npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-024-01286-7

Attention towards chemistry agnostic and explainable battery lifetime prediction

Fuzhan Rahmanian[1,2,3,4,5] ✉, Robert M. Lee[6], Dominik Linzner[6], Kathrin Michel[6], Leon Merker[1,2], Balazs B. Berkes[6], Leah Nuss[1,3,4,5] & Helge Sören Stein[3,4,5] ✉

Predicting and monitoring battery life early and across chemistries is a significant challenge due to the plethora of degradation paths, form factors, and electrochemical testing protocols. Existing models typically translate poorly across different electrode, electrolyte, and additive materials, mostly require a fixed number of cycles, and are limited to a single discharge protocol. Here, an **a**ttention-based **rec**urrent **a**lgorithm for **n**eural **a**nalysis (ARCANA) architecture is developed and trained on an ultra-large, proprietary dataset from BASF and a large Li-ion dataset gathered from literature across the globe. ARCANA generalizes well across this diverse set of chemistries, electrolyte formulations, battery designs, and cycling protocols and thus allows for an extraction of data-driven knowledge of the degradation mechanisms. The model's adaptability is further demonstrated through fine-tuning on Na-ion batteries. ARCANA advances the frontier of large-scale time series models in analytical chemistry beyond textual data and holds the potential to significantly accelerate discovery-oriented battery research endeavors.

Lithium-ion batteries (LIBs) enable the electrification of everything, yet there is a maze of challenges that must be navigated in order to optimize the batteries of the future[1–4]. Critical to the advancement of battery research is the rapid understanding of why and how some batteries degrade and what needs to be changed to prevent premature capacity fade[5]. Material degradation can occur due to numerous factors, including unpreventable solid electrolyte interphase growth, loss of active material, and other electrochemical phenomena[6]. However, investigating battery degradation is a time-consuming task, as non-linear capacity loss can occur over hundreds or thousands of cycles[7]. Another challenge in early lifetime prediction is the diversity of battery chemistries in the anode, cathode, and electrolyte, along with various form factors and testing protocols.

Battery lifetime can be evaluated through various methods, such as conventional cycling until the end of life (EOL) under constant current-constant voltage (CC–CV) conditions or cycling for a predetermined number of cycles. From these data, measures such as coulombic efficiency (CE) can be calculated[8] and correlated to more in-depth techniques such as electrochemical impedance spectroscopy (EIS)[9] to fundamentally assess the underlying degradation mechanisms. Accurate measurement of CE[10,11] does, however, require bespoke instrumentation and a considerable amount of time, i.e., cycling a battery for 1000 cycles at $1C/1D$ takes approximately 11 weeks. Reducing the required number of cycles by a factor of 10 while maintaining a high level of fidelity is, therefore, of great interest[12]. Machine Learning (ML) and deep learning (DL) can accelerate testing by lowering the number of cycles required to understand the underlying chemistries[13]. An example of predicting the EOL of batteries using initial discharge capacity curves was demonstrated by Severson et al.[3], who used regression models. They integrated data generation with data-driven models to forecast the lifetime of LFP/graphite cells based on $\Delta Q(V)$ and classified their longevity. In further work, Attia et al.[12] employed a Bayesian algorithm to accelerate the optimization of fast-charging protocols. By using early-cycle data for low-fidelity predictions, the approach enabled the optimization of high-fidelity experimental outcomes, thus significantly reducing the experimental duration from 500 to 16 days.

The most reliable models do not, however, merely predict just predict a quantity but also allow assessment of the model's uncertainty. Emblematic of this is the work by Tong et al.[14], who introduced ADLSTM-MC, a hybrid predictive model using adaptive dropout long short-term memory (LSTM) with Monte Carlo simulations. This approach, which requires minimal training data, enhances robustness through Bayesian-optimized dropout

[1]Helmholtz Institute Ulm, Applied Electrochemistry, Helmholtzstr. 11, 89081 Ulm, Germany. [2]Karlsruhe Institute of Technology, Institute of Physical Chemistry, Fritz-Haber-Weg 2, 76131 Karlsruhe, Germany. [3]Technische Universität München, School of Natural Sciences, Department of Chemistry, Lichtenbergstr 4, 85748 Garching, Germany. [4]Technische Universität München, Munich Data Science Institute, Walther-von-Dyck-Straße 10, 4, 85748 Garching, Germany. [5]Technische Universität München, Munich Institute for Robotic and Machine Intelligence, Georg-Brauchle-Ring 60-62, 80992 Munich, Germany. [6]BASF SE, Ludwigshafen, Germany. ✉e-mail: fuzhan.rahmanian@tum.de; helge.stein@tum.de

rates and improves the remaining useful life of two types of LIBs. In a correlative study[15], a recurrent autoregressive deep ensemble network with aleatoric and epistemic uncertainties was developed along with saliency analysis to assess the impact of input parameters on output prediction. This provided an intuitive understanding of feature importance. Another advantage of using DL algorithms is their ability to use raw data, which has gained interest in the estimation of battery State of Health (SOH). For instance, Yang et al.[16] developed a hybrid convolutional neural network architecture with parallel residual connections, which utilizes raw data across multiple dimensions. By incorporating attention mechanisms, their model achieves remarkable accuracy in predicting the early stages of degradation. These advances support the increased focus on more adaptive and generative modeling frameworks, of which recent efforts include reinforcement learning from human feedback (RLHF) and the prompt paradigm in Generative Artificial Intelligence (GAI) techniques regarded for their potential to unravel complex structure–activity relationships in material behavior[17]. Although these approaches are applied in battery research[18,19], their prominence is not as widespread as in other scientific fields. However, this lesser emphasis provides an opportunity for further exploration and discovery.

Beyond these early lifetime prediction models, sequence-to-sequence (Seq-to-Seq) models have been used to monitor battery lifetime and (SOH)[18,20,21]. They leverage intrinsic temporal dependencies in degradation data, providing high predictive accuracy and computational efficiency. Li et al.[20] developed a one-shot LSTM-based Seq-to-Seq framework that not only predicts future capacities but also identifies knee points in the degradation curve, maintaining stability even in the face of stochastic disturbances. Although Seq-to-Seq models demonstrate robust predictions, they also exhibit limitations in generalization and require large and diverse datasets to enhance performance[4].

Despite the promises made by ML and DL for lifetime predictions[22–24], these models, while robust, face challenges of precision and trustworthiness[25]. Existing models often focus on single-task learning, neglecting the potential benefits of multi-objective learning for various predictive settings[4]. In particular, data-driven approaches[26,27] tend to overlook the inherent variations between, for example, production batches or individual cells[28]. Such discrepancies, originating from manufacturing processes or aging mechanisms, can profoundly impact lifetime predictions. Addressing these variations requires integrating domain knowledge into the learning process to enhance the model's ability to adapt and accurately forecast across diverse conditions[27]. Furthermore, despite the assertions of recent studies that they are chemistry-agnostic[15,29], they often require enhanced explainability to optimize their effectiveness in various chemistry settings. Transfer learning offers a promising solution to the challenge of scarce data but requires more investigation for transparency and interpretability[30]. The acquisition of extensive datasets, essential for DL algorithms[31], remains a significant hurdle[26,32,33]. Nevertheless, innovative strategies, such as the use of common features in databases and the documentation of various chemistries and protocols[34], establish the foundation for more in-depth research[31]. Our goal is to develop a model characterized by its adaptable design and robustness, with the capability to provide both uncertainty quantification and explainability. The model's strength is underlined by its adaptability in dynamically fine-tuning to specific chemical domains. Such a model would be invaluable to the academic community and would find marketable applications in the real world[31], accelerating battery design and data collection based on active learning.

Results

Data resources

Developing a model that generalizes well necessitates a diverse and large dataset[26] that ideally covers a spectrum of chemistries and formats given high-dimensional correlations and cell variations[30,35], obtained from various laboratories and measured under different operating conditions[12]. Data diversity not only ensures an accurate representation of different cycling behaviors but also tames the irreducible uncertainty in the predictions while mitigating the risk of overfitting. However, the scarcity of large and comprehensive datasets[25] that include both high and low-performing cells creates a challenge for training generalized models, i.e., to overcome a positive bias[30,36]. Available data often exhibit noise, discontinuities, and varying formats that require extensive curation, adding a layer of complexity. Initiatives such as Battery Archive[37] or other cloud services[38] are therefore commendable in promoting Findable, Accessible, Interoperable, and Reusable (FAIR) data[39,40] handling in battery research[32,33].

In this study, we develop a model trained on ca. 17,400 batteries from BASF research laboratories that cover a diverse range of LIBs chemistries and multiple cycling protocols. Exposure of our model to such a wide variety of data enables robust generalization. Utilizing our pre-trained model on a set of unseen data, we effectively predict the early degradation trajectory. The ultimate test of our model, therefore, is to apply it to data from cells produced in a different location and with varying chemistries. Due to intellectual property constraints that prevent the authors from making the model trained on the BASF dataset openly accessible, we have retrained our model by leveraging a diverse array of publicly available datasets from respected institutions and research groups, including the Toyota Research Institute (TRI) in partnership with MIT and Stanford[41,42], NASA[43], the Center for Advanced Life Cycle Engineering (CALCE)[44], Karlsruhe Institute of Technology (KIT)[45], Hawaii Natural Energy Institute (HNEI)[46], and Sandia National Laboratories (SNL)[46]. Furthermore, we have incorporated data from our in-house cycled cells[47–50] with successful and failed experiments to further enrich model training and reduce bias. In Supplementary Section 1, we provide an overview of all datasets; we include a brief summary in Table 1 with an indication of which datasets were used during training and which remained completely unseen for model testing. This approach ensures a thorough understanding of the data sources, thus improving the transparency and reproducibility of our research.

Architecture overview

Central to this study is the Attention-based ReCurrent Algorithm for Neural Analysis with LSTM (ARCANA) model. This is an attention-based Seq-to-Seq architecture specifically engineered to assess early-stage battery degradation and perform lifecycle monitoring. The model demonstrates superior multi-output predictive capabilities, supported by its high modularity and dynamic adaptability. It is designed to utilize a flexible range of past battery cycle data, known as historical temporal segments, for input. In addition, the model includes predetermined parameters for future conditions, such as discharge rates and cycle numbers. These parameters are known in advance of the experiment, i.e., they are controlled by the measurement device and are referred to as encoded temporal segments. This dual capability offers multifaceted advantages, from cost and time savings to improved material selection and protocol optimization.

The ARCANA model is augmented with additional features such as the attention mechanism, which provides insight into the decision-making process of the model. This feature distinguishes between predictions based on underlying patterns and those arising from stochastic variability. Saliency analysis is additionally performed to emphasize the relative importance of each parameter through a computation of the absolute gradient of the model output relative to the input of the test set. It quantifies the sensitivity of the input parameters, revealing how minor variations significantly alter the output results[15], thus aligning the internal logic of the model with domain-specific knowledge. Adding another layer of robustness is uncertainty quantification, which is valuable not only for understanding the reliability of cycling protocols but also for assessing material performance across different battery chemistries.

As illustrated in the unified modeling language (UML) diagram (Fig. 1), the ARCANA model consists of four principal classes, each performing a different function, and is designed to accept raw data, thus negating the need for preliminary feature engineering. This design versatility extends to its operational modes with Naive Training for initial experiments, Dynamic Tuning for real-time adaptability via extensive hyperparameter optimization, Fine-Tuning for integration of a pre-trained model with selective

Table 1 | Collected cycling data for training and testing

Location	Cell form	Cell chemistry	Protocol charge\discharge	No. cell	Cycle range	Nominal capacity [Ah]	Usage
BASF	Coin	Heterogenous	Multimodal	17400	Multimodal	Multimodal	*M*(*B*) Train\Val
TRI[41]	Cylindrical commercial	LFP\graphite	*CC*1(*Q*1)*CC*2, *CC*–*CV*@1*C*, 4.2*V* *CC*@4*C*	124	169–2235	1.1	*M*(*P*) Train\Val
TRI[42]	Cylindrical commercial	LFP\graphite	*CC*1(20%)*CC*2(40%)*CC*3(60%)*CC*4(80%), *CC*–*CV*@1*C*, 4.2*V**CC*–*CV*@4*C*, 2*V*	233	100–862	1.1	*M*(*P*) Train\Val
CALCE[44]	Prismatic commercial CX2	LCO\graphite	*CC*–*CV*@0.5*C*, 4.2*V*, *CC*@(0.5*C*, 1*C*)	6	781–1082	1.35	Testing
CALCE[44]	Prismatic commercial CS2	LCO\graphite	*CC*–*CV*@0.5*C*, 4.2*V*, *CC*@0.5*C*	6	1701–2016	1.1	*M*(*P*) Train\Val
KIT[45]	Cylindrical commercial	NCA\graphite-Si	*CC*–*CV*@(0.25*C*, 0.5*C*, 1*C*), 4.2*V*, *CC*@1*C*	58	29–800	3.5	*M*(*P*) Train\Val
KIT[45]	Cylindrical commercial	NCM\graphite–Si	*CC*–*CV*@(0.25*C*, 0.5*C*, 1*C*), 4.2*V*, *CC*@1*C*	55	43–1277	3.5	*M*(*P*) Train\Val
KIT[45]	Cylindrical commercial	NCM+NCA \graphite	*CC*–*CV*@0.5*C*, 4.2*V*,*CC*@(1*C*, 2*C*, 4*C*)	9	912–1031	2.5	Testing
KIT[47]	Coin self-made	LNO\graphite	*CC*–*CV*@1*C*, 4.2*V*, *CC*@1*C*	43	82–505	0.004618	60% for $M(P)_f$, 40% Testing
KIT[48]	Coin commercial	LCO\graphite	*CC*–*CV*@1*C*, 4.25*V*, *CC*–*CV*@1*C*, 2.75*V*	26	150–600	0.045	*M*(*P*) Train\Val
KIT[49]	Coin self-made	NMC622\graphite	*CC*–*CV*@1*C*, 4.2*V*,*CC*@1*C*	11	228–501	0.00328	Testing
KIT[50]	Coin self-made	$Na_{0.9}[...]O_2$ \graphite	*CC*@1*C* *CC*@1*C* or C-rates test	44	40–140	0.00015	60% for $M(P)_{Na}$ and $M(B)_{Na}$, 40% Testing
NASA[43]	Cylindrical commercial	NCA\graphite	*CC*–*CV*@0.75*C*, 4.2*V*, *CC*@(0.5*C*, 1*C*, 2*C*)	34	24–196	2.0	*M*(*P*) Train\Val
HNEI[46]	Cylindrical commercial	LCO-NMC \graphite	*CC*–*CV*@0.5*C*, 4.3*V*, *CC*@1.5*C*	14	1102–1133	2.8	*M*(*P*) Train\Val
SNL[46]	Cylindrical commercial	LFP\graphite	*CC*–*CV*@0.5*C*, 4.2*V*,*CC*@(0.5*C*, 1*C*, 2*C*, 3*C*)	28	2621–19,174	1.1	*M*(*P*) Train\Val
SNL[46]	Cylindrical commercial	NCA\graphite	*CC*–*CV*@0.5*C*, 4.2*V*, *CC*@(0.5*C*, 1*C*, 2*C*)	24	463–7877	3.2	*M*(*P*) Train\Val
SNL[46]	Cylindrical commercial	NMC\graphite	*CC*–*CV*@0.5*C*, 4.2*V*, *CC*@(0.5*C*, 1*C*, 2*C*, 3*C*)	25	388–11,149	3.0	*M*(*P*) Train\Val

An overview of the collected cycling data utilized for training and testing. The model *M*(*B*), was trained with data provided by BASF, and the model *M*(*P*) was trained with publicly available data. The model $M(P)_f$ represents a fine-tuned version of *M*(*P*) for lithium-ion coin cell data. $M(P)_{Na}$ and $M(B)_{Na}$ models are fine-tuned *M*(*B*) and *M*(*P*), respectively, adapted for sodium coin cells.

gradient updating, and prediction for efficient inference. Through modularity, a logging mechanism ensures data integrity and traceability, adhering to FAIR data principles[40]. The open-source codebase uses the PyTorch library[51] for model development and the Optuna library[52] for hyperparameter optimization.

The encoder–decoder framework. The encoder (Fig. 2a) initiates the Seq-to-Seq model in the ARCANA framework by processing historical temporal segments of the past battery life cycles. Employing an LSTM network, it is designed to capture complex, non-linear relationships and time dependencies inherent in sequence data. The encoder processes the input tensor to accommodate sequences of different lengths, employing a padding mechanism that enables the LSTM to efficiently process these sequences without being constrained by their varying lengths. Within the LSTM, the temporal data is transformed into a tensor, constructing hidden and cell states that capture sequential information. A skip connection incorporates the initial input into the LSTM output, thus preserving crucial temporal features and stabilizing the learning process. Layer normalization, when applied to the LSTM output, not only accelerates convergence but also leads to robust performance, mitigating the challenges associated with long-sequence dependencies[53]. The encoder returns a rich latent representation of the historical data, consisting of the output tensor and the updated hidden and cell states, which are then utilized by the decoder to enable accurate forecasting in subsequent steps.

The decoder (Fig. 2a) takes on the task of generating future state predictions. It is initialized with the hidden and cell states from the encoder and begins by processing the most recent historical cycle data. The model then integrates its own previous predictions and known future conditions, such as the expected discharge current and the cycle number. These two inputs are temporally encoded to capture their positional relevance[54], ensuring that the decoder is informed of the predefined condition and the timing of each data point within the life cycle. The decoder employs an attention mechanism that can dynamically adjust sequence weights, identifying critical information at each prediction step. This approach overcomes the limitations of static-length vector representation in conventional encoder-decoder models[55], allowing the decoder to focus on the most relevant parts of historical data. The attention mechanism then computes a context vector associated with the encoder's output, which highlights the encoder sequences with the highest relevance to the current decoding task. This context vector, combined with the current input, forms a feature-rich tensor that is subsequently processed by an LSTM layer. Post-LSTM, the output layer is passed through a fully connected layer with a leaky ReLU activation function, crucial in maintaining network stability, and enhanced with a dropout layer placed to reduce overfitting risks. The culmination of this process is a decoder that generates forecasts for the 0.1, 0.5, and 0.9 quantiles. These provide a probabilistic range indicative of the inherent uncertainty and offer a statistical interpretation of the potential future states of the degradation profile.

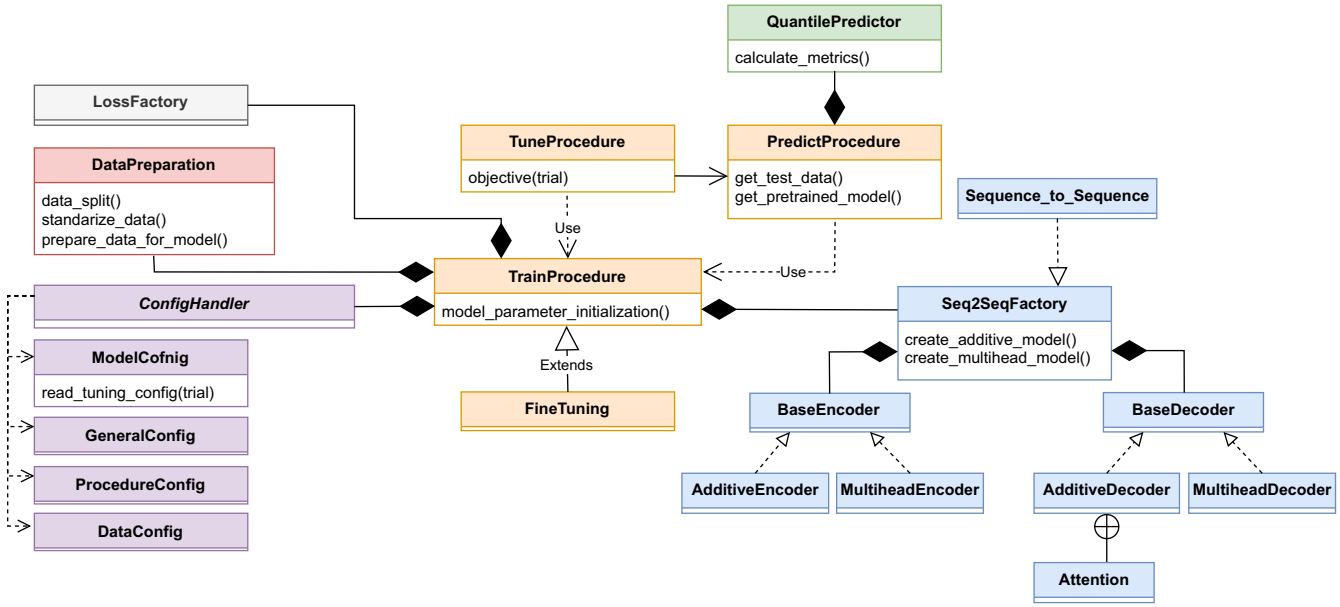

Fig. 1 | An UML diagram of the computational framework. The framework is designed around three principal class clusters. The first includes a `ConfigHandler` engineered to manage a comprehensive set of user-defined configurations and establishes a blueprint for handling various subconfigurations such as general settings, data properties, and model specifications. During hyperparameter optimization tasks, `ConfigHandler` interfaces with the Optuna optimization library to adaptively create and update the tuning configuration. The second key class structure includes `TrainProcedure`, which serves as an architectural template for the training process. Its attributes are employed throughout the computational pipeline, starting with data preparation and extending to the instantiation of specialized loss functions and Seq2Seq models via the `LossFactory` and `Seq2SeqFactory`. `FineTuning` is a specialized subclass that inherits from `TrainProcedure` while `TuneProcedure` and `PredictProcedure`, the latter of which uses the `QuantilePredictor`, are incorporated into the pipeline depending on the desired use case and settings. The tuning operates on single trials with a TPESampler when multiple runs are desired. Lastly, `Seq2SeqFactory` is engineered to govern the instantiation of encoder-decoder architectures. Depending on the user-defined configurations, it can orchestrate a multihead or an additive encoder-decoder mechanism. The inclusion of custom attention mechanisms within the architecture is handled by the `AdditiveDecoder` class or the `MultiheadDecoder`, conditional upon the configuration stipulations.

Seq-to-seq integration. In the broader Seq-to-Seq model, the encoder and decoder are orchestrated to facilitate the overall predictions, as can be seen in Fig. 2b. Here, the model processes the temporal data using a sliding window approach that enhances the ability to discern local patterns within long input sequences[54]. This technique allows for the integration of the last observed data or transitions to the decoder's self-generated predictions, supplemented with temporally encoded future conditions. During training, a dynamic teacher forcing strategy is employed, in which actual target outputs are used as inputs in lieu of previous predictions to promote model convergence, prediction fidelity, and generalizability in the model. This hybrid training strategy allows effective learning from the ground truth while gradually becoming equipped for self-guided predictions. At the end of the processing of this sequence, quantile-based predictions are collected into a stack of tensors, encapsulating a comprehensive forecast for subsequent decision-making processes. Thus, this forward pass provides a fine-grained, probabilistic understanding of the evolving battery life-cycle stages, with the potential to inform risk assessment and optimize operational efficiency.

Experimental configuration

This study evaluates the ARCANA architectural model through a two-stage experimental process. Our aim is to present findings that resonate across multiple disciplines, highlighting both the complexity and versatility of our approach. The first stage involved training model M with the coin cell dataset B from BASF. The resulting trained model is here denoted $M(B)$. We encoded predetermined parameters, including cycle number and discharge current, into temporal segments to capture past and future discharge conditions. The training used an additive attention mechanism in the ARCANA architecture for initial learning, with a detailed explanation in Section "Methods". In the second stage, the model M is re-trained from scratch (parameters available in Supplementary Table 1), with publicly available datasets as mentioned in Table 1 and denoted as $M(P)$. This entails various cell types, including 26 coin cells and 6 prismatic cells with Lithium–Cobalt–Oxide (LCO) cathodes, with the majority being cylindrical cells with Lithium–Iron–Phosphate (LFP), Nickel–Manganese–Cobalt (NMC), and Nickel–Cobalt–Aluminum Oxide (NCA) cathode materials. To address these cell chemistry variations, we introduced an additional predefined parameter, the nominal capacity of each cell in logarithmic format. This inclusion was critical for the model to effectively differentiate and interpret response characteristics[56]. The public dataset selected for $M(P)$ was significantly smaller, comprising 627 cell entries and accounting for only 3.35% of the total data size of the initial model $M(B)$. The dataset was distributed with 65% for training, 30% for validation, and 5% for testing.

To emphasize generalizability and test model performance, we incorporated four distinct test datasets, each sourced from different locations and created by various experts. The first two test sets, denoted (D_{LNO}) and D_{NMC}, comprise coin cell measurements made at the Institute of Physical Chemistry (IPC) of KIT, featuring the Lithium-Nickel-Oxide (LNO) and NMC materials, respectively. The third dataset consisted of cylindrical cells from the Institute of Applied Materials (IAM) of KIT, containing NMC blended with NCA cathode materials ($D_{NMC+NCA}$). The final dataset involved prismatic cells of the CALCE institute with LCO materials (D_{LCO}). The complete description of these cells is provided in the Supplementary Section 1. This approach in dataset selection and testing allowed an in-depth evaluation of $M(P)$ for its adaptability to various cell types and experimental setups.

The publicly available data for $M(P)$ presented distinctive challenges, as they included prematurely failed cells and high experimental noise, in contrast to the high-quality data used for training $M(B)$. These complexities required a change from an additive to a multihead attention mechanism in $M(P)$. We also encountered a wide range of cycles, from as few as 196 to as many as 19176. However, most of the tests we considered had fewer than

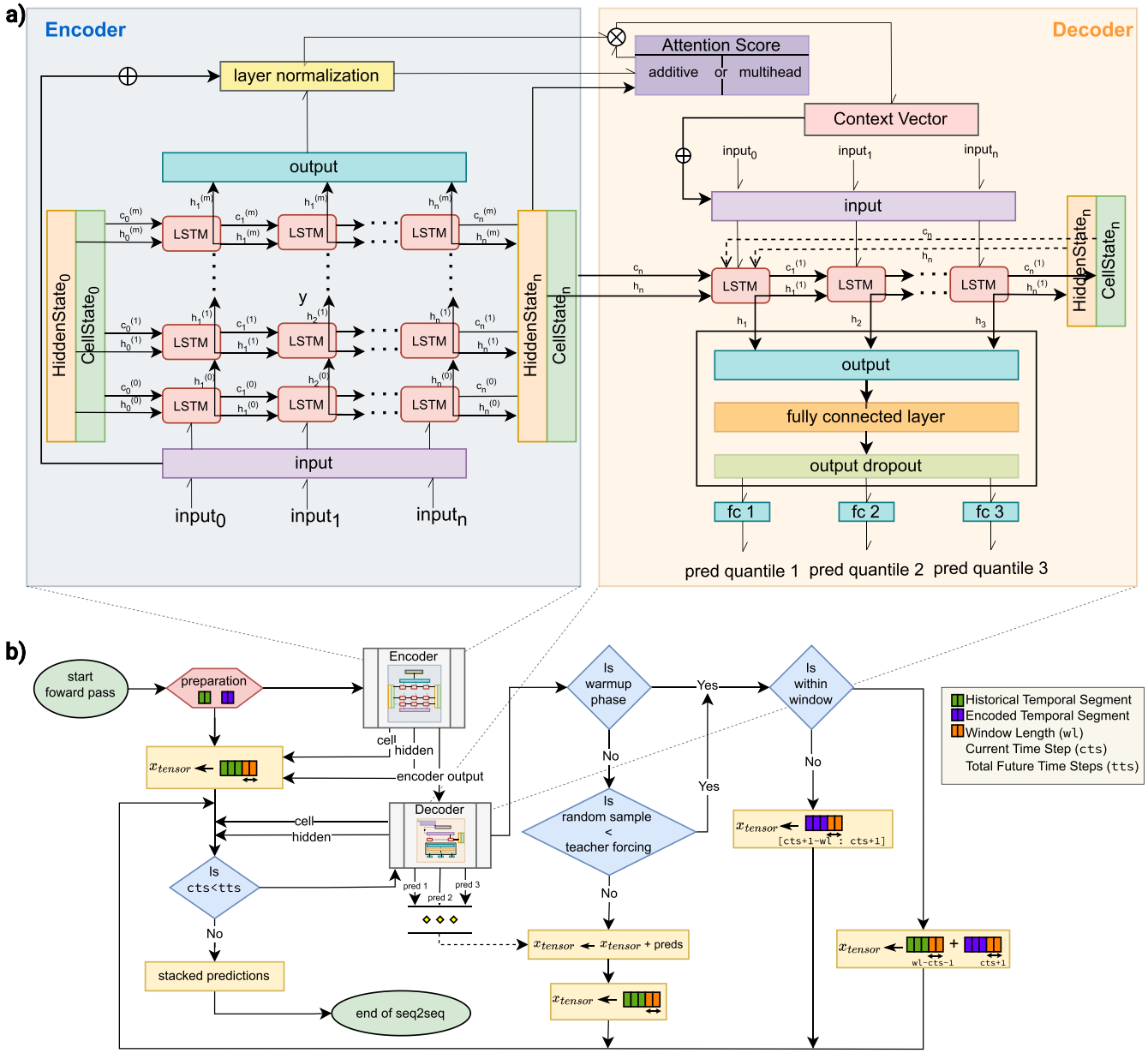

Fig. 2 | Architectural overview of Seq-to-Seq model. In this overview subfigure **a** depicts the detailed architecture of the encoder and decoder components. The LSTM-based encoder processes historical temporal segments to capture the intricate pattern of battery life cycles. It integrates skip-connection and layer normalization to preserve and stabilize essential key temporal features. The decoder is initialized with the encoder's final states and applies an attention mechanism to focus on relevant temporal features from the encoder output and enrich the context of its predictions. The attention-enhanced representations are combined with the initial decoder input and subsequently propagated through LSTM layers. A fully connected layer with leaky ReLU activation and a dropout layer—used solely during training and inactive during inference—for regularization follow the LSTM outputs. The model outputs are then fed into three separate fully connected layers for predicting a specific quantile of the future distribution based on the pattern learned during training, thus providing a probabilistic characterization of the forecast. Subfigure **b** illustrates the integrated Seq-to-Seq model flow, depicting the progression from encoding historical data to multi-output future forecasts. It highlights the sliding-window approach that underpins the model's capability to handle both the tail-end of historical data and the integration of self-generated forecasts with known future conditions. This process also captures the dynamic training process, which incorporates teacher forcing to enhance the predictive fidelity of the model.

500 cycles. This variability posed a potential risk of gradient instability and inconsistent learning in the training process. To mitigate the risk of poor convergence and the possibility of overfitting, we adopted a standardization approach in which all cells were limited to a maximum of 500 cycles, ensuring better balance in the training data and reducing bias, thus increasing reliability.

Both $M(B)$ and $M(P)$ focused on predicting three parameters, which were selected for their established significance in the existing literature and their availability across the datasets. They included discharge capacity, crucial for understanding the (SOH)[3], CE, as emphasized in studies by Burns et al.[57,58] as the key to understanding the impact of electrode additives and electrode materials on battery long-term performance, and the voltage drop during the relaxation phase between charging and discharging cycles. The last parameter is less explored but, as described by e.g. Zhu et al.[59], it offers valuable insights independent of the charging process. This parameter is easily calculated from cycling data, even if the studies where the data originated did not directly measure it. In this section, we evaluate our model's performance on various scenarios, focusing on the impact of data quality on model generalization and interpretability, investigating its adaptability to different chemistries, and deriving insights from attention mechanisms and saliency analysis.

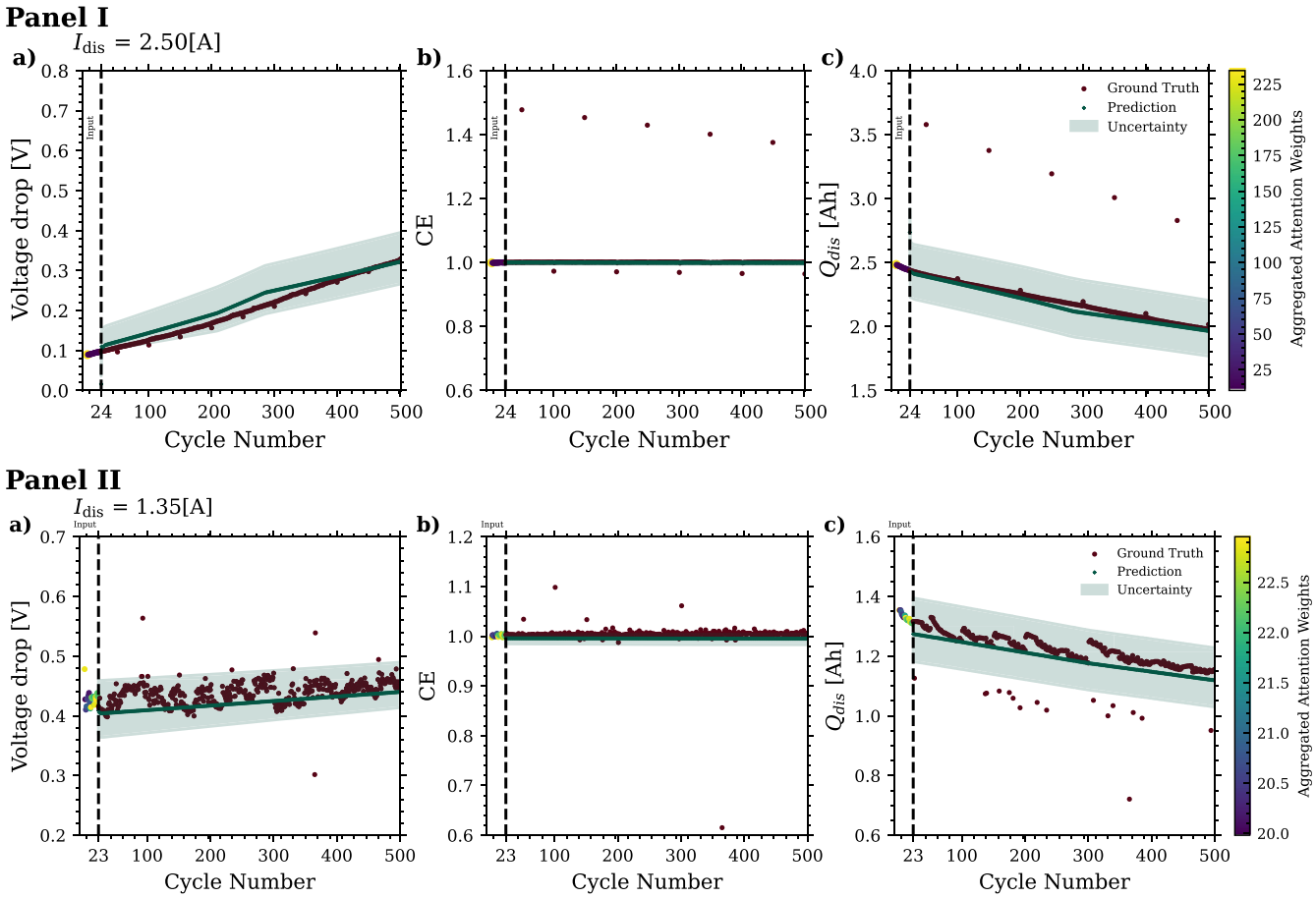

Fig. 3 | ARCANA's predictive performance on cylindrical sample cells. The performance of the proposed framework on two unseen datasets, namely cylindrical $D_{NMC+NCA}$ in Panel I and prismatic D_{LCO} in Panel II, when predicting battery behavior over 500 cycles for three predictors of Voltage drop [V] (**a**), CE (**b**) and Q_{dis} [Ah] (**c**). The uncertainty at the 10th and 90th percentiles effectively captures underlying data variability and highlights the model's predictive reliability and adaptability across diverse unseen datasets, demonstrating deep insight into data characteristics.

Model performance across battery types

The hyperparameters of $M(P)$ were selected using Optuna's hyperparameter tuning with 250 trials and are described in Supplementary Fig. 2, along with its training performance (Supplementary Fig. 3). The model generalization is evaluated on two datasets; cylindrical cells of $D_{NMC+NCA}$ and prismatic cells of D_{LCO}, neither of which were seen by the model during training. Here, the objective was to determine how effectively the model generalizes across different battery configurations despite the presence of noisy data.

As shown in Fig. 3, the model handles multidimensional predictions for both $D_{NMC+NCA}$ and D_{LCO} well. For $D_{NMC+NCA}$, it accurately forecasts up to 500 cycles based on 24 input cycles (see Panel I, Fig. 3) even though the extracted data exhibits occasional jumps despite the discharge current remaining constant throughout. Given that these unexpected jumps are not annotated in the original dataset, we have chosen to acknowledge their presence but not alter them for the sake of data integrity. Aggregated attention weights in early cycles indicate their importance for long-term forecasting. Emblematic is D_{LCO}, which starts from a 23-cycle profile (Panel II, Fig. 3); the model demonstrates robustness even in the presence of more complex noise patterns. Here, the attention weights are distributed not only in the initial cycles but also in later cycles, proving the necessity of incorporating an attention mechanism. Illustrating the model's generalization capabilities, a detailed analysis of Q_{dis} in Fig. 4 is presented. In both $D_{NMC+NCA}$ and D_{LCO}, there is good agreement between the model's predictions and actual values (Panel I & II, Fig. 4a), as complemented by the density graphs in Fig. 4b. For $D_{NMC+NCA}$, the predicted and actual densities closely overlap. For D_{LCO}, the predicted density is highly similar, with a slightly skewed distribution towards lower Q_{dis}. The better density distributions for $D_{NMC+NCA}$ are likely attributable to the larger proportion of cylindrical cells in the training data, which accounts for 94.9% of the total. A detailed evaluation of the uncertainty of the model $M(P)$ is provided in Fig. 4c–e for both datasets. Panel I & II of Fig. 4c evaluate the calibration by comparing the observed quantile proportions to the expected proportions under the assumption of a normal distribution. This continuous curve indicates the model's general performance across the entire probability distribution. The miscalibration area, quantified by the degree of deviation from the ideal diagonal line, represents the aggregate of discrepancies[60]. For $D_{NMC+NCA}$, the predicted distribution of Q_{dis} is well calibrated around the median but diverges at the tail, with calibration points showing underconfidence at higher quantiles. For D_{LCO}, the individual calibration points suggest a slight overconfidence in the 10th–50th percentile and underconfidence in the ranges 50th-90th and 10th-90th percentile. The miscalibration area for D_{LCO} is 0.16, which is slightly higher than $D_{NMC+NCA}$, likely due to noisier data. The overall calibration performance across both datasets is comparable. Figure 4e) shows a histogram of prediction interval quantiles, revealing the spread between the 10th and 90th percentiles and evaluating the concentration of its predictive distribution as indicated by sharpness. The lower values suggest higher confidence in the prediction[61]. For $D_{NMC+NCA}$, a bimodal distribution highlights variable prediction certainty across cycles, suggesting potential fluctuations in battery behavior. D_{LCO} shows two clusters of distributions, mostly around a central quantile with a sharpness of 0.19, indicative of consistent uncertainty. Figure 4d further supports these findings by illustrating the model's median prediction uncertainty and the variability of these predictions by interquartile range

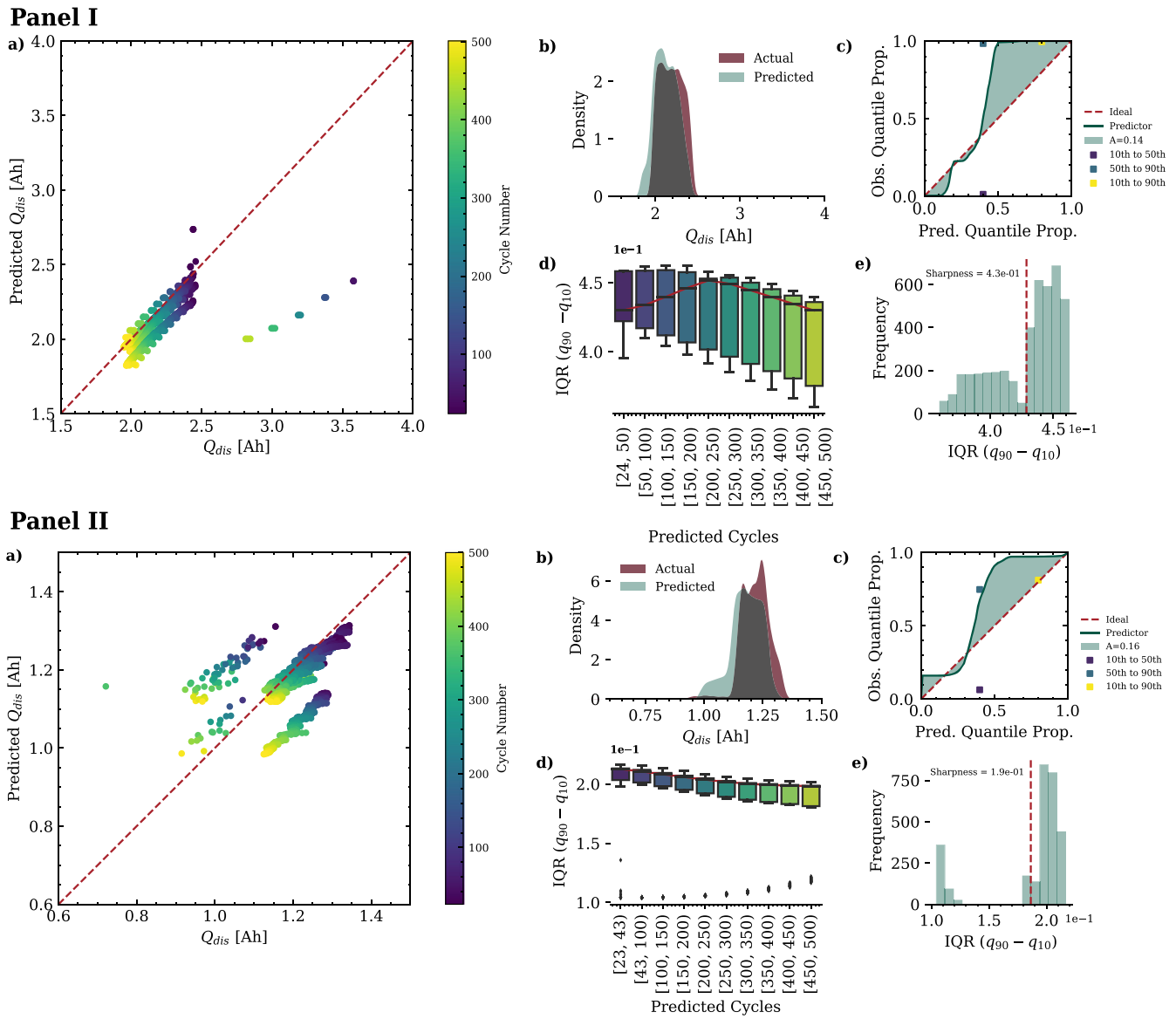

Fig. 4 | Comparative analysis of model predictions and its uncertainty and calibration for Q_{dis} in cylindrical sample cells. Analytical comparison for Q_{dis} for two datasets; $D_{NMC+NCA}$ (Panel I) and D_{LCO} (Panel II), where **a** depicts the relationship between predicted and actual values of Q_{dis}, with the diagonal dashed line indicating perfect prediction accuracy, **b** illustrates the density distributions of predicted versus actual Q_{dis}. The calibration plot in **c** assumes a normal distribution, where the mean and standard deviation are estimated from the 10th, 50th, and 90th percentiles of predictions. It depicts the cumulative proportion of actual Q_{dis} values that fall at or below the predicted quantile values rather than within symmetric intervals around the predictions. The ideal diagonal line represents perfect calibration with the shaded area indicating the degree of miscalibration, denoted *A*. The approximately diagonal trend of the calibration line up to the 0.5 quantile shows that data with residuals below the median are well described by the predictive distribution. The jump from 0.5 to 1 indicates that the predictive distribution extends further to positive values than the observed distribution of residuals; almost all test data are already covered by the predicted 0.6 quantiles for both datasets. However, the overall miscalibration areas for both datasets are quite similar, indicating that despite different patterns of over- and underconfidence at specific quantiles, the general calibration performance across both datasets is comparable. Box plots at **d** show the prediction intervals over multiple cycles, demonstrating the median and variability of the model prediction uncertainty over the battery's lifespan. **e** provides histograms that depict the quantile-based prediction interval width between the 10th and 90th percentiles as a measure of sharpness. The red dashed line indicates the sharpness as the mean interval width and shows the concentration of the predictive distributions that indicate narrower distribution and, consequently, higher confidence in predicting Q_{dis} for $D_{NMC+NCA}$ in Panel I. Further comparisons are in Supplementary Figs. 7, 8, 10, and 12.

(IQR). Here, $D_{NMC+NCA}$ in Panel I shows varying IQR, suggesting changes in model confidence over the lifespan. In contrast, D_{LCO} maintains a more uniform IQR, indicating steady prediction uncertainty and aligning with the model's attention on later cycles to contend with the increased complexity and noise. These metrics complement the information provided in Fig. 4c–e, serving as a benchmark for the model's reliability and its capacity to generalize within a precise estimate range.

The multi-output predictive capabilities of $M(P)$ are further highlighted by its performance in predicting the second parameter, voltage drop (Supplementary Fig. 4). The model exhibits strong prediction accuracy with both datasets. $D_{NMC+NCA}$ shows a smaller range of predictions over increasing cycles, and D_{LCO} shows a stable range with decreasing median intervals, while the calibration accuracy and the reliability of the predictions remain high across both datasets. The performance on the third predictor, CE (Supplementary Figs. 6 and 11), shows consistency and low prediction uncertainty, although the high measurement noise present in this dimension poses a challenge and makes convergence more demanding[62]. Additional examples are shown in Supplementary Figs. 5 and 9. The evaluation

metrics for $M(P)$ (Supplementary Table 2) demonstrate its predictive strengths for both $D_{NMC+NCA}$ and D_{LCO}. For the D_{LCO} dataset, the voltage drop is predicted with a root mean square error (RMSE) of 0.0335 and a mean absolute percentage error (MAPE) of 6.6052. However, $D_{NMC+NCA}$ outperforms CE with significantly lower error rates of 0.0256 and 0.2489 for the RMSE and MAPE, respectively. However, both datasets present higher error rates in the predicted discharge capacity. To counteract the impact of systematic noise, Median Absolute Error (medAE) is used along with MAE for a more robust error analysis. These metrics highlight $M(P)$'s versatile predictive capabilities in handling diverse dataset requirements for multiple features and long-term predictions[4,63].

We further examine $M(P)$'s performance on unseen coin cell datasets, D_{LNO} and D_{NMC}. The model predicts the voltage drop and CE well but shows limitations and high uncertainty when predicting the discharge capacity with an RMSE of 0.5827. This may stem from the low representation of coin cells in the training data, just 4.1% of the total. To alleviate this problem, we fine-tuned the decoder weights of $M(P)$ using the data of 17 coin cells from D_{LNO}, resulting in an updated model, $M(P)_f$. This fine-tuning process and training performance are detailed in Supplementary Figs. 13 and 14 and led to a substantial improvement in predicting Q_{dis}, dropping the RMSE to 0.0002, indicating a significantly enhanced precision. $M(P)_f$'s performance will be compared with $M(B)$, trained with the BASF dataset B, in the following section.

Model performance on coin cell data for generalization insights

While comparing the predictive performance of models $M(B)$ and $M(P)_f$ on subsets of unseen D_{LNO} (Supplementary Figs. 15 and 20) and D_{NMC} dataset (Supplementary Figs. 21 and 23), $M(P)_f$ demonstrates reliable predictive alignment for voltage drop, CE, and Q_{dis}. In contrast, $M(B)$ shows a divergent pattern in voltage drop predictions, which may be due to its training on data with inherently long relaxation time profiles compared to those in D_{LNO}, where measurements are taken shortly after state changes. However, it maintains consistency in CE predictions and adjusts Q_{dis} predictions in response to changes in the test protocol.

In our analysis of D_{LNO} for Q_{dis}, Fig. 5 demonstrates that $M(P)_f$ achieves high predictive fidelity. This is evident from the dense alignment of the predictions with the actual values in the scatter plot (Fig. 5a), and the significant overlap in distributions seen in the density plot (Fig. 5b). The model's precision is further highlighted by concentrated prediction intervals and a calibration curve that closely traces the diagonal (Fig. 5c–e). It achieves a high proportion of data points within the predictive bounds, indicative of accuracy, without excessively wide intervals that could decrease the utility of the predictions. Panel II for $M(B)$ also demonstrates a close tracking of the actual values, with a marginally broader prediction interval and higher miscalibrated area of 0.16 compared to $M(P)_f$'s of 0.022 (Panel I). Despite this variance, $M(B)$ maintains a reasonable estimate range. Qualitatively (Table 2), $M(P)_f$ achieves a lower MAPE (9.2285) for predicting voltage drop, indicating its capability for learning trends commonly observed in training datasets with short relaxation times during cycling. On the other hand, the $M(B)$ model demonstrates a notably lower MAPE in Q_{dis} (8.8914), showcasing its superior ability to capture proportional changes across a broader dataset. This performance illustrates the impact of prior knowledge and training data diversity on the learning outcomes of the models. Detailed analyses of additional predictive dimensions for D_{LNO} for both models and the complete dataset D_{NMC} are available in Supplementary Figs. 16–19, 22, 24, 25 and Supplementary Table 3. Despite the D_{LNO} data originating from another institute, the generalization of $M(B)$ highlights the potential of well-trained DL models to overcome the variability of data sources.

Adaptive chemical modeling

ARCANA has so far been demonstrated to generalize well across battery formats, electrolyte formulations, cathode chemistries, and cycling procedures for LIBs. The ultimate generalization would be achieved if the model could also be deployed to Na-ion batteries. Since the underlying degradation mechanism of Na-ion batteries is very different, we performed fine-tuning to test the adaptability of $M(B)$ and $M(P)$ to this distinct chemical domain[30,64]. These fine-tuned models are denoted $M(B)_{Na}$ and $M(P)_{Na}$, and are trained on Na-ion cycling data with CC-CV and pulse discharge settings. Details on the fine-tuning parameters and training performance for both models are available in Supplementary Figs. 26–29.

In Figs. 6 and 7, we evaluate the fine-tuned $M(B)_{Na}$ and $M(P)_{Na}$ models on an unseen C-rate test protocol (Figs. 6a and 7a). Both models demonstrate flexibility in adjusting to changes in C-rates, with voltage drop, CE, and Q_{dis} depicted in Figs. 6b–d and 7b–d. The model $M(B)_{Na}$ shows narrower prediction intervals, indicative of lower uncertainty and greater predictive robustness. This trend is consistent across all predictive dimensions, and the model is probably benefiting from the larger initial dataset on which it was trained, since it provided a richer learning environment for the model to become more 'protocol-agnostic'. Its precision is especially notable in predicting the voltage drop and CE estimations, closely following the ground truth despite the substantial experimental noise. The aggregated attention mechanism in $M(B)_{Na}$ (Fig. 7d) also appears more fine-tuned, with greater weights on the latest cycle data, which is consistent with its precise predictions. While $M(P)_{Na}$ is adaptable, it shows a marginally wider uncertainty (Fig. 6b–d).

Sensitivity analysis, as shown in Figs. 7e–g and 6e–g evaluates the input parameter influence on future predictions for $M(B)_{Na}$ and $M(P)_{Na}$. Both models demonstrate increased sensitivity to the most recent input data, i.e., cycles 7–9 in this provided example, aligned with their attention distributions, with cycle 9 receiving the highest attention. This increased emphasis on the last input cycles corresponds to the rapid degradation patterns in this sodium coin cell. As the model receives each successive cycle, the most recent data, here in cycle 9, becomes important in shaping its predictions, allowing the model to more accurately predict ongoing trends.

In Fig. 7, $M(B)_{Na}$ shows a greater overall sensitivity across input cycles, particularly for the dimensions of voltage drop and Q_{dis}. This is further illustrated in sensitivity profiles and cumulative plots (Fig. 7h–j), highlighting a refined input-response relationship and a lower uncertainty interval in the primary prediction (Fig. 7a–c). Such a distinct sensitivity indicates $M(B)_{Na}$'s ability to precisely identify and respond to subtle variations. Despite the high experimental noise and limited battery performance, the saliency and attention trends of both models remain remarkably similar. This suggests that both mechanisms are intrinsic to the model's architecture, enabling them to perform consistently in diverse scenarios.

To further substantiate our initial findings, the plots in Fig. 8, show both models' Q_{dis} predictions aligning well with the ground truth. $M(P)_{Na}$ exhibits a tighter clustering around the actual values, while $M(B)_{Na}$ exhibits a broader spread. The prediction intervals and the distribution of quantiles across the 10th and 90th percentile for both models confirm their consistency and calibrated confidence. Further assessments are found in Supplementary Figs. 30–32 and Supplementary Table 4. These evaluations provide insights into the model's robustness. The performance of $M(B)_{Na}$'s especially underscores the advantage of extensive and diverse pretraining datasets in enhancing model generalization across different battery chemistries.

Discussion

We demonstrated the chemistry-, format- and cycling procedure-agnostic ARCANA framework and its ability to reliably monitor battery life and SOH by utilizing multitask learning with an attention mechanism. ARCANA excelled across three predictive settings, demonstrating that augmenting the model with diverse knowledge streams enhances its generalization across virtually all variations possible in batteries, such as anode, cathode, electrolyte, and shuttle ion chemistry and format. The ARCANA model integrates uncertainty quantification and attention mechanisms for each and every cycle to elucidate the model's focus for each prediction and is essential for uncovering complex patterns associated with multiple factors. Further evaluation involves saliency and sensitivity assessments, allowing us to understand the impact of perturbation of input parameters on output predictions. By examining whether saliency and attention are directly

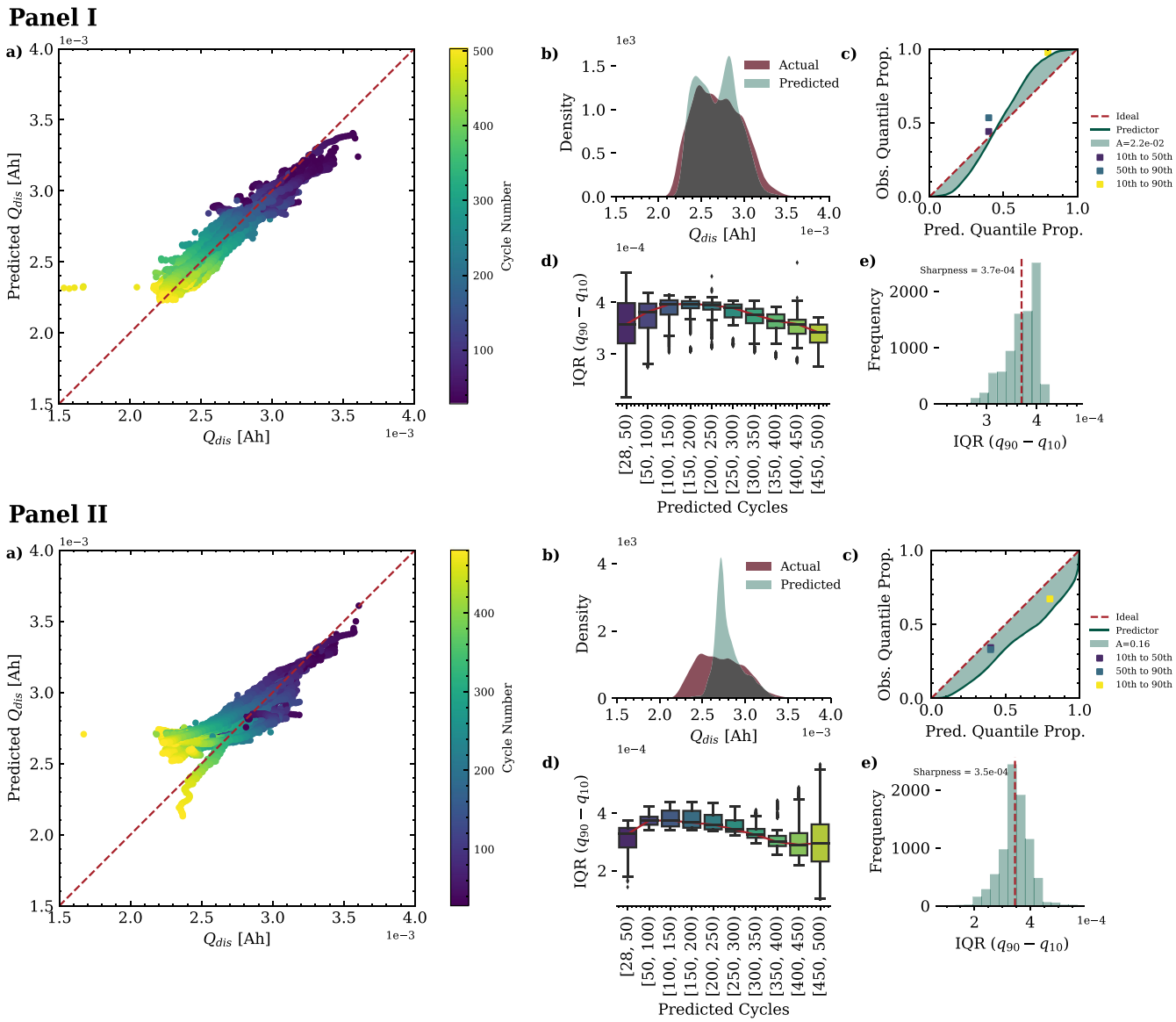

Fig. 5 | Performance analysis of $M(P)_f$ and $M(B)$ for Q_{dis} in coin sample cells. Performance of $M(P)_f$ (Panel I) and $M(B)$ (Panel II) on D_{LNO} for Q_{dis} prediction. Plot **a** illustrates the relationship between models' predictions and the actual Q_{dis} with the diagonal line representing perfect prediction accuracy, plot **b** compares the density distribution of actual and predicted Q_{dis}, plot **c** presents calibration curves that reflect the degree of alignment between predicted probabilities and observed frequencies under a normal distribution assumption. The discrete points on the calibration curve show the observed proportions of actual values that fall within three specific intervals based on the quantiles: between the 10th and 50th, 50th and 90th, and 10th and 90th percentiles. Model $M(P)_f$ shows a high level of calibration for predicting Q_{dis} of D_{LNO} samples with a minimal miscalibrated area of 0.022. The points for the 10th, 50th, 50th, and 90th percentiles lie close to the diagonal line, indicating a nearly perfect calibration for these intervals. $M(B)$ exhibits a slight overconfidence by deviating from the ideal line, with a miscalibration area of 0.16. The three calibration markers for $M(B)$ are all positioned just below the diagonal line, showing uniform overconfidence across these quantile ranges, yet they remain close to this line, indicating a generally well-calibrated model. Plots **d** show the prediction intervals across lifespan cycles, highlighting models' uncertainty over time, and plot **e** details the distribution of prediction intervals' quantiles between the 10th and 90th percentiles, which convey the models' prediction uncertainty; a distribution skewed towards the lower quantiles suggests a higher confidence in predictions at these quantiles. The sharpness, as a measure of mean interval width, is approximately similar for both models at 3.7×10^{-4} and 3.5×10^{-4} for $M(P)_f$ and $M(B)$, respectively. Together, these plots demonstrate the $M(P)_f$'s precision in capturing discharge capacity behavior and $M(B)$'s robust generalization.

correlated or orthogonal to each other, we gain a comprehensive understanding of input–output relationships, increasing the model's explainability and reliability in extrapolation. Incorporating raw data and failed experiments, as suggested in prior studies[4,36] is a deliberate strategy to teach our models to recognize variations across similar cell types and manufactures. This inclusion not only enables uncertainties to be quantified more accurately but also deepens reliability insights, reduces bias, and offers a more meaningful understanding of the data. A conceptually straightforward extension to this work would be to incorporate additional features, such as the rate of change of voltage with respect to capacity (dQ/dV)[34,65], and leverage different characterization methods, like spectroscopy, to enhance the predictive power of the models. This will not only enhance multi-feature predictions but also deepen the understanding of degradation processes[3,4,63].

We observed that $M(P)$, trained on public data, offers broader generalization across various battery types and protocols, albeit with increased uncertainty. $M(B)$, trained on a more extensive dataset, demonstrates a lower uncertainty. This further motivates the importance of data sharing and management. Our findings also reveal that fine-tuning the models with few labels significantly improves their generalization to different chemistries, especially for $M(B)$. The methodology outlined in this paper presents

Table 2 | Evaluation metrics for $M(P)_f$ and $M(B)$ using D_{LNO}

Metrics	$M(P)_f$			$M(B)$		
	Voltage drop [V]	CE	Q_{dis} [Ah]	Voltage drop [V]	CE	Q_{dis}[Ah]
RMSE	0.0703	0.0331	0.0002	0.1247	0.0588	0.0003
MAPE	9.2285	1.1922	20.7946	34.8638	4.4560	8.8914
MAE	0.0353	0.0076	0.0001	0.0867	0.0335	0.0002
medAE	0.0181	0.0021	0.0001	0.0513	0.0104	0.0001

Summary of the evaluation metrics for $M(P)_f$ and $M(B)$, tested on 26 unseen coin cells (D_{LNO}), using 27 initial cycles of historical data, to predict the cell behavior up to the 500th cycle. Note that the number of initial cycles was chosen randomly to resemble practical scenarios with limited initial data. The user can specify any preferred number of initial cycles in the provided configuration file, which is detailed at https://github.com/basf/ARCANA/blob/master/config/.

an opportunity for other researchers to create their own high-performance models. By retraining or fine-tuning with different datasets, researchers can tailor these predictive models to their specific experimental setups and desired outcomes. This flexibility allows for the exploration of different perspectives and approaches, facilitating the development of more accurate and specialized models. One could envision a model-sharing and transfer-learning community similar to those found today in the fields of computer vision and language modeling. Furthermore, the performance metrics explored here raise the tantalizing prospect of further improving model quality via a federated learning approach. This could enable researchers from diverse backgrounds and institutions to pool their data and expertise, leading to more powerful models.

The modular design of the ARCANA pipeline enables real-time monitoring of battery degradation profiles, promoting timely and cost-effective interventions. This proactive approach prevents prolonged sub-optimal testing conditions, improves the R&D process, and contributes to more informed material selection and protocol optimization. By automating data collection, processing, and analysis, researchers can streamline their experimental workflows and reduce human error. Furthermore, ML models can continuously learn from upcoming data, adapt to evolving experimental conditions, and provide real-time insights. This integration of ML and laboratory workflows has the potential to transform battery research, enabling researchers to make data-driven decisions, uncover insights more rapidly, and accelerate the pace of discovery.

Overall, we demonstrated that incorporating multitask learning with an attention mechanism creates a framework that can achieve chemistry agnosticism as envisioned by Battery 2030+[1] and the interesting fact that a DL architecture trained on a smaller, noisier, but more diverse dataset yields better generalization at the cost of higher uncertainty. We hope that the pipeline will emerge as an indispensable and transformative tool to bridge the gap between lab-scale research and commercial viability and will become essential for the development of applications and insightful predictive models in the energy storage field.

Methods

In the following section, some of the key components of the ARCANA framework are explained to underscore their contribution to the overall efficacy and reliability of the model. This includes an exploration of attention mechanisms, a teacher forcing scheduler, methods to quantify predictive uncertainty, a strategic early stopping protocol, a training procedure, and evaluation metrics.

Attention mechanism

Within the proposed ARCANA framework, two distinct attention mechanisms are implemented. The first, termed additive attention, is also known as Bahdanau attention[55]. This mechanism aligns the hidden state of the decoder h_t at each time step t with the hidden states of the encoder (h_s), thus producing a context vector that encapsulates the weighted relevance of each historical temporal segment from the past cycles. This vector provides a dynamically focused representation of the input sequence pertinent to the current decoding step. This mechanism is functional through a parameterized attention model. The model calculates an attention score e_{ts} (Eq. (1)) for each encoder state h_s given by:

$$e_{ts} = v^T \tanh(W_1 h_t + w_2 h_s) \tag{1}$$

where W_1 and W_2 are the weight matrices that transform the respective hidden states into a common feature space and v is a weight vector that projects the activated sum into a scalar score. Attention weights α_{ts} are then determined by normalizing these scores using the softmax function (Eq. (2)):

$$\alpha_{ts} = \frac{\exp(e_{ts})}{\sum_{k=1}^{T_e} \exp(e_{tk})} \tag{2}$$

here, T_e is the total number of time steps in the encoder sequence.

The context vector c_t results from aggregating the encoder hidden states, each weighted by its respective attention weight, as can be seen in Eq. (3), and can improve the model's capacity for handling Seq-to-Seq predictions[66].

$$c_t = \sum_{s=1}^{T_e} \alpha_{ts} h_s \tag{3}$$

Another attention mechanism that can be employed within the ARCANA architecture is multihead attention. This mechanism expands the model's capacity to focus on different positions of the input sequence simultaneously[67], which is crucial for capturing a wider range of dependencies inherent in battery lifetime data. This attention mechanism operates by projecting the decoder's hidden states and the encoder outputs, representing the past cycle's information, into multiple subspaces. This is formulated as: (Eq. (4))

$$\text{MultiHead}(Q, K, V) = \text{Concat}\left(\text{head}_1, \ldots, \text{head}_h\right) W^0 \tag{4}$$

$$\text{head}_i = \text{Attention}\left(QWi^Q, KWi^K, VWi^V\right) \tag{5}$$

where each head (head_i) captures different aspects of the input data and is computed as shown in Eq.(5). The operation applied in each head is defined by the attention of the scaled dot product and is presented in Eq. (6).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{6}$$

Here, Q, K, and V are the query, key, and value matrices, respectively. Q is generated from the hidden states of the decoder, while K and V are derived from the encoder outputs. This arrangement enables the decoder to integrate the current state information with historical data provided by the encoder. The parameter matrices W_i^Q, W_i^K, and W_i^V for each head i, along with the output weight matrix W^0, are optimized during the training process. These matrices are instrumental in transforming the input data into different representational subspaces to capture various aspects and dependencies within the data. The parameter d_k, representing the dimension of the key vectors, scales the dot product within the attention mechanism. In Eq. (6), the softmax function is applied to these scaled attention scores, which originate from the interactions between the query and key matrices. This process results in the production of a context vector, which integrates information from different representational subspaces and allows the model to consider multiple aspects of historical data[54,68].

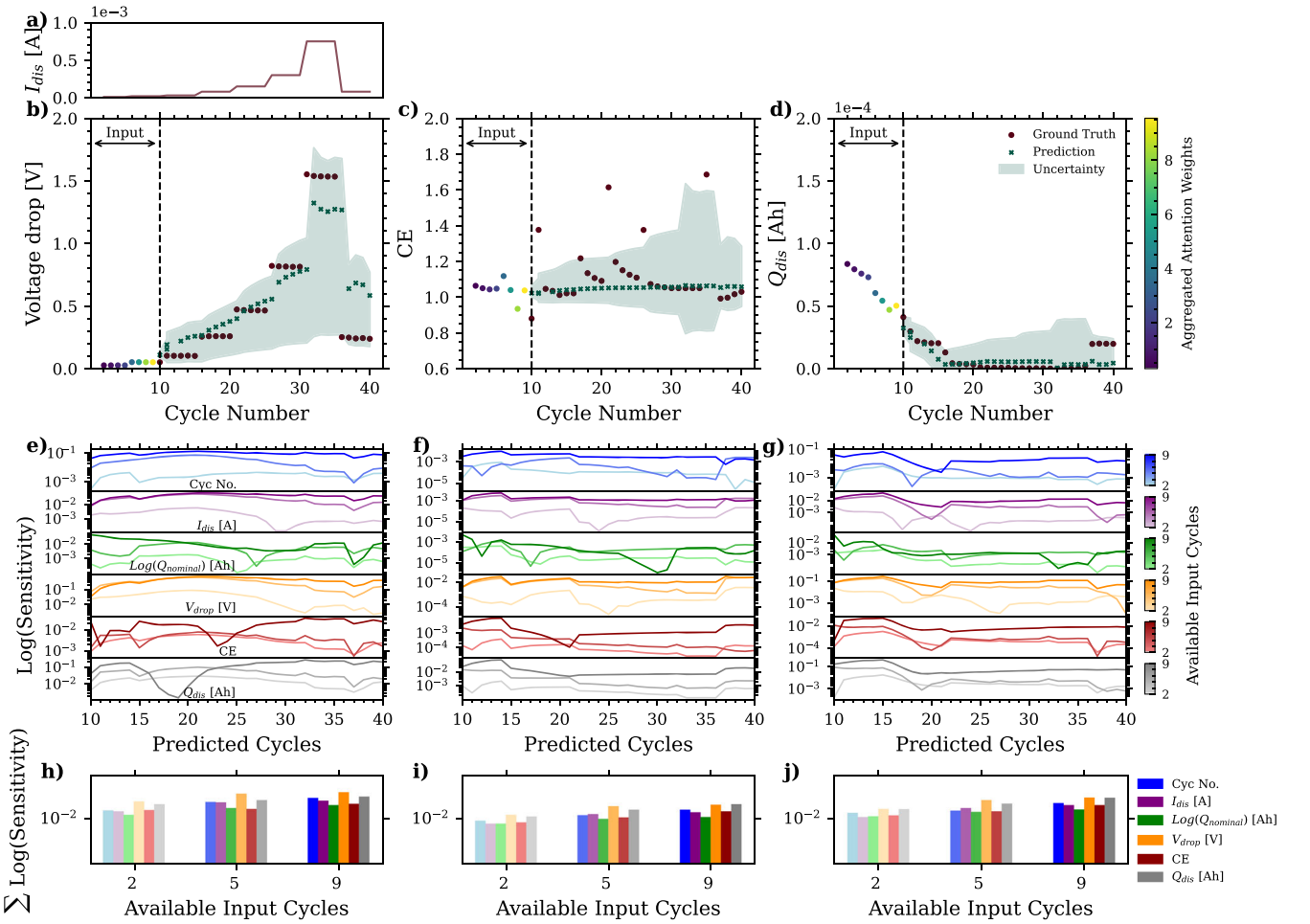

Fig. 6 | Analysis of $M(P)_{Na}$'s predictive accuracy and input sensitivity on Na-ion data. Plot **a** presents the C-rate profile for cycling one battery, while plots **b**–**d** compare the model's prediction to actual data, showing consistency and adaptability. Sensitivity to input parameters across predicted cycles is analyzed in plots **e**–**g** on a logarithmic scale. The color intensity in these plots denotes the specific cycles from which the input parameter originates. Plots **h**–**j** show the sum of the logarithmic contribution of each input parameter towards predicting future cycles with a selective representation of three past cycle data. These visualizations confirm the model's attentive adjustment to the latest available input data and its capacity for generalization, despite the high experimental noise and limited battery performance.

Teacher forcing

Teacher forcing optimizes the learning of temporal dependencies. By integrating the real data from previous time steps, the technique promotes rapid stabilization and convergence of the model. In the present study, the implementation of the teacher forcing strategy is applied through a calculated division of training epochs. This division is reflective of the model's incremental improvement in processing sequences with varying lengths over time by prioritizing shorter sequences at the early stages of training to ensure intensive guidance. This preferential focus ensures that the model does not prematurely plateau when learning to predict longer-term dependencies.

To quantitatively define this approach, the training period consisting of E epochs is divided into D equal segments s. Within the i-th segment, the teacher forcing ratio is adjusted through a decay parameter λ, which represents how quickly the training procedure switches from using real data as decoder inputs to using model predictions from the previous cycle, as depicted in Fig. 2b. The allocation of epochs per division d_i is calculated as can be seen in Eq. (7)

$$d_i = \text{round}\left(\frac{s \cdot e^{-\lambda i}}{\sum_{j=0}^{D-1} s \cdot e^{-\lambda j}} \cdot E\right) \tag{7}$$

Following this, the teacher forcing ratio for the t-th epoch in the i-th segment is linearly reduced from a starting ratio R_{start} to an ending ratio R_{end}, using the following equation, Eq. (8).

$$\begin{aligned} A &= \left(\frac{R_{start} - R_{end}}{d_i + \epsilon}\right) \\ R_{t_i} &= R_{start} - A \cdot t \end{aligned} \tag{8}$$

Here, R_{t_i} indicates the teacher forcing ratio at epoch t for the ith segment. The expression A represents the decrease per epoch in that segment. To ensure numerical stability and avoid division by zero, a small constant ϵ, set to 10^{-8}, is included in the calculation as indicated in Eq. (8). The teacher forcing ratio, as a probabilistic measure, represents the likelihood that the model will utilize the actual observation from the training data at a given prediction step. This approach modulates the ratio to facilitate a smooth transition from guided to self-generated sequence prediction. The adjusted ratios are indicative of the model's learning trajectory, enhancing its independent predictive accuracy across different sequence lengths.

Uncertainty quantification

The pinball loss, in this study, provides a robust metric for predicting a range of potential outcomes, rather than a single point estimation. This is an effective measure for forecasting scenarios where the impacts of over-prediction and underprediction are asymmetric[69]. It is defined for a set of quantiles $Q = \{q1, q2, q3\}$ where $q1 < q2 < q3$ and in this study, we select

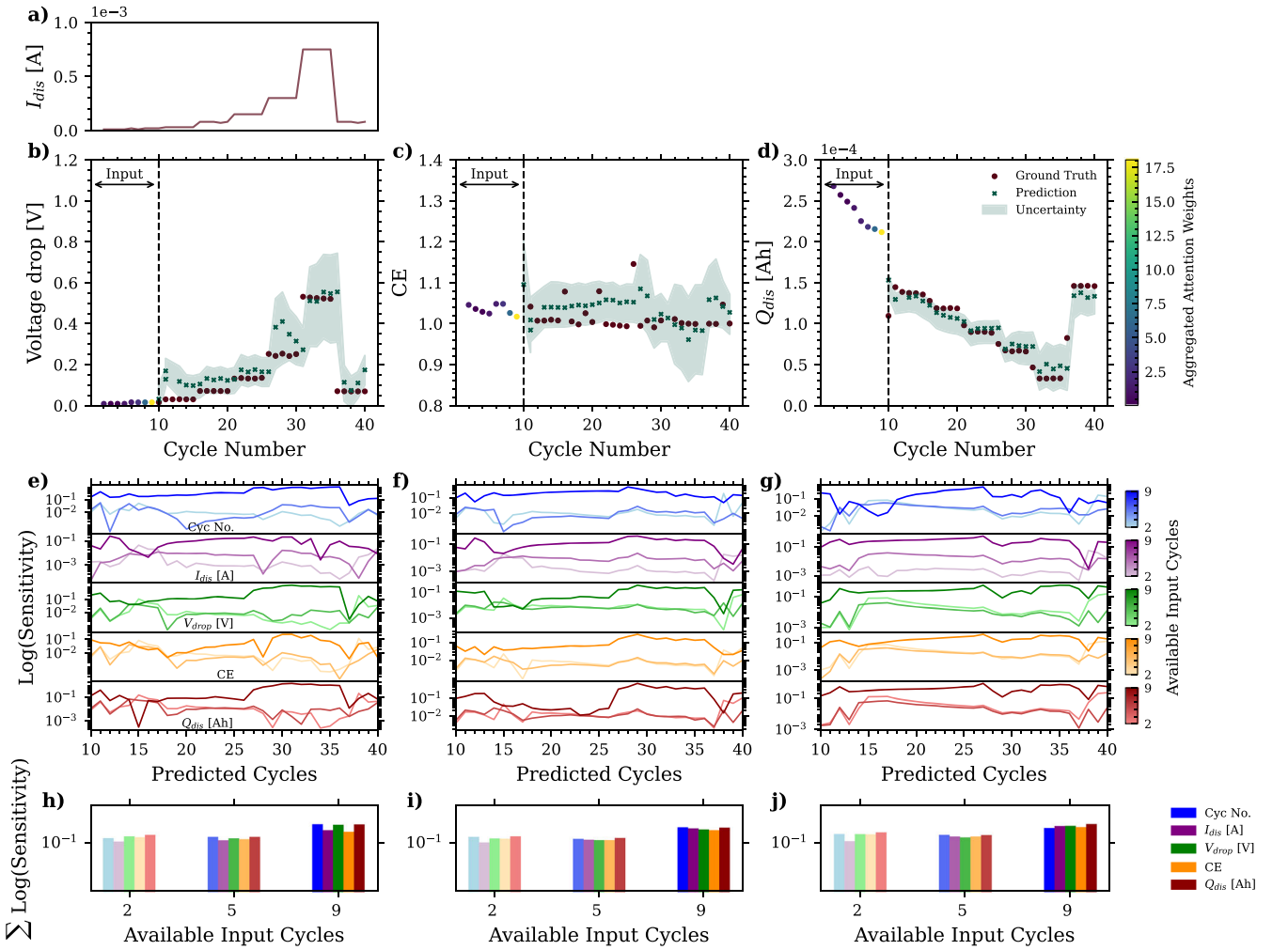

Fig. 7 | Evaluation of $M(B)_{Na}$'s predictive performance and input sensitivity on our in-house Na-ion data. Plot **a** shows the discharge current profile, while plots **b**–**d** depict the predictions for voltage drop, CE, and Q_{dis} against the ground truth. The color bar here shows the aggregated attention weights across the input data. Plots **e**–**g** provide a detailed logarithmic sensitivity analysis per predictive cycle for each input parameter, and plots **h**–**j** aggregate these sensitivities, highlighting the model's focus on different input cycles, especially the most recent ones, reflecting $M(B)_{Na}$'s protocol adaptability and robust response to experimental noise.

$Q = \{0.1, 0.5, 0.9\}$ corresponding to the 10th, 50th, and 90th percentiles, respectively. For a given predicted value $\hat{y}$ and the actual target value y, the pinball loss for a single quantile q is calculated as:

$$L_q(\hat{y}, y) = \begin{cases} (1-q)\cdot(\hat{y}-y) & \text{if } y < \hat{y} \\ q\cdot(y-\hat{y}) & \text{if } y \geq \hat{y} \end{cases} \tag{9}$$

In the implementation of this loss function, a mask is provided and applied to each quantile's loss to selectively evaluate certain predictions, allowing for the exclusion of outliers. The total pinball loss for multiple quantiles is then the sum of the individual losses for each quantile, averaged over all predictions, as shown in Eq. (10), reflecting the model's performance across the specified range of quantiles.

$$L(Q, \hat{Y}, Y) = \frac{1}{N}\sum_{i=1}^{N}\sum_{q\in Q} L_q(\hat{y}_{qi}, y_i) \tag{10}$$

Here, N is the number of observations, $\hat{Y}$ is a stack of vectors, with each vector containing the predictions for all observations at one of the specified quantiles, and Y is the vector of the true target values. Each element $\hat{y}_{qi}$ in $\hat{Y}$ denotes the predicted value for the ith observation at quantile q. This configuration not only facilitates efficient computation of the loss function across multiple quantiles and observations, but also captures the central tendency and variability of the predictions, making it a comprehensive loss function for probabilistic forecasting[69,70].

Early stopping

To optimize training, a rigorous early stopping approach is incorporated. This method was originally proposed by Prechelt et al.[71] and combines criteria to prevent overfitting while ensuring substantial training progress, especially in the presence of noisy data. Here, a dual-criteria strategy is implemented. The first criterion assesses the ratio between generalization loss (GL) and training progress, which is shown in Eq. (11), where E_{val} represents the validation error at the current epoch, $E_{min\ val}$ is the lowest validation error obtained up to the current epoch, and $E_{train\ strip}$ denotes the training errors within a recent sequence of epochs. This sequence, or strip, is a designated period in which the progress quotient (PQ) is measured. If the generalization-loss-to-progress-quotient-ratio (GL/PQ) surpasses a predefined value, it may indicate that further training will not be beneficial for the model's generalizability.

$$\begin{aligned} \text{GL} &= 100 \cdot \left(\frac{E_{val}}{E_{min\ val}} - 1\right) \\ \text{PQ} &= 1000 \cdot \left(\frac{\text{Mean}(E_{train\ strip})}{\text{Min}(E_{train\ strip})} - 1\right) \end{aligned} \tag{11}$$

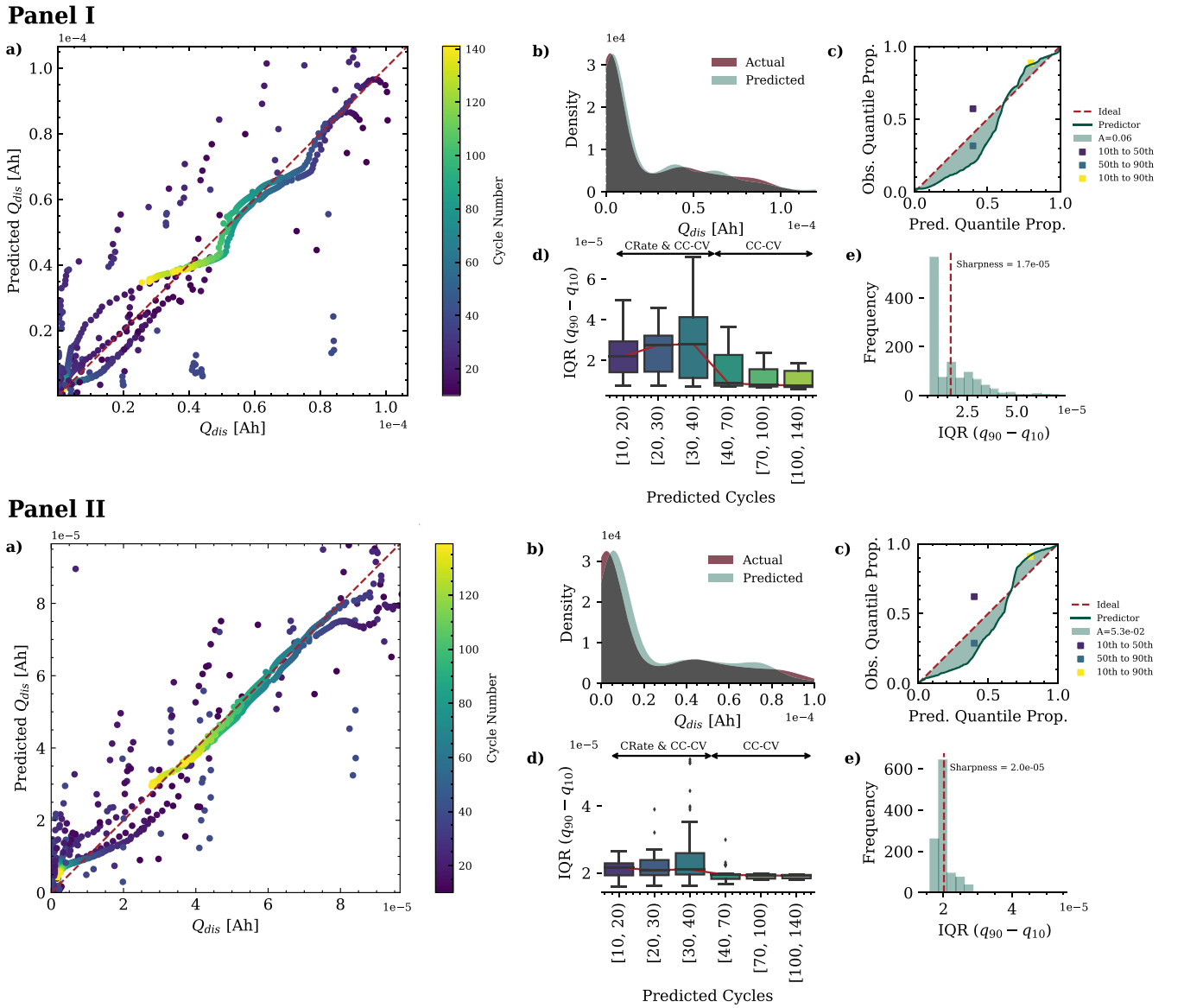

Fig. 8 | Comparative analysis of $M(P)_{Na}$ and $M(B)_{Na}$ on Q_{dis} prediction for Na-ion batteries. Prediciton analysis for $M(P)_{Na}$ (Panel I) and $M(B)_{Na}$ (Panel II) for Q_{dis} prediction of Na-ion batteries. The scatter plots **a** illustrate the models' alignment with actual measurements. Density plots **b** compare the distributions of predicted and actual values, demonstrating the models' accuracy in estimating Q_{dis}. Calibration plots in **c** depict how well the predicted probabilities match the observed outcomes against the benchmark line, with the discrete points representing the observed proportions of actual values that fall within three quantile intervals. Both models demonstrate a pattern of marginal overconfidence below the 70th percentile and a slight underconfidence above this percentile, as evidenced by the calibration points positions beneath and above the diagonal line, respectively. $M(P)_{Na}$ shows a larger area of divergence, $A = 0.06$, while $M(B)_{Na}$ presents a closer fit with a miscalibration of 0.053, highlighting both models' well-calibrated prediction capabilities across different chemistries. Boxplots **d** visualize the spread and consistency of prediction intervals across predicted cycles. Histograms in **e** represent the distribution of the quantile intervals of the models' prediction, highlighting uncertainty; these distributions indicate where, within the prediction range, the models' confidence is concentrated, with sharpness values of 1.7×10^{-5} for $M(P)_{Na}$ and 2.0×10^{-5} for$M(B)_{Na}$, demonstrating a precise estimation of uncertainty.

The second criterion implements a conventional check and is applied to monitor the trend in validation error. An increased trend over the epoch sequence suggests that overfitting could be occurring. Training is discontinued when both the ratio criterion and the error-trend criterion indicate that further training is unlikely to yield significant gains. In general, this strategy offers a control mechanism that aligns the duration of training with the achievement of a well-generalized model capable of accurate predictions.

Training procedure

Expanding on Seq-to-Seq integration, the training phase begins by initializing the data loaders for batch processing and configuring the parameters of the Seq-to-Seq model, the loss criteria, the optimizer, and a dynamic learning rate scheduler[62]. Hyperparameter optimization, through a series of trials using Optuna's[52] Tree-structured Parzen Estimator (TPE) Sampler, employs a probabilistic model to specify the most promising parameter configuration, navigating the search space while balancing exploration and exploitation within a complex and high-dimensional domain[72]. Training unfolds over several epochs, with each iteration starting with a reset of the model's hidden states and zeroing gradients to ensure clean computation for the forward pass. The pinball loss function is selected for its effectiveness in probabilistic forecasting, eliminating the need for a presumptive data distribution model[70], unlike traditional metrics[69], which are more sensitive to noise and anomalies. These asymmetric and non-parametric criteria assess

forecast accuracy by penalizing deviations from three targeted quantiles, namely 0.1, 0.5, and 0.9, enhancing robustness to outliers and the efficacy of LSTM-based networks[69]. At the same time, a masking technique[63] is implemented to filter out padding-induced distortions from the loss calculation, ensuring the integrity of the learning signal. Backpropagation follows loss computation, incorporating gradient clipping to prevent divergence and gradient explosion in recurrent network architectures. Additionally, learning rate adjustments encourage robust convergence. The validation phase alternates with training, where performance is assessed, and early stopping criteria are applied to mitigate overfitting. Optuna enhances optimization by pruning the less promising trials. Once the training is completed, the model parameters are saved and a comprehensive report is generated detailing the training results. The training procedure steps described are schematically depicted in Supplementary Fig. 1.

Evaluation metrics

For this study, the following metrics are implemented, including both average errors and variability of individual predictions, to evaluate the performance of the model. These metrics are RMSE (Eq. (12)) which provides a measure of the magnitude of prediction errors, MAPE (Eq. (13)), which measures the average magnitude of errors as a percentage, medAE (Eq. (14)) to capture the median error, reducing the influence of outliers, and mean absolute error (MAE) (Eq. (15)) which represents the mean absolute differences.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{13}$$

$$\mathrm{medAE} = \mathrm{median}(|y_i - \hat{y}_i| : i = 1, 2, \ldots, n) \tag{14}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{15}$$

Data availability

Open source data supporting the findings of this study are available online, with access details provided in Table 1 and can be found in the corresponding literature[41–48,50]. In addition, public pre-trained model weights can be accessed at https://doi.org/10.5281/zenodo.10293072.

Code availability

The ARCANA framework can be installed using `pip install arcana-batt` or cloned from https://github.com/basf/ARCANA.

Received: 5 January 2024; Accepted: 24 April 2024;
Published online: 10 May 2024

References

1. Amici, J. et al. A roadmap for transforming research to invent the batteries of the future designed within the european large scale research initiative battery 2030+. *Adv. Energy Mater.* **12**, 2102785 (2022).
2. Xu, Y., Ge, J. & Ju, C.-W. Machine learning in energy chemistry: introduction, challenges and perspectives. *Energy Adv.* **2**, 896–921 (2023).
3. Severson, K. A. et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy* **4**, 383–391 (2019).
4. Che, Y., Hu, X., Lin, X., Guo, J. & Teodorescu, R. Health prognostics for lithium-ion batteries: mechanisms, methods, and prospects. *Energy Environ. Sci.* **16**, 338–371 (2023).
5. Stein, H. S. Nonlinear potentiodynamic battery charging protocols for fun, education, and application. *ACS Eng. Au* **0**, 0 (2023).
6. Kabir, M. & Demirocak, D. E. Degradation mechanisms in li-ion batteries: a state-of-the-art review. *Int. J. Energy Res.* **41**, 1963–1986 (2017).
7. Attia, P. M. et al. "knees" in lithium-ion battery aging trajectories. *J. Electrochem. Soc.* **169**, 060517 (2022).
8. Yang, F., Song, X., Dong, G. & Tsui, K.-L. A coulombic efficiency-based model for prognostics and health estimation of lithium-ion batteries. *Energy* **171**, 1173–1182 (2019).
9. Rahmanian, F. et al. Conductivity experiments for electrolyte formulations and their automated analysis. *Sci. Data* **10**, 43 (2023).
10. Dahn, J., Burns, J. & Stevens, D. Importance of coulombic efficiency measurements in r&d efforts to obtain long-lived li-ion batteries. *Interface* **25**, 75 (2016).
11. Smith, A., Burns, J., Trussler, S. & Dahn, J. Precision measurements of the coulombic efficiency of lithium-ion batteries and of electrode materials for lithium-ion batteries. *J. Electrochem. Soc.* **157**, A196 (2009).
12. Attia, P. M. et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature* **578**, 397–402 (2020).
13. Adamu, H., Abba, S. I., Anyin, P. B., Sani, Y. & Qamar, M. Artificial intelligence-navigated development of high-performance electrochemical energy storage systems through feature engineering of multiple descriptor families of materials. *Energy Adv.* **2**, 615–645 (2023).
14. Tong, Z., Miao, J., Tong, S. & Lu, Y. Early prediction of remaining useful life for lithium-ion batteries based on a hybrid machine learning method. *J. Cleaner Prod.* **317**, 128265 (2021).
15. Rieger, L. H. et al. Uncertainty-aware and explainable machine learning for early prediction of battery degradation trajectory. *Digit. Discov.* **2**, 112–122 (2023).
16. Yang, Y. A machine-learning prediction method of lithium-ion battery life based on charge process for different applications. *Appl. Energy* **292**, 116897 (2021).
17. Liu, Y. et al. Generative artificial intelligence and its applications in materials science: current situation and future perspectives. *J. Materiomics* **9**, 798–816 (2023).
18. Gong, Q., Wang, P. & Cheng, Z. An encoder-decoder model based on deep learning for state of health estimation of lithium-ion battery. *J. Energy Storage* **46**, 103804 (2022).
19. Zhu, C., He, Z., Bao, Z., Sun, C. & Gao, M. Prognosis of lithium-ion batteries' remaining useful life based on a sequence-to-sequence model with variational mode decomposition. *Energies* **16**, 803 (2023).
20. Li, W. et al. One-shot battery degradation trajectory prediction with deep learning. *J. Power Sources* **506**, 230024 (2021).
21. Deng, Z., Lin, X., Cai, J. & Hu, X. Battery health estimation with degradation pattern recognition and transfer learning. *J. Power Sources* **525**, 231027 (2022).
22. Bhowmik, A. et al. Implications of the battery 2030+ Ai-assisted toolkit on future low-trl battery discoveries and chemistries. *Adv. Energy Mater.* **12**, 2102698 (2022).
23. Fichtner, M. et al. Rechargeable batteries of the future—the state of the art from a battery 2030+ perspective. *Adv. Energy Mater.* **12**, 2102904 (2022).
24. Strange, C. & Dos Reis, G. Prediction of future capacity and internal resistance of li-ion cells from one cycle of input data. *Energy and AI* **5**, 100097 (2021).
25. Ling, C. A review of the recent progress in battery informatics. *npj Comput. Mater.* **8**, 33 (2022).
26. Ng, M.-F., Zhao, J., Yan, Q., Conduit, G. J. & Seh, Z. W. Predicting the state of charge and health of batteries using data-driven machine learning. *Nat. Mach. Intell.* **2**, 161–170 (2020).
27. Liu, Y. et al. Data quantity governance for machine learning in materials science. *Natl Sci. Rev.* **10**, nwad125 (2023).
28. Baumhöfer, T., Brühl, M., Rothgang, S. & Sauer, D. U. Production caused variation in capacity aging trend and correlation to initial cell performance. *J. Power Sources* **247**, 332–338 (2014).

29. Roman, D., Saxena, S., Robu, V., Pecht, M. & Flynn, D. Machine learning pipeline for battery state-of-health estimation. *Nat. Mach. Intell.* **3**, 447–456 (2021).
30. Jha, S. et al. Learning-assisted materials development and device management in batteries and supercapacitors: Performance comparison and challenges. *J. Mater. Chem. A* **11**, 3904–3936 (2023).
31. Yao, Z. et al. Machine learning for a sustainable energy future. *Nat. Rev. Mater.* **8**, 202–215 (2023).
32. Dos Reis, G., Strange, C., Yadav, M. & Li, S. Lithium-ion battery data and where to find it. *Energy AI* **5**, 100081 (2021).
33. Wu, B., Widanage, W. D., Yang, S. & Liu, X. Battery digital twins: perspectives on the fusion of models, data and artificial intelligence for smart battery management systems. *Energy AI* **1**, 100016 (2020).
34. Li, X., Wang, Z. & Yan, J. Prognostic health condition for lithium battery using the partial incremental capacity and gaussian process regression. *J. Power Sources* **421**, 56–67 (2019).
35. Zhang, Y. et al. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nat. Commun.* **11**, 1706 (2020).
36. Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
37. De Angelis, V., Preger, Y. & Chalamala, B. R. Battery lifecycle framework: a flexible repository and visualization tool for battery data from materials development to field implementation. Preprint at osf.io/preprints/ecsarxiv/h7c24 (2021).
38. Li, W. et al. Digital twin for battery systems: cloud battery management system with online state-of-charge and state-of-health estimation. *J. Energy Storage* **30**, 101557 (2020).
39. Draxl, C. & Scheffler, M. Nomad: the fair concept for big data-driven materials science. *Mrs Bulletin* **43**, 676–682 (2018).
40. Wilkinson, M. D. et al. The fair guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 1–9 (2016).
41. Toyota Research Institute (TRI). Experimental data platform: project data-driven prediction of battery cycle life before capacity degradation. *data.matr.io* https://data.matr.io/1/projects/5c48dd2bc625d700019f3204 (2021).
42. Toyota Research Institute (TRI), Experimental data platform: Project closed-loop optimization of extreme fast charging for batteries using machine learning. *data.matr.io* https://data.matr.io/1/projects/5d80e633f405260001c0b60a (2019).
43. Saha, B. & Goebel, K. Nasa. *Prognostics Data Repository* https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository (2007).
44. Center for Advanced Life Cycle Engineering (CALCE), *University of Maryland* https://calce.umd.edu/data (2011).
45. Zhu, J. et al. Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation. *Zenodo* https://doi.org/10.5281/zenodo.6405084 (2022).
46. Battery Archive, Homepage of Battery Archive. https://www.batteryarchive.org, (2021).
47. Zhang, Merker, Sanin & Stein. Cycling data of 64 cells manufactured by autobass. *Zenodo* https://doi.org/10.5281/zenodo.7299473 (2022).
48. Merker, L. 2023 commercial coincell 45mah. *Zenodo* https://doi.org/10.5281/zenodo.10102627 (2023).
49. Merker, L. Inzepro inform 300 cycles cccv after eol. *Zenodo* https://doi.org/10.5281/zenodo.10102508 (2023).
50. Nuss, L., Merker, L., Zhang, B. & Stein, H. Formation and cycling data for Na-ion batteries from high-throughput synthesis, coating, and assembly. *Zenodo* https://doi.org/10.5281/zenodo.7981011 (2023).
51. Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process Syst.* **32**, 8024–8035 (2019).
52. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: a next-generation hyperparameter optimization framework. *DLP-KDD '19* 2623–2631 (2019).
53. Cooijmans, T., Ballas, N., Laurent, C., Gülçehre, Ç. & Courville, A. Recurrent batch normalization. Preprint at https://arxiv.org/abs/1603.09025 (2017).
54. Yoo, J., Kim, B., Lee, B., Song, J.-h & Kang, K. An artificial neural network using multi-head intermolecular attention for predicting chemical reactivity of organic materials. *J. Mater. Chem. A* **11**, 12784–12792 (2023).
55. Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. Preprint at https://arxiv.org/abs/1409.0473 (2016).
56. Smith, A. et al. Potential and limitations of research battery cell types for electrochemical data acquisition. *Batter. Supercaps* **6**, e202300080 (2023).
57. Burns, J. et al. Evaluation of effects of additives in wound li-ion cells through high precision coulometry. *J. Electrochem. Soc.* **158**, A255 (2011).
58. Burns, J. et al. Predicting and extending the lifetime of li-ion batteries. *J. Electrochem. Soc.* **160**, A1451 (2013).
59. Zhu, J. et al. Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation. *Nat. Commun.* **13**, 2261 (2022).
60. Guo, C., Pleiss, G., Sun, Y. & Weinberger, K. Q. On calibration of modern neural networks. Preprint at https://arxiv.org/abs/1706.04599 (2017).
61. Gneiting, T., Balabdaoui, F. & Raftery, A. E. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Series B Stat. Methodol.* **69**, 243–268 (2007).
62. Goldberg, Y. A primer on neural network models for natural language processing. *J. Artif. Intell. Res.* **57**, 345–420 (2016).
63. Li, W., Zhang, H., van Vlijmen, B., Dechent, P. & Sauer, D. U. Forecasting battery capacity and power degradation with multi-task learning. *Energy Storage Mater.* **53**, 453–466 (2022).
64. Chen, G., Song, Z., Qi, Z. & Sundmacher, K. Generalizing property prediction of ionic liquids from limited labeled data: a one-stop framework empowered by transfer learning. *Digit. Discov.* **2**, 591–601 (2023).
65. Bloom, I. et al. Differential voltage analyses of high-power, lithium-ion cells: 1. Technique and application. *J. Power Sources* **139**, 295–303 (2005).
66. Niu, Z., Zhong, G. & Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **452**, 48–62 (2021).
67. Ross, J. et al. Large-scale chemical language representations capture molecular structure and properties. *Nat. Mach. Intell.* **4**, 1256–1264 (2022).
68. Xu, C., Wang, Y. & Barati Farimani, A. Transpolymer: a transformer-based language model for polymer property predictions. *npj Comput. Mater.* **9**, 64 (2023).
69. Wang, Y. et al. Probabilistic individual load forecasting using pinball loss guided lstm. *Appl. Energy* **235**, 10–20 (2019).
70. Liu, B., Nowotarski, J., Hong, T. & Weron, R. Probabilistic load forecasting via quantile regression averaging on sister forecasts. *IEEE Trans. Smart Grid* **8**, 730–737 (2015).
71. Prechelt, L. Early stopping—but when? In *Neural Networks: Tricks of the Trade: Second Edition*, 53–67 (2012).
72. Bergstra, J., Bardenet, R., Bengio, Y. & Kégl, B. Algorithms for hyper-parameter optimization. *Adv. Neural Inf. Process Syst*. **24** (2011).

Acknowledgements

This work contributes to TUM.Battery, the Munich Data Science Institute, and the Munich Institute for Robotic and Machine Intelligence. This work contributes to the research performed at CELEST (Center for Electrochemical Energy Storage Ulm-Karlsruhe) and was partly funded by the German Research Foundation (DFG) under Project ID 390874152 (POLiS Cluster of Excellence). This project also received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 957189 (BIG-MAP). The project is part of BATTERY 2030+, the large-scale European research initiative for inventing sustainable batteries for the future, funded by the European Union's Horizon 2020 research

and innovation program under Grant Agreement No. 957213. HSS acknowledges funding from the German Research Foundation (DFG) under Project ID 390776260 (eConversion Cluster of Excellence).

Author contributions

K.M. and B.B. provided the comprehensive BASF dataset, and L.M. and L.N. conducted all cycling data for Li-ion and Na-ion batteries at KIT/IPC, respectively. Data assembly, data cleaning, model idea including the design architecture, code implementation, repository curation, training, evaluation, and package creation is conducted by F.R. R.L. and D.L. supervised the model development. K.M., B.B., H.S., R.L. and D.L. supervised this research. All authors reviewed the paper.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-024-01286-7.

Correspondence and requests for materials should be addressed to Fuzhan Rahmanian or Helge Sören Stein.