Contents

Preface: Shaping the Future of Catalysis Research with Artificial Intelligence *xi*
Valentine P. Ananikov and Mikhail V. Polynski

Part I Machine Learning Applications in Structural Analysis and Reaction Monitoring *1*

1 Computer Vision in Chemical Reaction Monitoring and Analysis *3*
Marc Reid
1.1 Introduction *3*
1.2 Fundamentals of Computer Vision in Chemistry *4*
1.2.1 Color Theory *4*
1.2.1.1 RGB Color Spaces *6*
1.2.1.2 CIE XYZ and xyY Color Space *6*
1.2.1.3 Hue, Saturation, and Value *8*
1.2.1.4 CIE–L*a*b* *9*
1.2.2 Digital Photography Basics *10*
1.2.2.1 How Does a Digital Camera Work? *10*
1.2.2.2 Image and Video: The "Crude Oil" that Fuels Computer Vision *12*
1.2.2.3 Algorithms and Computer Vision *14*
1.2.2.4 Color Distributions/Histograms *14*
1.2.2.5 Texture Analysis *14*
1.2.2.6 Convolution and Edge Detection *17*
1.3 Computer Vision and Machine Learning in Chemistry *19*
1.3.1 Single Image Applications *19*
1.3.1.1 High-throughput Digital Fingerprinting of Drug-like Compound Libraries *19*
1.3.1.2 Segmentation and Parallel Analysis of Nerve Agents *20*
1.3.1.3 Neural Networks for Identifying Glassware *24*
1.3.2 Video Analysis Applications *27*
1.3.2.1 Phase Separation and Partition Coefficients *27*

1.3.2.2 Reaction Kinetics *29*
1.4 Summary and Conclusion *31*
References *31*

2 Machine Learning Meets Mass Spectrometry: A Focused Perspective *35*
Daniil A. Boiko and Valentine P. Ananikov
2.1 Introduction *35*
2.2 Mass Spectrometry in the Machine Learning Era *36*
2.3 Mass Spectrometry Methods Landscape and Their Potential for Machine Learning Applications *38*
2.4 Representative Mass Spectrometry Applications of Machine Learning *41*
2.4.1 Sample Preparation *41*
2.4.2 Data Acquisition *42*
2.4.3 Data Preprocessing *43*
2.4.4 Data Analysis *43*
2.5 Protocol for Solving General Problems in Mass Spectrometry Using Machine Learning *45*
2.5.1 Data Source *45*
2.5.2 Spectra Representation *46*
2.5.3 Algorithm Development *47*
2.5.4 Metric Selection *48*
2.6 Summary and Conclusion *48*
References *49*

3 Application of Artificial Neural Networks in the Analysis of Microscopy Data *55*
Anna V. Matveev, Anna G. Okunev, and Anna V. Nartova
3.1 Introduction *55*
3.2 Deep Machine Learning for Image Analysis *57*
3.2.1 STM Image analysis *57*
3.2.2 TEM Image Analysis *60*
3.2.3 Comparison of Different Neural Networks *62*
3.3 iOk Platform for Automatic Image Analysis *62*
3.3.1 Web-service ParticlesNN *63*
3.3.1.1 Limitation of ParticlesNN *64*
3.3.1.2 Advantage of ParticlesNN *64*
3.3.2 Chat Bot DLgram *65*
3.3.2.1 Limitations of DLgram *67*
3.3.3 No Code ML *68*
3.3.3.1 Limitations of No Code ML *69*
3.3.3.2 Advantages of No Code ML *69*
3.3.4 Comparison of iOk Platform Services with Other Products *70*
3.4 Analysis of TEM Images of Heterogeneous Catalyst by iOk Platform *71*
3.4.1 Automated Analysis of Supported Catalyst TEM Images *71*
3.4.2 High-resolution TEM Images *72*
3.4.3 Single Site Analysis *74*

3.5 Practical Summary *76*
3.6 Future Prospects *77*
3.7 Acknowledgments *78*
References *78*

Part II Quantum Chemical Methods Meet Machine Learning *81*

4 Construction of Training Datasets for Chemical Reactivity Prediction Through Computational Means *83*
Thijs Stuyver and Javier Alfonso-Ramos
4.1 Introduction *83*
4.2 Oracle Design *84*
4.2.1 Compute Time – Accuracy Trade-off *85*
4.2.2 Implications of Optimizing for Multiple Criteria Simultaneously *87*
4.2.3 Benchmarking *88*
4.2.4 Reproducibility *90*
4.3 Sampling the Search Space *91*
4.4 Active Learning Strategies *93*
4.5 Automation Software for Accelerated Oracle Design *94*
4.5.1 autodE *94*
4.5.2 RMSD-PP-TS *95*
4.5.3 TS-tools *97*
4.6 Summary and Conclusion *99*
References *99*

5 Machine Learned Force Fields: Fundamentals, Their Reach, and Challenges *105*
Carlos A. Vital-José, Román J. Armenta-Rico and Huziel E. Sauceda
5.1 Introduction *105*
5.2 Fundamentals of Machine Learning *107*
5.3 Introduction to Neural Networks *110*
5.3.1 The Perceptron *110*
5.3.2 Multilayer Perceptron *112*
5.3.3 The Architecture of a Neural Network *112*
5.3.4 Optimization Algorithms *113*
5.4 Introduction to Kernel Methods *114*
5.5 Machine Learning in Chemical Reactions and Catalysis *115*
5.5.1 Selectivity Prediction *116*
5.5.2 Catalyst Design and Discovery *116*
5.5.3 Experimental Condition Optimizations *117*
5.5.4 Active Site Determination *117*
5.6 Overview and Trends in MLFFs *118*
5.6.1 Neural Network-based FF *118*
5.6.2 Kernel-based FF *119*
5.7 Neural Network-based Force Fields: The SchNet Case *120*

5.7.1 Atom-type Embeddings *120*
5.7.2 Interaction Blocks *121*
5.7.3 The Explicit SchNet Model for $T = 2$ *122*
5.8 Kernel-based Force Fields: The GDML Framework *124*
5.9 Summary and Concluding Remarks *126*
Acknowledgments *127*
References *127*

Part III Catalyst Optimization and Discovery with Machine Learning *131*

6 Optimization of Catalysts Using Computational Chemistry, Machine Learning, and Cheminformatics *133*
David Dalmau and Juan V. Alegre-Requena
6.1 Introduction *133*
6.2 Molecular Descriptors *135*
6.3 Databases *136*
6.4 Cheminformatics *139*
6.5 Automation of QM Protocols *141*
6.6 Automation of ML Protocols *143*
6.7 Concluding Remarks *145*
References *146*

7 Predicting Reactivity with Machine Learning *157*
Lauriane Jacot-Descombes and Kjell Jorner
7.1 Introduction *157*
7.2 Yield *160*
7.3 Activation Energy and Rate Constant *166*
7.4 Selectivity *172*
7.5 Turnover Frequency and Volcano Plots *179*
7.6 Summary and Conclusion *180*
References *183*

8 Predicting Selectivity in Asymmetric Catalysis with Machine Learning *195*
Pavel Sidorov
8.1 Introduction *195*
8.2 Particularities of Enantioselectivity Modeling *196*
8.2.1 Enantioselectivity as a Target Property *196*
8.2.2 Enantioselectivity Data *197*
8.2.3 Principles of Reaction Modeling *198*
8.3 Models for Enantioselectivity *201*
8.3.1 Models Using 2D Descriptors *201*
8.3.2 Models Using 3D Descriptors *204*

8.4 Summary and Outlook *207*
References *209*

9 Artificial Intelligence-assisted Heterogeneous Catalyst Design, Discovery, and Synthesis Utilizing Experimental Data *213*
Rasika Jayarathna, Seyed Majid Ghoreishian, Rahat Javaid, Azadeh Mehrani, Thossaporn Onsree, and Jochen Lauterbach
9.1 Introduction *213*
9.2 Machine Learning Process *216*
9.2.1 Data Generation *216*
9.2.1.1 Literature Data Extraction *217*
9.2.1.2 Laboratory Data Generation *218*
9.2.2 Machine Learning Model Development *221*
9.2.2.1 Data Cleaning *221*
9.2.2.2 Feature Generation/Engineering and Selection *221*
9.2.2.3 Model Selection and Evaluation *222*
9.3 AI-assisted Catalyst Design *222*
9.3.1 Design Rule Extraction via Data Analysis *223*
9.3.2 Design Rule Extraction via Model Interpretation *224*
9.4 AI-assisted Catalyst Discovery *228*
9.4.1 Initial Machine Learning Model *229*
9.4.2 Search Space Determination *230*
9.4.3 Catalyst Recommendation *230*
9.4.4 Catalyst Synthesis and Testing *232*
9.4.5 Iterative Process *232*
9.4.6 Selected Use Cases from the Literature *233*
9.5 AI-assisted Catalyst Synthesis *235*
9.6 Summary and Conclusion *238*
References *239*

Index *249*