Journal of Polymer Science

WILEY

JOURNAL OF POLYMER SCIENCE

RESEARCH ARTICLE OPEN ACCESS

AI-Driven Discovery of High Performance Polymer Electrodes for Next-Generation Batteries

Subhash V. S. Ganti[1,2] | Lukas Wölfel[1] | Christopher Kuenneth[1,2]

[1]Faculty of Engineering Science, University of Bayreuth, Bayreuth, Germany | [2]Bavarian Center for Battery Technology (BayBatt), University of Bayreuth, Bayreuth, Germany

Correspondence: Christopher Kuenneth (christopher.kuenneth@uni-bayreuth.de)

Received: 18 February 2025 | **Revised:** 22 May 2025 | **Accepted:** 29 May 2025

Funding: This work was supported by the Bayerisches Zentrum für Batterietechnik (BayBatt).

Keywords: batteries | data-fusion | energy-storage | meta-learning | multitask machine learning | organic materials

ABSTRACT

The use of transition group metals in electric batteries requires extensive usage of critical elements like lithium, cobalt, and nickel, which pose significant environmental challenges. Replacing these metals with redox-active organic materials offers a promising alternative, thereby reducing the carbon footprint of batteries by one order of magnitude. However, this approach faces critical obstacles, including the limited availability of suitable redox-active organic materials and issues such as lower electronic conductivity, voltage, specific capacity, and long-term stability. To overcome the limitations for lower voltage and specific capacity, a machine learning (ML) driven battery informatics framework is developed and implemented. This framework utilizes an extensive battery dataset and advanced ML techniques to accelerate and enhance the identification, optimization, and design of redox-active organic materials. In this contribution, a data-fusion ML coupled meta learning model capable of predicting the battery properties, voltage, and specific capacity for various organic negative electrodes and charge carriers (positive electrode materials) combinations is presented. The ML models accelerate experimentation, facilitate the inverse design of battery materials, and identify suitable candidates from three extensive material libraries to advance sustainable energy-storage technologies.

1 | Introduction

To reduce dependency on fossil fuels, electric batteries have become a prominent theme in modern scientific research [1, 2]. However, several studies assert that the demand for electric vehicles (EV) soars multifold, which is difficult to meet by transition metals only owing to the need for intensive mining and depleting resource [3, 4]. Currently, redox-active organic materials play an instrumental role in addressing the challenges posed by heavy reliance on fossil fuels and the skyrocketing demand for transition-metal elements like Li, Co, and Ni in conventional batteries [5–8]. Redox-active organic materials exhibit diverse chemistries, structures, and applications for energy storage and mobility [5]. They offer a wide range of characteristics like versatility, high-rate performance, and high theoretical capacity [9]. About 200 redox-active organics are currently being used as electrodes in batteries [10–16]. However, they are bound by limitations such as dissolution in the electrolyte, poor electrical conductivity, or low volumetric density [17–20]. To address these issues, we require novel organic materials with improved electrochemical performance [19, 21]. But the large chemical space of redox-active organic materials [22, 23] makes it very time consuming and expensive [22, 24] for conventional approaches such as high throughput experimentation or combinatorial chemistry to identify possible candidates. ML methods provide a platform to navigate vast chemical spaces for designing materials with suitable properties [25–29]. They are proven to be successful for material space explorations in the inorganic material space in the field of energy-storage applications (mainly batteries and super-capacitors) [30–33].

In these ML models, compositional, elemental, and structural parameters are taken as ML inputs, while, for example, the discharge capacity is predicted [30, 34].

For the unique and machine-readable representation of the chemical structures, Simplified Molecular Input Line Entry System (SMILES) strings are often used for encoding chemistries such as side chains, branches, rings, or chemical bonds [35]. Tools like polyBERT and Morgan fingerprints convert SMILES strings to vectors that serve as the numerical inputs to the ML models [36–38]. Several previous investigations have utilized these fingerprinting approaches for the prediction of diverse properties such as the HOMO-LUMO gap, atomization energy, redox potential, and so on [39–46]. Only limited research has focused on the prediction of experimental properties such as voltage for organic batteries [47]. Moreover, while data scarcity entails a significant limitation for training ML models on experimental properties [48, 49], advanced learning approaches like multitask, multifidelity, and transfer learning are contemporary solutions to address these challenges [50, 51].

In this contribution, we employ a ML approach as illustrated in Figure 1 to screen and identify high-voltage and high-specific capacity redox-active polymers for electrochemical applications. We utilize SMILES strings based notation to represent structures of organic materials (molecules and polymers) in our dataset. An important component of the workflow is the implementation of polyBERT as a fingerprinting tool, which transforms the SMILES strings into numerical representations suitable for ML model [36]. We develop a data-fusion MT-ML model followed by a meta learner to learn voltage and specific capacity for varying charge carriers (positive electrode materials). This multistep data-fusion methodology demonstrates improved generalizability and superior performance metrics, including better coefficient of determination (R^2) values and reduced RMSE (Root Mean-Square Error). Finally, we employ an inverse design methodology by creating three large libraries of polymeric materials. We propose new redox-active polymer candidates with maximum energy density, facilitating the efficient discovery and development of novel materials for battery applications.

2 | Results and Discussion

2.1 | Dataset

We utilize a dataset of a total of 771 data points for training our multi step data-fusion ML models as listed in Table 1. Our dataset contains both redox-active and non-redox active organic materials as negative electrodes with their electrochemical

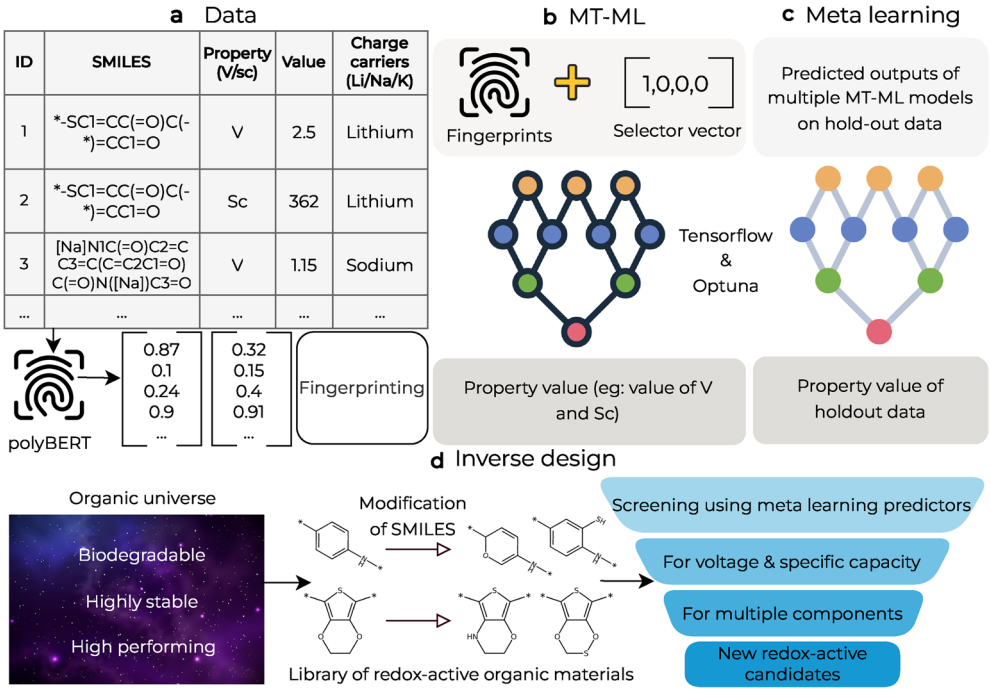

FIGURE 1 | Our battery-informatics workflow. (a) Data collection pipeline: We collect SMILES (Simplified molecular input line entry system) strings of organic negative electrodes with respect to different charge carriers (positive electrode materials) [35]. This is followed by conversion of SMILES strings to numerical format. (b) Architecture of our multitask machine learning (MT-ML) predictors. MT-ML models are trained to predict multiple properties with respect to varying different battery components. Variation of charge carriers (positive electrode materials), property and organic material classes (polymers/molecules) is represented using selector-vector. The fingerprints are concatenated with selector vector and are used as inputs to MT-ML model. Lighter version of gray represents inputs of the model and darker version represents output. (c) Meta learners trained on holdout dataset by taking outputs of multiple MT-ML models as the inputs. The outputs of meta-learner models is the property value (voltage and specific capacity). (d) Finally, the inverse design approach is employed. We take some reference organic materials that exhibit either higher performance for batteries or higher stability or biodegradability. We iteratively add redox-active moieties or replace elements and bonds at different positions of the organic materials to create a library of millions of organic materials. We screen for potential candidates with higher voltage and specific capacity by using the proposed meta-learner model.

TABLE 1 | Synopsis of the dataset for the battery property prediction models. The table shows the number of data points for different charge carriers (positive electrode materials) and polymer or molecule negative electrodes for voltage (V) and specific capacity (Sc).

Charge carrier (+ve)	Redox active		Nonredox active		Total
	V	Sc	V	Sc	
Polymer					
Al	2	4	0	0	6
K	1	1	0	0	2
Li	73	119	35	29	256
Mg	5	5	0	0	10
Na	10	7	29	35	81
Zn	2	3	0	0	5
Molecule					
Al	1	0	0	0	1
K	2	3	0	0	5
Li	60	165	35	29	289
Mg	1	2	0	0	3
Na	19	27	29	35	110
Zn	3	0	0	0	3
Total	179	336	128	128	771

properties voltage in volts and specific capacity in $mAhg^{-1}$. For non-redox active materials, the dataset comprises of 128 data points for both properties, for a total of 256 data points. These data points are equally divided for both properties for lithium and sodium charge carriers (positive electrode materials) (85 points for each property and positive electrode) for molecules and polymers respectively. To ensure our models accurately reflect chemical reality, we include non-redox active materials in the training dataset. This teaches the models to differentiate between materials that can and cannot undergo redox reactions. On the other hand, our redox-active materials dataset is more extensive, containing 336 data points for specific capacity and 179 data points for voltage measurements, totaling 515 data points. It is worth noting that, 66% of data points (338 our of 506) in specific capacity are theoretically calculated. By incorporating these values, the ML models learn inherent correlations between theoretical and measured discharge capacity. In this process, the theoretical capacity is encoded using the selector vector (see Method section), allowing the model to effectively integrate and interpret the relationship between theoretical predictions and experimental measurements. This allows for the prediction of the costly and labor-intensive measured values directly from theoretical data, significantly enhancing the model's overall performance. The discharge specific capacity data points are collected with respect to varying C-rates and active material content for the first cycle. The distribution of both the properties for redox-active negative electrodes are shown in the Figure 2.

2.2 | Model Performance

Utilizing our comprehensive dataset comprising of 771 data points, we train multitask and meta-learning models of the prediction of voltage and specific capacity. Figure 3 illustrates the comparative performance analysis of these models. The parity plots for MT-ML models, derived from the averaged results across all five-fold testing sets are shown in Figure 3a,c. The error bars indicate variance (2σ) of the predictions across the five-folds. MT-ML models demonstrate their efficiency in accurately predicting multiple electrochemical properties. A single MT-ML model is trained on multiple properties simultaneously, enhancing the predictive efficiency, addressing data scarcity and eliminating the necessity of training individual model for each property, as shown in Reference [50]. The MT-ML models achieves RMSE values of 0.43 V and $70.9\,mAhg^{-1}$ for voltage and specific capacity respectively. The voltage predictions of lithium charge carrier and polymeric negative electrodes are closer to the parity line in Figure 3a when compared to their molecular counterparts. On the other hand, prediction of specific capacity perform better for molecular negative electrodes and lithium charge carrier in Figure 3c. For charge carriers containing lesser data-points, the predictions are spread on the both sides of the parity line. Our multitask models exhibit robust performance across diverse electrode materials, successfully predicting properties for various charge carriers (positive electrode materials) and organic material classes (OMC).

Subsequently, we implement and train meta-learning techniques to enhance and bundle the predictive capabilities of our MT-ML framework for the deployment of our ML models. The resulting parity plots in Figure 3 demonstrate the exceptional generalization capacity of our meta-learner models, achieving remarkably high R^2 values of 0.99 and 0.95 for voltage and specific capacity, respectively. Our meta-learner model trains on the holdout dataset (20%) predictions (i.e., voltage and specific capacity) of MT-ML models and validates on the training dataset (80%). Our meta-learner models are similar to the Mixture-of-Experts (MoE) ML architecture, which dynamically assigns input-dependent weights to ML models and predict targeted output variables [52]. To mitigate overfitting risks, we implement dropout layers in both MT-ML and meta-learner models and run full hyperparameter optimization studies [53]. A comparative analysis between multitask (panel a and c) and meta-learner (panel b and d) in Figure 3 results reveals two significant improvements: (i) enhanced alignment of predictions with the parity line and (ii) substantial improvement of predictions for charge carriers (positive electrode materials) with less data-points. Notably, our models has these exceptional performance metrics irrespective of the inherent challenges of working with a relatively small dataset and multiple fidelity levels (incorporating both theoretical and experimentally measured capacity values across varying C-rates and active material loadings as encoded in the fingerprints). The strong improvement in predictive accuracy from multitask to meta-learner models showcases the efficiency of our multistep data-fusion approach in addressing the complexities in electrochemical property predictions.

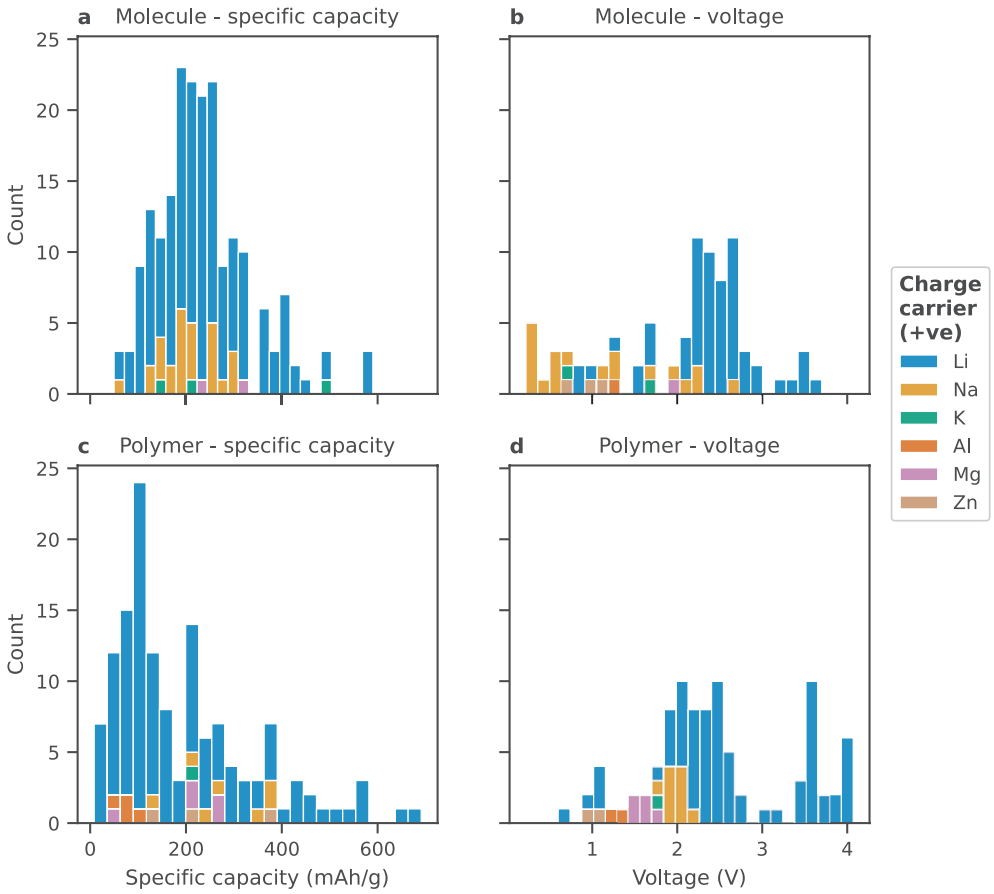

FIGURE 2 | Stacked histogram plots for redox active molecules (panel a and b) and polymers (panel c and d) for specific capacity and voltage when used as negative electrodes with respect to different charge carriers (positive electrode materials) in batteries as indicated in the legend.

2.3 | Inverse Design

In our inverse design methodology, we establish a search space from 12 selected reference organic polymers as shown in the Figure S2. Among those 12, two polymers are chosen for their high voltage, six for their structural stability and the remaining four based on their biodegradability. As illustrated in Figure 4, the polymer candidates are categorized into three distinct libraries based on their reference: high-voltage polymeric negative electrodes (Figure 4a), stable plastics (Figure 4b), and biodegradable polymers (Figure 4c). These candidate libraries serve as search space for potential replacements to existing organic electrodes that exhibit high voltage, specific capacity, or energy density. Through systematic structural modifications like strategic addition and substitution of redox-active moieties (see Figure S3) at both main-chain and side-chain positions, the 12 reference polymers are expanded into an extensive library of approximately 1.8 million candidates. Because the search space of 1.8 million candidates is derived from the initial candidate population of 12 polymers, the core structural motifs of the 12 initial candidates are retained. Using a different set of initial candidates would create an entirely different search space. The electrochemical properties of these candidates were predicted using our two-step data-fusion models, with the resulting property distributions as depicted in Figure 4. The observed property distributions exhibit systematic correlations that reflect fundamental chemical principles through the direct modification of molecular electronic structure, which occurs via incorporation of electron-withdrawing or electron-donating redox-active groups in organic materials [54]. Our screening process identifies several promising candidates with operating voltages between 4.1 and 4.5 V (see Figure S5), exceeding the maximum voltage of 4.07 V present in our training dataset. Analysis of the distribution patterns reveals varying characteristics of the three libraries. The high-voltage organic negative electrode library (Figure 4a) shows a notable concentration of candidates in the high-voltage regime with moderate specific capacity. Conversely, the biodegradable polymer library (Figure 4c) exhibits an opposite trend (see Figure S6). All-organic batteries manufactured from biodegradable redox-active polymers present a viable solution for disposable electronic devices, offering enhanced safety in post-usage disposal [55]. The stable plastics library (Figure 4b) achieves the highest energy density (see Figure S4), approaching 1600 $Wh\,kg^{-1}$, indicating that the incorporation of redox-active moieties into stable organic backbones improves multiple battery properties [15]. On the other hand, our model also identifies materials with insulating properties, characterized by near-zero values across both properties, highlighting its unbiased prediction capabilities. This comprehensive performance spectrum, ranging from high-performing candidates to insulators, demonstrates the model's robust predictive capacity across the different classes of materials. Each of the three libraries—stable plastics, biodegradable materials, and high-performing polymers were screened for suitable candidates in an average of 1.5 min (90 s) per library.

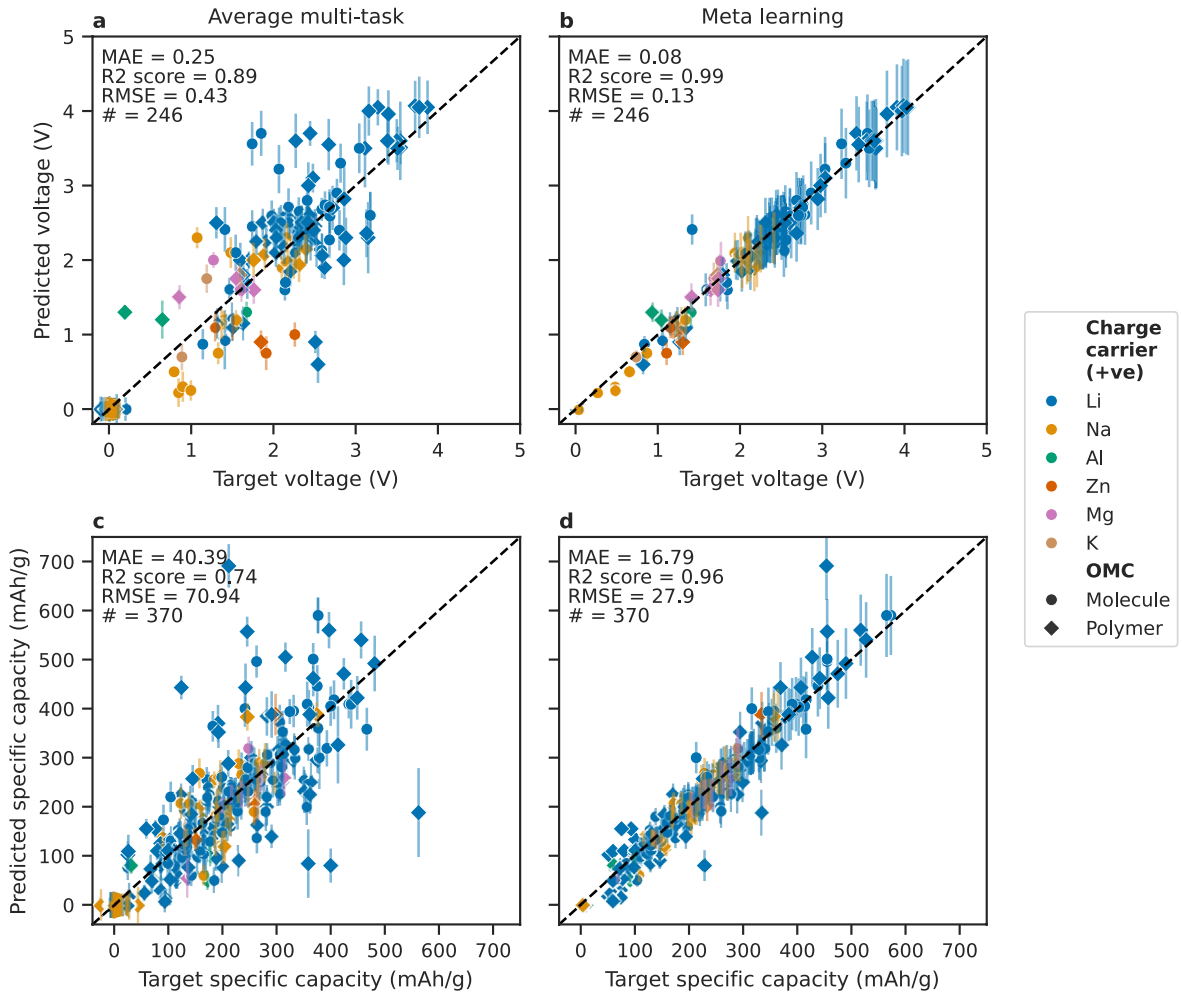

FIGURE 3 | Parity plots for the average of multitask (five models) and meta-learner models for voltage and specific capacity with respect to varying charge carriers (positive electrode materials) and organic material classes (OMC) used. Plots (b) and (d) represent meta-learning, whereas plots (a) and (c) represent the average of testing sets of five folds for MT-ML models.

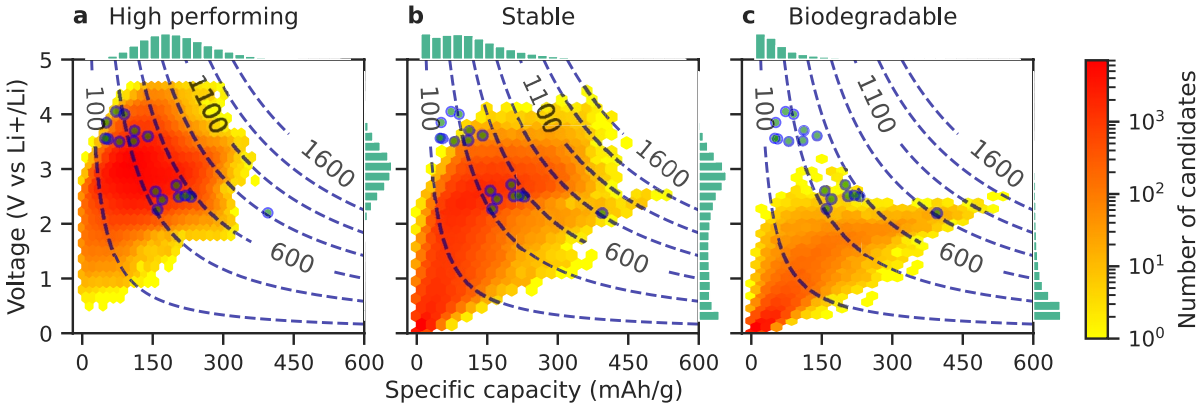

FIGURE 4 | Specific capacity versus voltage distribution for three libraries totaling to 1.8 million candidates. The voltage values represent average working potential versus lithium, while specific capacity values correspond to discharge capacity at 1C for the first cycle (50% active material loading). The overlaid dashed lines represent energy density ($Wh\,kg^{-1}$). The blue colored points represent experimental data-points versus lithium at 1C used for training our property predictor models.

The physiochemical analysis of top-performing polymer candidates showcases a trend in their molecular design; the strategic incorporation of sulfur and nitrogen hetero-atoms into either conjugated frameworks or fused ring systems influences the electronic structure and hence voltage. Superior specific capacity values are observed in materials with substantial presence of sulfur-based moieties, such as thiol and di-sulfide groups, owing to their participation in multielectron transfer reactions. Finally,

a combination of moieties containing multiple hetero-atoms, such as thiol groups, disulfide linkages, ester groups, carbonyls, and other groups within conjugated polymer-backbone assists in optimizing both voltage and specific capacity.

3 | Conclusion

Our investigation demonstrates the effectiveness of our data fusion ML model for predicting electrochemical properties across multiple domains and fidelity levels (active material loading), C-rate, property (voltage or specific capacity), organic material classification (molecule or polymer), and the charge carrier in positive electrode (Li, Na, K, Zn, Mg, and Al). The MT-ML framework is instrumental in addressing the critical challenge of data scarcity while enhancing the prediction accuracy simultaneously. Building upon this foundation, our meta-learning approach significantly improves model performance and generalization capabilities across various property domains and material classifications. Our framework shows significant computational efficiency, enabling ultra-fast prediction of electrochemical properties for large-scale material screening. This capability proves instrumental in the discovery of next-generation battery materials. By leveraging an inverse-design approach, our framework enables the rapid generation of candidate material libraries. Furthermore, it suggests both promising novel candidates and effective replacements for existing materials, significantly accelerating the screening process and reducing the time and resources required for experimental validation.

Our virtual library based inverse design approach involves structural modification of a defined set of initial candidate population. This enables a computationally efficient screening of new candidates, but confines the chemical space to the initial candidate population. In comparison, generative design strategies, such as Variational Auto-Encoders (VAE) [56, 57], General Adversarial Networks (GAN) [56, 57], and other de-novo methods possess capabilities to design novel molecular structures in an entirely unexplored chemical space without the requirement of an initial candidate population. However, these methods usually require large datasets and computationally expensive implementations [58]. While our ML models demonstrate exceptional property prediction performance, its reliance on 2.5D chemical structure fingerprints (derived from PSMILES strings) omits long distance structural information [59]. Consequently, our ML models do not account for redox-potential fluctuations arising from varying degrees of disorder, like crystallinity [60, 61].

Our research aims to enhance the accuracy and efficiency of the organic-battery informatics approach by integrating the influence of electrolytes and separators. Additionally, we are developing a novel fingerprinting approach for inorganic materials in batteries, which has the potential to significantly increase the number of available data points, leveraging the extensive databases available for batteries in the inorganic space. Our research has broader implications beyond theoretical advancement, facilitating the development of cost-effective and environmentally sustainable energy storage solutions. By paving the way for the discovery of high-performing organic battery materials, this work represents a significant step toward our goal of the identification of all-organic batteries using AI. These advancements could lead to batteries with improved capacity, longer lifespans, and reduced environmental impact, addressing critical challenges in the field of energy storage.

4 | Methods

4.1 | Data Preparation

Our dataset comprises 771 data points curated from peer-reviewed research publications, prominently focusing on review articles [5, 15, 62, 63]. We collected and curated the data from the review articles for different negative and charge carriers (positive electrode materials), voltages, and specific capacities. The UMAP (Uniform Manifold Approximation and Projection) plot of our dataset visualizing the distribution of different OMC (polymers and molecules) in a two-dimensional embedding space is shown in Figure S1 [64]. Prior to partitioning the dataset into training and testing subsets, we applied min-max scaling to normalize each target property (voltage and specific capacity) independently, linearly transforming the output values to a range based on their respective minimum and maximum values within the dataset.

4.2 | Fingerprinting

Fingerprinting involves conversion of chemical information represented by SMILES to machine-readable numerical format in form of vectors. For fingerprinting, we utilize the tool polyBERT [36], resulting in 600-dimensional vectors. Conversion to fingerprinting is followed by concatenation with selector vectors. Selector vector is an indicator of distinct domains in the data. They are five distinct numerical values, which encode key experimental and material parameters: active material content, C-rate, property type, OMC (molecule, polymer, or ladder polymer), and the charge carrier (e.g., Li, Na, K). An example of a selector vector used in the model is $[\mathbf{0.5}, \mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1}]$ indicating that the organic material is a polymer with 50% active material content, operating at 1C. Additionally, the property being predicted is voltage with respect to a sodium-positive electrode. Since, while calculating theoretical capacity, C-rate cannot be mentioned, we represented it using the number −1. When fingerprints are concatenated with the selector vectors, the input to our ML models yields a 605-dimensional numerical representations. This approach enables the representation of multiple fidelity levels within a single computational framework, thereby reducing generalization error [50].

4.3 | Model Architecture and Training

Our two-step data-fusion ML approach encompasses both MT-ML and meta-learning components. The dataset was randomly split, allocating 80% for MT-ML model development and reserving 20% for meta-learner training. The MT-ML training dataset is subdivided into 5 folds for cross validation, and five independent models are trained for each fold. The neural network-based MT-ML optimization is performed using the TensorFlow framework, supported by Optuna [65, 66], enabling systematic and

full hyper-parameter tuning across neural network architecture parameters (i.e., neuron count, number of layers, activation function, and dropout rate) and training parameters (i.e., initial learning rate and early stopping criteria). The hyper-parameter optimization for each fold is performed independently. The meta-learning step consists of a deployment-ready ensemble framework integrating insights deduced from all cross-validated models. The meta learning follows a two-step methodology: initially, predictions are generated for the held-out 20% dataset using the 5-fold cross-validated models, followed by the training of a neural network that uses these predictions as inputs to learn final property values. The model's generalization capability is tested against the original 80% training dataset. Comparison of cross-validated and meta learner models is shown in Tables S1 and S2. Uncertainty quantification is implemented for both multitask and meta-learning predictions using Monte Carlo dropout methodology, providing confidence intervals at the 95% level [67]. This meta-learning approach ensures comprehensive model generalization across the entire dataset while maintaining prediction reliability.

4.4 | Virtual Polymer Libraries

Our inverse design approach is based on the screening of three large virtual libraries of candidates, namely high-voltage, biodegradable, and stable plastic organic materials. The libraries are created through specific guidelines: We start with 12 synthesizable and highly researched candidates from the scientific literature (see Figure S2). Out of the 12 candidates, two polymers exhibit high voltage, six demonstrate high structural stability, and the remaining four are biodegradable, as indicated in Figure S2. Second, we iteratively replaced single atoms and chemical moieties with predefined chemical fragments (see Figure S3) at all individual positions, repeated this process iteratively to explore all permutations of chemical fragments at all the positions to create three vast candidate libraries totaling 1.8 million distinct PSMILES strings. Third, these distinct PSMILES strings are chemically evaluated using the cheminformatics tools, RDKit and PSMILES [36, 68]. The Python programming code for creating the 1.8 million candidates is found in our github repository https://github.com/kuennethgroup/organic_battery_predictor/blob/main/inverse_design.ipynb. Finally, the chemically valid PSMILES strings are fingerprinted using polyBERT, followed by property predictions using our trained meta-learner ML model. The candidates that exhibit the highest voltage, specific capacity, and energy density are displayed in Figures S4–S6. It is important to note that the synthesizability of the new candidates has not been evaluated.

5 | CO_2 Emission and Timing

Experiments are conducted using a computing cluster at the University of Bayreuth, with the carbon efficiency of $0.344\,kg\,CO_2eq\,kWh^{-1}$. A total of 10 h of computations were carried out on four A-100-40GB GPU's with a Thermal Design Power (TDP) of 250 W. The training of MT-ML and meta learner models producing emissions around $1.4\,kg\,CO_2eq$. In the inference mode, the computation per polymer is around 2 s combining fingerprinting, concatenation of selector vector and prediction from the trained model in total on one GPU, emitting around $57\,g\,CO_2eq$.

Acknowledgments

The authors sincerely thank the graduate school of the Bavarian Center for Battery Technology (BayBatt) for funding the ongoing research. The code used for training the MT-ML and meta learner models is available for academic use at https://github.com/kuennethgroup/organic_battery_predictor and Zenodo [69].

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the authors upon reasonable request.

References

1. M. Li, J. Lu, Z. Chen, and K. Amine, "30 Years of Lithium-Ion Batteries," *Advanced Materials* 30 (2018): 1800561, https://doi.org/10.1002/adma.201800561.

2. D. Larcher and J.-M. Tarascon, "Towards Greener and More Sustainable Batteries for Electrical Energy Storage," *Nature Chemistry* 7 (2015): 19–29, https://doi.org/10.1038/nchem.2085.

3. C. Xu, Q. Dai, L. Gaines, M. Hu, A. Tukker, and B. Steubing, "Future Material Demand for Automotive Lithium-Based Batteries," *Communications Materials* 1 (2020): 99, https://doi.org/10.1038/s43246-020-00095-x.

4. International Energy Agency, *The Role of Critical Minerals in Clean Energy Transitions* (OECD, 2021), https://doi.org/10.1787/f262b91c-en.

5. Y. Lu and J. Chen, "Prospects of Organic Electrode Materials for Practical Lithium Batteries," *Nature Reviews Chemistry* 4 (2020): 127–142, https://doi.org/10.1038/s41570-020-0160-9.

6. J. Kim, J. Ling, Y. Lai, and P. J. Milner, "Redox-Active Organic Materials: From Energy Storage to Redox Catalysis," *ACS Materials Au* 4 (2024): 258–273, https://doi.org/10.1021/acsmaterialsau.3c00096.

7. B. Esser, F. Dolhem, M. Becuwe, P. Poizot, A. Vlad, and D. Brandell, "A Perspective on Organic Electrode Materials and Technologies for Next Generation Batteries," *Journal of Power Sources* 482 (2021): 228814, https://doi.org/10.1016/j.jpowsour.2020.228814.

8. T. Bertaglia, C. M. Costa, S. Lanceros-Méndez, and F. N. Crespilho, "Eco-Friendly, Sustainable, and Safe Energy Storage: A Nature-Inspired Materials Paradigm Shift," *Materials Advances* 5 (2024): 7534–7547, https://doi.org/10.1039/D4MA00363B.

9. S. Lee, J. Hong, and K. Kang, "Redox-Active Organic Compounds for Future Sustainable Energy Storage System," *Advanced Energy Materials* 10 (2020): 2001445, https://doi.org/10.1002/aenm.202001445.

10. M. E. Bhosale, S. Chae, J. M. Kim, and J.-Y. Choi, "Organic Small Molecules and Polymers as an Electrode Material for Rechargeable Lithium Ion Batteries," *Journal of Materials Chemistry A* 6 (2018): 19885–19911, https://doi.org/10.1039/C8TA04906H.

11. P. Poizot, F. Dolhem, and J. Gaubicher, "Progress in All-Organic Rechargeable Batteries Using Cationic and Anionic Configurations: Toward Low-Cost and Greener Storage Solutions?," *Current Opinion in Electrochemistry* 9 (2018): 70–80, https://doi.org/10.1016/j.coelec.2018.04.003.

12. B. Häupler, A. Wild, and U. S. Schubert, "Carbonyls: Powerful Organic Materials for Secondary Batteries," *Advanced Energy Materials* 5 (2015): 1402034, https://doi.org/10.1002/aenm.201402034.

13. T. B. Schon, B. T. McAllister, P.-F. Li, and D. S. Seferos, "The Rise of Organic Electrode Materials for Energy Storage," *Chemical Society Reviews* 45 (2016): 6345–6404, https://doi.org/10.1039/C6CS00173D.

14. Z. Song and H. Zhou, "Towards Sustainable and Versatile Energy Storage Devices: An Overview of Organic Electrode Materials," *Energy & Environmental Science* 6 (2013): 2280, https://doi.org/10.1039/c3ee40709h.

15. Y. Liang, Z. Tao, and J. Chen, "Organic Electrode Materials for Rechargeable Lithium Batteries," *Advanced Energy Materials* 2 (2012): 742–769, https://doi.org/10.1002/aenm.201100795.

16. J. Wang, A. E. Lakraychi, X. Liu, et al., "Conjugated Sulfonamides as a Class of Organic Lithium-Ion Positive Electrodes," *Nature Materials* 20 (2021): 665–673, https://doi.org/10.1038/s41563-020-00869-1.

17. S. Lee, G. Kwon, K. Ku, et al., "Recent Progress in Organic Electrodes for Li and Na Rechargeable Batteries," *Advanced Materials* 30 (2018): 1704682, https://doi.org/10.1002/adma.201704682.

18. R. Shi, S. Jiao, Q. Yue, G. Gu, K. Zhang, and Y. Zhao, "Challenges and Advances of Organic Electrode Materials for Sustainable Secondary Batteries," *Exploration* 2 (2022): 20220066, https://doi.org/10.1002/EXP.20220066.

19. Z. Yu, Y. Wang, and Z. Luo, "Review on Redox-Active Organic Compounds for all-Organic Batteries," *Industrial & Engineering Chemistry Research* 63 (2024): 9619–9630, https://doi.org/10.1021/acs.iecr.4c00394.

20. G. Kwon, Y. Ko, Y. Kim, K. Kim, and K. Kang, "Versatile Redox-Active Organic Materials for Rechargeable Energy Storage," *Accounts of Chemical Research* 54 (2021): 4423–4433, https://doi.org/10.1021/acs.accounts.1c00590.

21. X. Li, Y. Wang, L. Lv, G. Zhu, Q. Qu, and H. Zheng, "Electroactive Organics as Promising Anode Materials for Rechargeable Lithium Ion and Sodium Ion Batteries," *Energy Materials* 2 (2022): 200014, https://doi.org/10.20517/energymater.2022.11.

22. J.-L. Reymond, "The Chemical Space Project," *Accounts of Chemical Research* 48 (2015): 722–730, https://doi.org/10.1021/ar500432k.

23. Y. Lu, Q. Zhang, and J. Chen, "Perspectives on the Redox Chemistry of Organic Electrode Materials in Lithium Batteries," *CCS Chemistry* 5 (2023): 1491–1508, https://doi.org/10.31635/ccschem.023.202302740.

24. S. P. Leelananda and S. Lindert, "Computational Methods in Drug Discovery," *Beilstein Journal of Organic Chemistry* 12 (2016): 2694–2718, https://doi.org/10.3762/bjoc.12.267.

25. C. Zhang, Y. Zhai, Z. Gong, et al., "Transfer Learning Across Different Chemical Domains: Virtual Screening of Organic Materials with Deep Learning Models Pretrained on Small Molecule and Chemical Reaction Data," *Journal of Cheminformatics* 16 (2024): 89, https://doi.org/10.1186/s13321-024-00886-1.

26. K. T. Butler, K. Choudhary, G. Csanyi, A. M. Ganose, S. V. Kalinin, and D. Morgan, "Setting Standards for Data Driven Materials Science," *npj Computational Materials* 10, no. 1 (2024): 231, https://doi.org/10.1038/s41524-024-01411-6.

27. G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, and A. Fazzio, "From DFT to Machine Learning: Recent Approaches to Materials Science–a Review," *Journal of Physics: Materials* 2 (2019): 032001, https://doi.org/10.1088/2515-7639/ab084b.

28. K. A. Shah, R. Q. Albuquerque, C. Brütting, and H. Ruckdäschel, "Machine Learning-Based Time Series Analysis of Polylactic Acid Bead Foam Extrusion," *Journal of Applied Polymer Science* 141 (2024): e56170, https://doi.org/10.1002/app.56170.

29. S. M. Pai, K. A. Shah, S. Sunder, R. Q. Albuquerque, C. Brütting, and H. Ruckdäschel, "Machine Learning Applied to the Design and Optimization of Polymeric Materials: A Review," *Next Materials* 7 (2025): 100449, https://doi.org/10.1016/j.nxmate.2024.100449.

30. P. Thakkar, S. Khatri, D. Dobariya, D. Patel, B. Dey, and A. K. Singh, "Advances in Materials and Machine Learning Techniques for Energy Storage Devices: A Comprehensive Review," *Journal of Energy Storage* 81 (2024): 110452, https://doi.org/10.1016/j.est.2024.110452.

31. A. Chen, X. Zhang, and Z. Zhou, "Machine Learning: Accelerating Materials Development for Energy Storage and Conversion," *InfoMat* 2 (2020): 553–576, https://doi.org/10.1002/inf2.12094.

32. Z. Ahmad, T. Xie, C. Maheshwari, J. C. Grossman, and V. Viswanathan, "Machine Learning Enabled Computational Screening of Inorganic Solid Electrolytes for Suppression of Dendrite Formation in Lithium Metal Anodes," *ACS Central Science* 4 (2018): 996–1006, https://doi.org/10.1021/acscentsci.8b00229.

33. Z. Shen, H. Liu, Y. Shen, J. Hu, L. Chen, and C. Nan, "Machine Learning in Energy Storage Materials," *Interdisciplinary Materials* 1 (2022): 175–195, https://doi.org/10.1002/idm2.12020.

34. G. Wang, T. Fearn, T. Wang, and K.-L. Choy, "Machine-Learning Approach for Predicting the Discharging Capacities of Doped Lithium Nickel–Cobalt–Manganese Cathode Materials in Li-Ion Batteries," *ACS Central Science* 7 (2021): 1551–1560, https://doi.org/10.1021/acscentsci.1c00611.

35. D. Weininger, "SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules," *Journal of Chemical Information and Computer Sciences* 28 (1988): 31–36, https://doi.org/10.1021/ci00057a005.

36. C. Kuenneth and R. Ramprasad, "polyBERT: A Chemical Language Model to Enable Fully Machine-Driven Ultrafast Polymer Informatics," *Nature Communications* 14 (2023): 4099, https://doi.org/10.1038/s41467-023-39868-6.

37. A. P. Bento, A. Hersey, E. Félix, et al., "An Open Source Chemical Structure Curation Pipeline Using RDKit," *Journal of Cheminformatics* 12 (2020): 51, https://doi.org/10.1186/s13321-020-00456-1.

38. D. Rogers and M. Hahn, "Extended-Connectivity Fingerprints," *Journal of Chemical Information and Modeling* 50 (2010): 742–754, https://doi.org/10.1021/ci100050t.

39. R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, "Quantum Chemistry Structures and Properties of 134 Kilo Molecules," *Scientific Data* 1 (2014): 140022, https://doi.org/10.1038/sdata.2014.22.

40. G. A. Pinheiro, J. Mucelini, M. D. Soares, R. C. Prati, J. L. F. Da Silva, and M. G. Quiles, "Machine Learning Prediction of Nine Molecular Properties Based on the SMILES Representation of the QM9 Quantum-Chemistry Dataset," *Journal of Physical Chemistry A* 124 (2020): 9854–9866, https://doi.org/10.1021/acs.jpca.0c05969.

41. Z. Wu, B. Ramsundar, E. Feinberg, et al., "MoleculeNet: A Benchmark for Molecular Machine Learning," *Chemical Science* 9 (2018): 513–530, https://doi.org/10.1039/C7SC02664A.

42. T. Zhang, K. Wang, K. Jing, et al., "A Ring2Vec Description Method Enables Accurate Predictions of Molecular Properties in Organic Solar Cells," *npj Computational Materials* 10 (2024): 268, https://doi.org/10.1038/s41524-024-01372-w.

43. K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-Chemical Insights From Deep Tensor Neural Networks," *Nature Communications* 8 (2017): 13890, https://doi.org/10.1038/ncomms13890.

44. C. Kuenneth, J. Lalonde, B. L. Marrone, C. N. Iverson, R. Ramprasad, and G. Pilania, "Bioplastic Design Using Multitask Deep Neural Networks," *Communications Materials* 3 (2022): 96, https://doi.org/10.1038/s43246-022-00319-2.

45. C. Kuenneth, W. Schertzer, and R. Ramprasad, "Copolymer Informatics With Multitask Deep Neural Networks," *Macromolecules* 54 (2021): 5957–5961, https://doi.org/10.1021/acs.macromol.1c00728.

46. S. S. Shukla, C. Kuenneth, and R. Ramprasad, "Polymer Informatics Beyond Homopolymers," *MRS Bulletin* 49 (2024): 17–24, https://doi.org/10.1557/s43577-023-00561-0.

47. R. P. Carvalho, C. F. Marchiori, D. Brandell, and C. M. Araujo, "Artificial Intelligence Driven In-Silico Discovery of Novel Organic Lithium-Ion Battery Cathodes," *Energy Storage Materials* 44 (2022): 313–325, https://doi.org/10.1016/j.ensm.2021.10.029.

48. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine Learning for Molecular and Materials Science," *Nature* 559 (2018): 547–555, https://doi.org/10.1038/s41586-018-0337-2.

49. Y. Chen, P. Long, B. Liu, et al., "Development and Application of Few-Shot Learning Methods in Materials Science Under Data Scarcity," *Journal of Materials Chemistry A* 12 (2024): 30249–30268, https://doi.org/10.1039/D4TA06452F.

50. C. Kuenneth, A. C. Rajan, H. Tran, L. Chen, C. Kim, and R. Ramprasad, "Polymer Informatics With Multi-Task Learning," *Patterns* 2 (2021): 100238, https://doi.org/10.1016/j.patter.2021.100238.

51. M. L. Hutchinson, E. Antono, B. M. Gibbons, S. Paradiso, J. Ling, and B. Meredig, "Overcoming Data Scarcity With Transfer Learning," (2017), https://arxiv.org/abs/1711.05099.

52. L. Morand and D. Helm, "A Mixture of Experts Approach to Handle Ambiguities in Parameter Identification Problems in Material Modeling," *Computational Materials Science* 167 (2019): 85–91, https://doi.org/10.1016/j.commatsci.2019.04.003.

53. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks From Overfitting," *Journal of Machine Learning Research* 15 (2014): 1929–1958.

54. J. Bitenc, K. Pirnat, O. Lužanin, and R. Dominko, "Organic Cathodes, a Path Toward Future Sustainable Batteries: Mirage or Realistic Future?," *Chemistry of Materials* 36 (2024): 1025–1040, https://doi.org/10.1021/acs.chemmater.3c02408.

55. M. H. Lee, J. Lee, S. Jung, et al., "A Biodegradable Secondary Battery and Its Biodegradation Mechanism for Eco-Friendly Energy-Storage Systems," *Advanced Materials* 33 (2021): 2004902, https://doi.org/10.1002/adma.202004902.

56. Q. Vanhaelen, Y. C. Lin, and A. Zhavoronkov, "The Advent of Generative Chemistry," *ACS Medicinal Chemistry Letters* 11 (2020): 1496–1505, https://doi.org/10.1021/acsmedchemlett.0c00088.

57. H. H. Loeffler, J. He, A. Tibo, et al., "Reinvent 4: Modern AI–Driven Generative Molecule Design," *Journal of Cheminformatics* 16 (2024): 20, https://doi.org/10.1186/s13321-024-00812-5.

58. D. M. Anstine and O. Isayev, "Generative Models as an Emerging Paradigm in the Chemical Sciences," *Journal of the American Chemical Society* 145 (2023): 8736–8750, https://doi.org/10.1021/jacs.2c13467.

59. L. Pattanaik and C. W. Coley, "Molecular Representation: Going Long on Fingerprints," *Chem* 6 (2020): 1204–1207, https://doi.org/10.1016/j.chempr.2020.05.002.

60. T. H. Le, Y. Kim, and H. Yoon, "Electrical and Electrochemical Properties of Conducting Polymers," *Polymers* 9 (2017): 150, https://doi.org/10.3390/polym9040150.

61. C. Liu, Z. G. Neale, and G. Cao, "Understanding Electrochemical Potentials of Cathode Materials in Rechargeable Batteries," *Materials Today* 19 (2016): 109–123, https://doi.org/10.1016/j.mattod.2015.10.009.

62. J. Kim, Y. Kim, J. Yoo, G. Kwon, Y. Ko, and K. Kang, "Organic Batteries for a Greener Rechargeable World," *Nature Reviews Materials* 8 (2022): 54–70, https://doi.org/10.1038/s41578-022-00478-1.

63. X. Judez, L. Qiao, M. Armand, and H. Zhang, "Energy Density Assessment of Organic Batteries," *ACS Applied Energy Materials* 2 (2019): 4008–4015, https://doi.org/10.1021/acsaem.9b00534.

64. L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," (2018), https://arxiv.org/abs/1802.03426.

65. M. Abadi, A. Agarwal, P. Barham, et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," (2015), https://www.tensorflow.org/.

66. T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-Generation Hyperparameter Optimization Framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage AK USA* Association for Computing Machinery, (2019), 2623–2631, https://doi.org/10.1145/3292500.3330701.

67. Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," (2015), https://arxiv.org/abs/1506.02142.

68. G. Landrum, P. Tosco, B. Kelley, et al., "rdkit/rdkit: 2025_03_2 (Q1 2025) Release," (2025), https://doi.org/10.5281/zenodo.15286010.

69. S. V. Ganti, C. Künneth, and L. Woelfel, "AI-Driven Discovery of High Performance Polymer Electrodes for Next-Generation Batteries," (2025), https://doi.org/10.5281/zenodo.14886803.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.