

Emerging Chemistry & Machine Learning



Cite This: *JACS Au* 2022, 2, 541–542



Read Online

ACCESS |

Metrics & More

Article Recommendations

We all are witnesses to the recent explosion of applications of machine learning (ML) in many branches of science. As a way to realize artificial intelligence (AI), ML itself has undergone three stages of progression, being deductive (1950s), knowledge-based (1980s), and data-driven (2000 to now). Undoubtedly, big data, i.e., the increasing accumulation of learnable data, has enabled numerous recent scientific achievements through ML, highlighting the above progression of ML. Nowadays, ML has achieved significant successes in many disciplines, including mathematics, physics, materials science, environmental science, biology and medicine, as well as chemistry. Specifically, ML has greatly boosted the measurement and characterization of chemical species and materials, the analysis and understanding of chemical data and simulation results, as well as the design and optimization of chemical reagents and reaction pathways.

How could chemistry benefit so profoundly from ML? First, ML allows researchers to predict on top of established knowledge, or even to foresee unseen systems, properties and scenarios to some degree by extrapolating beyond our existing knowledge. Second, ML removes the heavy reliance on empirical experience, chemical intuition, as well as repetitive manual labor and thus saves time and resources for more creative and innovative tasks. Third, ML excels in recognizing the intrinsic bias of an individual practitioner, which is favorable for bridging the gaps between experimental and theoretical studies. Finally, with ML, it is possible to learn and extract useful information from unsuccessful efforts. All these virtues have combined to bring forth fresh perspectives and even paradigm changes in many subdisciplines of chemistry, and they will likely make chemistry a more systematic, economic, predictive, and productive branch of science in the near future. At some point, we might see the outdated notion of “chem-is-try” revived in the era of AI, provided that ML enables far more intelligent and efficient ways to “try” than ever before.

While the development of chemistry can now be increasingly driven by ML, our ML techniques also evolve continuously, with user demands and chemical insights incorporated into their frameworks. The most significant limiting feature currently is the limited number of available data in chemistry. Unlike the scale of available data in other disciplines—billions or trillions—the amount of available data in chemistry is often only thousands or even hundreds of examples. As a result, appropriate ML algorithms have to be carefully selected when they are applied in chemical research. Furthermore, the descriptors used in ML should be carefully designed as well.

This Virtual Issue consists of 15 published Articles and Perspectives associated with ML selected from *JACS Au*. The

subjects of these works cover all branches of chemistry, including organic chemistry, inorganic chemistry, analytical chemistry, physical chemistry, and biochemistry, representing the emergence of the breadth of understanding as well as advanced utilization of ML for deep understanding of chemical processes.

Unsupervised ML algorithms that focus on clustering are suitable for categorizations of experiences. To analyze transmission electron microscopy (TEM) images of nanoparticles, T. Head-Gordon, A. P. Alivisatos, and co-workers developed the *AutoDetect-mNP* algorithm, where an unsupervised K-means image segmentation is the essential algorithm (DOI: [10.1021/jacsau.0c00030](https://doi.org/10.1021/jacsau.0c00030)). Remarkably, *AutoDetect-mNP*, with six shape descriptors, can effectively categorize different kinds of Au nanorods and recognize spheroidal impurities from only 20 TEM images that contain less than 1000 individual particles. In another work, H. H. Girault and co-workers demonstrate that, for the noninvasive monitoring of skin disorders, unsupervised hierarchical cluster analysis (HCA) and principal component analysis (PCA) are effective for the analysis of the matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectra (DOI: [10.1021/jacsau.0c00074](https://doi.org/10.1021/jacsau.0c00074)). They found that HCA could distinguish MALDI-TOF mass spectra measured for 66 skin regions from 9 volunteers into three typical skin conditions. Meanwhile, PCA can be used for monitoring the progression stage of skin disorders, which facilitates early diagnosis.

Supervised algorithms that focus on regression and classification are particularly useful for identification, decision making, and high precision prediction. In a Perspective, N. Boehnke and P. T. Hammond demonstrated that ML tools can gain mechanistic insight into drug delivery and thus benefit nanomedicine (DOI: [10.1021/jacsau.1c00313](https://doi.org/10.1021/jacsau.1c00313)). R. Gómez-Bombarelli, B. L. Pentelute, and co-workers used a convolutional neural network (CNN) model for the rational design of short cell-penetrating peptides (CPPs) that can covalently attach antisense oligonucleotides while having a limited number of toxic arginine residues (DOI: [10.1021/jacsau.1c00327](https://doi.org/10.1021/jacsau.1c00327)). They revealed that with rational augmentation of the antisense-peptide database, the CNN model predicts a CPP with 18 total residues and only one arginine residue.

Received: March 1, 2022

Accepted: March 1, 2022

Published: March 11, 2022



Subsequent *in vivo* testing confirmed the predicted CPP's efficiency for drug delivery with no kidney toxicity. T. M. Reineke and co-workers used SHapley Additive exPlanations (SHAP), as well as a linear causality model, to unveil the structure–function relationships between nine polyplex descriptors and the average treatment effect of different polymer delivery vehicles for plasmids (pDNA) and ribonucleoproteins (RNP) (DOI: 10.1021/jacsau.1c00467). They aimed for not only a predictive model but also an interpretive model, establishing useful design guidelines for efficient pDNA delivery and RNP delivery, respectively. S. Park and co-workers showed that the graph convolutional network (GCN) with chemical space vectors that take chromophore–solvent interactions into account can predict experimental optical spectra of dyes in different solvents and in the solid state (DOI: 10.1021/jacsau.1c00035). With the ML-trained model, a blue emitter was rationally designed and its optical and photophysical properties were confirmed experimentally. S. Chen and Y. Jung demonstrated that the message passing neural network algorithm can be adopted for the retrosynthesis of organic compounds (DOI: 10.1021/jacsau.1c00246). They emphasized that an extra global reactivity attention layer with descriptors including molecule graphs, atom features, and bond features can improve prediction accuracy, especially for the reactions including multiple products.

Z.-J. Zhao, J. Gong, and co-workers as well as J. Patrick Zobel and L. González summarized the recent processes of ML-boosted molecular simulations for reactions in operando conditions (DOI: 10.1021/jacsau.1c00355) and excited states (DOI: 10.1021/jacsau.1c00252), respectively, which help the understanding of underlying mechanisms of chemical processes. ML can also help to identify collective variables for reactions of macromolecules and their surrounding environments, such as functional conformational changes of proteins, as highlighted by X. Huang and co-workers (DOI: 10.1021/jacsau.1c00254). J. C. Grossman and co-workers used GCN and random forest algorithms to understand lithium adsorption behaviors on metallic two-dimensional materials (DOI: 10.1021/jacsau.1c00260). They found that, by considering the linear relationship between the lithium adsorption energy and the work function of substrates, the high accuracy and transferability of ML predictions aid the screening of high-voltage materials. Z. Li and co-workers adopted a Gaussian approximation potential in an iterative way to accelerate molecular dynamic simulations for chemical reactions on metallic surfaces (DOI: 10.1021/jacsau.1c00483). They found that, at high temperatures, i.e., those near the melting point of the substrate, the reactions are quite different from those predicted by the temperature-dependent partition function with the optimized structures at zero Kelvin, which should be attributed to the changes of the local chemical environment, atom mobility, and thermal expansion of the surface at high temperature.

The descriptors have to reflect the characteristics of the systems under study. With internal molecular coordinates as the descriptors, B. Jiang, R. J. Maurer, and co-workers showed that the embedded atom neural network (EANN) can accurately predict the potential energy surfaces (PESs) of adsorbed systems. Using these highly accurate ML-based PESs, the memory effects on electronic friction for the scattering of high vibrational state NO on Au(111) have been identified (DOI: 10.1021/jacsau.0c00066). With embedded density descriptors, B. Jiang, J. Jiang, and co-workers also showed

that EANN can precisely predict the transition electronic and magnetic dipole moments of a peptide moiety, which can generate accurate protein circular dichroism spectra of different configurations and thus allow monitoring of molecular details during the evolution of the secondary structures of proteins (DOI: 10.1021/jacsau.1c00449).

Overall, this Virtual Issue reflects only a small fraction for the surge of ML applications in all of chemistry. Despite the major advances already achieved, new developments of ML in chemistry remain essential. A standard process for searching for proper ML algorithms for different kinds of chemical problems is highly desirable. Unveiling the underlying physics of complex problems requires ever more sophisticated descriptors associated with molecular structures and properties. Last but not the least, standardization, digitization, and automation of chemistry is essential for enabling the rapid collection of high-quality data for ML in chemistry (DOI: 10.1021/jacsau.1c00303). It can be anticipated that, with further advancing of the methods and applications of ML, chemistry will be thrust into an unprecedented and fruitful adventure in the coming years.

Christopher W. Jones, Editor-in-Chief

Wasiu Lawal, Assistant Managing Editor

Xin Xu, Associate Editor  orcid.org/0000-0002-5247-2937

■ AUTHOR INFORMATION

Complete contact information is available at:
<https://pubs.acs.org/10.1021/jacsau.2c00142>

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.