THE JOURNAL OF PHYSICAL CHEMISTRY C
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCC

Article

Inquiry into the Appropriate Data Preprocessing of Electrochemical Impedance Spectroscopy for Machine Learning

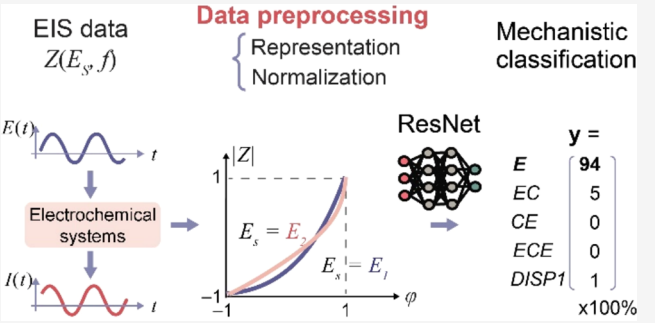
Jingwen Sun, Weitong Zhang, Yuanzhou Chen, Benjamin B. Hoar, Hongyuan Sheng, Jenny Y. Yang, Quanquan Gu, and Chong Liu*

Cite This: J. Phys. Chem. C 2025, 129, 1044−1051

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Electrochemical impedance spectroscopy (EIS) is an important analytical technique for the understanding of electrochemical systems. With the recent advent and burgeoning deployment of machine learning (ML) in EIS analysis, a critical yet hitherto unanswered question emerges: what is the appropriate manner to preprocess the EIS data for ML-based analysis? While the preprocessing of a model's input data is known to be critical for a successful deployment of the ML model, EIS is known to possess multiple classical venues of data representation, and moreover, a proper data normalization protocol for comparative EIS studies remains elusive. Here, we report the methodology and the outcomes that evaluate the efficacy of multiple data preprocessing methods in an ML-based EIS analysis. Within our proof-of-concept parameter space, plotting the input training data's impedance magnitude (|Z|) against phase angle (φ) while individually normalizing each EIS curve yields the highest accuracy and robustness in the correspondingly established residual neural network (ResNet) model. Rationalized by additional "importance" analysis of the input data, such a data representation method extracts information and hidden features more effectively. While the Nyquist plot is widely used in manual analysis, a different data representation of EIS data seems equally plausible for ML-based EIS analysis. Our work offers a protocol for future researchers to decide on the proper preprocessing method for different ML applications in electrochemistry on a case-by-case basis.



INTRODUCTION

Electrochemical impedance spectroscopy (EIS)[1] is a powerful technique for investigating material properties and analyzing complicated reaction processes in electrochemical systems including batteries,[2−4] sensors,[5−7] and catalysts.[8−10] Experimentally, EIS data are typically collected by monitoring the sinusoidal current $I(t)$ upon the application of a sinusoidal voltage $E(t)$ centered around the potential E_s over a series of frequencies f (Figure 1A).[1] While impedance $Z(f)$ is straightforwardly calculated as the ratio between output signal $E(t)$ and input signal $I(t)$ under certain E_s,[1,11,12] additional considerations have been devoted toward how to visualize $Z(E_s, f)$ for human researchers in EIS analysis. Because $Z(E_s, f)$ is generally a complex number with two degrees of freedom, there are different venues to visually represent $Z(f)$ under certain E_s with the Nyquist and Bode plots as the most prominent ones[1,11] (Figure 1B): The Nyquist plot displays $Z(f)$ under the Cartesian coordinate of $Z' = \mathrm{Re}(Z)$ and $-Z'' = -\mathrm{Im}(Z)$ as the horizontal and vertical axis, respectively, with f component embedded in the plot; the Bode plots display $Z(f)$'s amplitude |Z| and phase φ separately as a function as $\log_{10}(f)$, implicating the benefits of presenting $Z(E_s, f)$ in a polar coordinate of |Z| and φ. While Nyquist plot is the most common practice thanks to its straightforwardly distinctive semicircle features for human researchers,[1,13,14] the diversity of data visualization in current EIS data analysis not only illustrates $Z(E_s, f)$'s complexity and information richness but also showcases that each data representation method has its own fundamental merits in a manual EIS analysis.

With the advent of artificial intelligence transforming every corner of our society and research community, the issue of data representation in EIS analysis for a machine learning (ML) model warrants another reckoning if we consider the ML-based analysis augmenting the human researchers in the future. ML presents a promising solution for advancing mechanistic analysis in electrochemical systems,[15−19] leveraging ML's proficiency in uncovering hidden patterns and providing data-driven insights with minimal human interference. Recent advancements showcase ML's capability in automatically

Received: September 13, 2024
Revised: October 22, 2024
Accepted: October 24, 2024
Published: January 5, 2025

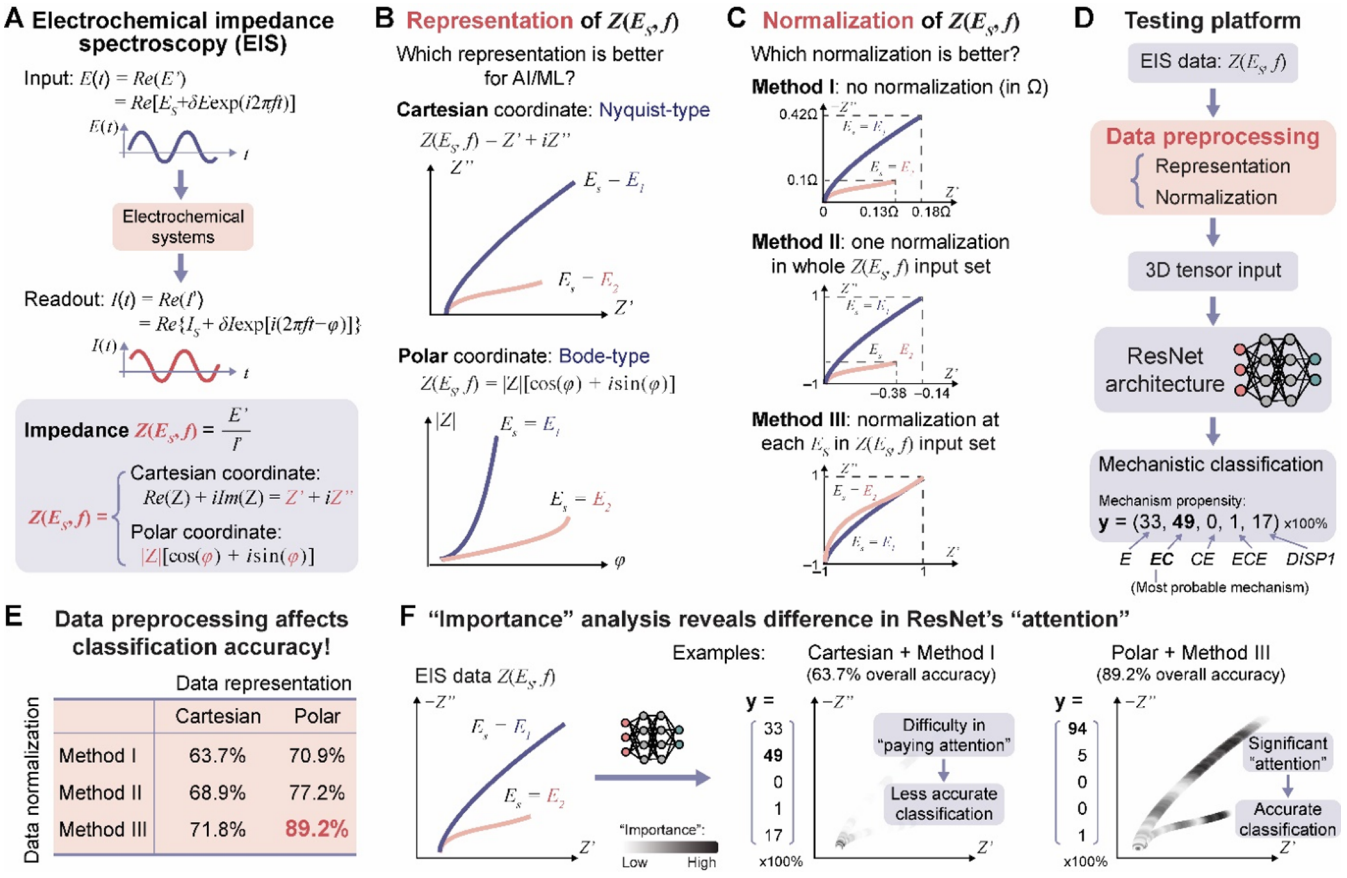


Figure 1. Data preprocessing of electrochemical impedance spectroscopy (EIS) in the world of machine learning (ML) and artificial intelligence (AI). (A), Short summary of the experimental and mathematical foundation of the EIS technique. $E(t)$ and $I(t)$, sinusoidal electrochemical potential and the measured electric current densities on electrodes, respectively; E_s and I_s, the steady-state electrochemical potential electric current densities, respectively; f, oscillating frequency of $E(t)$; Z, Z', and Z'', electrochemical impedance $Z(E_s, f)$ as well as its real and complex component, respectively; $|Z|$ and φ, magnitude and phase angle of the $Z(E_s, f)$, respectively. (B), Two types of representation in a comparative study of $Z(E_s, f)$ under different E_s values: the Nyquist-like Cartesian coordinate of Z' and $-Z''$ as the two axes; the Bode-like polar coordinate of $|Z|$ and φ as the two axes. (C), Three types of normalization in a comparative study of $Z(E_s, f)$ under different E_s values: raw data with no normalization at all (Method I), one normalization for each $Z(E_s, f)$ data point (Method II), and independent normalization at each E_s value (Method III). (D), We evaluate how the strategies of $Z(E_s, f)$ preprocessing, including both data representation and normalization, affect the classification accuracy by residual neural network (ResNet), within the same 5-mechanism parameter space that was defined in our prior work of ResNet-based voltammogram analysis.[17] The ResNet model reported here analyzes preprocessed $Z(E_s, f)$ and yields the vector $\mathbf{y} = \{y_i\}$ (i = 1 to 5), reflecting mechanism's propensity, among the 5 predefined mechanisms. (E), Summary of the average classification accuracies from ResNet based on different $Z(E_s, f)$ preprocessing strategies. N = 8. Accuracy was reported when evaluated with test set ("Test accuracy" in Table S1). (F), The "importance" analysis,[25] analogous to the one conducted to ResNet-based voltammogram analysis,[17] reveals differences in ResNet's attention under different preprocessing strategies, which correlates with the overall classification accuracies of the corresponding ResNet model. ResNet architecture is adapted from ref 17 (CC BY 4.0).

analyzing EIS data, specifically for the classification and parameter regression of equivalent circuit models.[20−24] Yet there is a dearth of discussion about the proper data representation of EIS data for ML-based analysis, with the seemingly ill-justified consensus of deploying Nyquist plot, namely, the Cartesian coordinate of $Z(E_s, f)$, as the EIS input format for ML models.[20,22,23]

The unique nature of ML architecture, which requires input data normalized to unity for the model's general applicability, poses additional challenges for inputting EIS data into an ML model. It is well documented that different normalization strategies of input data will lead to different levels of performance even for the same ML architecture.[26] In the context of EIS analysis, $Z(E_s, f)$'s sensitivity toward E_s alludes that it is common to study $Z(E_s, f)$'s dependence under different E_s values in a comparative analysis,[1] which is equivalently the decade-old reaction mechanism analysis done by human researchers.[27,28] Hence, it is critical to properly normalize multiple $Z(f)$ traces under various E_s values as the input data for ML-based comparative analysis. Are we supposed to normalize $Z(E_s, f)$? If so, how shall we normalize a $Z(E_s, f)$ data set of various E_s values (Figure 1C)? To date, such a topic has not been investigated, at least not in the public domain. Despite EIS's significance in electrochemistry[1,6,11] and the growing interest in ML-based EIS analysis,[20,22,23] the topic of $Z(E_s, f)$ preprocessing that includes proper representation and normalization remains surprisingly underexplored. It seems imprudent to ignore such a critical issue in the context of developing a next-generation ML-based EIS analysis.

We are intrigued by the scientific question of preprocessing EIS data in an ML-based analysis and are blessed with suitable expertise to inquire about this topic in a proof-of-concept fashion. We have recently developed ML models based on

https://doi.org/10.1021/acs.jpcc.4c06206
J. Phys. Chem. C 2025, 129, 1044−1051

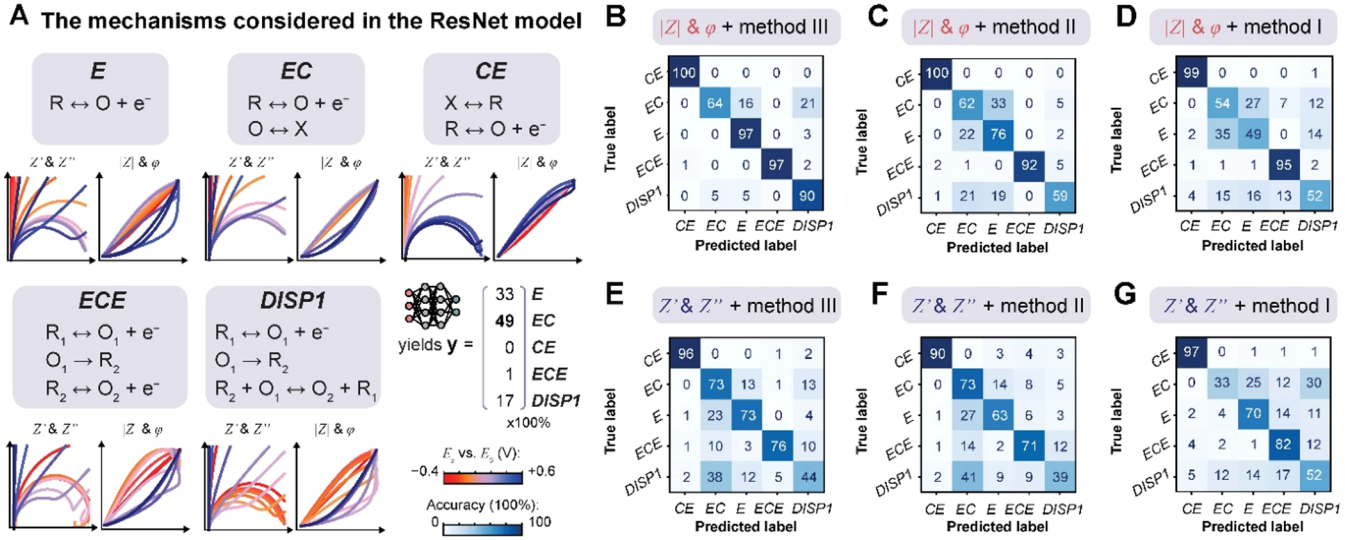


Figure 2. (A) Mechanism included in our ResNet models for EIS analysis. More detailed mathematical definitions of the mechanisms are available in Supporting Information Note 3. (B–G) Confusion matrix, a commonly deployed performance evaluation tool that represents the accuracy of a classification model, for the ML models trained by EIS data $\{Z(E_{s,i}, f_j)\}$ under different preprocessing approaches. (B–D) Training data with polar coordinate of |Z| and φ while normalizing $\{Z(E_{s,i}, f_j)\}$ individually at each $E_{s,i}$ (Method III, (B)), one normalization for each $\{Z(E_{s,i}, f_j)\}$ data point (Method II, (C)), and no normalization at all (Method I, (D)). (E–G) Training data with Cartesian coordinate of Z' and Z'' while normalization $\{Z(E_{s,i}, f_j)\}$ following Method III (E), Method II (F), and Method I (G). The medium ResNet model out of the 8 replicates for each preprocessing method is presented. ResNet architecture is adapted from ref 17 (CC BY 4.0).

residual neural networks (ResNet)[29] and Faster R-CNN (regional convolutional neural network) architecture[30] for the mechanistic investigation of cyclic voltammograms.[17,18] In particular, an ML model based on ResNet[17] has been trained by simulated voltammograms to classify electrochemical mechanisms among five of the most classic mechanisms in electrochemistry textbooks,[31] with an input data structure of three-dimensional tensor of 6-scan single-redox voltammograms ($\{v, i(E)\}_n$, n = 1 to 6). One noticeable feature in our study is the establishment of "importance" evaluation,[25] which illustrates where ResNet model prioritizes information extraction and decision making, offering a glimpse into ML's innerworkings for a comparative study between human and ML-based voltammogram analysis.[17] Those technical tools are duly applicable toward the inquiry of proper data preprocessing in ML-based EIS analysis.

Here, we present the protocol and results of our proof-of-concept inquiry into the proper preprocessing of EIS data in an ML-based analysis—a factor not addressed in our previous CV work where data representation concerns are minimal due to its simple illustration as current vs potential. In contrast, EIS data require complex representations and optimal normalization techniques due to significant variations in impedance amplitude across different potentials. Within the same 5-mechanism parameter space that were defined in our prior work of ResNet-based voltammogram analysis,[17] we strive to investigate how the classification performance of ResNet model trained by simulated $Z(E_s, f)$ data will depend on the preprocessing methods of $Z(E_s, f)$ in the training data (Figure 1D). Here, the preprocessing methods include both data representation: Nyquist-type Cartesian coordinate of Z' and Z'' versus. Bode-type polar coordinate of |Z| and φ as well as data normalization without any normalization (Method I), one normalization for each $Z(E_s, f)$ data point (Method II), and independent normalization at each E_s value (Method III) for 14 $Z(f)$ traces under different E_s values (Figure 1). Among different combinations of $Z(E_s, f)$'s representation and normalization methods, our results affirm the hypothesis that preprocessing method strongly affects the classification accuracy of the correspondingly trained ResNet model for ML-based EIS analysis (Figure 1E). We found that representing $Z(E_s, f)$ in the Bode-type polar coordinate of |Z| and φ with normalization at each E_s value (Method III) yields the highest accuracy and robustness in the correspondingly established ResNet model. An "importance" evaluation protocol, similar to the one deployed for ML-based voltammogram analysis,[17] reveals correlational hints if not causal rationales that help explain such differences in classification performance (Figure 1F). Our work serves as a reminder of the significance of proper data preprocessing for ML-based electrochemical analysis. The best $Z(E_s, f)$ preprocessing method reported here will at least serve as a starting point for future ML-based EIS research. The reported protocol and evaluation methodology toward optimal $Z(E_s, f)$ preprocessing is generally applicable for a wide range of scenarios in ML-based analysis in electrochemistry and beyond.

RESULTS AND DISCUSSION

Establishing a Simulated $Z(E_s, f)$ Data Set at Different E_s Values. We evaluate the strategy of $Z(E_s, f)$ preprocessing based on the classification accuracy of a ResNet model trained by the same $Z(E_s, f)$ data set following the corresponding preprocessing method. This requires the establishment of a $Z(E_s, f)$ data set upon which the ResNet is trained. As a proof of concept, we decided to establish a $Z(E_s, f)$ data set numerically simulated by finite-element method, following the practice and mechanistic definitions reported in our previous work of ResNet-based voltammogram analysis.[17] While our limited $Z(E_s, f)$ data set does not cover all of the possible application scenarios of EIS analysis, our findings reported here remain potentially applicable to a general ML model of EIS analysis. Moreover, we believe our work offers a general

https://doi.org/10.1021/acs.jpcc.4c06206
J. Phys. Chem. C 2025, 129, 1044–1051

protocol that evaluates methods of data preprocessing for electrochemical research, despite our limited $Z(E_s, f)$ data set from numerical simulation.

Our simulated $Z(E_s, f)$ data set includes five common homogeneous electrochemical mechanisms in textbooks[11,31]-(Figure 2A and original data in Figure S2): a reversible/quasi-reversible single-electron transfer with a redox potential E_0 and the oxidation as the forward reaction direction (*E* mechanism/step); an aforementioned *E* step followed by a homogeneous reversible chemical reaction (*C* step) (*EC* mechanism); an *E* step preceded by a reversible *C* step (CE mechanism); a system of two *E* steps connected by an irreversible rate-limiting irreversible *C* step with the second *E* step being more thermodynamically facile than the first one (*ECE* mechanism); and a net two-electron transfer that is similar to *ECE* yet the second *E* step is replaced by a solution disproportionation reaction (*DISP1* mechanism). The construction of numerical models of partial differential equations, boundary conditions, and initial conditions follows the mechanisms' definitions in textbooks (see the Supporting Information).[11,31] The parameters defining the reaction systems include but are not limited to concentrations, interfacial charge transfer rate constant in the *E* steps, and thermodynamic equilibrium constants of the *C* steps. The constraints and sampling approaches of these parameters follow our group's previous work[17] to ensure physical realism and class diversity. Such practice ensures a near-exhaustive sampling of each noted mechanism defined in the textbook kinetic zone diagrams.[11,31]

For each electrochemical system whose reaction mechanisms and parameters are numerically set following the aforementioned defections, the Z values were numerically simulated at 14 different E_s values between −0.4 and +0.6 V against the *E* step's redox potential E_0 with 50 or 100 mV intervals. At each E_s value, 500 f values, from 1 Hz to 100 kHz with a mostly even logarithmical distribution, were simulated. Each of the aforementioned 5 mechanisms include approximately 5000 entries of $Z(E_s, f) = \{Z(E_{s,i}, f_j)\}$ (i = 1 to 14; j = 1 to 500), whose reaction parameters were randomly sampled for a sufficient, balanced, and diverse training data set. The two-dimensional formalism of $Z(E_s, f)$ allows for a comparative EIS analysis of reaction mechanism at different E_s values, as researchers have done for decades,[27,28] which provides rich information toward mechanism classifications. Yet as exemplified in Figure 2A, the E_s-dependent $Z(f)$ responses among the 5 noted mechanisms, plotted in both Cartesian and polar coordinates, are difficult to visually differentiate, which calls for an ML-based EIS analysis that effectively performs a numerical fitting against all of the possible variations for each mechanism as defined in the textbooks.[15]

Protocols of Data Preprocessing and ML Model Training. Before being fed into the ResNet model for training and testing, the simulated data set of $\{Z(E_{s,i}, f_j)\}$ (i = 1 to 14; j = 1 to 500) was preprocessed under a variety of protocols that includes both representation and normalization. As noted earlier (Figure 1B), the representation of Nyquist-like Cartesian coordinate presents each $Z(E_s, f)$ data as a two-dimensional matrix as $\{Z'_j(E_{s,i}), Z''_j(E_{s,i})\}$ (i = 1 to 14; j = 1 to 500), while the representation of Bode-like polar coordinate presents each $Z(E_s, f)$ data as $\{\varphi_j(E_{s,i}), |Z|_j(E_{s,i})\}$ (i = 1 to 14; j = 1 to 500). Moreover, three different normalization protocols are implemented after the selection of the representation method (Figure 1C). There could be no normalization at all in $\{Z(E_{s,i}, f_j)\}$ and only the calculated values in the unit of ohms directly from numerical simulation (Method I in Figure 1C). Alternatively, a so-called "layer normalization" protocol[32] conducts one normalization of $\{Z(E_{s,i}, f_j)\}$ across all of the $E_{s,i}$ values within the same single data point. Different channels, i.e., $Z(f)$ traces under different E_s values are scaled together using a common normalization factor calculated by the maximum and minimum values among all of the channels (Method II in Figure 1C). Last, the protocol termed as "instance normalization"[33] will individually normalize $\{Z(E_{s,i}, f_j)\}$ data for each $E_{s,i}$ value within the data point (Method III in Figures 1C and S1A, see the Supporting Information). Here, Method II ("layer normalization") keeps the same relative magnitude of $Z(E_{s,i}, f_j)$ between different $E_{s,i}$ values as the original data (Method I, no normalization), while Method III ("instance normalization") adjusts for variations of magnitude across different $E_{s,i}$ values. All values are scaled into a specific range between −1 and 1 for both normalization Methods II and III. Afterward, the two-dimensional input matrix of $\{Z(E_{s,i}, f_j)\}$ (i = 1 to 14; j = 1 to 500) was transformed into a three-dimensional tensor $\{n, y, m\}$ (Figure S1B). In this tensor, n = 14 as it corresponds to the number of E_s values; m = 500 as it correlates to the number of f values; y = 3 with the channels storing the E_s values, as well as $\{Z'_j(E_{s,i})\ Z''_j(E_{s,i})\}$ or $\{\varphi_j(E_{s,i}), |Z|_j(E_{s,i})\}$ for the Cartesian or polar coordinates, respectively.

A series of ML models based on ResNet architecture,[29] more specifically ResNet with 18 residual learning layers (hence ResNet-18), are trained and validated by the same initial $\{Z(E_{s,i}, f_j)\}$ data set yet preprocessed under different combinations of data representation and normalization, largely following the previously reported protocol for ResNet-based analysis of cyclic voltammograms.[17] We chose ResNet for EIS analysis due to its ability to handle complex, high-dimensional data. Its deep architecture with residual skip connections prevents vanishing gradients, ensuring more efficient and robust training than basic neural networks or convolutional neural networks.[29] In contrast, simpler linear models may struggle with high dimensionality and nonlinear relationships in EIS data. To mitigate the effects of randomness stemming from data splitting and parameter initialization in the training process, 8 replicates of ResNet models (n = 8) were trained with about 3500 $Z(E_s, f)$ data points, randomly sampled for every training process, for each mechanism through 1000-epoch training. The size of 3500 $Z(E_s, f)$ data points for the training set is chosen because the ResNet model's accuracy, under the representation of $|Z|$ and φ, steadily increases with increasing $Z(E_s, f)$ numbers in the training set before plateauing beyond 3500, which indicates the model has achieved a stable performance level, suggesting that further increases in training set size may not yield significant improvements in accuracy (Figure S2A). The ResNet model is considered sufficiently trained after 1000 epochs because the cross-entropy loss that surrogates the ResNet's accuracy in the training process has asymptotically approaching zero after 1000 epochs of training (Figure S2B). The trained ResNet model yields an output vector $\mathbf{y} = \{y_1, y_2, y_3, y_4, y_5\}$ (Figure 1D), in which each y_i quantitatively represents the propensity as a surrogate for the statistical probability associated with the respective mechanisms of *E*, *EC*, *CE*, *ECE*, and *DISP1* shown in Figure 2A. The classification process designates the mechanism with the largest y component as the most probable one for the electrochemical system.

ML Model Performance is Affected by Data Preprocessing Methods. The methods of data preprocess-

https://doi.org/10.1021/acs.jpcc.4c06206
J. Phys. Chem. C 2025, 129, 1044−1051

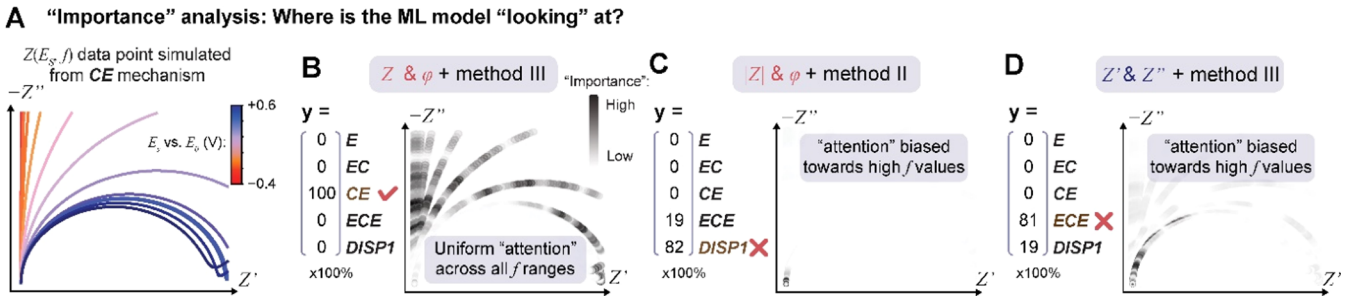


Figure 3. "Importance" analysis reveals ResNet's "attention" when yielding mechanistic assignments. (A), One example of our EIS data simulated based on the *CE* mechanism plotted with the Cartesian coordinate of Z' and Z'' (the Nyquist plot). (B), Classification result, and the "importance" analysis of ResNet model trained by EIS data with polar coordinate ($|Z|$ and φ) and normalizing individually at each $E_{s,i}$ (Method III). (C) Classification result and the "importance" analysis of ResNet model with polar coordinate and one normalization for each $Z(E_s, f)$ data point (Method II). (D), Classification result, and the "importance" analysis of ResNet model with Cartesian coordinate and normalization via Method III. All EIS data are represented in the Nyquist plot for illustration purposes despite differences in the deployed representation methods. Additional examples can be found in the Supporting Information.

ing significantly affect the resultant ResNet model's accuracy of mechanistic classification (Figures 1F and S2C and Table S1). When tested against EIS data unseen before by the trained models, the classification accuracy of ResNet model under a certain preprocessing method ("Test accuracy" in Table S1), averaged from 8 replicates to ensure sufficient statistical validity (n = 8), ranges from 63.7 ± 10.8% (Cartesian coordinate + Method I) to 89.2 ± 1.7% (polar coordinate + Method III). Generally speaking, the accuracies of the ML models trained with $\{Z(E_{s,i}, f_j)\}$ data set under Cartesian representation (63.7 to 71.8%) are lower than the ones under polar representation (70.9 to 89.2%). For models with the same EIS representation, the accuracies of trained models gradually increase from Method I (63.7 and 70.9% for Cartesian and polar coordinates, respectively, same below), Method II (68.9 and 77.2%), to Method III (71.8 and 89.2%). Within our defined space of electrochemical mechanisms and the corresponding training set, the optimal data preprocessing method involves representing $\{Z(E_{s,i}, f_j)\}$ in the polar coordinate of $|Z|$ and φ and individually normalizing $\{Z(E_{s,i}, f_j)\}$ data for each $E_{s,i}$ value (Method III). These results affirm our aforementioned hypothesis that data preprocessing is critical toward ML-based EIS analysis. Moreover, our results hint that the seemingly prevailing practice of representing EIS data under the Nyquist plot hence the Cartesian coordinate[20,22,23] may not be optimal, in contrast to strong preference of Nyquist plot in human-based EIS analysis.

Additional analysis of the accuracy difference among ResNet models from different preprocessings of training data suggests that the models of higher classification accuracy are better at discerning the mechanisms that are also prone to be misclassified based on human analysis. In machine learning, it is common to deploy the plot of the confusion matrix as a performance evaluation tool that represents the accuracy of a classification model. For each of the preprocessing methods, we plot the correspondingly trained ResNet model's confusion matrix from the medium model from the 8 replicates (Figure 2B−G), in which each row enlists the percentage of EIS data simulated based on a designed mechanism ("true label" in Figure 2B−G) that are classified into a specific mechanism ("predicted label" in Figure 2B−G). The *EC* mechanism seems to be the most difficult mechanism to classify, evident by the marginal accuracy improvement for the *EC* mechanism across all ResNet variants. Yet the difference in preprocessing strategy has a significant influence on other mechanisms, most notably for *E* and *DISP1* mechanism and to a lesser extent for *ECE* mechanism. Our observation that ResNet models are most easily confused among *E*, *EC*, and *DISP1*, also observed in our ML-based voltammogram analysis[17,18] is consistent with the results of human analysis that has been detailed in textbooks.[11,31] Because these ResNet models are trained by the same $\{Z(E_{s,i}, f_j)\}$ data set and the only variable here is the preprocessing strategy, such results suggest that different preprocessing strategy affects ML model's classification capability unevenly across different mechanisms. In a similar vein, our results hint that some additional mechanistic information could be easier to uncover with a suitable preprocessing strategy.

The significant dependence of ML model's classification accuracy on training data's preprocessing strategy inspired us to strive investigating the origin of such a dependence. As a valuable tool in machine learning, the "importance" analysis identifies the regions or features in the input data that have the most significant influence on the model's decision-making process by visualizing the relative magnitudes of gradient values of the model's output with respect to the input data.[25] Hence, in our case, we decided to use the "importance" analysis to evaluate where the ResNet model is "looking" at with the EIS data to provide a mechanistic classification in a multipotential comparative analysis. As shown in Figure 3, the gradient values were collectively normalized based on the highest value achieved among all of the ResNet models and plotted in the grayscale. Additionally, for easier visualization of "importance" on each simulated spectrum with varying magnitudes, all simulated spectra were individually normalized and plotted on the Cartesian coordinate of Z' and Z'' (Nyquist plot), irrespective of the preprocessing methods used for the input data. In principle, we could have flooded this report with the results of "importance" analysis from our training data set of ~10^4 $\{Z(E_{s,i}$, and $f_j)\}$ data points. Yet we have found a prevailing trend based on our manual inspections of "importance" analysis, with one example presented below and a few additional ones in the Supporting Information. As detailed below, such an "importance" analysis provides some insights into the difference created by preprocessing approaches for ML-based EIS analysis.

We found out that the preprocessing approach of the EIS input data notably impacts the resultant ResNet model's capability of extracting mechanistically relevant information. Figure 3 displays the example for a $\{Z(E_{s,i}, f_j)\}$ data point

https://doi.org/10.1021/acs.jpcc.4c06206
J. Phys. Chem. C 2025, 129, 1044−1051

simulated based on the *CE* mechanism. While the ML model trained by EIS data represented by polar coordinate and normalized via method III correctly classifies the mechanism as *CE* as 100% propensity (Figure 3B), the models from other preprocessing methods generate inaccurate classification: for example, 82% propensity of *DISP1* mechanism from the model with representation of polar coordinate and normalization via method II (Figure 3C), and 81% propensity of *ECE* mechanism from the model with representation of Cartesian coordinate and normalization via method III (Figure 3D). Concurrent with such a discrepancy in classification/accuracies, the "importance" analysis for those three models reveals drastic differences. The model trained from representation of polar coordinate and normalized via method III (Figure 3B) displays a distribution of "importance" across both low- and high-frequency regions, capturing information related to both kinetics and mass transfer, with a particular emphasis on the mass transfer-controlled low-frequency region. Conversely, models with incorrect classification (Figure 3C,D) focus their "importance" predominantly on the high-frequency region, with little if any "importance" on the low-frequency part. Similar conclusion could be drawn from other examples (Supporting Information) that models of correct mechanistic assignment tend to spread the "attention" across the whole frequency domains that cover both (electro)chemical kinetics and mass transport, while models of erroneous classification usually fail to spread the model's "attention" across a wide range of frequencies, neglecting the critical low-frequency information. The preprocessing of EIS data seems to strongly affect the ResNet model's ability to uncover features and extract mechanistic information across a wide range of frequencies, ensuring a balanced evaluation that encompasses both (electro)chemical kinetics and mass transport. This underscores the necessity of careful data preprocessing to achieve enhanced mechanistic interpretation in EIS analysis.

The construction of the EIS-analyzing model based on supporting vector classifier (SVC) models provides supplementary evidence affirming the broader validity of our conclusions on preprocessing methods (Figure S6). With the same Method III normalization, the SVC model trained with the $\{Z(E_{s,i}, f_j)\}$ data set under polar representation achieves 82.4% test accuracy, compared to 66.8% under Cartesian representation. As the corresponding confusion matrix shows (Figure S6), the SVC model trained under polar representation shows marginal accuracy improvements for 5 mechanisms, notably for *E* (68 to 90%), *ECE* (68 to 82%), and *DISP1* (48 to 70%), with slight enhancements for *EC* (67 to 73%) and *CE* mechanisms (93 to 97%). Although SVC models' performances were less satisfactory than ResNet models, the consistent accuracy improvement with polar representation supports our hypothesis on the critical role of data preprocessing in ML-based EIS analysis, validating the general applicability of our conclusions on preprocessing methods.

One note is that, while equivalent circuit model (ECM) is a classical and widely used technique in EIS analysis, it may oversimplify electrochemical processes by representing them with discrete electrical components,[1] potentially failing to capture the complexity of underlying physico-electrochemical phenomena, especially in cases with overlapping processes or intricate reaction mechanisms. Therefore, a direct comparison between our ML-based method and the ECM might not be entirely appropriate.

CONCLUSIONS

In this work, we inquired into the proper preprocessing of EIS data in an ML-based analysis, using the ResNet model and a classification task of five mechanisms as a proof of concept. Through a comprehensive evaluation of various preprocessing methods, we demonstrate that plotting the input training data's impedance magnitude ($|Z|$) against phase angle (φ), along with individual normalization of each EIS curve (Method III), yields optimal accuracy in the established ML model. Suggested by "importance" analysis, this preprocessing method effectively extracts hidden features and hence information from the EIS data. While the Nyquist plot remains a staple in manual analysis, our findings suggest that an ML-based EIS analysis may require an alternative data representation method. Our work provides a valuable protocol for researchers to select appropriate preprocessing methods tailored to specific ML applications in electrochemistry.

While mechanism classification using EIS is less common compared to techniques like CV, our study demonstrates that EIS can be extended for detailed mechanistic analysis across a range of potentials, aligning with its historical use in reaction mechanism analysis for complex systems.[27,28] Nearly two decades ago, it was suggested that reaction mechanism analysis could be efficiently performed using artificial neural networks in a pattern recognition mode for future EIS development.[34] Our model, leveraging ResNet and an enhanced data preprocessing technique, shows promising potential for EIS mechanistic interpretation. Although our current study focuses on proof-of-concept classification of classical homogeneous mechanisms, we foresee incorporating transfer learning and regression algorithms in future work to get more (electro)-chemical insights and extract quantitative parameters such as the ionic conductivity of electrolytes. Our current work presented here lays a foundation for further advancements of our approach, enhancing its practical utility and making it more versatile and useful for a wider range of applications.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.4c06206.

Protocols of the numerical simulations of electrochemical impedance spectroscopy spectra; detailed definitions of individual electrochemical mechanisms; protocols of establishing the ML models, and ML training and evaluation (PDF)

Additional examples of "importance" analysis (ZIP)

AUTHOR INFORMATION

Corresponding Author

Chong Liu − *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States; California NanoSystems Institute, University of California, Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0001-5546-3852; Email: chongliu@chem.ucla.edu

Authors

Jingwen Sun − *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0003-3657-7672

https://doi.org/10.1021/acs.jpcc.4c06206
J. Phys. Chem. C 2025, 129, 1044−1051

Weitong Zhang − *Department of Computer Science, University of California, Los Angeles, Los Angeles, California 90095, United States;* Present Address: School of Data Science and Society, University of North Carolina, Chapel Hill, Chapel Hill, North Carolina 27599, United States
Yuanzhou Chen − *Department of Computer Science, University of California, Los Angeles, Los Angeles, California 90095, United States*
Benjamin B. Hoar − *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States*
Hongyuan Sheng − *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States*
Jenny Y. Yang − *Department of Chemistry, University California, Irvine, Irvine, California 92697, United States;* orcid.org/0000-0002-9680-8260
Quanquan Gu − *Department of Computer Science, University of California, Los Angeles, Los Angeles, California 90095, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.4c06206

Author Contributions

C.L. and Q.G. supervised the project. J.S., B.B.H., C.C., and C.L. developed theoretical computational frameworks for the generation of simulated electrochemical impedance spectroscopy spectra. J.S. generated and sanitized the simulated data for model training/test. J.S. developed and implemented various data preprocessing strategies. W.Z. provided the scripts for establishing ResNet models for mechanistic analysis based on simulated cyclic voltammograms. J.S. adapted the algorithms with valuable help from W.Z. and built the ResNet models for this work. J.S. prepared the initial draft of the manuscript. All authors discussed the results of the project and assisted with manuscript preparation.

Notes

The authors declare no competing financial interest.
A provisional patent has been filed for J.W., Q.G., and C.L.

■ ACKNOWLEDGMENTS

J.Y.Y., Q.G., and C.L. acknowledge the support of National Science Foundation (NSF) (CHE-2247426). C.L. acknowledges the financial support from the Master of Applied Chemical Sciences program at UCLA Chemistry & Biochemistry.

■ ABBREVIATIONS

EIS, electrochemical impedance spectroscopy; ML, machine learning; ResNet, residual neural network; R-CNN, regional convolutional neural network; SVC, supporting vector classifier; ECM, equivalent circuit model

■ REFERENCES

(1) Orazem, M. E.; Tribollet, B. *Electrochemical Impedance Spectroscopy*; John Wiley & Sons, 2008.
(2) Westerhoff, U.; Kurbach, K.; Lienesch, F.; Kurrat, M. Analysis of lithium-ion battery models based on electrochemical impedance spectroscopy. *Energy Technol.* **2016**, *4* (12), 1620−1630.
(3) Gaberšček, M. Understanding Li-based battery materials via electrochemical impedance spectroscopy. *Nat. Commun.* **2021**, *12* (1), No. 6513.
(4) Moškon, J.; Gaberšček, M. Transmission line models for evaluation of impedance response of insertion battery electrodes and cells. *J. Power Sources Adv.* **2021**, *7*, No. 100047.
(5) Ruan, C.; Yang, L.; Li, Y. Immunobiosensor Chips for Detection of *Escherichia coli* O157:H7 Using Electrochemical Impedance Spectroscopy. *Anal. Chem.* **2002**, *74* (18), 4814−4820.
(6) Pejcic, B.; De Marco, R. Impedance spectroscopy: Over 35 years of electrochemical sensor optimization. *Electrochim. Acta* **2006**, *51* (28), 6217−6229.
(7) Magar, H. S.; Hassan, R. Y. A.; Mulchandani, A. Electrochemical Impedance Spectroscopy (EIS): Principles, Construction, and Biosensing Applications. *Sensors* **2021**, *21* (19), No. 6578.
(8) Lien, H.-T.; Chang, S.-T.; Chen, P.-T.; Wong, D. P.; Chang, Y.-C.; Lu, Y.-R.; Dong, C.-L.; Wang, C.-H.; Chen, K.-H.; Chen, L.-C. Probing the active site in single-atom oxygen reduction catalysts via operando X-ray and electrochemical spectroscopy. *Nat. Commun.* **2020**, *11* (1), No. 4233.
(9) Chen, L.; Xu, Q.; Oener, S. Z.; Fabrizio, K.; Boettcher, S. W. Design principles for water dissociation catalysts in high-performance bipolar membranes. *Nat. Commun.* **2022**, *13* (1), No. 3846.
(10) Shen, Z.; Jin, X.; Tian, J.; Li, M.; Yuan, Y.; Zhang, S.; Fang, S.; Fan, X.; Xu, W.; Lu, H.; et al. Cation-doped ZnS catalysts for polysulfide conversion in lithium−sulfur batteries. *Nat. Catal.* **2022**, *5* (6), 555−563.
(11) Bard, A. J.; Faulkner, L. R.; White, H. S. *Electrochemical Methods: Fundamentals and Applications*, 3rd ed.; John Wiley & Sons, 2022.
(12) Barsoukov, E.; Macdonald, J. R. *Impedance Spectroscopy: Theory, Experiment, and Applications*; John Wiley & Sons, 2018.
(13) Nara, H.; Yokoshima, T.; Osaka, T. Technology of electrochemical impedance spectroscopy for an energy-sustainable society. *Curr. Opin. Electrochem.* **2020**, *20*, 66−77.
(14) Lazanas, A. C.; Prodromidis, M. I. Electrochemical Impedance Spectroscopy—A Tutorial. *ACS Meas. Sci. Au* **2023**, *3* (3), 162−193.
(15) Sun, J.; Liu, C. What and how can machine learning help to decipher mechanisms in molecular electrochemistry? *Curr. Opin. Electrochem.* **2023**, *39*, No. 101306.
(16) Chen, H.; Kätelhön, E.; Compton, R. G. Machine learning in fundamental electrochemistry: Recent advances and future opportunities. *Curr. Opin. Electrochem.* **2023**, *38*, No. 101214.
(17) Hoar, B. B.; Zhang, W.; Xu, S.; Deeba, R.; Costentin, C.; Gu, Q.; Liu, C. Electrochemical Mechanistic Analysis from Cyclic Voltammograms Based on Deep Learning. *ACS Meas. Sci. Au* **2022**, *2* (6), 595−604.
(18) Hoar, B.; Zhang, W.; Chen, Y.; Sun, J.; Sheng, H.; Zhang, Y.; Yang, J.; Costentin, C.; Gu, Q.; Liu, C. Object-detecting deep learning for mechanism discernment in multi-redox cyclic voltammograms *ChemRxiv* 2023 DOI: 10.26434/chemrxiv-2023-r2v1k. (accessed October 15, 2024).
(19) Kennedy, G. F.; Zhang, J.; Bond, A. M. Automatically Identifying Electrode Reaction Mechanisms Using Deep Neural Networks. *Anal. Chem.* **2019**, *91* (19), 12220−12227.
(20) Zhao, Z.; Zou, Y.; Liu, P.; Lai, Z.; Wen, L.; Jin, Y. EIS equivalent circuit model prediction using interpretable machine learning and parameter identification using global optimization algorithms. *Electrochim. Acta* **2022**, *418*, No. 140350.
(21) Bongiorno, V.; Gibbon, S.; Michailidou, E.; Curioni, M. Exploring the use of machine learning for interpreting electrochemical impedance spectroscopy data: evaluation of the training dataset size. *Corros. Sci.* **2022**, *198*, No. 110119.
(22) Zhu, S.; Sun, X.; Gao, X.; Wang, J.; Zhao, N.; Sha, J. Equivalent circuit model recognition of electrochemical impedance spectroscopy via machine learning. *J. Electrochem. Anal.* **2019**, *855*, No. 113627.
(23) Doonyapisut, D.; Kannan, P.-K.; Kim, B.; Kim, J. K.; Lee, E.; Chung, C.-H. Analysis of Electrochemical Impedance Data: Use of Deep Neural Networks. *Adv. Intell. Syst.* **2023**, *5* (8), No. 2300085.
(24) Doonyapisut, D.; Kim, B.; Kim, J. K.; Lee, E.; Chung, C.-H. Deep generative learning for exploration in large electrochemical impedance dataset. *Eng. Appl. Artif. Intell.* **2023**, *126*, No. 107027.

(25) Simonyan, K.; Vedaldi, A.; Zisserman, A. In *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*, International Confernece on Learning Representations; ICLR, 2014.
(26) Huang, L.; Qin, J.; Zhou, Y.; Zhu, F.; Liu, L.; Shao, L.; Huang, L. Normalization Techniques in Training DNNs: Methodology, Analysis and Application. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45* (8), 10173−10196.
(27) Macdonald, D. D.; Real, S.; Smedley, S. I.; Urquidi-Macdonald, M. Evaluation of Alloy Anodes for Aluminum-Air Batteries: IV. Electrochemical Impedance Analysis of Pure Aluminum in at 25 °C. *J. Electrochem. Soc.* **1988**, *135* (10), 2410−2414.
(28) Gabrielli, C. *Indentification of Electrochemical Processes by Frequency Response Analysis*; Schlumberger Instruments, 1984.
(29) He, K.; Zhang, X.; Ren, S.; Sun, J. In *Deep Residual Learning for Image Recognition*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR); IEEE, 2016; pp 770−778.
(30) Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv, 2016 DOI: 10.48550/arXiv.1506.01497. (accessed October 15, 2024).
(31) Savéant, J.-M.; Costentin, C. *Elements of Molecular and Biomolecular Electrochemistry: An Electrochemical Approach to Electron Transfer Chemistry*, 2nd ed.; John Wiley & Sons, Inc., 2019.
(32) Ba, J. L.; Kiros, J. R.; Hinton, G. Layer normalization. arXiv, 2016 DOI: 10.48550/arXiv.1607.06450. (accessed October 15, 2024).
(33) Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. arXiv, 2017 DOI: 10.48550/arXiv.41607.08022. (accessed October 15, 2024).
(34) Macdonald, D. D. Reflections on the history of electrochemical impedance spectroscopy. *Electrochim. Acta* **2006**, *51* (8−9), 1376−1388.

https://doi.org/10.1021/acs.jpcc.4c06206
J. Phys. Chem. C 2025, 129, 1044−1051