ExLibris RapidILL

Rapid #: -20834426

CROSS REF ID: **1481420**

LENDER: **OUU (OU-Tulsa) :: Schusterman Library**

BORROWER: **TKN (University of Tennessee, Knoxville) :: TKN-Library**

TYPE:	Article CC:CCG
JOURNAL TITLE:	Nature Synthesis
USER JOURNAL TITLE:	Nature Synthesis
ARTICLE TITLE:	Empowering scientists with data-driven automated experimentation
ARTICLE AUTHOR:	Jonghee Yang, Mahshid Ahmadi
VOLUME:	N/A
ISSUE:	
MONTH:	
YEAR:	2023
PAGES:	N/A
ISSN:	2731-0582
OCLC #:	

Processed by RapidX: 5/31/2023 9:26:28 AM

https://doi.org/10.1038/s44160-023-00337-z

Empowering scientists with data-driven automated experimentation

Jonghee Yang & Mahshid Ahmadi

Automated experiments with integrated characterization techniques greatly accelerate materials synthesis and provide data to be used by machine learning algorithms. We reflect on the current use of data-driven automated experimentation in materials synthesis and consider the future of this approach.

In recent years, automated experiments have revolutionized the way scientists study materials. Advances in robotic systems have enabled the use of automation in chemical synthesis and materials processing, which has boosted the rate, accuracy, scope, and reproducibility of experiments carried out in a laboratory. When combined with rapid and precise characterization techniques, this greatly increases research throughput and accelerates overall research sequences while minimizing human efforts. Now, researchers can quickly, efficiently, and comprehensively assess the products from the various experimental batches in synthesis and compositional space. This high-throughput automated experimentation subsequently provides massive amounts of the multi-dimensional dataset associated with the physical and chemical properties, as well as functionalities of the materials, allowing for a collective and comprehensive exploration of physical and chemical phenomena[1–4].

Thanks to the rapid development of artificial intelligence, the implementation of machine learning (ML) algorithms opens a new avenue in materials exploration[5,6]. The ML-integrated experimental workflow interprets the vast dataset generated from automated experiments using the present algorithms, which would require a herculean effort for humans. Next, the ML probabilistically suggests plausible solutions to address the problems in the designed experiments (for example, to propose interesting points to be further explored and reveal latencies associated with parametric space)[2–4,7]. This benefits the decision-making process in the autonomous experimental sequence, thereby realizing the data-driven autonomous experimental workflow[2,4]. The closed-loop operation of this experimental cycle substantially accelerates the high-throughput explorations of materials, enabling us to further understand the comprehensive mechanistic insights based on physical and chemical principles, or efficiently optimize the synthesis parameters of materials with desired functionalities[8]. Autonomous experimental workflows have been performed to a high standard in organic chemistry, biochemistry, and pharmaceutical research in the past decade[2,8]. Recently, the advancements have

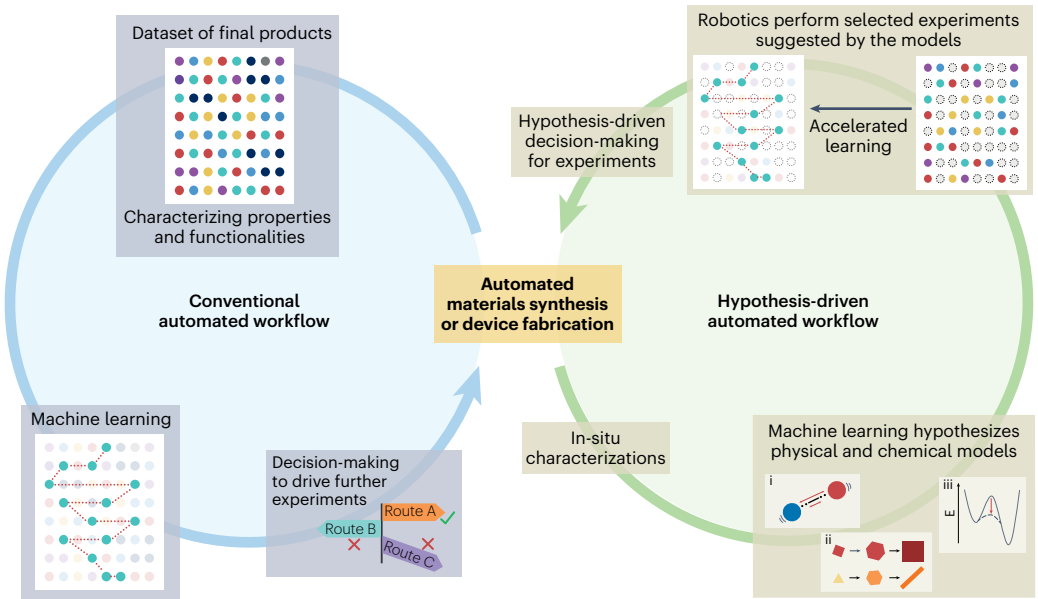

Fig. 1 | Operation workflow of data-driven automated experiments. A conventional workflow of automated experiments (left). The robotic platform produces a complete dataset of final products. From the dataset, ML algorithms make decisions to further drive the experiments. A hypothesis-driven automated experimental workflow (right). In-situ characterizations provide information about the reaction. Subsequently, ML algorithms hypothesize plausible physical and chemical models, such as i) changes in interatomic interaction, ii) nucleation with different crystal structures resulting in different shapes, and iii) changes in the thermodynamic barrier for reaction, in order to direct the robotics to perform selected batches suggested by the models.

extended to complicated materials systems, such as photocatalysts, perovskites, and colloidal nanocrystals[3,4,9]. These approaches have contributed to the understanding of structure–property relationships, provided fundamental insights associated with the synthesis reactions, and realized bespoke tailoring of the target functionalities via on-demand synthesis designs. Indeed, data-driven automated experiments are now revolutionizing the research workflow to be more efficient and effective.

A recent Review by Abolhasani and Kumacheva describes the significance and status quo of self-driving labs (SDLs) and provides guidelines and actions to encourage scientists to establish their own SDLs, in particular, encouraging collaborations between researchers in academia and industry[2]. In addition, limitations and bottlenecks of the current automated experimental platforms are identified. For example, in terms of robotics, there are substantial financial barriers to constructing automated experimental systems in order to explore user-designed tasks. Particularly, this may discourage early-career researchers eager to establish reliable and reproducible automated experimental systems to explore emerging and fascinating subjects in various material spaces. The use of open-source robotics modules for automated experiments or cloud labs can be a solution to notably mitigate such a desire–reality gap[2]. Here, standardization of hardware and the development of stable sample-transfer technologies will further relieve the researchers' concerns, stemming from the spatial discreteness between the synthesis, characterization, and processing modules in the alternative networked systems. In this regard, worldwide consensus as well as active collaborations are urgently recommended. In addition, a considerable understanding of mechanics, fluidics, and robotics is practically required for troubleshooting during operations, which is not trivial for scientists who majored in other disciplines, such as chemistry or materials science. The robotic systems should therefore provide user-friendly interfaces and guidelines so that even non-experts can easily work with laboratory workflows.

From the ML perspective, most ML algorithms implemented in the automated experimental workflow are highly specific to the parametric optimizations of the materials synthesis or processing sequence. As already mentioned, from a massive dataset collected via online, high-throughput characterizations, the ML algorithms can figure out the latencies that have not been identified or propose new experimental conditions as the next step to minimize the uncertainties associated with the chemical reactions, material processing, and so on. Then, the robotic experimental platforms self-operate based on the suggestions of ML algorithms in a closed loop. As a result, the ML models can provide comprehensive information regarding the parametric space of the subjects, revealing optimal conditions for the synthesis that manifest the best functionalities or performances of the materials or final devices (Fig. 1)[7]. These state-of-the-art ML algorithms and their combination with SDLs are unambiguously powerful in overall research workflows, enabling scientists to be free from time-consuming repetitive tasks and focus more on solving scientific questions[2]. However, the present ML algorithms do not directly provide fundamental principles or any scientific rationalization for the obtained results; they do not have any physical or chemical intuitions. Hence, SDLs can perform unnecessary tasks to produce data, which is beyond the minimum required to optimize parameters, discover functionalities, or disentangle scientific curiosities.

Lots of physical and chemical properties of precursors are independently and harmoniously associated with the chemical reaction for the materials synthesis. Therefore, subtle changes during the synthesis and processing can largely alter the functionalities of the resulting materials. For example, the functionalities of metal halide perovskites – a leading class of materials that can provide groundbreaking performances in optoelectronics – are not only strongly dependent on the details of processing parameters and compositional and chemical engineering, but also entangled with each parameter. So far, the ML algorithms can statistically predict the optimal conditions in fabrication or synthesis processes that realize the best performances as well as pursue figures of merit at the final stage (more specifically, device-level or product materials), based on the collected device assessment datasets. However, it remains challenging for the ML algorithms to predict the optimal conditions, particularly based on the considerations of the chemical and physical models and other phenomena associated with the materials – analogous to how humans drive the experiments. Attaining data- and physics-informed autonomous experimentation will be the next step for ML development[2] which can greatly impact the design of SDLs and accelerate materials discovery.

It should be noted that ML algorithms mimicking hypothesis-driven automated experiments, using a probabilistic approach to each plausible physical model, have been already implemented in characterization techniques[10]. However, for the synthesis of functional materials or multi-step device fabrication processing, it is extremely difficult to enumerate the physical and chemical models or predict plausible mechanisms. This is because multiple physical and chemical actions – taking place from the atomic and molecular level to the device level – are complicatedly entangled. As a possible strategy to overcome these complexities and realize advanced ML algorithms in materials synthesis, it is desirable to actively integrate fundamental, in-situ characterization techniques targeting the principle parameters in physics and chemistry with the SDL workflow. The resultant dataset associated with fundamental principles will subsequently allow us to formulate new ML algorithms with physio–chemical intuitions, scaling down the experimental batches with more feasible and necessary candidates. This in turn can realize a powerful SDL that wisely exploits the precious datasets obtained via data-driven automated experiments, thereby accelerating chemistry and materials science discoveries.

Jonghee Yang & Mahshid Ahmadi ✉
Institute for Advanced Materials and Manufacturing, Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN, USA.
✉e-mail: mahmadi3@utk.edu

Published online: 30 May 2023

References

1. Antami, K. et al. *Adv. Funct. Mater.* **32**, 2108687 (2022).
2. Abolhasani, M. & Kumacheva, E. *Nat. Synth.* https://doi.org/10.1038/s44160-022-00231-0 (2023).
3. Higgins, K., Valleti, S. M., Ziatdinov, M., Kalinin, S. V. & Ahmadi, M. *ACS Energy Lett.* **5**, 3426–3436 (2020).
4. Epps, R. W. et al. *Adv. Mater.* **32**, 2001626 (2020).
5. Liu, Z. et al. *Joule* **6**, 834–849 (2022).
6. Sun, S. et al. *Joule* **3**, 1437–1451 (2019).
7. Ahmadi, M., Ziatdinov, M., Zhou, Y., Lass, E. A. & Kalinin, S. V. *Joule* **5**, 2797–2822 (2021).
8. Shields, B. J. et al. *Nature* **590**, 89–96 (2021).
9. Higgins, K., Ziatdinov, M., Kalinin, S. V. & Ahmadi, M. *J. Am. Chem. Soc.* **143**, 19945–19955 (2021).
10. Ziatdinov, M. A. et al. *Adv. Mater.* **34**, 2201345 (2022).

Acknowledgements

The authors acknowledge support from the National Science Foundation (NSF), award no. 2043205, and the Alfred P. Sloan Foundation, award no. FG-2022-18275.

Competing interests

The authors declare no competing interests.