Application and Prospect of Machine Learning in the Field of Materials and Chemical Engineering

Fuzhi Wang

Suzhou University of Science and Technology, Suzhou 215009, Jiangsu, China

Abstract: With the rapid development of computer science, machine learning has gradually penetrated into various research fields and led to the development of related fields to different degrees. At recent stage, machine learning (ML) algorithms have been able to extensively search for variable relationships from a large amount of observed data, thus helping people to discover more laws that were once unnoticed. As an emerging approach, ML can efficiently utilize and process large-volume data generated from high-throughput experiments in the field of materials and chemicals. In recent years, with the continuous improvement of various algorithmic tools, more and more researchers have applied ML to materials and chemical research, which has greatly broadened the research ideas and methods. This paper briefly introduces the development of machine learning in recent years; summarizes the classification methods of common molecular descriptors and their types; introduces the classification, common algorithms and application scenarios of different types of ML; reviews the relevant research results in the field of materials prediction and inverse synthesis analysis of machine learning; puts forward related suggestions for the application of ML for research; and finally summarizes the prospects for the future development of machine learning applied to the field of materials and chemical engineering.

Keywords: machine learning, molecular descriptors, material prediction, inverse synthesis design

1. Introduction

In recent years, the continuous progress and maturity of Artificial Intelligence (AI) and Machine Learning (ML) technologies have injected tremendous energy into the rapid development of modern science and technology. Since McCarthy[1] proposed the concept of artificial intelligence in 1956, with the continuous enrichment of basic computer theories and the rapid improvement of performance, AI has made many remarkable achievements in the entire field of science and technology and even in the commercial field.Especially in 2016, AlphaGo's[2] defeat of Go world champion Lee Sedol marked the remarkable height to which artificial intelligence has advanced.

At the same time, machine learning, as the intersection of computer science and statistics, as well as the core of artificial intelligence and data science, researches and builds some special algorithms, so that the computer itself learns from the data and makes the next prediction[3]. In recent years, machine learning is having a significant impact on the field of science and technology, including but not limited to robotics and autonomous vehicle control, speech processing and natural language processing, neuroscience research, and computer vision applications, which makes machine learning for a range of data-intensive problems related to the industry to provide many fire-new and creative solutions.

However, with the vigorous development of cheminformatics[4], machine learning has made remarkable achievements in the research fields of density functional theory (DFT) calculation[5], quantum chemistry[6], molecular design[7], catalyst optimization[8], drug development[9], reaction prediction and inverse synthesis analysis[10], and has shown exciting application prospects. In the traditional field of chemical engineering research, mathematical modeling is extremely important for scientists to understand and design chemical engineering[11]. Octave Levenspiel[12] even pointed out that the development of chemical engineering mainly depends on the optimization and updating of models.

To sum up, as a sub-field of artificial intelligence research, machine learning continues to empower the fields of materials and chemical engineering. In the following content, we will introduce the basic principles and methods of machine learning, review the achievements made by machine learning methods in the materials and chemical engineering, and analyze how machine learning can be applied to various materials and chemical engineering research challenges, and look at future problems that may be solved using machine learning methods.

2. Analysis

Machine learning, which provides computer systems with the ability to learn and enhance experiences automatically and without specialized programming, has been called the most popular technology of the Fourth Industrial Revolution[13]

[14]. The main method to solve the problems is to abstract the actual problems into a mathematical model and solve them, and finally evaluate the model according to various indicators. Therefore, the steps of the application of machine learning in the field of materials and chemical engineering can be described as: building reaction data sets, characterizing data, training models, and analyzing results (As shown in Figure 1)
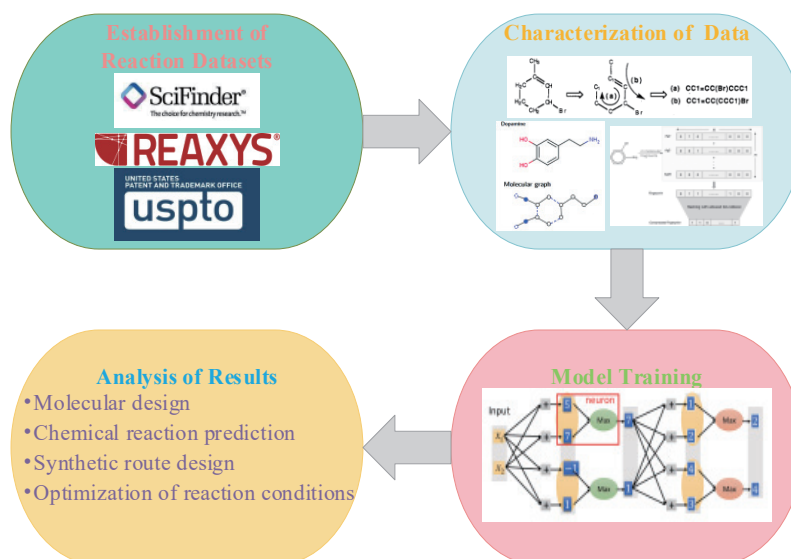

Figure 1. Workflows with machine learning algorithms applied in materials and chemical industry (Establishment of Reaction Datasets→Characterization of Data[15][16][17]→Model Training →Analysis of Results)

2.1 Establishment of reaction datasets

The operation of the machine learning model needs a large amount of data as support, and the model can be trained on the data. At present, the publicly available chemical databases SciFinder, Reaxys, PubChem, etc.[18] can be used for the preliminary screening of data. The screening, summary and sorting of available data sets often require a large amount of energy of scientific researchers, and are usually divided into 80% training set +20% verification set or 70% training set +20% verification set. Similarly, it is possible to characterize all the data before partitioning the data set, but not all the data can be used for model training, and the training set data cannot be used for model evaluation.

2.2 Characterize data

It is well known that the effectiveness of machine learning is often affected by the form of input data, common forms mainly include vectors, matrices, and images. However, the data in the field of materials and chemical engineering can not be used as the input of the model, and the data needs to be transformed to make it more suitable for the format of algorithm processing, which is called feature engineering. In the field of materials and chemical engineering, molecular information is usually described by researchers with molecular descriptors, which can be either the physical and chemical properties of molecules, or numerical indicators related to molecular structure derived according to algorithms.

Molecular descriptors can be categorized into quantitative and qualitative descriptors[19]. Quantitative descriptors include descriptors based on molecular graph theory, various theoretical or experimental spectral data (e.g. ultraviolet spectra), molecular composition (e.g. number of hydrogen bond donors, number of bonds), physical and chemical properties (e.g. ester-water distribution coefficients), molecular fields, and molecular shapes, etc. Qualitative descriptors are generally referred to as molecular fingerprints, which means that the structure, properties, fragmentation, or sub-structures of the molecules are represented by a certain kind of code. Commonly used molecular fingerprints include Daylight fingerprints, MACCS keys, MDL, public keys, and so on.

In addition, based on the data type of the descriptor, molecular descriptors can be classified as Boolean (e.g., chiral or not), Integer (e.g., ring number), Real Number (e.g., molecular weight), Vector (e.g., dipole moment), Tensor (e.g., electron polarizability), Scalar Field (e.g., electrostatic formula), and Vector Field (e.g., electrostatic potential gradient), and other types. Depending on the number of dimensions of the molecular structure required for the descriptor calculation, molecular descriptors can also be categorized as one, two, or three dimensional.

At the same time, there are other classification criteria for molecular descriptors for different computational systems,

for example, in Dragon software[20], according to the differences in the physical meaning of the descriptors can be divided into 20 modules such as Compositional Descriptors, Molecular Property Descriptors, Topological Descriptors, Geometrical Descriptors, etc., and the descriptors of each module represent different chemical information, as shown in Figure 2.

Common molecular descriptors include SMILES descriptors, Inchl descriptors, molecular fingerprints, molecular graphs, and quantum chemical descriptors, of which SMILES descriptors and Inchl descriptors do not require molecular spatial information and are suitable for large datasets; molecular fingerprints are more suitable for calculating similarities between molecules and describing molecular partial structural information; molecular graphs are suitable for use as the input to a graph neural network model and for occasions that require molecular spatial information; and quantum chemical descriptors are suitable for situations that require a precise description of the properties of the molecule.
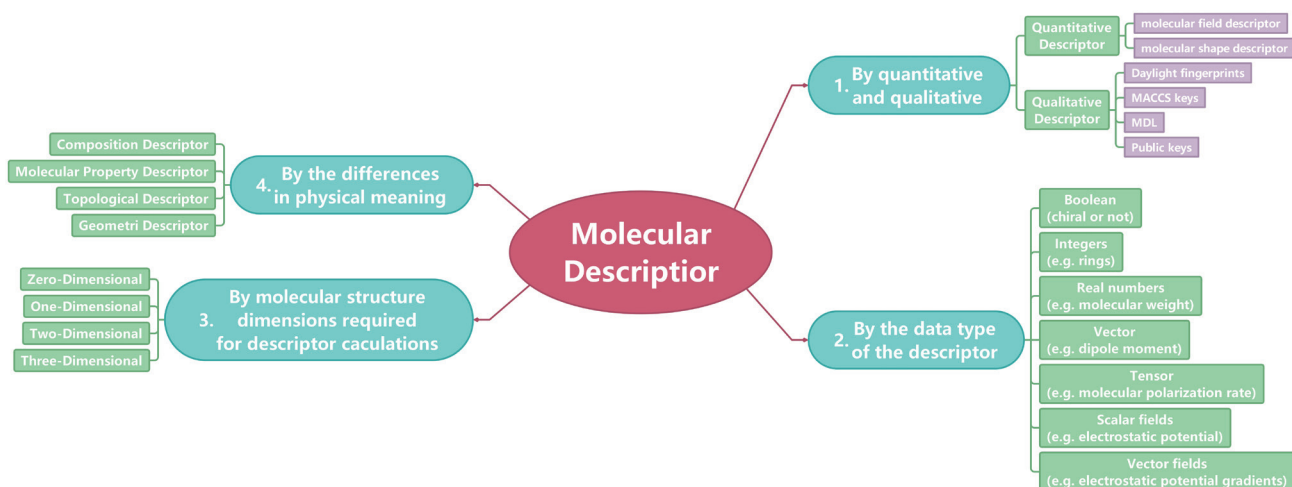

Figure 2. Different ways of categorizing molecular descriptors and their types

Among them, quantum chemical descriptors obtained by Gaussian, NWCHEM and other software calculations are usually used as one of the forms of input data for machine learning due to their ability to more accurately characterize the chemical-physical properties of molecules as well as the high reliability of the calculation results and can be used for the validation of chemical experimental results. For example, Li et al[21] summarized the mechanism and methodology of density-functional theory (DFT) calculations to validate the synthesis of axially chiral styrene, and they found that the nickel-catalyzed enantioselective three-component radical-transferring olefin reductive coupling could realize the synthesis of axially chiral styrene, and the whole reaction mechanism could be verified by finding the transition states in the reaction with the help of DFT calculations[22].

2.3 Model Training

The main methods that are widely used in the field of machine learning are Supervised Learning (SL), Unsupervised Learning (UL), Semi-Supervised Learning (SSL) and Reinforcement Learning (RL)[23].

Supervised learning is to train to get an optimal model based on the existing dataset and the known relationship between input and output results. Its main tasks include classification and regression, and common supervised learning algorithms include Support Vector Machine (SVM), Decision Trees (Decision Trees), Logistic Regression (Logistic Regression) and K-Nearest-Neighbors (KNN). As one of the most common machine learning algorithms, supervised learning can learn from labeled data and predict and generalize unseen data, and its main application scenarios include image recognition, natural language processing, speech recognition, and medical diagnosis.

Unsupervised learning is not clear about the relationship between data and features, and needs to get the relationship between data based on clustering and certain models. Its main categories include Clustering, Dimentionality Reduction, Association Rule Mining and Anomaly Detection. Common unsupervised learning algorithms include K-Means Clustering, Principal Component Analysis (PCA), Association Rule Mining and Anomaly Detection. It plays an important role in data mining, pattern recognition, feature learning and other application scenarios, which can obtain useful information from unlabeled data, provide effective preprocessing steps for other tasks, and help to understand and utilize the data better, and its main applications include clustering and calssificating, feature learning and dimensionality reduction, anomaly detection and association rule mining.

Semi-supervised learning lies between supervised and unsupervised learning and utilizes data containing both labeled and unlabeled data to construct a model, allowing the model to be better generalized to brand new data in the testing phase. The main categories are semi-supervised classification, semi-supervised regression, semi-supervised clustering, semi-supervised anomaly detection and semi-supervised learning in generative adversarial networks. Common semi-supervised learning algorithms include self-training, collaborative training, semi-supervised support vector machines, generative semi-supervised learning, semi-supervised deep learning and graph semi-supervised learning. Their main application scenarios include natural language processing, image recognition and computer vision, data clustering, medical imaging and diagnosis, robot control, image generation and data enhancement.

Reinforcement learning is for an agent to learn behavioral strategies through trial and error in an environment. By interacting with the environment, the agent adapts its behavioral strategies based on reward signals to achieve the goal of maximizing the cumulative rewards. The main categories include value-based reinforcement learning, policy-based reinforcement learning model-based reinforcement learning, deep reinforcement learning, and multi-agent reinforcement learning. Common reinforcement learning algorithms include Q-Learning, SARSA, DQN, A3C, PPO, TRPO. its potential for important applications in many intelligent and autonomous systems, and the main application scenarios include autonomous driving, robot control, gaming, medical therapy, speech recognition and natural language processing.

2.4 Analysis of Results

After the model validation is completed, researchers can compare the real experimental data with the predicted data to validate and assess the quality of the model, and analyze whether the model learning results can be widely used to solve practical problems[24]. Different model evaluation indexes are needed for different problems, for example, the evaluation of classification models in supervised learning can usually use ROC curve, PR curve, Accuracy and Recall; the evaluation of regression models can use Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Coefficient of Decision (R2), etc.

3. Review

3.1 Application of machine learning in the field of materials prediction

Materials machine learning is rapidly transforming many fields to rapidly design new materials with higher performance[25]. Materials have a major impact on the way high technology is developed, and impressive advances have been made in creating new materials due to the need for unique materials with the highest quality[26]. Machine learning provides us with a new strategy to find new materials, which involves iterative trials unlike traditional experimental methods. To provide direction for material design, machine learning makes new predictions by understanding the underlying patterns in the data[27]. Created many high performance alloy materials such as titanium alloys, nickel alloys, high entropy alloys (HEA), aluminum alloys, shape memory alloys, magnesium alloys, block metal glass alloys (BMG), nickel-based single crystal high temperature alloys, and copper alloys, and in the recent years, machine learning has been used to discover and predict many new materials with excellent structure and properties[28].

In machine learning, "material prediction" refers to the use of machine learning algorithms to predict the performance, physical-chemical properties, or a range of parameters of a material. This approach requires evaluating and learning from a large amount of material data to create models that can accurately predict the properties of new and untested materials. For example, Ju et al[29] proposed a goal-directed approach in 2020 to construct robust GA-MLR models to predict the PCE of N-P sensitizers, using data mining techniques and quantum chemical analysis for the design of N-cyclized perylene sensitizers, which showed potential for accelerating the discovery and design of other new energy materials, as well as photovoltaic components.Gao et al[30] in 2021 proposed a new search strategy combining machine learning and DFT computation to screen 5796 inorganic bis-calcite, and for the first time, eXtreme Gradient Boosting Regression (XGBR) algorithm was utilized to construct a machine learning (ML) model of calcite materials with robustness and prediction to obtain two new types of inorganic bis-calcite with a suitable bandgap and high stability of lead-free inorganic bicalcite: Na2MgMnI6, K2NaInI6. Cheng et al.[31] in 2022 integrated molecular modeling, machine learning, and process modeling and simulation to propose a multi-scale design framework of MOF-based membranes for the separation of CO2/CH4 mixtures. GCMC simulations were used to investigate the adsorption isotherms, isothermal heat of adsorption, and adsorption sites of CO2, CH4, and their mixtures in the MOF-based membrane IRMOF-1, and then MD simulations were conducted in an NVD system to investigate the self-diffusion rate and diffusion activation energy of CO2 and CH4 in the mixture under different operating conditions, and the membrane permeability, adsorption selectivity, diffusion selectivity, and membrane selectivity of IRMOF-1 were analyzed at the same time, and the effects of the operating conditions on the performance of IRMOF-1 in the separation

of the CO2/CH4 mixtures were explored by sensitivity analyses, and then the membrane performance of IRMOF-1 in the separation of the CO2/CH4 mixtures was investigated by the machine tool analysis of the artificial neural network (ANN) machine learning approach to construct predictive models of adsorption capacity and self-diffusivity for calculating permeability, providing a series of exciting ideas for the exploration and design of potential materials for related case applications.

3.2 Application of machine learning in the field of inverse synthesis

Inverse synthetic analysis is the most common method for designing synthetic routes by applying known reactions to iteratively decompose a molecule into potentially simpler and more easily synthesized precursors[32]. In recent years, with the development of artificial intelligence technology, computer-aided synthetic design (CASP) has allowed chemists to further consider more complex molecules and save a great deal of time and effort in designing synthetic experiments[33].

Lin et al[34] in 2020 proposed a template-free method independent of reaction templates, rules, or atom mappings for automated inverse synthesis route planning, used a multi-head attention-based Transformer architecture and demonstrated robustness in machine translation, and combined Monte Carlo tree searches and heuristic scoring functions to re-discover the inverse synthesis of four molecular routes.Wang et al[35] in 2021 proposed a template-free and Transformer-based single-step synthesis method called RetroPrime, which has significantly higher Top-1 accuracy than the traditional methods, reflecting the excellent generalizability and robustness of Transformer-based prediction.Liow et al[36] in 2022 used a reverse design framework to predict the optimal experimental design for the LiNixCo1-x-yMn1-x-y-zO2 (NCM) cathode with x + y + z = 1. K nearest neighbor (KNN), random forest (RF), and multiple imputations by chained equations (MICE) were employed to fill in the missing values of the experimental data, gradient boosting regression with decision trees (GBR) as forward modeling algorithms, tuning the hyperparameters of GBR by 10-fold cross-validation grid search, and using particle swarm algorithm (PSO)-based inverse design model to find the optimal experimental conditions for cathode NCMs, which has greatly facilitated the research on lithium-ion batteries.

3.3 Limitations and challenges in the application of machine learning

Despite the fact that the powerful data processing capabilities of machine learning can lead to more ideas and methods for research in the field of materials and chemical engineering, there are still many difficulties in the application of machine learning. The acquisition of high-quality labeled data and the integration of relevant domain knowledge into the models remains a great challenge, while more accurate predictions of complex material behaviors and intricate relationships can only be made by designing more sophisticated and clever algorithms[37].Wiktor Beker et al.[38] presented a study in 2022 based on the prediction of the reaction of the given substrate pair that is best suited to participate in the heteroaryl-heteroaryl or aryl-heteroaryl Suzuki coupling, although for more than 10,000 reaction statistics that appeared to be sufficient for successful ML, feed-forward and graph-convolutional neural network(NN) methods, as well as word-embedding and positive-unlabeled(PU) learning techniques were applied to develop the models after categorizing the solvents, bases, temperatures, and palladium sources, ultimately unfortunately all of them had low prediction accuracies.

It is possible that data-driven approaches may perform poorly for different reaction datasets[39], and to some extent this may reflect some generally unnoticed problems, e.g., in the case of poor reaction yield prediction, the limitations of reaction yield prediction may reflect the fact that reaction yields can vary significantly due to human or environmental factors, such as nuances in the experimental and manipulation skills of researchers, manual procedures, or even different times of the year[40]...... Therefore, despite the fact that machine learning possesses such powerful data analysis and prediction capabilities, we still cannot regard it as some kind of a lifesaver, nor can we simply take its results as authoritative. The application of machine learning can be regarded as a double-edged sword. Reasonable utilization of machine learning can certainly empower us in various aspects of research, but blind abuse of machine learning may sometimes confine our active thinking in the "black box".

4. Conclusion

To sum up, traditional research methods in materials and chemical engineering such as density functional theory (DFT) usually require long research and development cycles, which are not only expensive but also inefficient, and even difficult to keep up with the latest research progress. As one of the most effective methods that can replace repetitive experimental tests and DFT calculations[41], the powerful data processing capability of machine learning provides a new way to better solve the structures and properties of molecules, which is widely used in the materials and chemical engineering fields. In the last decade, machine learning has become a new tool in the chemical engineer's toolbox, and in the future, as machine learning algorithms continue to be optimized and updated they will undoubtedly be able to exaggerate the stock of data processing methods that can be used to solve classical chemical tasks. In order to make machine learning a more valuable and reliable

modeling method, simple and open access to data and models can be maintained in the chemical engineering community in the future; in addition to this, interpretable models should be built, new models for chemical applications are often inspired by existing algorithms, and investigating why a certain output is generated by a given input would benefit research in this field.

References

[1] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine, 27(4), 12.

[2] Silver, D., Huang, A., Maddison, C. et al. Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489 (2016).

[3] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260.

[4] Liu Yidi, Yang Qi, LI Yao et al. Application of Machine Learning in Organic Chemistry [J]. *Chinese Journal of Organic Chemistry*, 2019,40(11):3812-3827.

[5] Brockherde F, Vogt L, Li L, Tuckerman ME, Burke K, Müller KR. Bypassing the Kohn-Sham equations with machine learning. Nat Commun. 2017 Oct 11;8(1):872.

[6] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Phys. Rev. Lett. 109,059802.

[7] Sun T L, Pei J F. Drug design and drug information in the big data era (in Chinese). Chin Sci Bull, 2015, 60: 689–693, doi: 10.1360/N972014-01114

[8] Roger Monreal-Corona, Anna Pla-Quintana, Albert Poater.Predictive catalysis: a valuable step towards machine learning,Trends in Chemistry,2023,ISSN 2589-5974.

[9] LI Wei, YANG Jin-cai, HUANG Niu. Deep learning in drug design and discovery[J]. *Acta Pharmaceutica Sinica*, 2019, 54 (05): 761-767.

[10] Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature. 2018 Mar 28;555(7698):604-610.

[11] Maarten R. Dobbelaere, Pieter P. Plehiers, Ruben Van de Vijver, Christian V. Stevens, Kevin M. Van Geem.Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats.Engineering.Volume 7, Issue 9,2021,Pages 1201-1211,ISSN 2095-8099.

[12] Levenspiel O. Modeling in chemical engineering. Chem Eng Sci 2002;57(22– 23):4691–6.

[13] Sarker, I.H., Hoque, M.M., Uddin, M.K. et al. Mobile Data Science and Intelligent Apps: Concepts, AI-Based Modeling and Research Directions. Mobile Netw Appl 26, 285–303 (2021).

[14] Sarker, I.H., Kayes, A.S.M., Badsha, S. et al. Cybersecurity data science: an overview from machine learning perspective. J Big Data 7, 41 (2020).

[15] David Weininger. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 1988 28 (1), 31-36.

[16] Stefi, Nouleho, Barth, Dominique, David, Olivier, Quessette, Franck, Weisser, Marc-Antoine, Watel, Dimitri. Improving graphs of cycles approach to structural similarity of molecules. PLOS ONE, 2019/12/27(14).

[17] Chakravarti, Suman. Distributed Representation of Chemical Fragments. ACS Omega, 2018/03/31(3), 2825-2836.

[18] Warr WA. A Short Review of Chemical Reaction Database Systems, Computer-Aided Synthesis Design, Reaction Prediction and Synthetic Feasibility. Mol Inform. 2014 Jun;33(6-7):469-76.

[19] Ren Wei and Kong Dexin. On the correlation of molecule descriptors used in QSAR study [J]. *Computers and Applied Chemistry*, 2009, 26 (11): 1455-1458.

[20] Danishuddin, Asad U. Khan. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. Drug Discovery Today. Volume 21, Issue 8, 2016, Pages 1291-1302, ISSN 1359-6446.

[21] Zi-Hao Li, Quan-Zhe Li, He-Yuan Bai, Shu-Yu Zhang.Synthetic strategies and mechanistic studies of axially chiral styrenes.Chem Catalysis.Volume 3, Issue 6,2023,100594,ISSN 2667-1093.

[22] Quan-Zhe Li, Zi-Hao Li, Jun-Chen Kang, Tong-Mei Ding, Shu-Yu Zhang.Ni-catalyzed, enantioselective three-component radical relayed reductive coupling of alkynes: Synthesis of axially chiral styrenes.Chem Catalysis.Volume 2, Issue 11,2022,Pages 3185-3195,ISSN 2667-1093.

[23] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021).

[24] Sun Jie, Li Zihao, Zhang Shuyu. Machine learning applications in chemical synthesis and characterization of [J]. *Journal of Shanghai Jiao Tong University*, 2023, 57 (10): 1231-1244.

[25] Xiujuan Liu, Pengcheng Xu, Juanjuan Zhao, Wencong Lu, Minjie Li, Gang Wang.Material machine learning for alloys: Applications, challenges and perspectives.Journal of Alloys and Compounds.Volume 921,2022,165984,ISSN 0925-8388

[26] Md Hosne Mobarak, Mariam Akter Mimona, Md. Aminul Islam, Nayem Hossain, Fatema Tuz Zohura, Ibnul Imtiaz, Md Israfil Hossain Rimon.Scope of machine learning in materials research—A review.Applied Surface Science Advances.Volume 18,2023,100523,ISSN 2666-5239.

[27] Logan Ward, Stephanie C. O'Keeffe, Joseph Stevick, Glenton R. Jelbert, Muratahan Aykol, Chris Wolverton.A machine learning approach for engineering bulk metallic glass alloys.Acta Materialia.Volume 159,2018,Pages 102-111,ISSN 1359-6454.

[28] Yan Zhang, Cheng Wen, Changxin Wang, Stoichko Antonov, Dezhen Xue, Yang Bai, Yanjing Su.Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models.Acta Materialia. Volume 185,2020,Pages 528-539,ISSN 1359-6454.

[29] Lifei Ju, Minjie Li, Lumin Tian, Pengcheng Xu, Wencong Lu.Accelerated discovery of high-efficient N-annulated perylene organic sensitizers for solar cells via machine learning and quantum chemistry.Materials Today Communications.Volume 25.2020,101604,ISSN 2352-4928.

[30] Zhengyang Gao, Hanwen Zhang, Guangyang Mao, Jianuo Ren, Ziheng Chen, Chongchong Wu, Ian D. Gates, Weijie Yang, Xunlei Ding, Jianxi Yao.Screening for lead-free inorganic double perovskites with suitable band gaps and high stability using combined machine learning and DFT calculation.Applied Surface Science.Volume 568,2021,150916,ISSN 0169-4332.

[31] Xi Cheng, Yangyanbing Liao, Zhao Lei, Jie Li, Xiaolei Fan, Xin Xiao. Multi-scale design of MOF-based membrane separation for CO2/CH4 mixture via integration of molecular simulation, machine learning and process modeling and simulation. Journal of Membrane Science. Volume 672, 2023, 121430, ISSN 0376-7388.

[32] Feng-Yu Zhang, Sicheng Liu, Anwei Huang, Yi-Ning Li, Xiao-Yan Liu, Peng Zhang, A Theoretical Analysis of the Differential Chemical Reaction Results Caused by Chirality Induction, Molecules, 10.3390/molecules28176286, 28, 17, (6286), (2023).

[33] Florian Häse, Loïc M. Roch, Alán Aspuru-Guzik.Next-Generation Experimentation with Self-Driving Laboratories. Trends in Chemistry.Volume 1, Issue 3,2019,Pages 282-291,ISSN 2589-5974.

[34] Lin, Kangjie and Xu, Youjun and Pei, Jianfeng and Lai, Luhua.Automatic retrosynthetic route planning using template-free models.Chem. Sci.Volume 11,Issue 12,2020,Pages 3355-3364.

[35] Xiaorui Wang, Yuquan Li, Jiezhong Qiu, Guangyong Chen, Huanxiang Liu, Benben Liao, Chang-Yu Hsieh, Xiaojun Yao.RetroPrime: A Diverse, plausible and Transformer-based method for Single-Step retrosynthesis predictions. Chemical Engineering Journal.Volume 420, Part 1,2021,129845,ISSN 1385-8947.

[36] Chi Hao Liow, Hyeonmuk Kang, Seunggu Kim, Moony Na, Yongju Lee, Arthur Baucour, Kihoon Bang, Yoonsu Shim, Jacob Choe, Gyuseong Hwang, Seongwoo Cho, Gun Park, Jiwon Yeom, Joshua C. Agar, Jong Min Yuk, Jonghwa Shin, Hyuck Mo Lee, Hye Ryung Byon, EunAe Cho, Seungbum Hong. Machine learning assisted synthesis of lithium-ion batteries cathode materials. Nano Energy. Volume 98, 2022, 107214, ISSN 2211-2855.

[37] Hu Y, Buehler M J. Deep language models for interpretative and predictive materials science[J]. *APL Machine Learning*, 2023, 1(1).

[38] Wiktor Beker, Rafał Roszak, Agnieszka Wołos, Nicholas H. Angello, Vandana Rathore, Martin D. Burke, and Bartosz A. Grzybowski. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. Journal of the American Chemical Society 2022 144 (11), 4819-4827.

[39] Skoraczyński, G., Dittwald, P., Miasojedow, B. et al. Predicting the outcomes of organic reactions via machine learning: are current descriptors sufficient?. Sci Rep 7, 3582 (2017).

[40] Emami, F.S., Vahid, A., Wylie, E.K., Szymkuć, S., Dittwald, P., Molga, K. and Grzybowski, B.A. (2015), A Priori Estimation of Organic Reaction Yields. Angew. Chem. Int. Ed., 54: 10797-10801.

[41] Md Hosne Mobarak, Mariam Akter Mimona, Md. Aminul Islam, Nayem Hossain, Fatema Tuz Zohura, Ibnul Imtiaz, Md Israfil Hossain Rimon. Scope of machine learning in materials research — A review. Applied Surface Science Advances, Volume 18, 2023, 100523, ISSN 2666-5239.