Note: Page numbers in *italics* and **bold** refers to figures and tables respectively.

#### а

ab initio simulations 133 accelerated oracle design autodE 94-95 automation software for 94-98 RMSD-PP-TS 95-97, 96 TS-tools 97-98, 97 accuracy trade-off 85-87 acquisition function (AF) 231 activation energy 166-172 active learning 93-94, 228, 229, 233, 233 active site determination 117-118 Adam. see adaptive moment estimation (Adam) adaptive gradient algorithm 113 adaptive moment estimation (Adam) 113 AF. see acquisition function (AF) AFM. see atomic force microscopy (AFM) AHO. see asymmetric hydrogenation of olefins (AHO) AI-assisted catalyst design 222–223 design rule extraction via data analysis 223-224 design rule extraction via model interpretation 224-228 AI-assisted catalyst discovery 228-229, 238 active learning/sequential learning 228

catalyst recommendation 230-232 catalyst synthesis and testing 232 initial machine learning model 229-230 iterative process 232 search space determination 230 selected use cases from literature 233-234 AI-assisted catalyst synthesis 235–238 AIMNet-NSE 136 algorithm development 47-48 angular second moment (ASM) 14 ANN. see artificial neural networks (ANN) AQME toolkit 138, 138, 142-143, 143 artificial neural networks (ANN) 112, 116 ASM. see angular second moment (ASM) asymmetric catalysis 195-196, 199 asymmetric hydrogenation of olefins (AHO) 198 asymmetric synthesis 195 atom pair FP 203 atomic force microscopy (AFM) 56 atom-type embeddings 120–121, 122 augmented vs. standard connectivity matrix 202 autodE (open-source Python package) 94–95 automated analysis of supported catalyst TEM images 71-72, 72 automatic image analysis, iOk Platform for 62-71

Artificial Intelligence in Catalysis: Experimental and Computational Methodologies, First Edition. Edited by Valentine P. Ananikov and Mikhail V. Polynski. © 2025 Wiley-VCH GmbH. Published 2025 by Wiley-VCH GmbH.

Avalon FP 203 average steric occupancy (ASO) descriptor 173, 174, 206, 207

## b

Bag of Bonds (BoB) 168, 180 Bayer filter. see color filter Behler-Parrinello Neural Network model 118-119 benchmarking 88-89, 89 BERT. see bidirectional encoder representations from transformers (BERT) bidirectional encoder representations from transformers (BERT) 161, 161 black-box models 225 BoB. see Bag of Bonds (BoB) Bravais-Inspired gradient domain machine learning (BIGDML) model 126 Buchwald-Hartwig cross-coupling reactions 163, 164

### С

Cambridge Structural Database (CSD) 137 Cascade Mask-RCNN neural network 62 catalyst 55. see also catalyst optimization cobalt 181 design and discovery 116-117 iridium 181 rhodium 181 Ru-based NH<sub>3</sub> synthesis 234 synthesis 232 testing 232 used, in asymmetric catalysis 199 catalyst optimization cheminformatics 139-141 computational chemistry 133, 134 databases 136-139 ML protocols, automation of 143-145 molecular descriptors 135-136 QM protocols, automation of 141-143 catalyst recommendation acquisition function 231 expected improvement 231, 232 model uncertainty 230

probability of improvement 231, 232 upper confidence bound 231 CC. see coupled cluster (CC) CGR. see condensed graph of reaction (CGR) Chat Bot DLgram 65-68, 66, 67 limitations of 67-68, 69 chemical descriptors 163, 164 chemical reactivity prediction 83 accelerated oracle design, automation software for 94-98 active learning strategies 93-94 oracle design 84-91 search space sampling 91–93, 92 cheminformatics 139–141 format conversion and data interpretation 140-141 molecular descriptor generation 141 multipurpose chemical software 139 reaction discovery 140 tools and application 140 universal toolkits 140 chemoselectivity 172, 177 chromaticity 6 components 21 CIE-L\*a\*b\* color space 9, 10, 10 CIE XYZ color space 6-8, 7 circular substructures (CircuS) 175, 203, 204 CircuS. see circular substructures (CircuS) CNNs. see convolutional neural networks (CNNs) color distributions/histograms 14, 15 color filter 11 color matching functions 5 color spaces 5 color theory 4, 5 CIE-L\*a\*b\* 9, 10, 10 CIE XYZ color space 6-8, 7 hue, saturation, and value 8, 9 RGB color spaces 6, 7 xyY color space 6-8, 7 CoMFA. see comparative molecular field analysis (CoMFA) comparative molecular field analysis (CoMFA) 159, 174, 206

computational chemistry 133 current state of 134 evolution of 134 computer vision 3, 3, 4 algorithms and 14 chemistry, fundamentals in 4-19 color theory 4–10, 5 digital photography basics 10-19 high-throughput digital fingerprinting 19-20, 20 and machine learning in chemistry 19 neural networks 24-25, 25, 26 phase separation and partition coefficients 27-28, 28 reaction kinetics 29-31, 30 segmentation and parallel analysis of nerve agents 20-24, 21-23 single image applications 19 video analysis applications 27 condensed graph of reaction (CGR) 162,203 for asymmetric reactions 204 cone angle ( $\theta$ ), definition of 135 conformational sampling 140 ContraScope. see substrate scope contrastive learning (ContraScope) convolution, of computer vision 17-19, 18, 19 convolutional layers, of neural networks 113 convolutional neural networks (CNNs) 17, 24, 26 cost-effective molecular simulation 140 coupled cluster (CC) 105 cross-validation (CV) 200 crystallization process 236 CSD. see Cambridge Structural Database (CSD) CV. see cross-validation (CV)

# d

data 157 acquisition 42–43 analysis 43–45, 44 cleaning 221 extraction 217

generation 216 HT catalyst testing 219 HT reactors 219-221 literature data extraction 217 preprocessing 43 robotic assisted high-throughput catalyst synthesis 218 solution-based chemical methods 218 source 45-46 thin film deposition methods 218-219 databases 136-139 AQME 138, 138 FAIR principles 136 ioChem-BD repository 139 KRAKEN 136-138, 137 Open Reaction Database 139 OSCAR 137-138, 137 decision tree algorithm 223-224 deep learning (DL) 36-37, 167 deep machine learning for image analysis 57 comparison of neural networks 62 STM image analysis 57-60, 58, 58-60 TEM image analysis 60-62, 61 delta ( $\Delta$ ) learning 87, 167–168, 172 dense layers, of neural networks 113 density functional theory (DFT) 86, 105, 135, 215 design rule extraction via data analysis 223-224 via model interpretation 224-228 DEtection TRansformer (DETR) 77 DETR. see DEtection TRansformer (DETR) DFT. see density functional theory (DFT) Diels-Alder cycloadditions 178 Diels-Alder reaction 166 differential reaction fingerprint (DRFP) 163 digital camera, work of 10-12, 11 digital object identifiers (DOIs) 139 digital photography 10 algorithms and computer vision 14 color distributions/histograms 14, 15 convolution and edge detection 17-19, 18, 19

digital photography (*Cont.*) digital camera, work of 10–12, *11* image and video 12–14, *13* texture analysis 14, 16–17, *16*, *17* dimethyl sulfoxide (DMSO) 19 directed message-passing neural network (D-MPNN) *176* DL. *see* deep learning (DL) D-MPNN. *see* directed message-passing neural network (D-MPNN) DMSO. *see* dimethyl sulfoxide (DMSO) DOIs. *see* digital object identifiers (DOIs) DRFP. *see* differential reaction fingerprint (DRFP) drug-like compound libraries 19–20, *20* 

### е

edge detection, of computer vision 17-19, 18, 19 EDX. see energy-dispersive X-ray spectroscopy (EDX) EELS. see electron energy loss spectroscopy (EELS) EI. see expected improvement (EI) EI-Q-MS method 40 electron energy loss spectroscopy (EELS) 77 electronic lab notebooks (ELNs) 181 electrospray ionization (ESI) 36, 38 ELNs. see electronic lab notebooks (ELNs) enantiomeric excess (e. e.) 197 enantiomeric ratio (e. r.) 196 enantioselectivity data 197-198 prediction of 205 as target property 196-197 3D descriptors, models using 204-207 2D descriptors, models using 201-204 energetic profile similarity (EPSim) 180, 181 energetic span model 179 energy-dispersive X-ray spectroscopy (EDX) 77 EPSim. see energetic profile similarity (EPSim) EquiReact 170, 171

error minimization 114 ESI. *see* electrospray ionization (ESI) ESI-QQQ-MS method 40 Euclidean distances 23 experimental condition optimizations 117 experimental data active learning 233 HTE 160 ML models 216 surrogate machine learning models 216

# f

face selectivity 178 false negative (FN) predictions 59 false positive (FP) predictions 59 filter-generating networks 121–122 findable, accessible, interoperable, and reusable (FAIR) principles 136, 181 fingerprints (FP) 203 finite impulse response (FIR) filter 27 Fisher–Tropsch catalysts 214 fixed-length vectors 46 format conversion and data interpretation 140–141 formula assignment 43 FP. *see* fingerprints (FP)

## g

GANs. see generative adversarial networks (GANs) GAP. see Gaussian approximation potential (GAP) Gaussian approximation potential (GAP) 106 Gaussian mixture model (GMM) 224 Gaussian process regression (GPR) 167, 229-230 GCNN. see graph convolutional neural network (GCNN) GDML. see gradient-domain machine learning (GDML) generative adversarial networks (GANs) 77 geometric harmonics (GH) 236 GH. see geometric harmonics (GH)

glass-box methods 225 glassware, neural networks for identifying 24-25, 25, 26 GLCM. see gray level co-occurrence matrix (GLCM) GMM. see Gaussian mixture model (GMM) GPR. see Gaussian process regression (GPR) gradient descent 113 gradient-domain machine learning (GDML) 106 framework 124-126, 125 graph convolutional neural network (GCNN) 162 graph molecular maps 136 gray-box methods 225 gray level co-occurrence matrix (GLCM) 14, 16-17

## h

HAADF-STEM. see high-angle annular dark-field scanning TEM (HAADF-STEM) half-wave potential (HWP) 227 Hammett approach 199, 204 HDNN. see high-dimensional neural network architecture (HDNN) HER. see hydogen evolution reaction (HER) Hessian kernel 124-125 heterogeneous catalysts 213 TEM images by iOk platform, analysis of 71-76 high-angle annular dark-field scanning TEM (HAADF-STEM) 74 high-dimensional neural network architecture (HDNN) 105 highest occupied molecular orbital and lowest unoccupied molecular orbital (HOMO-LUMO) 137 high-resolution TEM images 72-73, 73 high-throughput digital fingerprinting, of drug-like compound libraries 19-20, 20 high-throughput experimentation (HTE) 160, 161, 165, 197, 214, 215

high-throughput screening techniques 133 HOMO–LUMO. *see* highest occupied molecular orbital and lowest unoccupied molecular orbital (HOMO–LUMO) HTE. *see* high-throughput experimentation (HTE) Hue, Saturation, and Value (HSV) color space 8, 9 HWP. *see* half-wave potential (HWP) hydrogen atom transfer (HAT) reactions 91–92, 169 hydrogen evolution reaction (HER) 233 hyperparameters 222

## i

IDPi catalysts 204 enantioselectivity, prediction of 205 IJP. see inkjet printing (IJP) image in computer vision 12-14, 13 convolution 18 deep machine learning for image analysis 57-62 infrared filter 11 inkjet printing (IJP) 219 In Silico Design and Data Analysis (ISIDA) 175, 203 interaction blocks 121-122 ioChem-BD repository 139 iOk platform analysis of TEM images of heterogeneous catalyst by 71-76 for automatic image analysis 62-71,64 Chat Bot DLgram 65-68, 66, 67 comparison with other products 70-71, 70, 70, 71 No Code ML 68-69 ParticlesNN web-service 63-65 ionization source 38 ISIDA. see In Silico Design and Data Analysis (ISIDA) itemset mining analysis 224 iterative process 232

## j

joint entropy 29-30

## k

Kennard–Stone algorithm 173 kernel-based force fields 119–120, 124–126, *125* kernel methods 114–115 kernel ridge regression (KRR) 115 kernel trick 115 KRAKEN database 136–138, *137 kraken* library 198 KRR. *see* kernel ridge regression (KRR)

## l

large language models (LLMs) 47–48, 118, 146 LFERs. *see* linear free energy relationships (LFERs) linear free energy relationships (LFERs) 159 linear regression with ridge regularization 137 LLMs. *see* large language models (LLMs)

## m

machine learning (ML) 83, 84, 157 activation energy 166-172 active site determination 117-118 algorithms 133 catalyst design and discovery 116-117 in chemical reactions and catalysis 115, 116 in chemistry 19-31 data 157 data cleaning 221 data generation 216-221 databases 136 experimental condition optimizations 117 experimental data 216 feature generation/engineering and selection 221-222 fundamentals of 107-109, 108, 109 generalizability issue 182 mass spectrometry in 36-38, 37

mass spectrometry methods landscape and potential for 38-41, 39 model development 221–222 model selection and evaluation 222 models 208 model training 157 MT toolkit 140 prediction 157 protocol for solving general problems in mass spectrometry using 45-48 protocols automation 143-145 QSAR models 159 rate constants 166-172 representations 157 representative mass spectrometry applications of 41-45, 42 selectivity 172-179 selectivity prediction 116 turnover frequency (TOF) 179-180 uncertainty of predicted catalyst 230 volcano plot 160, 179-180 yield 160-166 machine learning-based force fields (MLFFs) 119, 105-106 kernel methods 114–115 neural networks 110-113 trends in 118 machine learning interatomic potentials (MLIPs) 118 MAE. see mean absolute error (MAE) magnetron sputtering (MS) 219 MALDI-TOF technique 40 marginal probability 29 mass spectrometry (MS) 35 algorithm development 47-48 data acquisition 42-43 data analysis 43-45, 44 data preprocessing 43 data source 45-46 landscape and potential for machine learning applications 38-41, 39 in machine learning era 36-38, 37 metric selection 48 protocol for solving general problems in 45

representative mass spectrometry applications of machine learning 41-45, 42 sample preparation 41–42 spectra representation 46-47 mean absolute error (MAE) 89 mean squared error (MSE) 26 message-passing neural networks (MPNN) 106 metric selection 48 MFF. see multiple fingerprint features (MFF) microscopy 55-56 analysis of TEM images of heterogeneous catalyst 71-76 deep machine learning for image analysis 57-62 iOk platform for automatic image analysis 62-71 mikimo program 180 MIL. see multi-instance learning (MIL) ML. see machine learning (ML) MLFFs. see machine learning-based force fields (MLFFs) MLIPs. see machine learning interatomic potentials (MLIPs) model interpretation methods design rule extraction via 224-228 random forest model 226 molecular descriptors 135–136, 141 molecular-directed message-passing neural network (D-MPNN) 161, 162 molecular fingerprints 135 MORFEUS database 138, 138, 141 Morgan fingerprints 135, 203 MPNN. see message-passing neural networks (MPNN) MS. see magnetron sputtering (MS); mass spectrometry (MS) MSE. see mean squared error (MSE) multi-instance learning (MIL) 206 multilayer perceptron 112 multiple fingerprint features (MFF) 164, 203 multipurpose chemical software 139

multivariate linear regressions, based on stereo-electronic descriptor 179 mutual information (M.I.) analysis 29

#### n

natural language processing (NLP) 217 nerve agents, segmentation and parallel analysis of 20-24, 21-23 neural network potentials (NNP) 86 neural networks 62, 110 architecture of 112-113 force fields 118-119, 120-124, 121 for identifying glassware 24-25, 25, 26 multilayer perceptron 112 optimization algorithms 113 perceptron 110-112, 110-111 NLP. see natural language processing (NLP) NMR. see nuclear magnetic resonance (NMR) NNP. see neural network potentials (NNP) No Code ML 68 advantages of 69 limitations of 69 nuclear magnetic resonance (NMR) 136 nudged elastic band (NEB) approach 94

#### 0

OCM. see oxidative coupling of methane (OCM) on-demand descriptor database generators AQME 138, 138 MORFEUS 138, 138 one-hot encoding representation 168, 169 Open Babel (tool) 140-141 Open Reaction Database 139 optimization algorithms, of neural networks 113 oracle design 84, 85 accuracy trade-off 85-87 benchmarking 88-89, 89 optimizing for multiple criteria, implications of 87-88 reproducibility 90–91, 90 ORR. see oxygen reduction reaction (ORR) OSCAR database 137-138, 137

oxidative coupling of methane (OCM) 224 oxygen reduction reaction (ORR) 227, 228

#### р

Pacific Northwest National Laboratory (PNNL) 225 PAI. see promoter affinity index (PAI) parallel analysis, of nerve agents 20-24, 21-23 ParticlesNN web-service 63 advantage of 64-65 limitation of 64 partition coefficients 27-28, 28 Pd-catalyzed Suzuki-Miyaura reactions 135 perceptron 110-112, 110-111 perceptual uniformity 6 PES. see potential energy surface (PES) phase separation 27-28, 28 photography 14 physicochemical descriptors 165-166 PI. see probability of improvement (PI) PIE. see probe interaction energies (PIE) PNNL. see Pacific Northwest National Laboratory (PNNL) potential energy surface (PES) 86, 105 predictive models 158 probe interaction energies (PIE) 206 promoter affinity index (PAI) 226-227

## q

QM protocols automation AQME 142–143, *143* autodE 142 fully automated 141 multi-seeding 141 QSAR. *see* quantitative structure-activity relationship (QSAR) quantitative structure-activity relationship (QSAR) 135, 159 workflow *199* 

## r

rate constants 166–172 R-CNNs. *see* region-based convolutional neural networks (R-CNNs) RDKit 136.141 RDkit-specific FP 203 reaction discovery 140 reaction kinetics 29-31, 30 reaction modeling ML application 199 principles of 198-201 validation sets 200 reactivity cliff 164, 164 reactivity predictions 158, 196 recurrent layers, of neural networks 113 red, green, blue (RGB) array 12 color spaces 6, 7 region-based convolutional neural networks (R-CNNs) 25 region proposal networks (RPNs) 25 regions of interest (ROI) 3 regioselectivity 172, 178 representer theorem 114–115 reproducibility 90–91, 90 R-Mask CNNs 25 RMSD. see root mean square distances (RMSD) RMSD-PP-TS (open-source Python package) 95-97, 96 RMSE. see root mean squared error (RMSE) RMSProp. see root mean square propagation method (RMSProp) ROBERT program 144, 144, 145 RoboChem 146 ROI. see regions of interest (ROI) root mean squared error (RMSE) 26 root mean square distances (RMSD) 95 root mean square propagation method (RMSProp) 113 RPNs. see region proposal networks (RPNs)

## S

Sabatier principle 160 SC. *see* single crystals (SC) scanning electron microscopy (SEM) 56 scanning probe microscopy (SPM) 56

scanning tunneling microscopy (STM) 56 image analysis 57-60, 58, 58-60 SchNet 106, 119, 120, 121 atom-type embeddings 120–121, 122 explicit SchNet model 122–124 interaction blocks 121-122 search space determination 230 sampling 91-93, 92 segmentation, of nerve agents 20-24, 21-23 selectivity ASO descriptor 173 C-H functionalization 176-177 ExtraTrees algorithm 172 face selectivity 178 multivariate linear regressions based on stereo-electronic descriptors 179 neural network algorithm 173 prediction 116 of reaction 172 RDKit descriptors 175 regioselectivity prediction 176 SEM. see scanning electron microscopy (SEM) Semiempirical Extended Tight-Binding (xTB) 136, 140 sequential learning 228 Shapley additive explanations (SHAP) values 207, 225 simplified molecular input line entry system (SMILES) 83, 161 single crystals (SC) 118 single-interface nanoparticles (SINPs) 236-238 single site analysis 74-76, 75, 76, 76 SINPs. see single-interface nanoparticles (SINPs) SMILES. see simplified molecular input line entry system (SMILES) smooth overlap of atomic positions (SOAP) 119 SOAP. see smooth overlap of atomic positions (SOAP) space-time yield  $(STY_{HA})$  226

spectra representation 46–47 SPM. *see* scanning probe microscopy (SPM) sRGB color space 6 standard *vs.* augmented connectivity matrix 202 stereo-electronic descriptors 172, 174, 178 STM. *see* scanning tunneling microscopy (STM) stochastic gradient descent 113 substrate scope contrastive learning (ContraScope) 162 supervised learning 107, 143 symmetric-gradient domain machine learning (sGDML) model 126

### t

telecentric lens system 19 TEM. see transmission electron microscopy (TEM) TEP. see Tolman's electronic parameter (TEP) texture analysis, of computer vision 14. 16-17. 16. 17 thermochemistry analysis 139-140 3D descriptors, modeling with comparative molecular field analysis 206 Hammett approach 204 multi-instance learning 206 QC calculations 206 SHapley Additive exPlanations values approach 207 time-of-flight (TOF) mass analyser 40 TOF. see turnover frequency (TOF) Tolman's electronic parameter (TEP) 135, 135 TON. see turnover number (TON) transition state (TS) 158, 168 transmission electron microscopy (TEM) 56, 107 automated analysis of 71-72, 72 of heterogeneous catalyst 71 high-resolution 72-73, 73 image analysis 60-62, 61 single site analysis 74-76, 75, 76, 76 tree-based methods 199

tristimulus color theory 5 true positive (TP) predictions 59 TS. *see* transition state (TS) TS-tools (open-source Python package) 97–98, 97 turnover frequency (TOF) 56, 160, 179–180 turnover number (TON) 56, 179 2D descriptors, modeling with cinchona alkaloid-derived phase-transfer catalysts 202 fingerprints (FP) 203 molecular graphs 201, 202 prediction of enantioselectivity 202

## и

UCB. *see* upper confidence bound (UCB) United States Patent Office (USPTO) 83 Universal ML toolkits 144 unsupervised learning 107 USPTO. *see* United States Patent Office (USPTO)

### V

validation. *see also* cross-validation (CV) training and test set separation techniques 201 Vaska's complex 169 video, in computer vision 12–14, *13* visualization 6 volcano plot 160, 179–180, *181* 

## W

waterfall algorithm 23–24 white-box algorithm 166

#### X

xTB. *see* Semiempirical Extended Tight-Binding (xTB) xyY color space 6–8, 7

## у

yield HTE 160, 161, 165 prediction 162–165 of reaction 160 YOLOV5. *see* You Only Look Once (YOLOV5) Young-Helmholtz theory 4 You Only Look Once (YOLOV5) 77

## Ζ

zeolite imidazole frameworks (ZIFs) 227 ZIFs. *see* zeolite imidazole frameworks (ZIFs)