nature reviews chemistry

https://doi.org/10.1038/s41570-025-00740-4

Supplementary information

Developing machine learning for heterogeneous catalysis with experimental and computational data

In the format provided by the authors and unedited

Machine learning for the experimental and computational development of heterogeneous catalysis

Carlota Bozal-Ginesta[1,2‡*], Sergio Pablo-García[2‡], Changhyeok Choi[2], Albert Tarancón[1,4*], Alán Aspuru-Guzik[2,5,*]

[1] Nanoionics and Fuel Cells group, Catalonia Institute for Energy Research, Jardins de Les Dones de Negre 1, 08930 Sant Adrià de Besòs, Barcelona, Spain

[2] Department of Chemistry, University of Toronto, Lash Miller Chemical Laboratories, 80 St George Street, Toronto, ON M5S 3H6, Canada; Department of Computer Science, University of Toronto, 40 St. George Street, Toronto, ON, M5S 2E4 Canada

[3] Vector Institute for Artificial Intelligence, 661 University Ave. Suite 710, Toronto, ON, M5G 1M1 Canada

[4] ICREA, Passeig Lluís Companys 23, Barcelona 08010, Spain

[5] Department of Materials Science & Engineering, University of Toronto, 184 College St., Toronto, ON, M5S 3E4 Canada; Department of Chemical Engineering & Applied Chemistry, University of Toronto, 200 College St., Toronto, ON, M5S 3E5 Canada; Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), 66118 University Ave., Toronto, M5G 1M1 Canada; Acceleration Consortium, 80 St George St, Toronto, M5S 3H6 Canada

[‡] These authors have contributed equally

* Corresponding authors: Carlota Bozal-Ginesta (carlota.bozalginesta@empa.ch), Albert Tarancón (atarancon@irec.cat), Alán Aspuru-Guzik (aspuru@utoronto.ca)

Additional Box. Abbreviations for the machine learning methods cited in this review.

ANN: Artificial Neural Network
ABR: Adaptative Boosting Regressor.
BLR: Bayesian Linear Regressor
BO: Bayesian Optimization
DNN: Deep Neural Network
DSTAR: DFT and structure-free active motif-based representation
DTR: Decision Tree Regression
ETC: Extra Tree Classifier
ETR: Extra Tree Regression
GA: Genetic Algorithms
GBR: Gradient Boosting regression
GNN: Graph Neural Network
GPR: Gaussian Process Regression
KNR: K-Neighbor Regression
KRR: Kernel Ridge Regression
LASSO: Least Absolute Shrinkage and Selection Operator
LGBM: Light Gradient Boosting Machine
LR: Linear Regression
MLPR: Multilayer Perceptron Regression
MLM: Multi-Linear Model
MLR: Multi-Linear Regression
OLS: Ordinary Least Squares
PLS: Partial Least Squares
RFR: Random Forest Regression
RRF: Regularized Random Forest
SHAP: SHapley Additive exPlanations
SISSO: Sure Independence Screening and Sparsifying Operator
SVR: Support Vector Regression
t-SNE: t-Distributed Stochastic Neighbour Embedding

Table S1. Comparison of results from machine-learning-driven approaches based on computational parameters. The abbreviations HER, OER, ORR stand for hydrogen evolution reaction, oxygen evolution reaction and oxygen reduction reaction, respectively; RMSE, RMSD and MAE stand for root mean square error, root mean square deviation and mean absolute error, respectively.

Method	Descriptor / Parameter	Type of prediction pattern	Material / Device / Reaction	Samples	Result / Achievement	Reference
Active learning with KRR	d-band width of the muffin-tin orbital theory and electronegativity	Predict CO binding energy and suggest new catalysts (e.g. Cu3Sc@Cu, Cu3Y@Cu)	CO_2 reduction	263 fcc(100) surface of alloys	CO: 0.05 eV (RMSE)	Noh et al., 2018[1]
Active learning with ensembles	Atomic number, Pauling electronegativity, the number of coordinated atoms with the adsorbate, Voronoi tesselation, median adsorption energy between the adsorbate and pure elemen84t	Identify 131 candidate surfaces for CO2 reduction and 258 candidate surfaces for HER	CO_2 reduction, HER	1,499 different intermetallic combinations across 31 different elements (~20,000 for each adsorbate)	CO: 0.46 eV, H: 0.41 eV (RMSE)	Tran et al., 2018[2]
DNN	d-band, work function, radius, IP, EA, EN	Predict CO binding energy	CO_2 reduction	fcc(100)-terminated alloys	CO: 0.13 eV (RMSE)	Ma et al., 2015[3]
KNR, RFR, SVR, GBR, XGBR,	Atomic and composite features	Gibbs free energy for change of CO adsorption	CO_2 electroreductio n	171 SAC samples on different transition metals	Best, RF: RMSE: 0.1876 eV R^2: 0.875	Chen et al., 2020[4]
GBR, KRR, SVR, GPR, DTR, ETC, RFR, ABR, MLPR,	Atomic, electronic, geometric and adsorbate descriptors	CO* binding energy	CO_2 reduction	fcc(100) facets 263 samples surface alloys	Best, GBR: RMSE 0.07 eV, MAE: 0.048 eV, R^2: 0.982	Yang et al., 2020[5]
ANN, GPR, RFR, PCA.	Atomic, adsorbate and surface descriptors from online databases, and surface and electronic descriptors from DFT	Carbon, hydrogen and oxygen adsorption	Atomic adsorption energies	fcc(111) monometallic surfaces	Best GPR: RMSE: 0.36 eV (C), 0.16 eV (H) and 0.35 eV(O).	Tomacruz et al., 2022[6]
Active learning. t-SNE	Atomic and geometrical descriptors	Experimental performance	CO_2 electro-reduction	~4,000 adsorption sites of Cu containing alloys.	Experimental synthesis of better catalysts.	Zhong et al., 2020[7]
High-dimensional neural network potential (using DNN)	Atomic and geometrical descriptors	Phonon dispersion and N_2 dissociation barriers on Ru(0001)	Gas-Surface Dynamics	25000 DFT data points. Nitrogen over surface and disordered Ru(0001)	RMSE: 0.038 eV	Shakouri et al., 2017[8]
Active learning, SISSO	Adsorption enthalpies, atomic, bulk surface and connectivity features	Prediction of adsorption energies of relevant	CO methanation and OER	1387 for transition metal and 684 for transition metal oxides	0.18 eV for both transition metal and transition	Andersen et al., 2021[9]

		adsorbates on doped transition metal and transition metal oxides			metal oxides datasets	
PCA, GPR	Compound compositional properties and DFT-calculated bulk electronic features	Adsorption energies of *O and *OH. Ultimately experimental performance.	OER	358 DFT data points of ABO_3-type cubic perovskites	0.46 eV	Li et al., 2020 [10]
MLR, DT, ANN	Geometrical and structural descriptors	Molar catalytic activity, hydrogenation/combustion selectivity and molar thermodynamic stability.	Hydrogenation, combustion	4258 theoretical and experimental data of symmetric and 4258 of non symmetric Pt nanocrystals	Best, ANN: R^2: ~0.99 for symmetric geometries and R^2: ~0.93 for asymmetric geometries.	Fernandez et al., 2017[11]
ANN, GNN, MLM	Proxy descriptors obtained from ANN models based on CGCNN[12] and PaiNN[13], which use geometric and atomic features	Oxygen evolution reaction intermediate binding energies	OER	19000 bulk perovskite structures from already published databases and in-house DFT	R^2: 0.872	Lunger et al., 2023[14]
Active learning, GPR	DFT relaxation data	Pourbaix diagrams for ammonia synthesis in IrO_2 and MoS2	Ammonia synthesis reaction	500 configurations for IrO_2, and 126 for MOS_2	DFT evaluations, IrO_2: 93 manual, 20 active learning. MoS_2: 36 active learning	Ulissi et al., 2016[15]
SISSO	Atomic, bulk, surface and site descriptors	C, CH, CO, H, O, OH binding energies	OER, HER	884 DFT samples: 344 for pristine metals, 281 for SA alloys and 259 for AB alloys	RMSE: 0.18 eV	Andersen et al., 2019[16]
RRF	Atomic and surface descriptors, mainly geometrical	Hydrogen binding energy	HER	55 DFT Ni2P(0001) structures	RMSE: 0.09 eV	Wexler et al., 2018[17]
Unsupervised ILS. ETC and ETR.	Structural features based on atomic, crystallographic and topological descriptors	Molar catalytic activity proxies	OER, HER, ORR	1300 surfaces of Pt nanoparticles from molecular dynamics trajectories. Class 1: 100 microstructures, 200 facets. Class 2: 500 microstructures, 200 facets	Class 1: microstructures R^2: 0.978 RMSE: 0.02, facets R^2: 0.987 RMSE: 0.02. Class 2: microstructures R^2: 0.888 RMSE: 0.06, facets R^2: 0.978 RMSE 0.02	Parker et al., 2020[18]
ANN	Atomic features, neighbour features, connectivity distance	H and CO binding energies	General catalysis	20,771 surfaces for CO adsorption and 22,361 for H	MAE: 0.15 eV	Back et al., 2019[19]

				adsorption		
ANN	Geometry features *via* cartesian coordinates proxy	DFT local optimization	Pt structure	Pt9: 21,970 DFT relaxation steps. Pt13: 19,742 DFT relaxation steps	Pt_9 RMSE: 0.131-0.262 eV Pt_{13} RMSE: 298 eV	Zhai et al., 2016[20]
BLR	DFT energies of metal surfaces	Binding energies of NO, O and N on different-sized RhAu alloys	NO decomposition.	> 1000 DFT calculations	Error depending on the particle size. MAE large clusters: 0.10-0.15 eV, small clusters: 0.20-0.40 eV	Jinnouchi et al., 2017[21]
GNN	Molecular graph structure	DFT binding energies	Adsorption of organic molecules on pristine metals	2,853 adsorption energies of small organic molecules Target 30 big molecules test set	Small organic molecules: MAE: 0.18 eV Big molecules: MAE: 0.48 eV, RMSE: 0.53 eV, R^2: 0.83	Pablo-García et al. 2023[22]
DNN	Geometric features or electronic features	DFT binding energies	CO_2 electro-reduction	~250 ab initio adsorption energies on bimetallic alloys composed by VII and IB group metals	Geometric features, RMSE: 0.12 eV Electronic features, RMSE: 0.13 eV	Li et al., 2017[23]
DNN	Shape effects (morphology and size) and alloy effects (composition and configuration)	Neural network potential for DFT energies. Ultimately targeting experimental performance of oxygen reduction reaction	ORR	44,884 DFT atomic local environments for PtFeCu nanoparticles	RMSE: 10meV · $atom^{-1}$	Chun et al, 2021[24]
DNN	Symmetry function vectors from cartesian coordinates	Neural network potential for DFT energies. Ultimately targeting oxygen reduction reaction	ORR	33,877 DFT samples for PtNi, PtCu, CuNi and PtCuNi	RMSE: < 3meV· $atom^{-1}$	Kang et al, 2018[25]
GA, ETR, RFR	Synthesis and structural features	Catalytic performance through proxies	HER, OER, ORR	690 disordered platinum nanoparticles generated from classical MD	S_{fac} predictions, MAE: 0.24, RMSE: 0.33, R^2: 0.889 C_{20} prediction, MAE: 0.22, RMSE: 0.32, R^2: 0.897	Sun et al, 2018[26]
Active learning. ABR, GBR, KNR, LASSO, PLS, RFR, Ridge, Elastic-Net	Atomic, structural and electronic	Hydrogen evolution reaction activity using Gibbs free energy as proxy	HER	1125 H-binding energies on MM'XT2-type MXenes from DFT	Best, GBR: MAE: 0.358 eV, R^2: 0.826	Abraham et al, 2023[27]

DNN	Atomic and structural	Neural network potentials for DFT energies	CuAu nanoparticles	24,995 DFT calculations for alloy bulk and surface slab	MAE: 4.34 meV· $atom^{-1}$ RMS: 6.81 meV· $atom^{-1}$	Artrith et al, 2015[28]
MLR	Proxy descriptors based on group additivity and type of binding	DFT formation enthalpies	Thermochemical features of lignin monomers	591 DFT calculations	MAE: 2.81 kcal/mol in $\Delta H_{f,298}$, 1.07 cal/(mol K) in ΔS_{298} and 0.25 cal/(mol K) in $C_{p,300}$	Gu et al., 2016[29]
OLS	Atomic and structural parameters	DFT binding energies for *OH and *O and different binding sites of surface	ORR	1869 DFT calculation of high-entropy alloys (IrPdPtRhRu) s: 871 for *OH and 998 for *O	RMSD: 0.063 eV for *OH binding and 0.076 eV for *O binding.	Batchelor et al., 2019[30]
ANN, KRR, SVR, RFR	Atomic and structural features. Experimental and DFT	Gibbs free energy of hydrogen adsorption (ΔG_{H^*})	HER	299 DFT calculations of MXene materials, M being a transition metal and X being C or N	Best, RFR: RMSE: 0.18 eV, R^2: 0.96	Zheng et al., 2020[31]
ANN	Interatomic distances to compute two body and three body terms	CO binding energy and HOCO binding energy	CO_2 reduction	2488 DFT calculation of AuNps and dealloyed Au surfaces: 1384 for CO and 1104 for HOCO	CO, RMS:E 0.0521 eV HOCO: RMSE: 0.0614 eV	Chen et al., 2019[32]
ANN, unsupervised K-means clustering.	Synthetic IR spectra	Binding type and generalized coordination number	Pt structure	1090 DFT calculations of CO and NO coverage on platinum surfaces	RMSE: ~0.055%	Lansford et. al., 2020[33]
ANN	Theoretical XANES	XANES prediction	XANES autoencoder	~1253 theoretical XANES for the formation of Pd hydrade on Pd nanoparticles	MSE (reconstruction): ~0.003	Routh et al., 2021[34]
Bootstrapped Projected Gradient Descent	Atomic and structural descriptors	Binding energy of CO	CO_2 electrorreduction	298 metal alloys	MAE: 0.104 eV	Pankajakshan et al., 2017[35]
Manifold learning, PCA, GPR	Atomic, thermodynamic and electronic properties	Kinetic parameters, binding and binding and reaction energies	General catalysis	420 DFT calculations	*	Dasgupta et al., 2020[36]
GA	Structural candidates	$TiCl_4$-capped $MgCl_2$ structural determination	Ziggler-Natta catalyst	N/A	Structural candidates	Takasao et. al., 2021[37]
PCA, ANN, LGBM	Atomic, structural and electronic properties	ANN: candidate probability	ORR	500 samples of Boron-doped single atom	Binding energies, LGBM, RMSE:	Zafari et. al., 2020[38]

				catalyst	0.11 eV	
ANN, active learning	DFT calculation	Adsorbate configurations, Most active site prediction, associated CO binding energy	CO_2 electroreductio n	583 DFT relaxations of bimetallic NP: (training) 70000 DFT single-points (test)	RMSE 0.2 eV	Ulissi et. al., 2017[39]
RFR	Structural properties	CO adsorption energy	Binding energy of CO	Thiolated Ag-alloyed Au nanoclusters: Au_{25}: 4500 isomers $Au_{36}(SR)_{24}$: 1360 isomers $Au_{133}(SR)_{52}$: 1898 isomers	Au_{25}, R^C: 0.77869, MAE: 0.13196, RMSE: 0.17348 $Au_{36}(SR)_{24}$, R^2: 0.65388, MAE: 0.1856, RMSE: 0.21582 $Au_{133}(SR)_{52}$, R^2: 0.75063, MAE: 0.11882, RMSE: 0.17128 Log used for the predictions	Panapitaya et. al., 2018[40]
ANN	Atomic and structural properties	DFT energies and forces	Water coverage	9258 Cu and Cu2O surface structures. 1035 used as test set	Energies RMSE: 0.9 meV/atom Forces RMSE: 66.3 meV/Bohr	Natarajan et al., 2016[41]
ANN	Atomic and structural and atomic properties	DFT adsorption energies as proxy for thermodynamic and coverage predictions	Oxygen interactions. General catalysis.	11925 Pd bulk and surface (111) structures	RMSE: 0.009 eV/atom	Boes et al., 2017[42]
Transfer Learning	Structural, atomic, electronic and macroscopic properties. DFT and experimental	Band gaps of crystalline compounds and enthalpies, activation energies and rate determining step (RDS) for catalysed NO reduction	NO Reduction	Crystalline solids: 69640 DFT calculations. 1449 experimental measurements. NO reduction: 4000 DFT calculations	Crystalline solids: Best Latent Variable (Exp) [F1 Score]: 0.3, NO reduction: Best: Activation energies: Difference Learning [RMSE]: 0.85 eV, RDS series [F1 Score]: 0.93	Hutchinson et al., 2017[43]
KRR	Structural properties	Strained coordinations as proxy for mass activity	OER	812 platinum core-shell nanoparticles with Au, Ag, Pd, Cu and Ni as core metals	Strain on single atoms: MAE: Pt@Ag: 0.0015, Pt@Au: 0.0015, Pt@Pd: 0.0007, Pt@Cu: 0.0064 Pt@:Ni: 0.0159	Rück et al., 2020[44]

LASSO	Structural and electronic properties	O, OOH and OH binding energies using ηOER	HER, OER	48 DFT calculations of transition metal dichalcogenides	R^2: 0.83	Ge et al., 2020[45]
BLR	Structural and atomic properties	Binding energy of N, O, NO on and formation energies	NO decomposition	DFT of RhAu alloy nanoparticles data: Binding energies: N: 973 samples O: 1261 samples NO: 648 samples Formation energies: 1487 samples	Binding energies (surface, nanoparticle): N, MAE: (65, 87) meV, R^2: 0.99, O, MAE: (57, 77) meV, R^2: 0.99, NO, MAE: (68, 66), R^2: (0.99, 0.98) Formation energy (surface, nanoparticle): (6, 23) meV, R^2: (1.00, 0.97)	Jinnouchi et al., 2017[46]
LR, MLR, ANN, SISSO	Structural, atomic and thermodynamic properties	Activation energies	O_2 and N_2 dissociation	315 DFT transition states on pristine metallic surfaces	Best: ANN, MAE: 0.19 eV	Singh et al., 2019[47]
LR, RFR, SVR	Structural, atomic and thermodynamic properties	Activation energies	Methane dehydrogenation	788 DFT transition states on Pd(100)	Best: RF, Mean score: 0.94	Takahashi et al., 2018[48]
KRR, LASSO, Elastic Net	Functional forms of binding energy (E_{bind}) of a metal atom on a support and cohesive energy (E_c) of the bulk metal	Activation energies for diffusion	Diffusion activation energies	99 DFT calculations of single atom transition metals on graphene and metal oxides support	Best: Elastic net, RMSE: 0.198 eV, R^2: 0.96	Su et al., 2020[49]
ANN	Structural and electronic properties	CO and O_2 adsorption energies	CO and O_2 binding	91 DFT calculations on Au_2 and Au_{10} clusters.*	CO, RMSE: 0.163 eV, R^2: 0.806, O_2: RMSE: 0.257, R^2: 0.873	Davran-Candan et al., 2010[50]
GPR, PCA	Group additivity fingerprints	Intermediate energies as proxy for reaction path search	Syngas	99 DFT calculation on Rh(111)	Uncertainty: 0.05 eV	Ulissi et al., 2017[51]
OLR, RFR, GBR, ETR	Atomic, electronic and thermodynamic properties	Binding energies of CH_4, CH_3, CH_2, CH, C and H	Methane dehydrogenation	46 samples, handbook properties and DFT of Cu-based alloys	Best ETR and RFR, RMSE: 0.24 eV	Toyao et al., 2018[52]
ANN	Structural and thermodynamic properties.	CO_2 and H_2 uptake.	CO_2 reduction	164 silica zeolites	R^2, CO_2: 0.86016, H_2: 0.95767	Thornton et al., 2015[53]
ANN	Structural coordinates	Potential energy surface	CO_2 reduction	20000 DFT calculations of	RMSE: 12.4 meV/atom	Artrith et al., 2014[54]

				bimetallic Au/Cu nanoparticles.	MAE: 6.0 meV/atom	
LS-CGCNN-ens	Label site (LS) representation	CO and H binding energies	CO_2 reduction	~40,000 alloys	CO: 0.116 eV, H: 0.085 eV (MAE)	Gu et al., 2020[55]
LR, RFR, LASSO, Elastic net, KRR, SVR, GPR, KNN, RFR, ANN	d-band features, local electronegativity and atomic features	CO and OH binding energies	Methanol oxidation	~1,000 alloy surfaces	Best, RF: ~ 0.2 eV of RMSE for CO and OH	Li et al., 2017[56]
Modified DSTAR	Atomic features and active motif representations	Binding energy of CO_2 reduction intermediates	CO_2 reduction	DFT calculated binding energies of 2,271 (H), 1,938 (CO), 1,678 (CHO) and 2314 (OH) on transition metal nitrides	0.269 ~ 0.375 eV of MAE	Yohannes et al., 2023[57]
RFR, ETR, LGBR, GBR, LASSO, SVR	Atomic features and the number of atoms in the first and second nearest atoms in the same and sublayer	CO and OH binding energies	CO_2 reduction	GASpy dataset[2] (18,455 and 21,317 for CO and H)	Best, GBR, CO: 0.120 and H: 0.105 eV (MAE)	Mok et al., 2021[58]
ANN	Atomic and structural properties	H adsorption energy	HER	2460 H adsorption energies on (111) alloys surfaces	0.156 eV	Chen et. al., 2022[59]
Active learning, GPR, PCA	Chemical and structural information	Synthesizable IrO_3 polymorphs, formation enthalpies as proxy	OER	956 structurally unique IrO_2 and IrO_3 polymorphs (from 3800 AB_2 and AB_3 prototypes)	75 synthesizable IrO_3 polymorphs	Flores et. al., 2020[60]
KRR	Density of states	CH_4 activation barrier as proxy for microkinetic simulation	Methane oxidation	199 Projected DOS for three atoms in rutile oxide. 39 as the test set	MAE: 0.08 eV	Bang et. al., 2022[61]
ANN	Atomic and structural properties	H, N_2, N_2H, NH, NH_2 binding energies	N_2 reduction reaction.	3040 DFT surface simulations	MAE: 0.23 eV RMSE: 0.12	Kim et. al., [62]
ANN	DFT structure descriptors of different Pd-C-H compositions	DFT energy, force and stress	Acetylene semihydrogenation.	30,138 DFT Pd surface structures sampled using stochastic surface walking.	Good accuracy of NN with respect to DFT; Energy: 2.0 meV, Force: 0.058 eV/Å, Stress: 0.388 GPa Experimental validation	Li et al., 2020[63]
GNN	Atomic and bonding features.	Energy of formation per atom (Ef/atom). Proxy for electrochemical stability	OER	36,000 mixed metal oxides	MAE: 0.77 eV, RMSE: 0.108 eV, R^2: 0.95	Abed et al., 2024[64]
LR	DFT binding energies	OH and H binding energies.	ORR	3317 binding sites.	RMSD of 0.115 and 0.132 eV	Batchelor et al., 2021[65]

					for *OH and *O	
LR	DFT binding energies, explore catalytic activity	O binding energies	ORR	450 DFT samples	*	Svane et al., 2022[66]
GBR	DFT energies and atomic properties.	OH binding energies.	ORR	10,440 Ir-Pt-Ru-Rh-Ag-Fe candidates	R^2: 0.961 RMSE: 0.112 eV	Wan et al., 2022[67]
GNN, BO	Multi-objective BO. GNN trained with structural properties.	GNN targets OH and O binding energies as a surrogate for Bayesian optimization	ORR	In domain dataset: 19,955 data points: AgIrPdPtRu, 5,039 AuOsPdPtRu, 7,461 CuPtReRhRu, 7,455 Out of domain dataset: 4,020 data points, Composition-divers 2,138 (RuRhPdIrPt). Component-diversity subset: 1,882 data points (100 random 5-element combinations).	GNN, MAE (out-of-domain dataset), OH: 0.039 eV, O 0.051 eV	Xu et al, 2024[68]
GPR, BO	Structural properties	Current density	ORR	150 points sampled from the chemical space. 1000 random samples used for evaluating the acquisition function.	Acceleration on the discovery of quinary HEAs for ORR.	Pedersen et al., 2021[69]
Active learning with Gaussian approximation potentials (GAP) and Minima Hopping	SOAP descriptors (Smooth Overlap of Atomic Positions), two-body and many-body terms in GAP	Adsorption energies and global minimum geometries of adsorbates on catalyst surfaces	Syngas and ethanol synthesis	18 adsorbate/surface combinations (13 on Rh(111), 5 on Rh(211))	Efficiently found global minimum adsorption geometries with high accuracy using minimal DFT calculations.	Jung et al., 2023[70]
LLM, GNN	Structural descriptors, GNN embeddings	Adsorption energies of configurations	General adsorption/binding energies	ML relaxed adsorption energies. Training: Combination of approximately 476,000 structures (460,000 from OC20 and 16,000 from OC20-Dense). Validation: 920 ML-relaxed structures with DFT energies	MAE: 0.346 eV R^2: 0.882	Ock et al., 2024[71]
GNN, ensemble learning	Structural and atomic properties.	Classify configurational stability of surface–adsorbate systems from initial structures	General stability	Surface–adsorbate systems from the Open Catalyst 2020 (OC20) Training set: 460,328 structures Validation set: 99,854 structures Test set: 99,828 structures	Ensemble, F1: 0.922, AUROC: 0.906 MCC: 0.633	Noh et al., 2023[72]

KRR	Coulomb matrix, bag of bonds, and spectrum of london and axilrod-teller-muto potential (SLATM)	Metal/substrate oxidative addition process energy	C–C cross-coupling reactions.	7054 DFT calculations of organometallic compounds.	Best, SLATM, MAE: 2.61 kcal/mol	Meyer et al., 2018[73]

* Not properly reported

Table S2. Comparison of results from machine-learning-driven approaches based on experimentally-measured parameters.

Method	Descriptor / Parameter	Type of prediction - pattern	Material / Device / Reaction	Number of samples	Result / Achievement	Reference
ANN (back propagation, 3 layers)	Al_2O_3:SiO_2 ratio; Cu loading Reaction temperature [O_2] and [NO]	Reaction conversion yield of NO to N_2	NO decomposition into N_2 and O_2; Cu-ZSM-5 zeolite catalyst; Bed flow reactor	32, 100, 60	MSE=1-5% conversion*	Sasaki et al., 1995[74]
RFR, GBR, XGBoost, SHAP, Pearson Correlation	Catalyst properties (metal content, covalent radius of metal, promoter content and e, molecular weight of support, surface area); synthesis conditions (calcination temperature and time); reaction conditions (T, P, GHSV)	Methanol space-time yield	Thermocatalytic CO_2 hydrogenation to methanol; atalyst: Cu, Pd, In_2O_3, ZnO-ZrO_2	1201	R^2=0.81, RMSE= of 0.11 g_{MeOH} h^{-1} $gcat^{-1}$ to predict experimental data GHSV, P and metal content important	Suvarna et al., 2022[75]
RFR, GBR, SHAP, Pearson Correlation, clustering	Catalyst descriptors (crystal radius, Pauling's electronegativity, valence s/p/d/f electrons, specific heat, thermal conductivity cohesive and allow formation energy); reaction conditions (T, GHSV, H_2/CO ratio)	Higher alcohols space-time yield	Rh-Mn-P/SiO_2 catalyst (P=16 promoters) for higher alcohol synthesis from syngas	189 on 14 promoters 36 on 5 promoters (validation set)	R^2=0.76 against validation set; promoter's cohesive energy and alloy formation energy important	Suvarna et al., 2022[76]
ANN	Feed composition (C, H, N, O, ash content); operational conditions (T, P, time, solid content, catalyst ratio); catalyst descriptors (molecular mass, valence, ionization energy, electron affinity, conductivity)	H_2, CH_4, CO_2 and CO yields	Water gasification for H_2-rich syngas production	718	$R^2 > 0.85$	Li et al., 2021[77]
PCA, DT, RFR, Bayesian Machine Scientist	PCA components of 272 DFT descriptors	Adsorption energies (E_{ads}) of Br and CH; Experimental conversion, selectivity and yield	CH_2Br_2 hydrodebromination; SiO_2-supported catalysts based on Fe, Co, Ni, Cu, Ru, Rh, Ag, Ir and Pt	9	2 PCA components necessary to predict E_{ads}; mathematical equation established to predict experimental values from theoretical energies	Saadun et al., 2020[78]
PCA, Bayesian Machine Scientist	592 DFT descriptors (72 adsorption energies and geometries on 8 metallic surfaces)	Experimental conversion, selectivity and yield and rate	Hydrodehalogenation of CH_2X_2 (X = Br,Cl) to 3 main products SiO_2-supported catalysts based on Fe, Co, Ni, Cu, Ru, Rh, Ag, Ir and Pt	9	Mathematical equation established to predict experimental values from theoretical energies	Pablo-García et al., 2022[79]

PCA, ANN	Catalyst descriptors (metal loading and binding energy of C; promoter electronegativity, covalent and ionic radius, lowest and highest oxidation state, redox potential, MW, and loading; support redox potential, electronegativity, and highest and lowest oxidation state; support metal 1st ionization energy; entire support and support element MW); reaction conditions (calcination time and T; reaction T; H_2, H_2O, CO_2 and CO feeds; time on stream and F/W)	Reaction rate	Low-temperature water gas shift reaction with a primary metal (7) + a promoter (x24) + a support (x18)	2228	MSE reaches 0.2 after 30% of experimental data used; most important descriptor I T, followed by F/W gas feed and C binding energy; the worst descriptors are related to promoters and support metals	Smith et al., 2022[80]
Linear regression, t-test	Elemental composition; reaction temperature; reactant partial pressure ratio (p_{CH4}/p_{O2}); division in 10 groups based on 4 hypotheses	C_2 (ethane, ethylene) yield	Oxidative coupling of methane	1208	p-values < 0.005	Schmack et al., 2019[81], Palkovits, 2020[82], Zavyalova et al., 2011[83]
ANN (RBF), DTR	Elemental composition	C_2 yield	Oxidative coupling of methane	<1870	Experimental validation	Kondratenko et al., 2015[84], & Pirro et al., 2019[85]
PCA, ANN	Loading of Ni, mesoporous silica SBA-15, SiO_2, **γ** and α alumina, Bo, Mo, Zr, Ce; T, P, S:C ratio, Time-on-stream, catalyst reduction temperature	Methane conversion	Methane steam reforming	200	$R^2 > 0.99$ The most influential factors were time-on-stream, T and support SBA-15	Arcotumapathy et al., 2012[86]
ANN, LR, PCA	XRD (PCA components); elemental composition; lattice Parameter; phase distribution	Polarization resistance (at 450 and 650ºC) and activation energy	Oxygen electrocatalysts $Ln_{0.58}Sr_{0.4}Fe_{0.8}Co_{0.2}O_{3-\delta}$ (Ln = $La_{1-x-y-z}Pr_xSm_yBa_z$)	24	R^2>0.9 to predict electrochemical properties from composition and XRD*	Serra et al., 2011[87]
Heuristic for Variable Selection, ANN	Synthesis and characterization variables (theoretical Ti and SiR_3 loadings, type of silylating agent, hydrophobicity); molecular descriptors of the silylating agents: constitutional (atomic Sanderson electronegativities, atomic polarizabilities, electrotopological state, aromatic ratio, and number of bonds which can rotate); geometrical descriptors (average geometric distance degree, spin ratio, spherocity, asphericity, Petitjean shapes, and aromacity); and molecular properties (unsaturated index, and hydrophilic factor)	Initial epoxidation reaction rate	Olefin epoxidation with a support + a silylating agents SiR_3 (R x4) + TiO_2	58	R^2=0.91 max with 4 descriptors (3-fold CV)	Baumes et al.,2010[88]
SV, classification trees	Epoxidation: [Ti], [OH], [trimethylammonium (TMA)], [cetylTMA]; isomerization: type of support, type of enhancer, content of enhancer, type of promoter and content of promoter (as integers)	Epoxide yield (2 classes); isopentane yield (2 classes)	Cyclohexene epoxidation catalysts based on mesoporous titanium silicate materials; N-pentane paraffin into branched isomers with a metal-oxide based support, an acidity enhancer and 2 promoters	26 (epoxidation); 30 (isomerization)	0.92 recognition rate max. (5,10-fold CV with SVR) for epoxidation; 0.85 recognition rate max. (5,10,15-fold CV with SVR) for isomerization	Baumes et al., 2006[89,90]
PCA, Kohonen networks, clustering,	[Ti], [OH], [trimethylammonium (TMA)], [cetylTMA]; XRD spectrum	Epoxide yield	Cyclohexene epoxidation catalysts based on	63	100 ± 4% best prediction accuracy	Corma et al., 2005[91,92]

logistic regression equation, ANN, DTR, GA			mesoporous titanium silicate materials		of training and test data (5-fold CV)	
ANN	Elemental composition	CO conversion (number and class)	Water-shift reaction with a support, a metal oxide, a metal and an alkaline additive (6 elements max.)	344	R^2=0.8 max. for classification; Poor for regression	Baumes et al., 2004[93]
ANN	Molar composition of thirteen elements	C_2H_6 and O_2 conversion; C_2H_4 yield; C_2H_4, CO_2, and CO selectivity	Oxidative dehydrogenation of ethane	329	Mean of MAE = 11.27	Corma et al., 2002[94]
ANN, GA, PCA, clustering (hierarchical, k-means)	56 descriptors of physico-chemical properties (oxide formation enthalpy, possible coordination numbers, ionization energies, electronegativities, catalysts, average and variance of these values in multi-component catalysts); 17 discrete variables for the synthesis conditions (synthesis and calcination type, support, precursors and additives	Oxidation of propene and selectivity in 21 products in 5 T	Oxidation of propene by multicomponent catalysts	467	0.96 max. ratio of correct predictions to observation standard deviations on the test subset (with NN); In >50% of the cases, the optimization results are better than random	Farrusseng et al., 2005 and 2009[95,96]
ANN (MLPR and radial basis function)	Catalyst metal loading; feed ratio; reaction temperature	H_2 and CO yields; CH_4 and CO_2 conversion	Hydrogen-rich syngas from methane dry reforming over $Ni/CaFe_2O_3$ catalysts	27	R^2=0.86-.097	Hossain et al., 2016[97]
ANN	T; P; GHSV; Gas ratio	CO_2 conversion; Product selectivity (C_1, CO, C_2-C_4, C_{5+})	CO_2 hydrogenation in a microchannel reactor using a K-promoted iron-based catalyst	17	MSE > 99.9%	Sun et al., 2018[98]
Simple and multiple OLS regression, forward and backward elimination, penalized methods, Latent variable regression	14 theoretical material descriptors (d and e_g electrons, optimality of e_g, oxidation state, optimality of tolerance factor, bond lengths, Madelung potentials, ionization energy, Hubbard U, charge-transfer energy, magnetic moment)	Relative oxygen evolution reaction activity	Perovskite electrocatalysts for oxygen evolution reaction	101 activities in 51 perovskites	Best model based on Least Angle Regression (MAE <0.5), with the number of d electrons, charge-transfer energy (covalency), optimality of e_g occupancy and structural factors play an important role	Hong et al., 2016[99]
ANN	Catalyst design (calcination T and time, dopant oxidation number, dopant ionic radius, dopant concentration, specific surface area, substrate molecular weight); reaction conditions (peak wavelength emitted by light source, substrate initial concentration, photocatalyst loading, photon flux)	Bandgap; reaction rate	Photodegradation of aqueous organics by metal ion doped TiO_2 powders	<700	Correlation Coefficient > 0.92	Chesterfield et al., 2009[100]
XGBoost, SHAP	Total catalyst loading; elemental composition; T and P_{EtOH}/P_{Ar}	Ethanol conversion; acetaldehyde, butene, ethylene and 1,3-butadiene yield.	Ethanol-to-butadiene catalysts comprising up to 14 elements	276	24 catalysts discovered with 1,3-butadiene yield above 60%	Jayakumar et al., 2023 [101]
ANN (MLPR)	Proportion of elements in the catalyst	Yield of propene	Oxidative dehydrogenation of propane to propene	226	MAE = 0.3%	Holeňa et al., 2003[102]
Linear ensemble, GPR, RFR	Elemental composition	Overpotential	Quaternary electrocatalysts for oxygen evolution reaction	2121	Research can be accelerated by up to a factor of 20	Rohr et al., 2020[103]

					compared to random acquisition; MAE < 0.04V after >100 learning cycles	
Bayesian ANN (Gemini)	Elemental composition; price	Overpotential	Electrocatalyst for oxygen evolution reaction	2121	20 times less expensive evaluations than at random to achieve an overpotential of 0.4V	Hickman et al., 2021[104]
LR, SVR, KRR, RFR	Catalyst design (metal atomic number and amount, support type); experimental conditions (T, Ar, CH_4 and O_2 flow)	C_2 yield; Selectivities of CO, CO_2, C_2H_3, C_2H_6	Oxidative coupling of methane	12708	R^2 = 0.70-0.88 with RFR	Nguyen et al., 2020[105]
RFR, LR, SVR	Activation T and time; impregnation ratio; BET surface area	Pore volume; mesoporosity	Mesoporous activated carbon as catalyst support for Fischer-Tropsch synthesis	36	R^2 = 0.88 and 0.97 for pore volume and mesoporosity respectively with RFR	Teimouri et al., 2023[106]
OLS and PLS regression, GPR, GBR	Elemental information (group, bulk Wigner–Seitz radius, atomic number, atomic mass, period, electronegativity, ionization energy, enthalpy of fusion, density at 25 º)	d-band center	Metal catalysts and their pairwise bimetals	121	RMSE < 0.5 eV with GBR	Takigawa et al., 2016[107]
ANN, decision trees, SVR	Composition (cathodic and anodic Electrocatalyst elements, composition, and loading); operating conditions (humification, T, fuel type, fuel alcohol concentration, fuel and oxidant flow rate, O_2 content in the oxidant, cathode P)	CO conversion	Water gas shift reaction over Pt and Au based catalysts	4360	R^2 =0.905, RMSE=10.05 with NN, with the reaction T, the support type being the most important	Odabaşı et al., 2022[108]
decision trees, RFR	Catalyst variables (element and weight of elements A, B and X in ABX3 material, preparation method, calcination time and T, promoter loading method, element and weight); reaction conditions (light type, catalyst weight, alcohol percentage)	Hydrogen production rate; bandgap	Photocatalytic water splitting over perovskites	540	R^2 = 0.79, RMSE = 1194 µmol $gcat^{-1}h^{-1}$ for the hydrogen production rate with NN R^2 = 0.61, RMSE = 0.51 eV for the bandgap with NN	Can et al., 2019[109]
Decision trees	Components (cathode catalyst elements; support, membrane and cell type; backing layer, and catholyte and anolyte type and max. molar concentration); operating variables (CO_2 flux, pH, applied potential, catalyst loading, and catholyte and anolyte flux)	Maximum rate (classes); maximum faradaic efficiency (classes)	Electrocatalytic reduction of CO_2	471	69-86% and 83-92% classification accuracy for the max. faradaic efficiency and rate, respectively*	Günay et al., 2018[110]
ANN, decision trees	Preparation variables (preparation method, metal and support type, calcination T and time, reduction T, time and volume percentage of H_2); operating variables (reaction T, P, W/F, volume percentage of CH_4, CO_2, Ar, He and N_2 in reaction stream, time on stream)	CH_4 conversion	Dry reforming of methane	5521	R^2 =0.89, RMSE=8.66 with NN Reaction T, metal and support the most important	Şener et al., 2018[111]
Decision trees	Preparation variables (preparation method; base metal, promoter, reactor and support type; calcination T and time; reduction T, time and volume percentage of H_2, Ar, He and N_2); operating variables (P, T, W/F, Time on stream, volume percentage of CH_4, H_2O, H_2, and CO_2)	CH_4 conversion (classes)	Steam reforming of methane	5508	83-90% classification accuracy*	Baysal et al., 2017[112]

ANN, decision trees	Composition (cathodic and anodic electrocatalyst elements, composition, and loading; ion conducting polymer membrane material) Operating conditions (humification, T, fuel type, fuel alcohol concentration, fuel and oxidant flow rate, O_2 content in the oxidant, cathode P)	Current density -voltage (JV) curve; maximum power density	Direct alcohol fuel cells with different designs and operating conditions	271	R^2 >0.90 with NN for the JV curve, with the current density and the elemental content of the electrocatalyst being the most important	Tapan et al., 2016[113]
ANN, decision trees	Preparation variables (preparation method, catalyst composition and loading, support type and loading, calcination time and T, alcohol type, and fatty acid composition); operating variables (reaction T and time, alcohol:oil ratio, stirring speed, cycle number)	Fatty acid conversion	Fatty acid conversion	1324	R^2 = 0.87 with NN and the most important variables being the reaction time and catalyst loading	Tapan et al., 2016[114]
ANN Decision trees	Promoter type; Reaction T; H_2O and CO_2 content	CO conversion	CO oxidation over $Au/MO_x/Al_2O_3$ catalysts with promoters M = Mg, Mn, Co, Ce, Fe, Ni	252	RMSE = 1.60% with NN	Günay et al., 2013[115]
ANN, decision trees, clustering	Preparation variables (metal, promoter and support type, preparation method, calcination time and T); operating variables (reaction T, volume percentage of CO_2, H_2, O_2, H_2O, and CO, time on stream, F/W)		CO oxidation over noble metal catalysts	4585	R^2 =0.874, RMSE=12.23 with NN and the promoter type and amount being the most important	Günay et al., 2013[116]
ANN	Preparation variables (Au load, preparation method, calcination T and time, support type, promotor type and load); operating variables (reaction T, volume percentage of CO, O_2, H_2, H_2O and CO_2, time on stream)	Reaction rate	CO oxidation kinetics over Au based catalysts	882	R^2 = 0.922 Reaction T and catalyst calcination conditions found to be the most important	Günay et al., 2013[117]
ANN	Preparation variables (Cu loading, preparation method, support, promoter type and loading, calcination time and T); operating variables (reaction T, W/F, Time on stream, volume percentage of CO_2, H_2, O_2, H_2O, and CO)	CO conversion	CO oxidation over Cu-based catalysts	1337	R^2 =0.944, RMSE=8.69	Günay et al., 2011[118]
ANN	Electronegativity of metal ions; partial charge of oxygen ions; valence; coordination number; ionic radius; 4^{th} ionization potential; enthalpy of formation	Yield of CO_2 (oxidation of butane); selectivities of styrene, benzaldehyde, benzene, toluene, CO and CO_2 (dehydrogenation of ethylbenzene)	Oxidation of butane with lanthanide oxides; oxidative dehydrogenation of ethylbenzene on promoted SnO_2 catalysts	12 (oxidation of butane); 18 (dehydrogenation of ethylbenzene)	>3% absolute error in the yield (oxidation of butane)*; averaged absolute error of 1.5% (dehydrogenation of ethylbenzene)*	Hattori et al., 1995[119]
PCA, PLS Regression	4 descriptors based on Slater-type orbitals; reaction T; catalyst amount; O_2 : n-butane	Butane conversion; 1,3-butadiene selectivity	Oxidative dehydrogenation of butane to 1,3-butadiene ¡ on bimetallic mixed oxides supported on alumina	15	R^2 = 0.865 for butane conversion and 0.610 butadiene selectivity	Madaan et al., 2016[120]
Orthogonal PLS	4 descriptors based on Slater-type orbitals; Reaction T	Yield	Catalytic hydrogenation of 5-ethoxymethylfurfural with M/Al_2O_3 (M = Au, Cu, Ni, Ir, Od, Pt, Rh, Ru)	>24	R^2 > 0.8 and RMSE > 7.5%	Ras et al., 2012[121]
PCA	Yields of 8 products	Effect of solvent, T and support	Glucose valorization via catalytic hydrogenation of furfural derivatives (x3 T, x8	112	2 principal components	Ras et al., 2010[122]

			main metals, x6 promotor metals, x2 solvents, x2 supports)		explain 68% of the overall variance	
Reshef algorithm	50 material properties	Exchange current density	Hydrogen evolution with metals	38	Bulk modulus and melting point have the strongest correlation to the exchange current density	Leonard et al., 2013[123]
ANN Response Surface Methodology	Glycerol %; catalyst loading; Pt%; pH	Amount of produced H_2	Glycerol valorization to H_2 with Pt/TiO_2 photocatalysts	29	R^2 = 0.9913 and MAE = 5.9310 µmol with NN	Estahbanati et al., 2017[124]
RFR, GPR, Lasso	DFT-based descriptors (binding energies, reaction energies, lattice parameters, electronegativities, transition states, activation energies)	Total activity (experimental); reforming activity and selectivity (experimental)	Ethanol reforming with bimetallic catalysts	78 (+7 experimental)	C-C bind scission reaction as key step; RMSE = 0.001 mL and 2.4% for the reforming activity and selectivity respectively	Artrith et al., 2020[125]
ANN	Catalyst preparation (calcination temperature, cation atomic weight and number, support active surface weight fraction and flow path, acid loading, catalyst weight fraction); reaction conditions (time, T; feedstock concentration, carbon and bond number; alcohol:feedstock ratio, alcohol concentration)	Conversion in esterification and transsterification	Biodiesel production with 12-Tungstophosphoric acid	26	R^2=0.99	Baroi et al., 2014[126]
LR, SVR, RFR	Content of straw; heating rate; pyrolytic temperature	Methylene blue number; iodine number	Preparation of straw activated carbon	30	$R^2 > 0.9$*; Random Forest the best performing method*	Jiang et al., 2019[127]
PCA, clustering, ANN, Classification tree	3179 attributes, which include the concentrations of 60 elements from the periodic table, 19 attributes which are related to the synthesis method, and 3100 attributes which are taken from tabulated data	5 clusters out of 120 parameters including conversions, selectivities to 21 products, temporal behavior, and carbon mass balance at each T	Oxidation of propene with oxygen	467	Prediction rate >0.5 with a NN	Klanner et al., 2004[128]
Random Forest Classification	Catalyst composition (Preparation method, cations and support type and amount); reaction conditions (T, contact time, CH_4 and O_2 pressures)	C_2 yield	Oxidative coupling of methane	1868	Average score of 73 ± 1%	Takahashi et al., 2018[129]
PCA, K-means, GPR	16 physicochemical properties of the element used as additive	Temperature at a CO conversion rate of 50%	Ni/Al_2O_3 catalyst for CO methanation with 1 additive	63	A strong effect of Re addition discovered; Thermal properties have the largest effect	Han et al., 2017[130]

* Not properly reported

Table S3. Data used to plot **FIG. 2a,b** compiling the data in **Tables S1-S2**, organized by the ML method and the modeling problem type -regression or classification- corresponding to the prediction of continuous values and to the prediction/classification of discrete values, respectively. The label *Artificial Neural Networks* includes the methods ANN, DNN, GNN, MLPR and LLM; *Random Forests & Decision Trees* includes DTR, ETR, RFR and RRF; and *Linear Regression* includes BLR, KRR, LASSO, LR, MLM, and MLR; *Gradient Boosting Regression* includes ABR, GBR and LGBM, as defined in **Box S1** and identified in **Tables S1-S2**.

Type R=Regression C=Classification	**Method**	**Experimental**			**Computational**		
		Publi-cations	**Average number of datapoints in a dataset**	**Datapoints**	**Publi-cations**	**Average number of datapoints in a dataset**	**Datapoints**
R	Gaussian Process Regression	3	87,3	63, 78, 121	7	392,3	956,99,420,500,358,263,150
R	Artificial Neural Networks	28	940,0	32,100,60,718,2228,1870,200,24,58,63,344,329,467,27,17,700,226,2121,4360,5521,271,1324,252,4585,882,1337,12,18,29,26	36	45137,5	2853,19000,4258,19000,20771,22361,21970,19742,299,1384,1104,1090,1253,500,583,9258,11925,315,91,164,20000,2460,3040,30138,263,25000,250,44884,33877,24995,40000,36000,23975, 476000
R	Random Forests & Decision Trees	9	1883,1	30,78,540,36,12708,2121,9,225,1201	7	922,7	55,171,263,1125,299,4500,46
R	Support Vector Regression	5	2865	30,36,12708,26,30,4360	6	6882,2	171,263,299,788,18455,21317
R	Linear Regression	11	1309,8	1208,24,63,101,12708,36,121,15,24,78,30	25	3689,3	263,4258,19000,1000,1125,591,1869,299,7054,812,48,973,1261,648,315,788,99,46,1000,18455,21317,190,3317,450,7054
R	Gradient Boosting Regression	3	515,7	1201,225,121	8	6539.6	171,263,1125,46,18455,21317,10440,500

C	PCA	11	319	9,9,2228,200,24,63,467,15,112,63	6	466,6	956,99,500,420,358
C	Clustering	6	900,5	189,36,63,467,4585,63	1	1090	1090
C	Random Forests & Decision Trees	14	1686,3	1868,4585,252,1324,271,5508,5521,471,540,4360,63,9,26,30,467	1	266,7	100,500,200

Table S4. Data used to plot **FIG. 2c** compiling the data in **Tables S1-S2** as a function of the reaction type.

Type	Reaction type	Dataset size
Computational	Water splitting	1499,2071,358,19000,884,55,100,200,500,200,44884,33877,690,1125,871,998,299,500,812,48,315,91,2460,956,36000,3317,450,10440,23975, 150
Computational	CO and CO_2 reactivity	263,1499,171,263,4000,2071,250,1384,1104,298,70000,4500,1360,1898,91,99,164,20000,40000,2271,1938,1678,2314,18455,21317,18
Computational	Nitrogen reactivity	25000,500,1000,69640,4000,973,1261,648,1487,315,3040
Computational	Methane reactivity	788, 46, 199, 1387, 684
Computational	Other	4258,4258,20771,22361,21970,19742,2853,24995,591,1090,1253,420,9258,11925,99,1000, 30138,476000,7054
Experimental	Water splitting	540,101,2121, 38,29
Experimental	CO and CO_2 reactivity	1201,225,252, 718,344,4360,4585,882,1337,189,36,718,2228,17,471
Experimental	Methane reactivity	200,27,12708,5521,1208,5508,1868,63
Experimental	Other	58,36,271,1324,15,24,26,30,9,63,329,467,700,276,226,121,12,18,112,78,467,32,100,60

Table S5. Data used to plot **FIG. 3a-d**, based on the information gathered in **Tables S1-S2**. In papers reporting more than one model based on different datasets, the average R^2 and average number of datapoints per dataset are calculated. The label *CO and CO_2 reactivity* includes any catalysis with CO_2 and CO as substrates like syngas reactions, except where methane is involved; *Methane reactivity* includes all the studies where methane is either reactant or product; *Water splitting* includes HER, OER, and ORR.

Data type	Reaction type	Average number of datapoints	Number of datapoints	Input	Average R^2	Output	Ref.
Computational	CO and CO_2 reactivity	171	171	atomic	0,875	adsorption	4
Computational	CO and CO_2 reactivity	263	263	atomic, electronic, structural	0,982	binding	5
Computational	Other	4258	4258	structural	0,96	activity	11
Computational	Water splitting	19000	19000	atomic, structural	0,872	binding	14
Computational	Water splitting	250	1000	atomic, composition	0,95775	activity	18
Computational	Other	2853	2853	structural, composition	0,83	binding	22
Computational	Water splitting	690	690	synthesis, composition	0,893	activity	26
Computational	Water splitting	1125	1125	atomic, composition and electronic	0,826	adsorption	27
Computational	Water splitting	299	299	atomic, composition	0,96	adsorption	31
Computational	CO and CO_2 reactivity	1629	3258	composition	0,702255	adsorption	40
Computational	Water splitting	48	48	composition, electronic	0,83	binding	45
Computational	Nitrogen reactivity	1092,25	4369	composition, atomic	0,9866666667	binding, formation energy	46
Computational	Other	99	99	thermodynamic	0,96	diffusion activation energy	49
Computational	CO and CO_2 reactivity	91	91	composition, electronic	0,8395	adsorption	50
Computational	CO and CO_2 reactivity	164	164	composition, thermodynamic	0,86016	adsorption	51
Computational	Water splitting	36000	36000	atomic, composition, structural	0,95	formation energy	64

Computational	Water splitting	10440	10440	atomic, composition, thermodynamic	0,961	binding	68
Computational	Other	476000	476000	structural	0,882	adsorption	71
Experimental	CO and CO_2 reactivity	1201	1201	synthesis, reaction, atomic	0,81	space time yield	75
Experimental	CO and CO_2 reactivity	225	225	atomic, reaction	0,76	space time yield	76
Experimental	CO and CO_2 reactivity	718	718	atomic, reaction, composition	0,85	yield	77
Experimental	Methane reactivity	200	200	composition, reaction	0,99	conversion	86
Experimental	Water splitting	24	24	structural, composition	0,9	polarization resistance	87
Experimental	Other	58	58	synthesis, atomic	0,91	rate	88
Experimental	CO and CO_2 reactivity	344	344	composition	0,8	conversion	93
Experimental	Methane reactivity	27	27	reaction	0,915	yield, conversion	97
Experimental	Methane reactivity	12708	12708	composition, reaction	0,79	yield, selectivity	105
Experimental	Other	36	36	structural, reaction	0,925	porosity	106
Experimental	CO and CO_2 reactivity	4360	4360	composition, reaction	0,905	conversion	108
Experimental	Water splitting	540	540	atomic, synthesis, composition, reaction	0,7	rate, bandgap	109
Experimental	Methane reactivity	5521	5521	synthesis, composition, reaction	0,89	conversion	111
Experimental	Other	271	271	composition, reaction	0,9	current density	113
Experimental	Other	1324	1324	synthesis, composition, reaction	0,87	conversion	114
Experimental	CO and CO_2 reactivity	4585	4585	synthesis, composition, reaction	0,874	conversion	116
Experimental	CO and CO_2 reactivity	882	882	synthesis, composition, reaction	0,922	rate	117
Experimental	CO and CO_2 reactivity	1337	1337	synthesis, composition, reaction	0,944	conversion	118
Experimental	Other	15	15	electronic, reaction	0,865	conversion, yield selectivity	120
Experimental	Other	24	24	electronic, reaction	0,8	yield	121
Experimental	Water	29	29	reaction	1	yield	124

	splitting						
Experimental	Other	26	26	synthesis, composition, atomic, reaction	0,99	conversion	126
Experimental	Other	30	30	reaction	0,9	porosity	127

References

1. Noh, J., Back, S., Kim, J. & Jung, Y. Active learning with non- *ab initio* input features toward efficient CO_2 reduction catalysts. *Chem. Sci.* **9**, 5152–5159 (2018).
2. Tran, K. & Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nat. Catal.* **1**, 696–703 (2018).
3. Ma, X., Li, Z., Achenie, L. E. K. & Xin, H. Machine-learning-augmented chemisorption model for CO_2 Electroreduction Catalyst Screening. *J. Phys. Chem. Lett.* **6**, 3528–3533 (2015).
4. Chen, A., Zhang, X., Chen, L., Yao, S. & Zhou, Z. A Machine learning model on simple features for CO_2 reduction electrocatalysts. *J. Phys. Chem. C* **124**, 22471–22478 (2020).
5. Yang, Z., Gao, W. & Jiang, Q. A machine learning scheme for the catalytic activity of alloys with intrinsic descriptors. *J. Mater. Chem. A* **8**, 17507–17515 (2020).
6. Tomacruz Jan Goran T., Pilario Karl Ezra S., Remolona Miguel Francisco M., Padama Allan Abraham B., & Ocon Joey D. A machine learning-accelerated density functional theory (ML-DFT) approach for predicting atomic adsorption energies on monometallic transition metal surfaces for electrocatalyst screening. *Chem. Eng. Trans.* **94**, 733–738 (2022).
7. Zhong, M. *et al.* Accelerated discovery of CO_2 electrocatalysts using active machine learning. *Nature* **581**, 178–183 (2020).
8. Shakouri, K., Behler, J., Meyer, J. & Kroes, G.-J. Accurate neural network description of surface phonons in reactive gas–surface dynamics: N_2 + Ru(0001). *J. Phys. Chem. Lett.* **8**, 2131–2136 (2017).
9. Andersen, M. & Reuter, K. Adsorption enthalpies for catalysis modeling through machine-learned descriptors. *Acc. Chem. Res.* **54**, 2741–2749 (2021).
10. Li, Z., Achenie, L. E. K. & Xin, H. An adaptive machine learning strategy for accelerating discovery of perovskite electrocatalysts. *ACS Catal.* **10**, 4377–4384 (2020).
11. Fernandez, M., Barron, H. & Barnard, A. S. Artificial neural network analysis of the catalytic efficiency of platinum nanoparticles. *RSC Adv.* **7**, 48962–48971 (2017).
12. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
13. Schütt, K. T., Unke, O. T. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. Preprint at http://arxiv.org/abs/2102.03150 (2021).
14. Lunger, J. R. *et al.* Atom-by-atom design of metal oxide catalysts for the oxygen evolution reaction with machine learning. Preprint at http://arxiv.org/abs/2305.19930 (2023).
15. Ulissi, Z. W., Singh, A. R., Tsai, C. & Nørskov, J. K. Automated discovery and construction of surface phase diagrams using machine learning. *J. Phys. Chem. Lett.* **7**, 3931–3935 (2016).
16. Andersen, M., Levchenko, S. V., Scheffler, M. & Reuter, K. Beyond scaling relations for the description of

catalytic materials. *ACS Catal.* **9**, 2752–2759 (2019).
17. Wexler, R. B., Martirez, J. M. P. & Rappe, A. M. Chemical pressure-driven enhancement of the hydrogen evolving activity of Ni_2P from nonmetal surface doping interpreted via machine learning. *J. Am. Chem. Soc.* **140**, 4678–4683 (2018).
18. Parker, A. J., Opletal, G. & Barnard, A. S. Classification of platinum nanoparticle catalysts using machine learning. *J. Appl. Phys.* **128**, 014301 (2020).
19. Back, S. *et al.* Convolutional neural network of atomic surface structures to predict binding energies for high-throughput screening of catalysts. *J. Phys. Chem. Lett.* **10**, 4401–4408 (2019).
20. Zhai, H. & Alexandrova, A. N. Ensemble-average representation of Pt clusters in conditions of catalysis accessed through GPU accelerated deep neural network fitting global optimization. *J. Chem. Theory Comput.* **12**, 6213–6226 (2016).
21. Jinnouchi, R., Hirata, H. & Asahi, R. Extrapolating energetics on clusters and single-crystal surfaces to nanoparticles by machine-learning scheme. *J. Phys. Chem. C* **121**, 26397–26405 (2017).
22. Pablo-García, S. *et al.* Fast evaluation of the adsorption energy of organic molecules on metals via graph neural networks. (2022) doi:10.26434/chemrxiv-2022-m719x.
23. Li, Z., Ma, X. & Xin, H. Feature engineering of machine-learning chemisorption models for catalyst design. *Catal. Today* **280**, 232–238 (2017).
24. Chun, H. *et al.* First-principle-data-integrated machine-learning approach for high-throughput searching of ternary electrocatalyst toward oxygen reduction reaction. *Chem. Catal.* **1**, 855–869 (2021).
25. Kang, J. *et al.* First-principles database driven computational neural network approach to the discovery of active ternary nanocatalysts for oxygen reduction reaction. *Phys. Chem. Chem. Phys.* **20**, 24539–24544 (2018).
26. Sun, B., Barron, H., Opletal, G. & Barnard, A. S. From process to properties: correlating synthesis conditions and structural disorder of platinum nanocatalysts. *J. Phys. Chem. C* **122**, 28085–28093 (2018).
27. Abraham, B. M., Sinha, P., Halder, P. & Singh, J. K. Fusing a machine learning strategy with density functional theory to hasten the discovery of 2D MXene-based catalysts for hydrogen generation. *J. Mater. Chem. A* **11**, 8091–8100 (2023).
28. Artrith, N. & Kolpak, A. M. Grand canonical molecular dynamics simulations of Cu–Au nanoalloys in thermal equilibrium using reactive ANN potentials. *Comput. Mater. Sci.* **110**, 20–28 (2015).
29. Gu, G. H. & Vlachos, D. G. Group Additivity for thermochemical property estimation of lignin monomers on Pt(111). *J. Phys. Chem. C* **120**, 19234–19241 (2016).
30. Batchelor, T. A. A. *et al.* High-entropy alloys as a discovery platform for electrocatalysis. *Joule* **3**, 834–845 (2019).
31. Zheng, J. *et al.* High-throughput screening of hydrogen evolution reaction catalysts in MXene materials. *J. Phys. Chem. C* **124**, 13695–13705 (2020).
32. Chen, Y., Huang, Y., Cheng, T. & Goddard, W. A. Identifying active sites for CO_2 reduction on dealloyed gold surfaces by combining machine learning with multiscale simulations. *J. Am. Chem. Soc.* **141**, 11651–11657 (2019).
33. Lansford, J. L. & Vlachos, D. G. Infrared spectroscopy data- and physics-driven machine learning for characterizing surface microstructure of complex materials. *Nat. Commun.* **11**, 1513 (2020).
34. Routh, P. K., Liu, Y., Marcella, N., Kozinsky, B. & Frenkel, A. I. Latent representation learning for structural characterization of catalysts. *J. Phys. Chem. Lett.* **12**, 2086–2094 (2021).
35. Pankajakshan, P. *et al.* Machine learning and statistical analysis for materials science: stability and

transferability of fingerprint descriptors and chemical insights. *Chem. Mater.* **29**, 4190–4201 (2017).
36. Dasgupta, A., Gao, Y., Broderick, S. R., Pitman, E. B. & Rajan, K. Machine learning-aided Identification of single atom alloy catalysts. *J. Phys. Chem. C* **124**, 14158–14166 (2020).
37. Takasao, G. *et al.* Machine learning-aided structure determination for $TiCl_4$–capped $MgCl_2$ nanoplate of heterogeneous Ziegler–Natta catalyst. *ACS Catal.* **9**, 2599–2609 (2019).
38. Zafari, M., Kumar, D., Umer, M. & Kim, K. S. Machine learning-based high throughput screening for nitrogen fixation on boron-doped single atom catalysts. *J. Mater. Chem. A* **8**, 5209–5216 (2020).
39. Ulissi, Z. W. *et al.* Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO_2 reduction. *ACS Catal.* **7**, 6600–6608 (2017).
40. Panapitiya, G. *et al.* Machine-learning prediction of CO adsorption in thiolated, Ag-alloyed Au nanoclusters. *J. Am. Chem. Soc.* **140**, 17508–17514 (2018).
41. Natarajan, S. K. & Behler, J. Neural network molecular dynamics simulations of solid–liquid interfaces: water at low-index copper surfaces. *Phys. Chem. Chem. Phys.* **18**, 28704–28725 (2016).
42. Boes, J. R. & Kitchin, J. R. Neural network predictions of oxygen interactions on a dynamic Pd surface. *Mol. Simul.* **43**, 346–354 (2017).
43. Hutchinson, M. L. *et al.* Overcoming data scarcity with transfer learning. Preprint at https://doi.org/10.48550/arXiv.1711.05099 (2017).
44. Rück, M., Garlyyev, B., Mayr, F., Bandarenka, A. S. & Gagliardi, A. oxygen reduction activities of strained platinum core–shell electrocatalysts predicted by machine learning. *J. Phys. Chem. Lett.* **11**, 1773–1780 (2020).
45. Ge, L. *et al.* Predicted optimal bifunctional electrocatalysts for the hydrogen evolution reaction and the oxygen evolution reaction using chalcogenide heterostructures based on machine learning analysis of in silico quantum mechanics based high throughput screening. *J. Phys. Chem. Lett.* **11**, 869–876 (2020).
46. Jinnouchi, R. & Asahi, R. Predicting catalytic activity of nanoparticles by a DFT-aided machine-learning algorithm. *J. Phys. Chem. Lett.* **8**, 4279–4283 (2017).
47. Singh, A. R., Rohr, B. A., Gauthier, J. A. & Nørskov, J. K. Predicting chemical reaction barriers with a machine learning model. *Catal. Lett.* **149**, 2347–2354 (2019).
48. Takahashi, K. & Miyazato, I. Rapid estimation of activation energy in heterogeneous catalytic reactions via machine learning. *J. Comput. Chem.* **39**, 2405–2408 (2018).
49. Su, Y.-Q. *et al.* Stability of heterogeneous single-atom catalysts: a scaling law mapping thermodynamics to kinetics. *npj Comput. Mater.* **6**, 144 (2020).
50. Davran-Candan, T., Günay, M. E. & Yıldırım, R. Structure and activity relationship for CO and O_2 adsorption over gold nanoparticles using density functional theory and artificial neural networks. *J. Chem. Phys.* **132**, 174113 (2010).
51. Ulissi, Z. W., Medford, A. J., Bligaard, T. & Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nat. Commun.* **8**, 14621 (2017).
52. Toyao, T. *et al.* Toward effective utilization of methane: machine learning prediction of adsorption energies on metal alloys. *J. Phys. Chem. C* **122**, 8315–8326 (2018).
53. Thornton, A. W., Winkler, D. A., Liu, M. S., Haranczyk, M. & Kennedy, D. F. Towards computational design of zeolite catalysts for CO_2 reduction. *RSC Adv.* **5**, 44361–44370 (2015).
54. Artrith, N. & Kolpak, A. M. Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: a combination of DFT and accurate neural network potentials. *Nano Lett.* **14**, 2670–2676 (2014).

55. Gu, G. H. *et al.* Practical deep-learning representation for fast heterogeneous catalyst screening. *J. Phys. Chem. Lett.* **11**, 3185–3191 (2020).
56. Li, Z., Wang, S., Chin, W. S., Achenie, L. E. & Xin, H. High-throughput screening of bimetallic catalysts enabled by machine learning. *J. Mater. Chem. A* **5**, 24131–24138 (2017).
57. Yohannes, A. G. *et al.* Combined high-throughput DFT and ML screening of transition metal nitrides for electrochemical CO_2 reduction. *ACS Catal.* **13**, 9007–9017 (2023).
58. Mok, D. H. & Back, S. Atomic structure-free representation of active motifs for expedited catalyst discovery. *J. Chem. Inf. Model.* **61**, 4514–4520 (2021).
59. Chen, L. *et al.* A universal machine learning framework for electrocatalyst innovation: a case study of discovering alloys for hydrogen evolution reaction. *Adv. Func. Mater.* **32**, 2208418 (2022).
60. Flores, R. A. *et al.* Active learning accelerated discovery of stable iridium oxide polymorphs for the oxygen evolution reaction. *Chem. Mater.* **32**, 5854–5863 (2020).
61. Bang, G. J., Gu, G. H., Noh, J. & Jung, Y. Activity trends of methane oxidation catalysts under emission conditions. *ACS Catal.* **12**, 10255–10263 (2022).
62. Kim, M. *et al.* Artificial Intelligence to Accelerate the Discovery of N_2 Electroreduction Catalysts. *Chem. Mater.* **32**, 709–720 (2020).
63. Li, X.-T., Chen, L., Wei, G.-F., Shang, C. & Liu, Z.-P. Sharp increase in catalytic selectivity in acetylene semihydrogenation on Pd achieved by a machine learning simulation-guided experiment. *ACS Catal.* **10**, 9694–9705 (2020).
64. Abed, J. *et al.* Pourbaix machine learning framework identifies acidic water oxidation catalysts exhibiting suppressed ruthenium dissolution. *J. Am. Chem. Soc.* **146**, 15740–15750 (2024).
65. Batchelor, T. A. A. *et al.* Complex-solid-solution electrocatalyst discovery by computational prediction and high-throughput experimentation. *Angew. Chem. Int. Ed.* **60**, 6932–6937 (2021).
66. Svane, K. L. & Rossmeisl, J. Theoretical optimization of compositions of high-entropy oxides for the oxygen evolution reaction. *Angew. Chem. Int. Ed.* **61**, e202201146 (2022).
67. Wan, X. *et al.* Machine-learning-assisted discovery of highly efficient high-entropy alloy catalysts for the oxygen reduction reaction. *Patterns* **3**, 100553 (2022).
68. Xu, W., Diesen, E., He, T., Reuter, K. & Margraf, J. T. Discovering high entropy alloy electrocatalysts in vast composition spaces with multiobjective optimization. *J. Am. Chem. Soc.* **146**, 7698–7707 (2024).
69. Pedersen, J. K. *et al.* Bayesian optimization of high-entropy alloy compositions for electrocatalytic oxygen reduction. *Angew. Chem. Int. Ed.* **133**, 24346–24354 (2021).
70. Jung, H., Sauerland, L., Stocker, S., Reuter, K. & Margraf, J. T. Machine-learning driven global optimization of surface adsorbate geometries. *npj Comput. Mater.* **9**, 1–8 (2023).
71. Ock, J., Badrinarayanan, S., Magar, R., Antony, A. & Farimani, A. B. Multimodal language and graph learning of adsorption configuration in catalysis. Preprint at https://doi.org/10.48550/arXiv.2401.07408 (2024).
72. Noh, J. & Chang, H. Data-driven prediction of configurational stability of molecule-adsorbed heterogeneous catalysts. *J. Chem. Inf. Model.* **63**, 5981–5995 (2023).
73. Meyer, B., Sawatlon, B., Heinen, S., von Lilienfeld, O. A. & Corminboeuf, C. Machine learning meets volcano plots: computational discovery of cross-coupling catalysts. *Chem. Sci.* **9**, 7069–7077 (2018).
74. Sasaki, M., Hamada, H., Kintaichi, Y. & Ito, T. Application of a neural network to the analysis of catalytic reactions Analysis of NO decomposition over Cu/ZSM-5 zeolite. *Appl. Catal. A: Gen.* **132**, 261–270 (1995).
75. Suvarna, M., Araújo, T. P. & Pérez-Ramírez, J. A generalized machine learning framework to predict the space-

time yield of methanol from thermocatalytic CO_2 hydrogenation. *Appl. Catal. B: Environ.* **315**, 121530 (2022).
76. Suvarna, M., Preikschas, P. & Pérez-Ramírez, J. Identifying descriptors for promoted rhodium-based catalysts for higher alcohol synthesis via machine learning. *ACS Catal.* 15373–15385 (2022) doi:10.1021/acscatal.2c04349.
77. Li, J., Pan, L., Suvarna, M. & Wang, X. Machine learning aided supercritical water gasification for H_2-rich syngas production with process optimization and catalyst screening. *Chem. Eng. J.* **426**, 131285 (2021).
78. Saadun, A. J. *et al.* Performance of metal-catalyzed hydrodebromination of dibromomethane analyzed by descriptors derived from statistical learning. *ACS Catal.* **10**, 6129–6143 (2020).
79. Pablo-García, S. *et al.* Generalizing performance equations in heterogeneous catalysis from hybrid data and statistical learning. *ACS Catal.* **12**, 1581–1594 (2022).
80. Smith, A., Keane, A., Dumesic, J. A., Huber, G. W. & Zavala, V. M. A machine learning framework for the analysis and prediction of catalytic activity from experimental data. *Appl. Catal. B: Environ.* **263**, 118257 (2020).
81. Schmack, R. *et al.* A meta-analysis of catalytic literature data reveals property-performance correlations for the OCM reaction. *Nat. Commun.* **10**, 441 (2019).
82. Palkovits, S. A primer about machine learning in catalysis – a tutorial with code. *ChemCatChem* **12**, 3995–4008 (2020).
83. Zavyalova, U., Holena, M., Schlögl, R. & Baerns, M. Statistical analysis of past catalytic data on oxidative methane coupling for new insights into the composition of high-performance catalysts. *ChemCatChem* **3**, 1935–1947 (2011).
84. Kondratenko, E. V., Schlüter, M., Baerns, M., Linke, D. & Holena, M. Developing catalytic materials for the oxidative coupling of methane through statistical analysis of literature data. *Catal. Sci. Technol.* **5**, 1668–1677 (2015).
85. Pirro, L. *et al.* Descriptor–property relationships in heterogeneous catalysis: exploiting synergies between statistics and fundamental kinetic modelling. *Catal. Sci. Technol.* **9**, 3109–3125 (2019).
86. Arcotumapathy, V., Siahvashi, A. & Adesina, A. A. A new weighted optimal combination of ANNs for catalyst design and reactor operation: Methane steam reforming studies. *AIChE J.* **58**, 2412–2427 (2012).
87. Serra, J. M. & Vert, V. B. Quaternary mixture designs applied to the development of multi-element oxygen electrocatalysts based on the $Ln_{0.58}Sr_{0.4}Fe_{0.8}Co_{0.2}O_{3-\delta}$ system ($Ln=La_{1-x-y-z}Pr_xSm_yBa_z$): Predictive modeling approaches. *Catal. Today* **159**, 47–54 (2011).
88. Baumes, L. A., Serna, P. & Corma, A. Merging traditional and high-throughput approaches results in efficient design, synthesis and screening of catalysts for an industrial process. *Appl. Catal. A: Gen.* **381**, 197–208 (2010).
89. Baumes, L. A., Serra, J. M., Serna, P. & Corma, A. Support vector machines for predictive modeling in heterogeneous catalysis: A comprehensive introduction and overfitting investigation based on two real applications. *J. Comb. Chem.* **8**, 583–596 (2006).
90. Serra, J. M., Chica, A. & Corma, A. Development of a low temperature light paraffin isomerization catalysts with improved resistance to water and sulphur by combinatorial methods. *Appl. Catal. A: Gen.* **239**, 35–42 (2003).
91. Corma, A., Serra, J., Serna, P. & Moliner, M. Integrating high-throughput characterization into combinatorial heterogeneous catalysis: unsupervised construction of quantitative structure/property relationship models. *J. Catal.* **232**, 335–341 (2005).
92. Corma, A. *et al.* Optimisation of olefin epoxidation catalysts with the application of high-throughput and

genetic algorithms assisted by artificial neural networks (softcomputing techniques). *J. Catal.* **229**, 513–524 (2005).
93. Baumes, L., Farrusseng, D., Lengliz, M. & Mirodatos, C. Using artificial neural networks to boost high-throughput discovery in heterogeneous catalysis. *QSAR Comb. Sci.* **23**, 767–778 (2004).
94. Corma, A., Serra, J. M., Argente, E., Botti, V. & Valero, S. Application of artificial neural networks to combinatorial catalysis: modeling and predicting ODHE catalysts. *ChemPhysChem* **3**, 939–945 (2002).
95. Farrusseng, D. *et al.* Design of discovery libraries for solids based on QSAR models. *QSAR Comb. Sci.* **24**, 78–93 (2005).
96. Farrusseng, D., Clerc, F., Mirodatos, C. & Rakotomalala, R. Virtual screening of materials using neuro-genetic approach: Concepts and implementation. *Comput. Mater. Sci.* **45**, 52–59 (2009).
97. Hossain, M. A., Ayodele, B. V., Cheng, C. K. & Khan, M. R. Artificial neural network modeling of hydrogen-rich syngas production from methane dry reforming over novel Ni/$CaFe_2O_4$ catalysts. *Int. J. Hydrogen Energy* **41**, 11119–11130 (2016).
98. Sun, Y., Yang, G., Wen, C., Zhang, L. & Sun, Z. Artificial neural networks with response surface methodology for optimization of selective CO_2 hydrogenation using K-promoted iron catalyst in a microchannel reactor. *J. CO2 Util.* **24**, 10–21 (2018).
99. 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).
100. Chesterfield, D. & Adesina, A. A. Evidence-based design and optimisation of titania photocatalysts via artificial neural network analysis. *J. Chem. Eng. Japan* **42**, S185–S198 (2009).
101. Jayakumar, T. P., Suresh Babu, S. P., Nguyen, T. N., Le, S. D. & Taniike, T. Exploration of ethanol-to-butadiene catalysts by high-throughput experimentation and machine learning. Preprint at https://chemrxiv.org/engage/chemrxiv/article-details/6461dfe2a32ceeff2dade21d (2023).
102. Holeňa, M. & Baerns, M. Feedforward neural networks in catalysis. *Catal. Today* **81**, 485–494 (2003).
103. Rohr, B. *et al.* Benchmarking the acceleration of materials discovery by sequential learning. *Chem. Sci.* **11**, 2696–2706 (2020).
104. Hickman, R. J., Häse, F., Roch, L. M. & Aspuru-Guzik, A. Gemini: Dynamic bias correction for autonomous experimentation and molecular simulation. Preprint at http://arxiv.org/abs/2103.03391 (2021).
105. Nguyen, T. N. *et al.* High-throughput experimentation and catalyst informatics for oxidative coupling of methane. *ACS Catal.* **10**, 921–932 (2020).
106. Teimouri, Z., Abatzoglou, N. & Dalai, A. K. Design of a renewable catalyst support derived from biomass with optimized textural features for Fischer Tropsch synthesis. *Renew. Energy* **202**, 1096–1109 (2023).
107. Takigawa, I., Shimizu, K., Tsuda, K. & Takakusagi, S. Machine-learning prediction of the d-band center for metals and bimetals. *RSC Adv.* **6**, 52587–52595 (2016).
108. Odabaşı, Ç., Günay, M. E. & Yıldırım, R. Knowledge extraction for water gas shift reaction over noble metal catalysts from publications in the literature between 2002 and 2012. *Int. J. Hydrogen Energy* **39**, 5733–5746 (2014).
109. Can, E. & Yildirim, R. Data mining in photocatalytic water splitting over perovskites literature for higher hydrogen production. *Appl. Catal. B: Environ.* **242**, 267–283 (2019).
110. Günay, M. E., Türker, L. & Tapan, N. A. Decision tree analysis for efficient CO_2 utilization in electrochemical systems. *J. CO2 Util.* **28**, 83–95 (2018).
111. Şener, A. N., Günay, M. E., Leba, A. & Yıldırım, R. Statistical review of dry reforming of methane literature

using decision tree and artificial neural network analysis. *Catal. Today* **299**, 289–302 (2018).
112. Baysal, M., Günay, M. E. & Yıldırım, R. Decision tree analysis of past publications on catalytic steam reforming to develop heuristics for high performance: A statistical review. *Int. J. Hydrogen Energy* **42**, 243–254 (2017).
113. Tapan, N. A., Günay, M. E. & Yildirim, R. Constructing global models from past publications to improve design and operating conditions for direct alcohol fuel cells. *Chem. Eng. Res. Des.* **105**, 162–170 (2016).
114. Alper Tapan, N., Yıldırım, R. & Erdem Günay, M. Analysis of past experimental data in literature to determine conditions for high performance in biodiesel production: Determining conditions for high performance in biodiesel production. *Biofuels, Bioprod. Bioref.* **10**, 422–434 (2016).
115. Günay, M. E. & Yildirim, R. Modeling preferential CO oxidation over promoted Au/Al_2O_3 catalysts using decision trees and modular neural networks. *Chem. Eng. Res. Des.* **91**, 874–882 (2013).
116. Günay, M. E. & Yildirim, R. Knowledge extraction from catalysis of the past: A case of selective CO oxidation over noble metal catalysts between 2000 and 2012. *ChemCatChem* **5**, 1395–1406 (2013).
117. Günay, M. E. & Yildirim, R. Developing global reaction rate model for CO oxidation over Au catalysts from past data in literature using artificial neural networks. *Appl. Catal. A: Gen.* **468**, 395–402 (2013).
118. Günay, M. E. & Yildirim, R. Neural network analysis of selective CO oxidation over copper-based catalysts for knowledge extraction from published data in the literature. *Ind. Eng. Chem. Res.* **50**, 12488–12500 (2011).
119. Hattori, T. & Kito, S. Neural network as a tool for catalyst development. *Catal. Today* **23**, 347–355 (1995).
120. Madaan, N., Shiju, N. R. & Rothenberg, G. Predicting the performance of oxidation catalysts using descriptor models. *Catal. Sci. Technol.* **6**, 125–133 (2016).
121. Ras, E.-J., Louwerse, M. J. & Rothenberg, G. New tricks by very old dogs: predicting the catalytic hydrogenation of HMF derivatives using Slater-type orbitals. *Catal. Sci. Technol.* **2**, 2456 (2012).
122. Ras, E.-J., McKay, B. & Rothenberg, G. Understanding catalytic biomass conversion through data mining. *Top Catal.* **53**, 1202–1208 (2010).
123. Leonard, K. C. & Bard, A. J. Pattern recognition correlating materials properties of the elements to their kinetics for the hydrogen evolution reaction. *J. Am. Chem. Soc.* **135**, 15885–15889 (2013).
124. Estahbanati, M. R. K., Feilizadeh, M. & Iliuta, M. C. Photocatalytic valorization of glycerol to hydrogen: Optimization of operating parameters by artificial neural network. *Appl. Catal. B: Environ.* **209**, 483–492 (2017).
125. Artrith, N., Lin, Z. & Chen, J. G. Predicting the activity and selectivity of bimetallic metal catalysts for ethanol reforming using machine learning. *ACS Catal.* **10**, 9438–9444 (2020).
126. Baroi, C. & Dalai, A. K. Review on biodiesel production from various feedstocks using 12-tungstophosphoric acid (TPA) as a solid acid catalyst precursor. *Ind. Eng. Chem. Res.* **53**, 18611–18624 (2014).
127. Jiang, W., Xing, X., Li, S., Zhang, X. & Wang, W. Synthesis, characterization and machine learning based performance prediction of straw activated carbon. *J. Clean. Prod.* **212**, 1210–1223 (2019).
128. Klanner, C. *et al.* The development of descriptors for solids: teaching “catalytic intuition” to a computer. *Angew. Chem. Int. Ed.* **43**, 5347–5349 (2004).
129. Takahashi, K., Miyazato, I., Nishimura, S. & Ohyama, J. Unveiling hidden catalysts for the oxidative coupling of methane based on combining machine learning with literature data. *ChemCatChem* **10**, 3223–3228 (2018).
130. Han, X. *et al.* Using data mining technology in screening potential additives to Ni/Al_2O_3 catalysts for methanation. *Catal. Sci. Technol.* **7**, 6042–6049 (2017).