followed by the reconstruction of catalyst surface structures, which is involved in OER activity and stability.[39] Nevertheless, the findings in this paper demonstrated that the OER activity can be described using the parameters for initial bulk structures, which can be characterized *via* a general XRD measurement and obtained from materials databases, such as the ICSD. Therefore, this study presents versatile and easily-available structural descriptors for OER electrocatalysis on multimetal oxides. The novel findings presented in this study provide valuable insights into OER electrocatalysis on multimetal oxides and contribute to the rational and rapid design of highly active catalysts.

In conclusion, a data-driven ML approach was used to analyze the OER activities and structural parameters of 154 types of multimetal oxide collected from published literature. Data visualization using t-SNE demonstrated that the data were distributed based on the ionic radius of each A-site metal and structural category. This indicates that the atomic configurations of the A-site influence the OER activity. A quantitative comparison using ML revealed that the structural parameters involving A-site metals hold greater importance for the OER activity compared with those involving only B-site metals. This finding challenges the conventional belief in the field. Thus, this study provides a new route in the field of multimetal oxide OER electrocatalysts.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge the support by the Tokyo Tech Academy for Convergence of Materials and Informatics (TAC-MI) adopted by the WISE Program of MEXT. Part of this work was supported by JSPS KAKENHI Grant Number 20K15087 (Y. S.).

Notes and references

1 Y. Sugawara, S. Sankar, S. Miyanishi, R. Illathvalappil, P. K. Gangadharan, H. Kuroki, G. M. Anilkumar and T. Yamaguchi, *J. Chem. Eng. Jpn.*, 2023, **56**, 2210195.
2 S. Anantharaj and S. Noda, *Energy Adv.*, 2022, **1**, 511–523.
3 S. Yagi, I. Yamada, H. Tsukasaki, A. Seno, M. Murakami, H. Fujii, H. Chen, N. Umezawa, H. Abe, N. Nishiyama and S. Mori, *Nat. Commun.*, 2015, **6**, 8249.
4 K. Sugahara, K. Kamata, S. Muratsugu and M. Hara, *ACS Omega*, 2017, **2**, 1608–1616.
5 Y. Sugawara, K. Kamata, E. Hayashi, M. Itoh, Y. Hamasaki and T. Yamaguchi, *ChemElectroChem*, 2021, **8**, 4466–4471.
6 L. Vallez, S. Jimenez-Villegas, A. T. Garcia-Esparza, Y. Jiang, S. Park, Q. Wu, T. M. Gill, D. Sokaras, S. Siahrostami and X. Zheng, *Energy Adv.*, 2022, **1**, 357–366.
7 I. C. Man, H.-Y. Su, F. Calle-Vallejo, H. A. Hansen, J. Martinez, N. G. Inoglu, J. Kitchin, T. F. Jaramillo, J. K. Nørskov and J. Rossmeisl, *ChemCatChem*, 2011, **3**, 1159–1165.
8 H. Y. Li, Y. B. Chen, S. B. Xi, J. X. Wang, S. N. Sun, Y. M. Sun, Y. H. Du and Z. C. J. Xu, *Chem. Mater.*, 2018, **30**, 4313–4320.
9 Y. Sugawara, H. Kobayashi, I. Honma and T. Yamaguchi, *ACS Omega*, 2020, **5**, 29388–29397.
10 Y. Sugawara, K. Kamata, A. Ishikawa, Y. Tateyama and T. Yamaguchi, *ACS Appl. Energy Mater.*, 2021, **4**, 3057–3066.
11 E. Lv, J. Yong, J. Wen, Z. Song, Y. Liu, U. Khan and J. Gao, *Energy Adv.*, 2022, **1**, 641–647.
12 J. Suntivich, K. J. May, H. A. Gasteiger, J. B. Goodenough and Y. Shao-Horn, *Science*, 2011, **334**, 1383–1385.
13 W. T. Hong, K. A. Stoerzinger, Y. L. Lee, L. Giordano, A. Grimaud, A. M. Johnson, J. Hwang, E. J. Crumlin, W. L. Yang and Y. Shao-Horn, *Energy Environ. Sci.*, 2017, **10**, 2190–2200.
14 W. T. Hong, R. E. Welsch and Y. Shao-Horn, *J. Phys. Chem. C*, 2016, **120**, 78–86.
15 E. Y. Konysheva, X. X. Xu and J. T. S. Irvine, *Adv. Mater.*, 2012, **24**, 528–532.
16 A. Grimaud, K. J. May, C. E. Carlton, Y. L. Lee, M. Risch, W. T. Hong, J. G. Zhou and Y. Shao-Horn, *Nat. Commun.*, 2013, **4**, 2439.
17 X. Cheng, E. Fabbri, M. Nachtegaal, I. E. Castelli, M. El Kazzi, R. Haumont, N. Marzari and T. J. Schmidt, *Chem. Mater.*, 2015, **27**, 7662–7672.
18 I. Yamada, A. Takamatsu, K. Asai, T. Shirakawa, H. Ohzuku, A. Seno, T. Uchimura, H. Fujii, S. Kawaguchi, K. Wada, H. Ikeno and S. Yagi, *J. Phys. Chem. C*, 2018, **122**, 27885–27892.
19 C. Bloed, J. Vuong, A. Enriquez, S. Raghavan, I. Tran, S. Derakhshan and H. Tavassol, *ACS Appl. Energy Mater.*, 2019, **2**, 6140–6145.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 24 mis Gortheren 2023. Downloaded on 10/12/2025 01:57:44.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

20 D. Q. Guan, J. Zhou, Y. C. Huang, C. L. Dong, J. Q. Wang, W. Zhou and Z. P. Shao, *Nat. Commun.*, 2019, **10**, 3755.
21 X. Li, H. T. Zhao, J. Liang, Y. L. Luo, G. Chen, X. F. Shi, S. Y. Lu, S. Y. Gao, J. M. Hu, Q. Liu and X. P. Sun, *J. Mater. Chem. A*, 2021, **9**, 6650–6670.
22 H. Yamada, C. Liu, S. Wu, Y. Koyama, S. H. Ju, J. Shiomi, J. Morikawa and R. Yoshida, *ACS Central Sci.*, 2019, **5**, 1717–1730.
23 A. Ishikawa, K. Sodeyama, Y. Igarashi, T. Nakayama, Y. Tateyama and M. Okada, *Phys. Chem. Chem. Phys.*, 2019, **21**, 26399–26405.
24 C. Li, D. R. D. Leal, S. Rana, S. Gupta, A. Sutti, S. Greenhill, T. Slezak, M. Height and S. Venkatesh, *Sci. Rep.*, 2017, **7**, 5683.
25 X. F. Ma, Z. Li, L. E. K. Achenie and H. L. Xin, *J. Phys. Chem. Lett.*, 2015, **6**, 3528–3533.
26 K. Suzuki, T. Toyao, Z. Maeno, S. Takakusagi, K. Shimizu and I. Takigawa, *ChemCatChem*, 2019, **11**, 4537–4547.
27 J. Y. Peng, D. Schwalbe-Koda, K. Akkiraju, T. Xie, L. Giordano, Y. Yu, C. J. Eom, J. R. Lunger, D. J. Zheng, R. R. Rao, S. Muy, J. C. Grossman, K. Reuter, R. Gomez-Bombarelli and Y. Shao-Horn, *Nat. Rev. Mater.*, 2022, **7**, 991–1009.
28 B. C. Weng, Z. L. Song, R. L. Zhu, Q. Y. Yan, Q. D. Sun, C. G. Grice, Y. F. Yan and W. J. Yin, *Nat. Commun.*, 2020, **11**, 3513.
29 X. Jiang, Y. Wang, B. Jia, X. Qu and M. Qin, *ACS Omega*, 2022, **7**, 14160–14164.
30 Y. Sugawara, S. Ueno, K. Kamata and T. Yamaguchi, *ChemElectroChem*, 2022, **9**, e202101679.
31 K. Momma and F. Izumi, *J. Appl. Crystallogr.*, 2011, **44**, 1272–1276.
32 R. Kronberg, H. Lappalainen and K. Laasonen, *J. Phys. Chem. C*, 2021, **125**, 15918–15933.
33 L. van der Maaten, *J. Mach. Learn. Res.*, 2014, **15**, 3221–3245.
34 R. D. Shannon, *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.*, 1976, **32**, 751–767.
35 S. M. Lundberg and S. I. Lee, A Unified Approach to Interpreting Model Predictions, *31st Annual Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA, 2017.
36 J. C. Liang and X. Zhu, *J. Phys. Chem. Lett.*, 2019, **10**, 5640–5646.
37 J. J. Goings and S. Hammes-Schiffer, *ACS Central Sci.*, 2020, **6**, 1594–1601.
38 W. J. Yin, B. C. Weng, J. Ge, Q. D. Sun, Z. Z. Li and Y. F. Yan, *Energy Environ. Sci.*, 2019, **12**, 442–462.
39 C.-Z. Yuan, S. Huang, H. Zhao, J. Li, L. Zhang, Y. Weng, T.-Y. Cheang, H. Yin, X. Zhang and S. Ye, *Energy Adv.*, 2023, **2**, 73–85.

© 2023 The Author(s). Published by the Royal Society of Chemistry