Supporting information

for

Data-Driven Design of Electrocatalysts: Principle, Progress, and Perspective

Shan Zhu[1], Kezhu Jiang[1], Biao Chen[2#], and Shijian Zheng[1#]

[1] State Key Laboratory of Reliability and Intelligence of Electrical Equipment, School of Materials Science and Engineering, Hebei University of Technology, Tianjin 300401, P. R. China
[2] School of Materials Science and Engineering, Tianjin University, Tianjin, 300072, P. R. China
[#]Corresponding Authors

We have constructed a database of different types of catalyst systems and their data-driven analysis methods extracted from the literature (Table S1). This database contains more than just electrocatalysts. Each paper is indexed by its DOI, which makes it easy for readers to search. Some of the abbreviations involved in the table are as follows: K-nearest neighbor regression (KNN), random forest regression (RF), support vector machine (SVM), gradient boosting regression (GBR), extreme gradient boosting regression (XGB), gaussian process regression (GPR), neural network (NN), convolutional neural network (CNN), smooth overlap of atomic positions (SOAP), sure independence screening and sparsifying operator (SISSO), principal component analysis (PCA), shapley additive explanation (SHAP), t-distributed stochastic neighbor embedding (t-SNE), least absolute shrinkage and selection operator (LASSO).

Table S1. Catalytic material systems from the literature and their data-driven analysis methods

DOI	Published year	Active materials	Supports	Catalyst system/Devices	Output feature/Analysis target	Algorithm	Database
10.1016/j.cplett.2004.07.097	2004	PtRu	None	Proton exchange membrane fuel cell	Free formation energy	Bayesian-guided stochastic surface walking-NN	DFT calculation
10.1021/acs.accounts.1c00153	2021	Transition metals and transition metal oxides	None	CO methanation and OER	Faradaic efficiency, overpotential, current density	GBR, RF, etc.	Data mining from published papers
10.1002/cjce.23756	2019	Metal	SiO_2, TiO_2, Al_2O_3	SO_2 conversion reaction	Formation energies and binding energies	SOAP	DFT calculation
10.1002/cctc.202100495	2021	Mn	Na_2WO_4/SiO_2	Oxidative coupling reaction of methane	Degradation index	NN	Experimental test
10.1021/acs.jpclett.0c01128	2020	NRR catalysts	None	NRR, HER	Selectivity	SVM, RF, etc.	Experimental test
10.1016/j.jpowsour.2019.226916	2019	Pt	Commercial supports	Fuel cell	Faradaic efficiency and total current density	NN	Calculation
10.1021/jacs.1c00339	2021	Cu-based catalysts	None	CO_2RR	Specific activity	RF	Zeolite Structures Database of IZA

10.1021/acs.jpcc.0c05964	2020	Metal and nonmetal elements	N-doped graphene	CO_2RR	XANES spectra	NN, PCA	Theoretical XANES calculated by FEFF9 code
10.1039/d0ta08004g	2020	BAC	N-doped graphene	ORR	Thermodynamic stability	Genetic algorithm	DFT calculation
10.1021/acs.jctc.9b00420	2019	Metal	None	Carbon reforming	The feature of nanoparticles	ML-accelerated genetic algorithms	PtAu 147 atom icosahedron simulation
10.1039/d1cp02198b	2021	Nanoparticles	None		Surface segregation energies	GBR	DFT calculation
10.1002/cctc.202101481	2022	SAC	None	Ethanol dehydrogenation	Spectroscopic signal	Automatic target generation process pre-conditioned Joint-nonnegative matrix factorization	Electron energy-loss spectroscopy (EELS) dataset
10.1002/smtd.202100035	2021	Pd	TiO_2		SO_2 conversion, catalytic productivity, etc.	NN	Data mining from 31 published papers
10.1021/acs.iecr.8b04801	2019	NiPt	None	Ammonia decomposition	Power density, output voltage, etc.	NN, XGB, DT, etc.	Data mining from 64 published papers; Experimental test
10.1038/s41524-021-00514-8	2021	Alloy	None	HER	Power density of devices	NN, DT, XGB, SHAP, etc.	Experimental test of 126 records corresponding to different membrane electrode assembly preparation and test

							parameters
10.1021/jacs.0c11261	2021	Pt nanowire	None	HER	Peak power density	XGB	339 original reports from experimental tests
10.1021/acs.jcim.1c00726	2021	Metal nanoparticles	None	HER	Overpotential	NN, SVM, KNN, CNN, etc.	Published data sets with about 6000 samples
10.1021/acs.jpcc.1c05470	2021	Pt(111)	None		Overpotential	Sequential learning	High throughput experiments
10.1021/acs.jpcc.1c02890	2021	Pt-based alloy	None	ORR	Overpotential	RF	DFT calculation
10.1039/d1nr07661b	2022	$Pd_{40}Ni_{10}Cu_{30}P_{20}$	None	HER	Overpotential	CNN	Materials Project
10.1002/er.7602	2022	Pd	Carbon nanotubes	Formic acid fuel cell	O binding energy	Regression algorithms	DFT calculation
10.1021/acsami.1c23221	2022	Pt-based catalysts	Commercial supports	Fuel cell	O adsorption energy	CNN	DFT calculation
10.1038/s41598-021-81403-4	2021	Cu	Zeolite	Direct oxidation of methane to methanol	NO conversion rate	GBR, SVM, XGB, KNN, etc.	Data mining from over 2000 published papers
10.1021/acs.jpclett.1c01851	2021	Pt-based alloy	None		NH3 production rates	RF	Thermodynamic data from Materials Project.

10.1021/acs.jpclett.9b03678	2020	Mn	Na_2WO_4/SiO_2	Oxidative coupling reaction of methane	Nanoparticles	NN	DFT calculations
10.1039/d0cy01751e	2021	Metal and oxide	None	Low-temperature oxidative coupling of methane (OCM)	Nanoparticle sizes	Unsupervised learning	TEM results from experiments
10.1021/acscatal.9b02416	2019	IrO_X	None	OER	CO_2 conversion rates	KNN, DT, SVM, etc.	Data mining from 100 published papers
10.1016/j.ces.2021.116902	2022	Ru	$(BaO)_x(CaO)_y(Al_2O_3)_z$	Ammonia synthesis	Mass activity	GBR, SVM	Experimental test
10.1021/acs.chemmater.9b03686	2020	Metal	None	NRR	Lattice constants and formation energies	Genetic algorithm NN	DFT calculation
10.1016/j.matt.2020.11.010	2021	Au	None	CO_2RR	Interpret the EXAFS	NN	Molecular dynamics simulation
10.1039/c9ta13404b	2020	SAC	N-doped graphene	ORR, OER, HER	Intermediate binding energies	GPR	DFT calculation
10.1002/adfm.202100547	2021	Co SAC	N-doped carbon	HER	Infrared spectrum of CO	NN	DFT calculation
10.1021/acs.jpclett.9b01428	2019	Materials in the Materials Project	None		Hydrogen yield	PCA, NN, SHAP	Data mining from 46 published papers

		database					
10.1002/cjoc.202100352	2021	$Au_{13}Pt_{42}$	None	ORR	HER activity	Adaptive Neuro-Fuzzy Inference optimized by Gray Wolf Optimization, NN, SVM, etc.	DFT calculation
10.1016/j.jmat.2021.02.005	2021	MOF	None	CO_2 fixation	Half-Wave Potential	PCA, NN, SHAP, etc.	Data mining from published papers
10.1021/jacs.8b08800	2018	AgAu alloy nanoclusters	None		H binding energy	NN	DFT calculation
10.1063/1.5126597	2019	CuO	None	Methanation reaction	H binding energy	Genetic algorithm	Experimental test
10.1038/s41467-021-22204-1	2021	Pt	None		Adsorption energy of *H	ML-based method BEEF-vdW	DFT calculation
10.1021/acsami.1c20289	2021	Pt	Commercial supports	Fuel cell	Adsorption energy of *H	SVM NN	CASTEP calculation
10.3390/en13092182	2020	Alloy	None		Adsorption energy of *H	RF, etc.	DFT calculation
10.1002/aic.17041	2020	Pt(111)	None		Fingerprint of XANES	Extra trees regressor and classification	Ab initio XANES simulations
10.1021/acsami.1c06714	2021	CuPd alloy	Al_2O_3		Faradaic efficiencies	DT, RF, XGB, etc.	Molecular structure data, experimental test

10.1039/d1re00396h	2022	Binary alloy	None	Ethanol dehydrogenation	Energy barriers and *CO adsorption energy	GPR	DFT calculation
10.1021/acscombsci.0c00102	2020	Alloy	None	HER	Energy	NN	Data from stochastic surface waling (SSW) method
10.1021/acs.jpclett.0c00634	2020	Alloy, Metal	None	CO_2RR	Diffusion coefficient	Classification algorithm	Experimental test
10.1002/anie.202006928	2020	Non-precious metals catalysts	Commercial supports	Fuel cell	Current density	SVM, etc.	Experimental test
10.1021/acs.jpcc.8b09284	2018	Metal	None		Cu coordination number	NN	XANES spectra
10.1021/acscatal.9b04343	2020	High entropy alloys	None	CO_2RR	CO conversion rate	DT, NN	Experimental test
10.1021/acs.jpclett.1c01319	2021	Alloy	None	Multi-reactions	CO conversion	NN, SVM, etc.	Data mining from published papers
10.1021/acs.chemmater.6b04229	2017	Alloy	None	CO2RR	CO and H binding energy	Crystal graph convolutional neural network	Materials Project
10.1021/acs.jpca.9b07569	2020	Transition metals	None		Cleavage energy	CNN	DFT calculation including 36 kinds of elements
10.1021/jacs.1c04624	2021	Fe-based catalysts	None	Fischer-Tropsch synthesis	Classification of Au nanoparticles	Extra trees regressor, t-SNE	1300 Au nanoparticles generated from molecular dynamics (MD) trajectories

10.1016/j.cej.2021.131285	2021	Metal-based catalysts	None	Supercritical water gasification (SCWG)	Binding energies of CH_x, NH_x, OH_x, and OOH adsorbates	SVM, RF, Elastic net, etc.	Alchemical perturbation DFT (APDFT) calculation
10.1039/d1ta06772a	2021	Mn-Ce-M(M=Co, Cu, and Fe) mixed oxide system	None	Selective catalytic reduction (SCR) of NOx	Catalytic activity and selectivity	LAASO	VASP calculation
10.1039/d0ta04615a	2020	(100) surface of Alloy	None	HER	Carbon deposition resistance	PCA, SVM, GPR, etc.	Experimental test
10.1039/d1ta09184k	2022	Alloy	None	ORR	Calculation accuracy	Active Learning	Kinetic Monte Carlo (KMC) calculation
10.1039/c7sc03422a	2018	Alloy	None	CO_2RR	C_2H_6 yield	SVM	Experimental test
10.1021/acs.jcim.8b00657	2019	Ag-based alloy	None		C_2 yield	XGB, SHAP, etc.	Data mining (4759 OCM reaction data)
10.1021/acs.jpclett.0c03465	2021	Mn, Ti, Pd,	Na_2WO_4/SiO_2	Methane oxidation	C_2 yield	Gaussian mixture model	Experimental test of 58 OCM catalysts
10.1039/d0sc06502a	2021	Au nanoparticles	TiO_2	CO oxidation reaction	Binding energy of H*, O*, OH* and OOH*	RF	DFT calculation
10.1039/c9ta07651d	2020	Alloy and SAC	None	Ethanol decomposition, non-oxidative	Binding energy of CO/H	CNN, graph representation of bulk crystals	Materials Project

				dehydrogenation			
10.1021/acsomega.1c01726	2021	Pt	None		Binding energies	GBR, etc.	CatApp, DFT calculation
10.3390/molecules26216362	2021	Metal active sites	None	OER	Binding energies	GBR	DFT calculation
10.1021/acs.jpclett.7b02010	2017	RhAu Alloy	None	NO decomposition	Average charge difference	NN	DFT calculation of 3388 adsorption free energies
10.1021/acscatal.1c02029	2021	Ag	None	Ethene epoxidation	Atomic structure	Deep learning	Atomic electron tomography data
10.1038/s41598-021-00031-0	2021	Metal	N-doped graphene	HER	Aggregation energies and *O adsorption energy	GPR, etc.	DFT calculation
10.1021/acscatal.0c02158	2020	Pd	None	Selective acetylene hydrogenation	Adsorption energy; Formation energy	ReQM, NN	ReQM calculation
10.1021/acs.jpclett.1c03395	2021	Pt	None	HER	Adsorption energy of *O and *C	GBR	Calculation
10.3389/fchem.2019.00444	2019	Ni_5P_4	None	HER	Adsorption energy of *H, *N and *NH_x	CNN	High-throughput DFT computation for nitrogen adsorption on various compositions of CoMoFeNiCu HEA

					(111) surfaces		
10.1021/acscatal.9b01985	2019	(Ni-Fe-CoCe)O_x	Commercial supports	Water oxidation	Adsorption energy of *H, *O, *CO	NN, Crystal Graph	DFT calculation and Materials Project
10.1021/jacs.1c02471	2021	PdAg alloy	TiO2	Acetylene selective hydrogenation	Adsorption energy of *H and *CO	SSW-NN, LASP	Potential energy surface PES data obtained from the SSW global PES sampling
10.1021/jacs.0c06401	2020	Pd, Ag	Ag		Adsorption energy of *H	SOAP, KRR	DFT calculation
10.1021/acs.jctc.9b00986	2020	Pt(111)	None		Adsorption energy of *H	NN	DFT calculation
10.1038/s41524-019-0181-4	2019	PtAu	None		Adsorption energy of C intermediates	NN	247 C4 species；29 smaller C2 and C3 species
10.1021/acs.nanolett.1c01516	2021	Au nanoparticles	Amorphous silicon nitride		Adsorption energy of *OOH, *O, and *OH	RF	DFT calculation
10.3390/cryst11091035	2021	High Entropy Oxide	None		Adsorption energy of *OH	GBR	DFT calculation
10.1021/acs.jpcc.0c04432	2020	SAC	N-doped graphene	HER	Adsorption energy of *CO, *OH	NN	DFT calculation
10.1021/acs.jp	2021	High entropy	None	Ammonia	Adsorption energy of	GPR	DFT calculation of the

clett.1c01242		alloy		decomposition	*CO and *H		HEAs' (111) facets
10.1021/acs.jpcc.1c09082	2021	Ru-based catalysts	None		Adsorption energy of *CO	RF	DFT calculation
10.1021/acscatal.0c02089	2020	Pt-based alloy	None	Ethanol Reforming	Adsorption energy of *CO	Regression algorithms	DFT calculation
10.1021/acs.chemmater.7b01860	2017	Metal in zeolites	Zeolites	NRR, CO2RR	Adsorption energy of *CO	GBR	DFT calculation
10.1039/c7ta01812f	2017	Bimetallic alloy	None	Methane oxidation	Adsorption energy of *H	SISSO	DFT calculation
10.1039/c9sc05999g	2020	Oxides	None	OER	Adsorption energy	XGB, SISSO	Calculation of adsorption energies containing 506 datasets on TMs, 178 datasets on TM clusters, 281 datasets on SAAs, 258 datasets on ABs, and 1178 datasets on HEAs
10.1063/5.0009129	2020	Au nanoparticles	None	HER, HOR, ORR	Adsorption energy	NN with 36453 network parameters	DFT calculation
10.1038/s41467-020-15340-7	2020	Pt	None	CO adsorption	Adsorption energy	Lasso, SVM, GB, etc.	Calculation
10.1021/jacs.9b04956	2019	Au nanoparticles	None	CO2RR	Adsorption energy	KRR, GP, SVM, etc.	DFT calculation

10.1002/slct.201902627	2019	Additives of Ni/Al_2O_3	Al_2O_3	CO methanation	Adsorption energy	kinetic Monte Carlo (KMC), CNN, SOAP, t-SNE	DFT calculation
10.1016/j.cherd.2012.08.017	2013	Au, Au/MO_x (M = Mg, Mn, Co, Ce, Fe, Ni)	Al_2O_3	CO oxidation	Adsorption energy	Kernel Ridge Regression, Active Learning	DFT calculation
10.1021/acscatal.1c01473	2021	ZIF-derived catalysts	Carbon	ORR	Adsorption energy	IsoMap, GPR, PCA	DFT calculation
10.1021/acs.jpcc.0c01492	2020	SAC	None	Multi-reactions	Adsorption energy	GPR，KRR, etc.	DFT calculation
10.1038/s41467-020-20342-6	2021	Alloy	None		Adsorption energy	Genetic algorithm	DFT calculation
10.1002/cctc.202000536	2020	SAC	N-doped graphene		Adsorption energy	CNN	High-throughput dataset containing 37,000 adsorption energies on 2000 unique bimetallic alloy surfaces
10.3389/fenrg.2021.609070	2021	Metal		CO_2RR	Adsorption energy	Bootstrapped Projected Gradient Descent	DFT calculation of 298 alloys with 13 descriptors
10.1021/acscatal.1c01018	2021	Pt	None	HER	Adsorption energy	KRR, GBR, RFR, etc.	Computational Materials Repository (CMR)
10.1039/c9cy02070e	2020	Cu SAC and alloy	None		Adsorption energy	SOAP, GPR	DFT calculation of 540 sites
10.1016/j.ijhy	2014	Pt/Au-based	Al_2O_3, MgO,	Water gas	Adsorption energy	NN	DFT calculation

dene.2014.01.160		catalysts	etc.	shift reaction			
10.1021/acs.jpcc.6b12800	2017	Pt nanocluster	None	Adsorption of CO	Adsorption Energy	LASSO, SVM, NN, PCA, etc.	DFT calculation
10.1021/jacs.7b07139	2017	Fluid catalytic cracking particles	Commercial supports	Fluid catalytic cracking	Adsorption Energies of H, N2, N2H, NH, and NH2	CNN	DFT calculation of 3040 surfaces in 465 catalysts
10.1021/acsaem.0c01466	2020	Fe, Zn	None	ORR	*CO binding energy	NN, KRR	DFT calculation of 263 alloys
10.1016/j.cattod.2016.04.013	2017	Transition metals	None	CO_2RR	*CO adsorption energy	NN	DFT calculation
10.1021/acs.jcim.9b00550	2019	Alloy	None		Gibbs free energy change of CO adsorption	XGBR, KNR, RF, SVM, GBR, etc.	VASP calculation
10.1039/C8CP03801E	2018	Pt-based alloy	None	ORR	Adsorption enthalpy	SISSO	Data mining from published papers
10.1021/acsami.1c16696	2021	Cu/Co/Ni/Zn/Mg-based alloy	None	CO_2RR	CO adsorption energy and HOCO transition state formation energy	NN	DFT calculation
10.1021/acsami.1c22309	2021	Transition metals	None	OER, ORR	Electrocatalytic activity	NN	Ab-initio molecular dynamics (AIMD) and DFT calculation

Table S2. Typical machine learning software or data analysis platform

Name	Website	Description
Python	https://www.python.org/	A basic programming language.
Scikit-learn	https://scikit-learn.org/	The machine learning tools are based on Python.
Julia	https://julialang.org/	An open-source programming language for data science.
PyTorch	https://pytorch.org/	An optimized tensor library for deep learning.
TensorFlow	https://www.tensorflow.org/	A machine learning platform.
Jupyter Notebook	https://jupyter.org/	A web service for interactive computing.
Orange	https://orangedatamining.com/	An open source for machine learning and data visualization.
Weka	https://waikato.github.io/weka-wiki/	An open-source for machine learning.
Matlab	https://www.mathworks.com	A business mathematical software for data analysis.
Theano	https://pypi.org/project/Theano/	A Python library for fast numerical calculation.