ESM Cloud Toolkit: A Copilot for Energy Storage Material Research

Jing Xu(许晶)[1,2], Ruijuan Xiao(肖睿娟)[1,2*], and Hong Li(李泓)[1,2]

[1]*Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

(Received 3 January 2024; accepted manuscript online 17 April 2024)

Searching and designing new materials play crucial roles in the development of energy storage devices. In today's world where machine learning technology has shown strong predictive ability for various tasks, the combination with machine learning technology will accelerate the process of material development. Herein, we develop ESM Cloud Toolkit for energy storage materials based on MatElab platform, which is designed as a convenient and accurate way to automatically record and save the raw data of scientific research. The ESM Cloud Toolkit includes multiple features such as automatic archiving of computational simulation data, post-processing of experimental data, and machine learning applications. It makes the entire research workflow more automated and reduces the entry barrier for the application of machine learning technology in the domain of energy storage materials. It integrates data archive, traceability, processing, and reutilization, and allows individual research data to play a greater role in the era of AI.

DOI: 10.1088/0256-307X/41/5/054701

1. Introduction. Due to the depletion of fossil fuels, the intensification of environmental pollution, and the increasing demand for energy, it is urgent to utilize renewable clean energy such as solar and wind energy. Among various energy storage technologies, rechargeable battery is one of the most promising methods. Lithium ion batteries have been widely used in electrical vehicle, smart mobile phone and electricity grid, etc.[1,2] However, the energy density, rate performance, safety, and cost are still the key issues limiting the applications of batteries.[3] From a historical perspective, breakthroughs in batteries are regularly related to the discovery of new materials, such as $LiFePO_4$[4] and $LiCoO_2$.[5] As the foundation of batteries, it is crucial to search for and modify new active/non-active materials to improve the comprehensive performance through both experimental and theoretical investigations.

With the continuous accumulation of scientific research data and the development of artificial intelligence (AI) technology, the combination of AI and big data has been regarded as "the fourth paradigm of science",[6] and it showcases the powerful capacity when handle various tasks such as image recognition and natural language processing. These techniques have been applied in computational/experimental material science research.[7–10] The high quality and quantity datasets are key role in training generalized and robust machine learning models. At present, the academia data includes research publications,[11,12] datasets,[13] academic profiles, and other information associated with the academic community. One of the main origination of these data is the output produced during the scientific exploration by individual researchers. The materials science electronic laboratory platform (MatElab),[14] developed by the Condensed Matter Physics Data Center in Institute of Physics Chinese Academy of Sciences, is a flexible web application for scientific data managements and post processing. It provides electronic lab notebook service includes abundant features such as research data recording, logging, archiving, and sharing. In addition, the open application programming interface (API) allows the users to interact with MatElab or connect experimental equipment with MatElab through code. In order to collect and archive individual research data, the MatElab provides a convenient and precise solution.[15] As an example, Wang *et al.*[16] developed an automated data acquisition scheme of thin film growth based on MatElab, and Shi *et al.* employed this digital platform to assist and document the solid state synthesis process of materials. Compared to other scientific data management platforms such as elabFTW and LabFolder, MatElab, which is specifically designed for the physics discipline, comprehensively documents the entire process of data generation, processing, and reuse. Additionally, its API provides users with flexible control over individual scientific data. In this work, we implemented new tools, ESM Cloud Toolkit, based on the MatElab platform for storing, analyzing, and utilizing the computational and experimental data in energy materials research, establishing a comfortable and efficient big data-based toolkit for development of new energy storage materials.

The Energy Storage Materials Cloud Toolkit (denoted as ESM Cloud Toolkit) based on MatElab (Electronic Laboratory for Material Science, denoted as MatElab) is a series of automated workflows for energy storage material research. This toolkit implements three major functions currently: computational data analysis and collection, machine learning applications, and experimental data analy-

*Corresponding author. Email: rjxiao@iphy.ac.cn
© 2024 Chinese Physical Society and IOP Publishing Ltd

sis. For individual research groups, our toolkit enables the analysis of research data more efficiently, allowing researchers to focus more on solving scientific problem rather than tedious data archiving, and to prevent the serious issues of data loss and tampering. In addition, the reliable and fruitful datasets are collected standardly, which is the base of reutilization through AI techniques. This toolkit can make the combination of experimental data and computational data closer, reducing the entry barrier of machine learning applications by offering a simple interface, and enabling scientific data to play a greater role in the era of big data.

2. The Workflow of ESM Research. The Development of energy storage correlated materials generally can be summarized as a basic process, which consists of *investigation*, *exploration*, and *reutilization* (Fig. 1). After proposing a new idea, for example, the specific configuration of active materials[17] or the exploration for the optimal ratio of doping ions,[18] a comprehensive literature survey (*investigation*) is needed to check whether anyone else explored the similar idea and the corresponding conclusions. Usually, the *investigation* requires time-consuming and labor-intensive manual searching of literature. However, With the powerful application of natural language processing (NLP) in text classification and Question Answering systems, some literature search methods powered by AI emerged, such as the MatScholar,[19] created by research groups at Lawrence Berkeley National Laboratory. The survey's results may help us determine the methods or strategies adopted in the subsequent steps. The *exploration* can be conducted experimentally, computationally and theoretically. These types of methods are combined in research. The experimental results require theory to deduce the reason, and the results of theoretical simulations require experimentation to verify. Analyzing and interpreting the results obtained from the *exploration* step can lead to refine the initial idea and conduct further *investigation*. Ultimately, the results of experiments or computational calculations will be presented in the form of charts or text. As research progresses, a substantial amount of unstructured research data accumulates. A logical consideration is how to extract additional knowledge from this existing data, a process referred to as data *reutilization*. With wider applications and developments of machine learning techniques, it turns out that powerful models can be trained based on the big datasets. Hence, the experimental and computational data should be collected in a standardized manner and facilitate to learn more knowledge from the datasets. The model inferring from the collected datasets further guides the *exploration*. These three interrelated parts (investigation, exploration, and reutilization) together constitute the workflow of energy storage material research, and automating this workflow will accelerate development process of energy storage materials.

The ESM Cloud Toolkit makes the data flows easier and more efficient, the storage, analysis, reutilization of data can be completed in this toolkit, which reduces the tedious individual efforts and improves the scientific researchers over data greatly. Based on MatElab, the functions, such as calculation data collection, experimental data analysis, and machine learning prediction, are integrated, and online data processing and reutilization can be achieved through the exquisite graphical user interface (GUI).

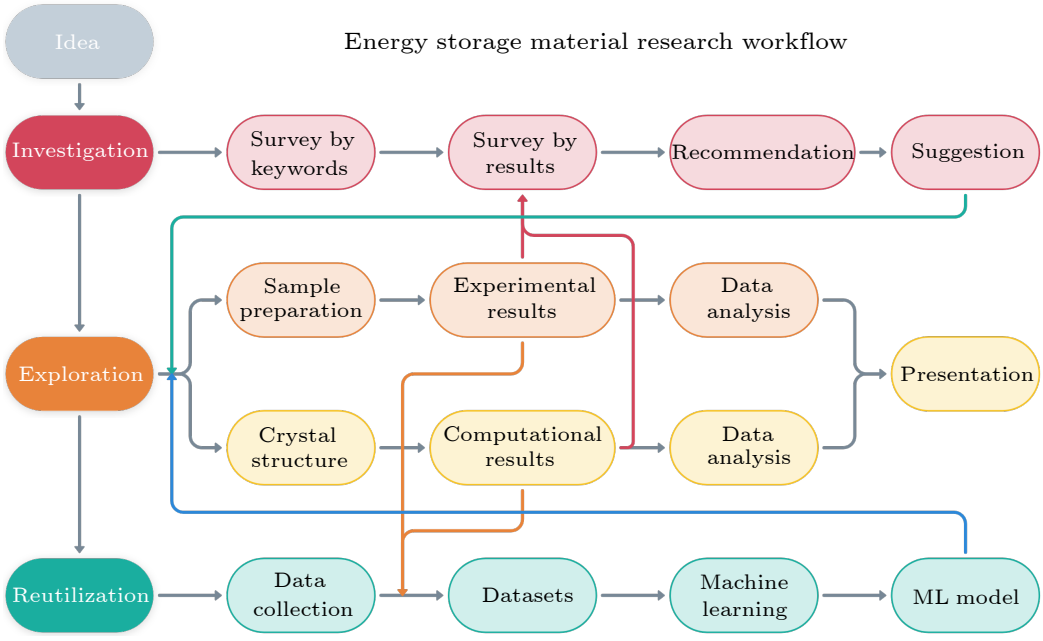

Fig. 1. The general workflow of energy storage material research. After proposed an *idea*, *investigation* and *exploration* are employed primarily, and then the large amount of data (experimental data or computational data) can be collected for *reutilization*. The ESM Cloud Toolkit implemented in MatElab is aimed to make the entire workflow more automated.

3. Main features of ESM Cloud Toolkit. According to the general workflow of energy storage material research, we developed three main functions in ESM Cloud Toolkit, which is computational results collection and archive, machine learning applications and experimental data analysis. All functions can be employed with similar ways by filling the setup modules and running the corresponding plugins. We introduce the details of each function in the following.

3.1 Computational Results Collection. Computational material science is significant in research of energy storage materials and the high-throughput computation plays a crucial role in screening new active materials (such as cathode,[20] and electrolyte[21]), predicting the performance and explaining the inherent mechanisms of rechargeable batteries.[22] A large amount of virtual crystal structures are designed for different scientific issues such as the configurations of Na ions in fast ion conductors[23] and the distribution of transition metal ions in NASICON electrolytes.[24] First principles methods are employed to determine the properties of various virtual crystal structures[25] and the data generated by this way is valuable because of its high accuracy and computational cost. In order to summarize and archive the simulation data, including the high-throughput results, we integrated the function of data archive in the ESM Cloud Toolkit, which can automatically traverse the folders which include the computational results and send the data to storage server, extract key calculation parameters or files, and issue warnings for data with some problems such as energy non-convergent in the form of a single record. Relevant information is also simultaneously uploaded to the MatElab platform. For example, in order to determine the distribution of Na ions in $Na_3Y_2Cl_9$ structures [as shown in Fig. 2(a)], we enumerate a series of different distribution of Na ions and calculate the corresponding energies. Figure 2(b) is a display of the record of computational data generated by inversely designing of new $Na_3Y_2Cl_9$ fast ion conductors.[21] The target description, related chemical elements, computational types, and setup parameters are recorded in this single record. It is worth noting that the form of the datasets collected and uploaded to this record is convenient for further exploration through these data. The data can also be transferred into other formats, e.g., the Atoms format (an instantiated class in ASE python library[26]) to storage the structure optimizing results, after packaging and compressing relative files.

There are many benefits to archive the computational data using automatic data uploading function in ESM Cloud Toolkit. First, most blanks required to fill in this record are automatically generated based on the computational results and do not require manual filling. At present, only the target description is filled manually and this situation may be alleviated with the broaden applications of large language models. Another noteworthy advantage is the standardized recording of data. Everyone has himself/herself own way of recording data, and even individuals use different ways to record data every time. For individuals, even if the data recorded within a few days are similar, it still causes difficulties for subsequent batch analyzing records. The automatic data uploading function creates standardized record and fills the "basic information" module about the description of the simulation task automatically, which can prevent erroneous and confusing recording. Standardized records are also beneficial for individuals to trace their own data and share their records with others, preventing data loss or malicious tampering, thus forming a dedicated and trustworthy dataset for themselves or individual groups and being employed in the subsequent big-data based analysis and utilization.

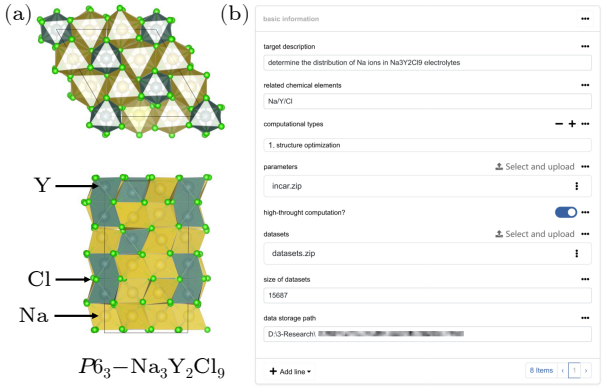

Fig. 2. (a) Top view and side view of the $Na_3Y_2Cl_9$ crystal structure with $P6_3$ space group. (b) The screenshot captured from electronic notebook for $Na_3Y_2Cl_9$ related scientific data on MatElab. The basic information of density functional theory calculations, such as the scientific target, related chemical elements, computational type (structure optimization or static calculation), and the significant setup file (INCAR) are extracted automatically and generated a recording report on MatElab.

3.2 Machine Learning Application. The collection and archive of computational results provide us reliable datasets, and the *reutilization* of these datasets can generate some more comprehensive results or representative models to help us explore and predict unknown data in the future. Therefore, the second main function of ESM Cloud Toolkit is a group of machine learning plugins, which can apply machine learning techniques to explore the datasets through GUI on MatElab platform.

A general machine learning workflow can be divided to 4 steps, which are data collection, feature generation, model training, and prediction[27–29] [Fig. 3(a)]. We developed four plugins to realize the machine learning application simple and flexible [Fig. 3(b)]. It is worth noting that the whole process of machine learning can be achieved on GUI, which eliminates the need for any setup of programming environment or codes. As for data collection, in other words, preparing datasets, the automatic computational results collection and archive function mentioned above can make the preparation of datasets efficiently. However, these datasets need to be further processed to be suitable for machine learning, typically referred to as feature engineering. For example, it is important to predict the total energy and atomic forces of crystal structure in computational materials science. However, the density functional theory or empirical force field face the speed and accuracy trade-off, and the machine learning interatomic potentials

(MLIP) can alleviate this challenge.[30] The Cartesian coordinates of atoms in a structure unit cell is not suitable for machine learning model because it is not consistent with the obvious physical symmetries. Therefore, many descriptors based on inner coordinates are designed to represent crystal structures.[31–33] Some popular descriptors generation methods are integrated into the *descriptor generation* plugin and users can run it by a simple single click. After representing crystal structure with descriptors, the machine learning model can be trained based on different algorithms. In order to simplify the model training process, we developed the plugin to realize model training on MatElab platform. For researchers who are not familiar with programming can also easily use machine learning techniques by plugins and for researchers who have a lot of knowledge about machine learning, they can also train some models through flexible parameter configuration via ESM Cloud Toolkit. Calling the trained model to predict unknown data is one of the most important advantage of machine learning models. The ESM Cloud Toolkit is convenient for users when the trained model is deployed on MatElab. As long as one can log in to the MatElab platform, one can call the model anytime, anywhere on any device. It is important to note that data quality significantly influences the performance of machine learning models.[34] The generation conditions of scientific research data are used to integrate the datasets according to their precision. This strategy avoids the impact of mixing data with different accuracies on machine learning performance. The advantage of MatElab and ESM Cloud Toolkit is to record the complete process of generating scientific research data and it is beneficial for constructing datasets with the same accuracy level.

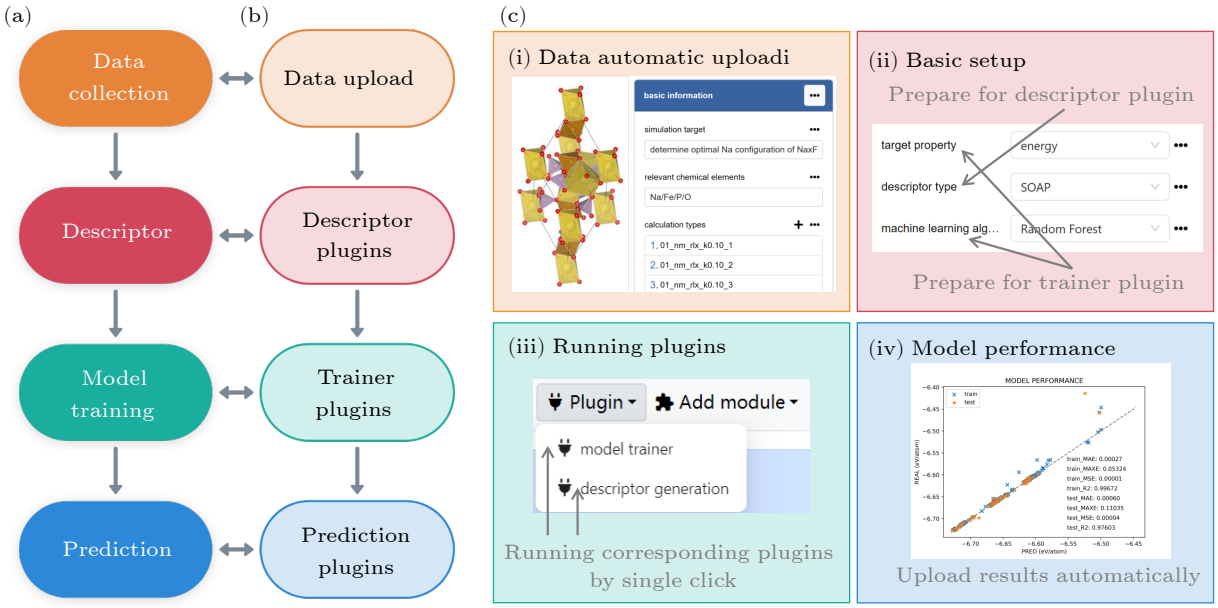

Fig. 3. (a) The four main steps in machine learning applications. (b) Four features in the ESM Cloud Toolkit enable machine learning techniques easily applied for scientific issues in energy-storage material domain. (c) Demonstration of the four features mentioned in (b). The datasets of $Na_xFe_2(PO_4)_3$ can be uploaded to electronic lab notebook automatically, and single click the plugin running icons after filling up the setup parameters blanks. Finally, the final model and corresponding performance are uploaded automatically.

In order to provide a clear demonstration of the functions of machine learning plugins group, we explore the relationship between crystal structures and total energies in $Na_xFe_2(PO_4)_3$ ($x = 1, 2, 3$) datasets. NASICON materials are promising candidates as both solid electrolytes and cathodes in solid-state batteries,[35] and the voltage of operating batteries can be simulated by calculating total energies of ground state of active materials with different working ion contents. The dataset of $Na_xFe_2(PO_4)_3$ ($x = 1, 2, 3$) is obtained when we study the operating voltage of this specific material and this dataset is abbreviated as NxFP dataset. The transformation from original first-principles computational results to standard NxFP dataset is via automatic data uploading function in ESM Cloud Toolkit. Figure 3(c)i displays the representative crystal structure and the corresponding record. Before the descriptors generation, some basic setup is necessary, which includes the designation of descriptor type, machine learning algorithms, and so on, as shown in Fig. 3(c)ii. After basic setup, we can generate the specific descriptors and train the model via running plugins in ESM Cloud Toolkit on MatElab. The final results, including the model and performance of model, will be uploaded to the same record automatically as shown in Figs. 3(c)iii and 3(c)iv. In the demonstration of machine learning task about predicting the total energies of crystal structures, the mean absolute error of model on training set and testing set reach 0.27 meV/atom and 0.60 meV/atom, respectively.

The trained model is deployed via ESM Cloud Toolkit on MatElab and the model prediction can be done anywhere and anytime. The controllable workflow we display will make the model training on specific datasets more convenient and easier, which can promote the machine learning techniques applying in energy storage materials. This approach is not constrained by the research field and the nature of the target properties, making it easily applicable to any problem with a suitable dataset.

In addition to explore the relationship between crystal structures and total energies, more challenging task, such as pretrained neural network fine-tuning, is also integrated to the ESM Cloud Toolkit. Machine learning interatomic potentials (MLIPs) based on deep learning techniques are powerful tools, which enable us to simulate physical phenomena with larger size in both time and space and exhibit the impressive performance in reproducing and predicting the thermodynamic and kinetic properties of materials.[36] Training an MLIP from scratch needs a great amount of expensive first-principles data, thus fine-tuning a pretrained model for similar tasks to solve the specific task is a better strategy. This strategy not only greatly accelerates training speed but also reduces the required training data.[37,38] The process of MLIP fine tuning can be easily completed using the ESM Cloud Toolkit. For example, the newly released pretrained model named CHGNet[38] is integrated into ESM Cloud Toolkit, and researchers can call archived computational data to complete CHGNet fine-tuning task with GUI. Figure 4 exhibits an example of fine-tuning CHGNet pretrained model on $Li_xNi_{10}CoMnO_{24}$ with different Li content dataset via ESM Cloud Toolkit. The dataset can be uploaded to electronic lab notebook (ELN) in MatElab and some basic training parameters should be setup via filling the corresponding blanks [Fig. 4(a)]. Submitting training task by single click model trainer plugin and task status can be monitored by running trainer status plugin [Figs. 4(b) and 4(c)]. After training the model, the final results (ML model and corresponding performance) are returned into record as shown in Fig. 4(d). By simply setting some parameters and running the corresponding plugins, the fine-tuning process of the pretrained model can be completed even without extensive background knowledge of MLIP or without any coding skills. The obtained MLIP can be stored in MatElab and adopted for the atomic simulations in the future.

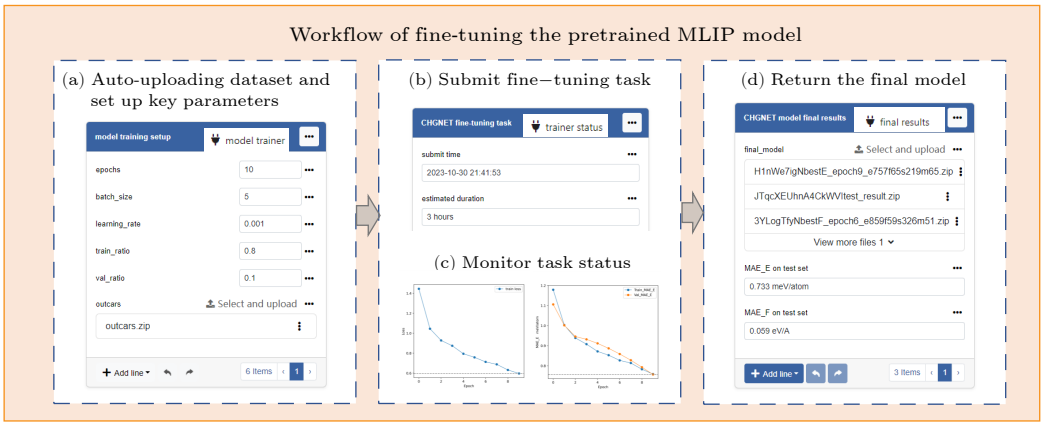

Fig. 4. The workflow of fine-tuning the pretrained models. The demonstration is re-training the CHGNet model by specific datasets, and the value of the loss function is continuing decline during the training process, which means that the task is running normally. The obtained MLIP model and its performance are uploaded to MatElab after running final results plugin.

3.3 Experimental Data Process. Thanks to the rapid development of computational power and advanced simulation method, we have collected a large amount of high-throughput computational results as well as data from previous calculations targeting specific systems, such as $Na_3Y_2Cl_9$,[23] and a systematic workflow of data archive and reutilization are designed. In addition to theoretical exploration, experimental data can be processed by ESM Cloud Toolkit as well. The theoretical results should be verified by experimental results, and the novel physical phenomenon require the theoretical explanations. Therefore, raw experimental data analysis, archive, display and backup are significant functions, which are concerned when developed ESM Cloud Toolkit. As a representative instance, we show the powerful and easy-use capacity of ESM Cloud Toolkit for experimental data.

Electro-chemical tests, such as cyclic test and electrochemical impedance spectroscopy (EIS), are crucial in characterizing the performance of battery devices, for instance, the stability and dynamic properties. Figure 4 exhibits comparison of original data analysis pipeline and automatic data analysis pipeline in ESM Cloud Toolkit, and the steps required for data processing are greatly reduced to 3 main steps, which are uploading all data files, filling parameters setup blanks, and running plugins by single click. The right part of orange panel in Fig. 5 displays the final results (images) via experimental data processing function in ESM Cloud Toolkit. Furthermore, in order to share latest results with other researchers, a simple but interesting function in ESM Cloud Toolkit is to generate

slides automatically, which collect the critical results of specific record and present in the format of a PowerPoint file. These features can greatly reduce the workload of duplicate data processing, enabling data backup, processing, and result saving to be completed in a single record, and the entire data flow to be fully recorded. It is worth noting that this data recording method shortens the (physical and practical) distance between theoretical results and experimental data, making it convenient for data comparison and collaborative processing. It is particularly crucial for building a complete computational and experimental database in the future.

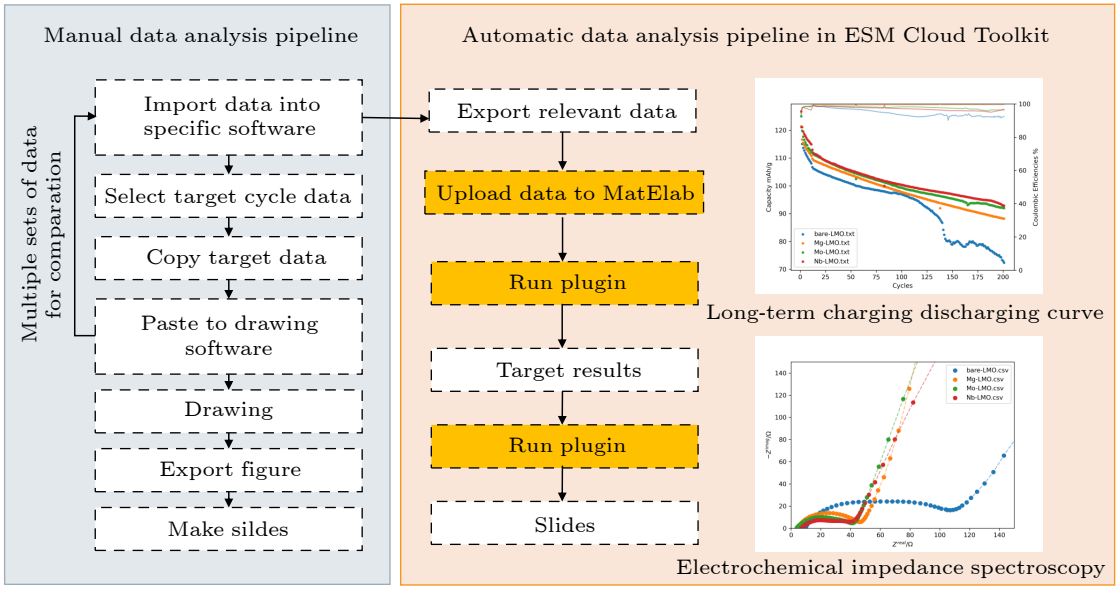

Fig. 5. Comparison of data analysis pipeline between manual and automatic data process. It is evident that the manual workload required for automated processes is greatly reduced. The right part of orange panel is demonstration of long-term charging discharging curve and electrochemical impedance spectroscopy generated by automatic data analysis pipeline via ESM Cloud Toolkit.

4. Conclusion and Perspectives. The core purpose for ESM Cloud Toolkit is to make the entire research chain of *investigation*, *exploration*, and *reutilization* automatically, and the relevant multiple functions integrated in the ESM Cloud Toolkit greatly promote the automation of research and greatly reduce the workload of researchers in both experimental and computational areas. Importantly, this workflow integrates data archive, data processing, and data reutilization, enabling scientific data to play a greater role in the big data era. The functionalities offered by this free and open-source toolkit, specifically designed for energy storage materials development, stem from practical research in the field of energy storage materials. This unique focus gives it a competitive edge over other materials development platforms. The way computational data and experimental data are recorded and processed in the same record also closely link theoretical calculation and experimental research. The GUI also makes the entire workflow easier to advance. There is no need for each researcher to configure the programming environment, install software, etc. on their own computer, and it can be all centralized on a cloud server. In addition, the key concepts in ESM Cloud Toolkit is scalable and similar workflows can also be developed in other material science fields, allowing the big data plus machine learning techniques to be a copilot for daily research.

As a demonstration, the powerful and useful functions of ESM Cloud Toolkit based on MatElab are attractive and we will integrate more functions in the future. The various parts of entire *investigation*, *exploration*, and *reutilization* workflow will be more intelligent and the relationships between different parts of the research more closely related and mutually reinforcing, as shown in Fig. 1. In addition, the integration with hardware is equally important, achieving automatic generation of specified records on MatElab after implementing experiments, and automatic data processing. The deep integration of data backup, data processing, and data utilization on MatElab via ESM Cloud Toolkit will promote the development of energy storage materials. The wide establishment of physical laboratory is an extremely significant milestone in the development of physics in 19th century,[39] and it is rational to believe that the combination of automatic physical laboratory and machine learning techniques will be an ongoing revolution in materials science. The main goals of MatElab and ESM Cloud Toolkit is to drive this revolution in the domain of condensed matter physics and energy storage materials.

Data Availability Statement. The data, source code and demonstration videos that support the findings of this work are available at https://github.com/xujingZR/ESM-Cloud-Toolkit.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant Nos. 52022106 and 52172258), and the Informatization

Plan of Chinese Academy of Sciences (Grant No. CAS-WX2021SF-0102).

References

[1] Dunn B, Kamath H, and Tarascon J M 2011 *Science* **334** 928

[2] Marom R, Amalraj S F, Leifer N, Jacob D, and Aurbach D 2011 *J. Mater. Chem.* **21** 9938

[3] Goodenough J B and Kim Y 2010 *Chem. Mater.* **22** 587

[4] Padhi A K, Nanjundaswamy K S, and Goodenough J B 1997 *J. Electrochem. Soc.* **144** 1188

[5] Kang K, Meng Y S, Bréger J, Grey C P, and Ceder G 2006 *Science* **311** 977

[6] Lombardo T, Duquesnoy M, El-Bouysidy H, Årén F, Gallo-Bueno A, Jørgensen P B, Bhowmik A, Demortière A, Ayerbe E, Alcaide F, Reynaud M, Carrasco J, Grimaud A, Zhang C, Vegge T, Johansson P, and Franco A A 2022 *Chem. Rev.* **122** 10899

[7] Wang Y, Wu S, Shao W, Sun X, Wang Q, Xiao R, and Li H 2022 *J. Materiom.* **8** 1038

[8] Gong S, Xie T, Zhu T, Wang S, Fadel E R, Li Y, and Grossman J C 2019 *Phys. Rev. B* **100** 184103

[9] Salgado J E, Lerman S, Du Z, Xu C, and Abdolrahim N 2023 *npj Comput. Mater.* **9** 214

[10] Lu L, Han X, Li J, Hua J, and Ouyang M 2013 *J. Power Sources* **226** 272

[11] Gilligan L P J, Cobelli M, Taufour V, and Sanvito S 2023 *npj Comput. Mater.* **9** 222

[12] Ling C 2022 *npj Comput. Mater.* **8** 33

[13] Liu Y, Wang S, Nolan A M, Ling C, and Mo Y 2020 *Adv. Energy Mater.* **10** 2002356

[14] Electronic Laboratory for Material Science

[15] Wang D, Zhou M, Li Y L, He L, Sun W X, Lin Z F, Li J H, Zhu Y, Chen Q H, Zhang B H, Yuan J, and Weng H M 2023 *Front. Data Comput.* **5** 29 (in Chinese)

[16] Wang D, Zhou M, Huang D, Li Y, Lin Z, Liu J, Zhu T, Zhu Y, Li M, Xiao R, Yuan J, and Weng H 2023 *Chin. Sci. Bull.* **69** 1164

[17] Wang J, He T, Yang X, Cai Z, Wang Y, Lacivita V, Kim H, Ouyang B, and Ceder G 2023 *Nat. Commun.* **14** 5210

[18] Zhu Z, Chu I H, Deng Z, and Ong S P 2015 *Chem. Mater.* **27** 8318

[19] Materials science literature search powered by advanced NLP algorithms.

[20] Lee E and Persson K A 2014 *Adv. Energy Mater.* **4** 1400498

[21] Liang J W, van der Maas E, Luo J, Li X N, Chen N, Adair K R, Li W H, Li J J, Hu Y F, Liu J, Zhang L, Zhao S Q, Lu S G, Wang J T, Huang H, Zhao W X, Parnell S, Smith R I, Ganapathy S, Wagemaker M, and Sun X L 2022 *Adv. Energy Mater.* **12** 2103921

[22] Wu Y, Xu J, Lu P, Lu J, Gan L, Wang S, Xiao R, Li H, Chen L, and Wu F 2023 *Adv. Energy Mater.* **13** 2301336

[23] Xu J, Wang Y, Wu S, Yang Q, Fu X, Xiao R, and Li H 2023 *ACS Appl. Mater. Interfaces* **15** 21086

[24] Liu Y, Rong X, Bai R, Xiao R, Xu C, Zhang C, Xu J, Yin W, Zhang Q, Liang X, Lu Y, Zhao J, Chen L, and Hu Y S 2023 *Nat. Energy* **8** 1088

[25] Saal J E, Kirklin S, Aykol M, Meredig B, and Wolverton C 2013 *JOM* **65** 1501

[26] Ask H L, Jens J M, Jakob B *et al.* 2017 *J. Phys.: Condens. Matter* **29** 273002

[27] Liu Y *et al.* 2020 *Energy Storage Mater.* **31** 434

[28] Liu Y *et al.* 2017 *J. Materiom.* **3** 159

[29] Butler K T *et al.* 2018 *Nature* **559** 547

[30] Behler J 2021 *Chem. Rev.* **121** 10037

[31] Behler J 2011 *J. Chem. Phys.* **134** 074106

[32] Willatt M J, Musil F, and Ceriotti M 2019 *J. Chem. Phys.* **150** 154110

[33] Drautz R 2019 *Phys. Rev. B* **99** 014104

[34] Liu Y *et al.* 2023 *Natl. Sci. Rev.* **10** nwad125

[35] Knauth P 2009 *Solid State Ionics* **180** 911

[36] Zhang L, Han J, Wang H, Car R, and E W 2018 *Phys. Rev. Lett.* **120** 143001

[37] Zhang D, Bi H R, Dai F Z, Jiang W R, Zhang L F, and Wang H 2022 arXiv:2208.08236v3 [physics.chem-ph]

[38] Deng B, Zhong P, Jun K, Riebesell J, Han K, Bartel C J, and Ceder G 2023 *Nat. Mach. Intell.* **5** 1031

[39] Hu S H 2023 *Physics* **52** 857 (in Chinese)