IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 February 2024, accepted 13 March 2024, date of publication 21 March 2024, date of current version 28 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3380452

SURVEY

Comprehensive Review of Machine Learning, Deep Learning, and Digital Twin Data-Driven Approaches in Battery Health Prediction of Electric Vehicles

A. PRAVIN RENOLD AND NEERAJ SINGH KATHAYAT

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu 600127, India

Corresponding author: A. Pravin Renold (pravinrenold.a@vit.ac.in)

ABSTRACT This paper presents a comprehensive survey of machine learning, deep learning, and digital twin technology methods for predicting and managing the battery state of health in electric vehicles. Battery state of health estimation is essential for optimizing the battery usage, performance, safety, and cost-effectiveness of electric vehicles. Estimating the state of health of a battery is a complex undertaking due to its dependency on multiple factors. These factors include battery characteristics such as type, chemistry, size, temperature, current, voltage, impedance, cycle number, and driving pattern. There are drawbacks to traditional methods, such as experimental and model-based approaches, in terms of accuracy, complexity, expense, and viability for real-time applications. By employing a variety of algorithms to discover the nonlinear and dynamic link between the battery parameters and the state of health, data-driven techniques like machine learning, deep learning, and data-driven digital twin technologies can get beyond these restrictions. Data-driven methods can also incorporate physics and domain knowledge to improve the explainability and interpretability of the results. This paper reviews the latest advancements and challenges of using data-driven techniques for battery state of health estimation and management in electric vehicles. The paper also discusses the future directions and opportunities for further research and development in this field. The survey scope spans publications from the year 2021 to 2023.

INDEX TERMS Machine learning models, deep learning models, data-driven methods, lithium-ion batteries, digital twin technology, battery health prediction, electric vehicles.

I. INTRODUCTION

Electric vehicles (EVs) are not only a promising solution to the environmental and economic challenges posed by conventional gasoline vehicles, but also a significant opportunity for innovation and growth in the automotive sector. According to the International Energy Agency (IEA), Electric vehicles constituted 4.6% of worldwide automobile sales and comprised 1% of the global automotive inventory in the year 2020, despite the Covid-19 pandemic [1]. As per IEA, global electric vehicle stock could reach 145 million by 2030 in the current policies scenario, or 230 million under sustainable development. EVs provide several benefits, including the potential to decrease Greenhouse Gas (GHG) emissions, improve air quality, enhance energy security, diversify energy sources, and reduce fuel and maintenance costs for consumers [2]. EVs can also enable smart grid integration, vehicle-to-grid services, and demand response management, which can enhance the reliability and efficiency of the power system [3].

However, the performance of electric vehicle batteries degrades over time, which can lead to reduced driving range and increased maintenance costs. To address this issue, researchers have developed machine learning models to predict the State of Health (SOH) of electric vehicle batteries. These models can help vehicle owners and manufacturers to

The associate editor coordinating the review of this manuscript and approving it for publication was Junho Hong.

optimize battery usage and reduce costs. Battery SOH is a measure that evaluates the extent of deterioration and available capacity in a battery. It signifies the contrast between the condition of a brand new battery and that of a previously used battery, typically presented as a percentage of the battery's original capacity [4]. Battery SOH estimation is essential for battery health management and second-life utilization. However, many of the current methods are developed under ideal laboratory conditions and do not account for the complex and dynamic operational environments of electric vehicles. Therefore, researchers have proposed various methods to estimate the battery SOH under realistic EV conditions, such as data-driven models, machine learning algorithms, and regional capacity analysis [4], [5], [6].

Estimating the battery SOH is crucial for EVs as it can affect the performance, safety, and cost-effectiveness of the battery and the vehicle. It can help EV owners to know the reliability, range, and performance of their vehicles, and buyers and sellers of used EVs to accurately value the product and have increased confidence in the EV's worth, longevity, and range. Additionally, it can help determine if the battery is suitable for reuse and repurposing or if it should be sent directly to recycling when the EV is retired [4]. Predicting the SOH under dynamic conditions is challenging due to the complex and nonlinear interactions between the battery parameters and the operating environment, which may vary significantly depending on the driving patterns, charging strategies, and ambient temperature. These factors can influence the electrochemical and thermal processes within the battery, resulting in different aging mechanisms and degradation rates [33], [34].

Experimental methods rely on direct measurements of battery parameters, such as capacity, resistance, and impedance, to evaluate the SOH. However, these methods are often time-consuming, costly, and intrusive, and may not be suitable for online and real-time applications. Model-based methods use mathematical or physical models to describe the electrochemical and thermal processes of battery degradation, and to infer the SOH from the model parameters. However, these methods require accurate knowledge of the battery characteristics and operating conditions, and may not capture the complex and nonlinear dynamics of battery aging [35].

Data-driven models, a form of artificial intelligence, have the ability to learn from data and experience without requiring explicit programming. They can play a major role in predicting the battery SOH in EVs, which is a measure of the battery degradation and remaining capacity. Forecasting SOH of batteries can assist maximize battery efficiency, safety, and performance while lowering the cost and environmental effect of electric vehicles. But estimating the SOH of a battery can be difficult since it depends on a lot of variables, including the battery type, chemistry, size, temperature, current, voltage, impedance, cycle count, and driving pattern. The accuracy, complexity, expense, and viability of traditional methods such as experimental and model-based approaches are all constrained when it comes to real-time application. By employing data-driven methods like support vector regression, neural networks, and deep learning to understand the nonlinear and dynamic relationship between the battery parameters and the SOH. Data-driven models can also incorporate physics and domain knowledge to improve the explainability and interpretability of the results.

However, data-driven models also face some challenges, such as data quality, availability, and security, as well as model validation, verification, and generalization [36]. Therefore, further research and development efforts are necessary to enhance the performance, efficiency, and intelligence of data-driven models for estimating battery SOH.

The survey in [42] investigates data-driven models for battery SOH estimation, with a focus on their application in Battery Management Systems (BMS). It covers key parameters like SOH, Remaining Useful Life (RUL), and State of Charge (SOC), offering contributions such as an overview of feature extraction methods, a comprehensive survey of state-of-the-art models from 2018 to 2022, and discussions on challenges and future research directions. The work categorizes data-driven SOH models, including Neural Network (NN) Models, Gaussian Process Regression (GPR) Models, Ensemble Learning (EL) Models, Transfer Learning (TL) Models, and Conventional Models (CM).

The work [43] examines data-driven techniques for assessing health and predicting the lifespan of lithium-ion batteries, crucial for electric vehicles and diverse applications. It covers battery ageing mechanisms, stress factors, and provides a proxy for conditions accelerating ageing. Additionally, it explores state-of-health estimation methods using various parameters and surveys analytical models and machine learning for health prediction. The work highlights challenges and future directions in data-driven battery health management, including ageing mechanism identification, self-improving models, and diagnosis and prognosis at module and pack levels.

The authors in [44] review methods for assessing battery SOH and SOC in electric vehicles, encompassing the history, chemistries, challenges in battery management, and diverse state estimation techniques. It also explores open-access datasets, critiques existing works, notes limitations and research gaps, and suggests future research directions. The work categorizes SOH estimation into three types: Experimental methods, providing accurate but time-consuming and offline measurements; Model-Based methods, offering fast, online estimation dependent on model accuracy and complexity; and Machine learning methods, utilizing algorithms for flexible SOH prediction, requiring large, high-quality datasets for training and validation.

The methods for estimating Li-ion battery health in electric vehicles, encompassing battery operation, characteristics, degradation factors, and various modeling techniques are discussed in [45]. It surveys state-of-the-art methods for estimating SOH and predicting RUL, comparing their performance and suggesting improvements. The paper discusses current model limitations, proposes future research directions, and categorizes SOH estimation techniques into experimental-based, model-based, and data-driven methods, providing examples such as empirical methods, neural networks, support vector machines, and gaussian process regression for data-driven estimation.

The work [46] examines various data-driven approaches for forecasting the state of health, state of charge, and remaining useful life of lithium-ion batteries, extensively employed in electric vehicles and diverse applications. It presents a thorough overview of available battery-cycling test databases, detailing characteristics like cell chemistry, capacity, voltage, cycle, temperature, and file format. The review paper categorizes and elucidates studies utilizing data-driven methods, classifying them based on the predictive models employed, including deep learning, machine learning, and hybrid methods. Current challenges and future trends in battery health estimation and prediction, such as data quality and quantity, model accuracy and robustness, online adaptation and transfer learning, and digital twins, are discussed.

The work [47] reviews the major data-driven methods for SOH estimation, such as Extended Kalman Filter (EKF), Particle Filter (PF), Autoregressive Integrated Moving Average (ARIMA), Extreme Learning Machine (ELM), Support Vector Machine (SVM), Relevance Vector Machine (RVM), and Long Short-term Memory (LSTM).The study compares the performance of these methods on a real-world electric vehicle dataset, using three metrics: Mean Absolute Percentage Error (MAPE), training time, and computation time. Additionally, the work discusses the advantages and limitations of each method in terms of accuracy, confidence interval, nonlinearity, robustness, computation complexity, data sparsity, and generalization. The work provides insights and recommendations for future research on data-driven SOH estimation for lithium-ion batteries in EVs.

Though there are works available in a similar domain, our work differs from the above-mentioned survey works in the following ways:

- The focus is on recent methods, challenges, and future directions, including machine learning, deep learning, and digital twin technology. Our work specifically addresses a gap in the existing literature by focusing on data-driven techniques within the digital twin domain for electric vehicles.
- A comparative analysis of the performance, advantages, and disadvantages of different models and algorithms, using various datasets and metrics, is provided. This emphasis on comparison may offer a more comprehensive understanding of the strengths and weaknesses of different approaches in battery health estimation.
- We discuss the importance of data quality, availability, and security, as well as model validation, verification, and generalization, for developing reliable and robust machine learning models for battery SOH prediction.

The main contributions of the work are as follows:

- The paper presents a review of recent studies (2021-2023) that employed data-driven methods for estimating the SOH of lithium-ion batteries in electric vehicles.
- The paper highlights the significance of various prediction methods, analyses the advantages, drawbacks and presents the results obtained using each of these methods.
- The paper also discusses the challenges and future directions of data-driven methods for battery health assessment.

This paper is organized as follows: Section II provides a review of predicting and managing battery health in EVs, categorizing approaches under machine learning, deep learning, and data-driven digital twin. Section III discusses the characteristics of public datasets. Future directions related to SOH prediction are discussed in Section IV. Finally, Section V concludes the paper.

II. REVIEW OF PREDICTING AND MANAGING THE BATTERY HEALTH IN ELECTRIC VEHICLES

This section reviews the recent methods and challenges of using data-driven techniques, such as machine learning, deep learning and digital twin, to estimate and optimize the battery SOH in EVs.

A. BATTERY HEALTH PREDICTION USING MACHINE LEARNING

Battery health prediction through machine learning is a research area dedicated to the creation and assessment of data-centric models that can forecast the SOH, SOC, or RUL of lithium-ion batteries. These predictions are made possible by employing diverse machine learning methodologies, including neural networks, support vector machines, random forests, and other relevant techniques.

In the operational framework [8], the Autoregressive (AR) model utilizes battery voltage and current data to estimate the SOH. Leveraging its capability to capture temporal dependencies, the AR model analyzes historical information, discerning trends indicative of the battery's health. The Relevance Vector Machine (RVM) complements the AR model, crucially enhancing SOH estimation accuracy. The RVM's primary role is to address errors in the AR model's output compared to the actual SOH value, refining overall precision. This compensation involves dynamic adjustments to weights and relevance vectors, optimizing agreement between the modeled and actual SOH through an adaptive learning process, ultimately contributing to a more reliable battery health estimation.

The methodology employed in [9] integrates the Two-Step Noise Reduction Method, Domain-Specific Features, and Stacking Ensemble Learning. The Two-Step Noise Reduction Method employed in this study utilizes a moving average filter and a wavelet transform to effectively reduce noise present in the battery data. This method aims to enhance the overall data quality by mitigating disturbances and inconsistencies. Additionally, the methodology incorporates Domain-Specific Features, a set of characteristics derived from domain knowledge and battery physics. Examples of these features include discharge time, discharge energy, and discharge capacity, providing valuable insights into the battery's behavior and performance. Lastly, the approach integrates Stacking Ensemble Learning such as linear regression, support vector regression, and random forest regression into a meta-learner. This ensemble learning strategy contributes to improved prediction accuracy and generalization by leveraging the diverse strengths of individual models. The combination of these methods forms a comprehensive approach to battery data analysis, addressing noise reduction, feature engineering, and predictive modeling within a unified framework. Careful consideration is required in terms of model evaluation due to complexity and computational resource requirements.

The approach in [10] integrates Empirical Mode Decomposition (EMD) with the Particle Filter (PF) algorithm to predict the SOH of lithium-ion batteries. It utilizes EMD to decompose battery capacity data into Intrinsic Mode Functions (IMFs), capturing trends and fluctuations. The Particle Filter algorithm is then applied to estimate SOH based on the extracted IMFs, providing robust SOH prediction with uncertainty representation.

The work in [11] presents a method for predicting the SOH of lithium-ion batteries using variational mode decomposition (VMD) and dung beetle optimization-support vector regression (DBO-SVR).The VMD decomposes the SOH data into components, which are then individually modeled and predicted using SVR. The DBO algorithm optimizes the SVR parameters to enhance prediction accuracy. The predicted values of the components are combined to obtain the final SOH prediction. This methodology aims to improve prediction accuracy, stability, and robustness compared to other methods, as demonstrated through experiments on the NASA dataset.

In [12] a data-driven method for SOC estimation of lithium-ion batteries using optimized random forest regression algorithm is proposed. The method uses battery voltage and current sensor data as input features and employs Principal Component Analysis (PCA) to reduce dimensionality. The method also uses a synthetic feature, SOCxI, to capture the relationship between current magnitude and state of charge. The method achieves high accuracy with errors typically within $\pm2\%$ and not exceeding $\pm3\%$ for cells with more than 80% SOC. The drawbacks include challenges related to data quality, a limited investigation of resistance increase, dependency on specific datasets, and a trade-off between model size and accuracy.

The work in [14] introduces an online battery health diagnosis method for electric vehicles using DTW-XGBoost, focusing on predicting the SOH through real-time charging data. The methodology comprises three key steps: feature extraction using DTW-based clustering, XGBoost prediction for SOH modeling, and the development of an online diagnosis platform. Feature extraction involves dynamic time warping for aggregating voltage, current, and temperature data, followed by dimensionality reduction and generation of 24 features per charging process. XGBoost is employed for training a predictive model, optimizing parameters for minimal RMSE and MAE. The online diagnosis platform integrates acquisition control, modeling analysis, and application service modules to collect real-time BMS data, predict SOH, and offer various services based on prediction results, including health trend prediction and charging operation analysis.

A model called the Modified Support Vector Machine (M-SVM) [15] is suggested for precise assessment of the SOH of lithium-ion batteries in electric cars. The M-SVM is a regression-based method used for estimating the SOH of lithium-ion batteries in electric vehicles. M-SVM is designed to estimate and predict the SOH of batteries based on input features, taking into account the continuous nature of the health parameter.M-SVM modifies the SVM algorithm to perform regression by fitting a curve that best represents the relationship between input features and the continuous output (SOH).The NASA Li-ion battery ageing dataset, which consists of 34 cylindrical cells that were cycled between 70% and 80% of their initial capacity while being exposed to different temperatures and discharge rates, is used to assess the model. The M-SVM model's performance is contrasted with that of various machine learning techniques, including Neural Network (NN) and Linear Regression (LR).

The work in [16] presents an approach for estimating the SOH of electric vehicle battery packs. The proposed method uses Catboost, a machine learning algorithm that performs regression or classification tasks by using gradient boosting on decision trees. Catboost can automatically handle categorical features without the need for manual encoding or preprocessing. Additionally, Catboost uses ordered boosting, a technique that reduces overfitting by introducing randomness into the gradient boosting process. The proposed method is validated by using a year-long operation dataset of an electric taxi to develop and test the method.

Reference [17] proposes a framework for accurate SOH estimation in real-world EVs through parameters optimization, addressing challenges in capturing complex operating conditions and achieving precise predictions. The work employs the PSO-ELM (Particle Swarm Optimization-Extreme Learning Machine) model for SOH estimation. Extreme Learning Machine (ELM) as the base learner, employing a single-hidden layer feedforward neural network. This type of network randomly assigns weights and biases between the input and hidden layers while analytically determining weights to the output layer. Optimization of ELM parameters, including the number of hidden nodes, activation function, and learning rate, is achieved through Particle Swarm Optimization (PSO).

Reference [18] proposes a method for precisely estimating the state of charge and SOH of lithium-ion batteries using dual filters and the Interacting Multiple Model (IMM) algorithm. The IMM algorithm combines state hypotheses from multiple filter models to enhance state estimates for targets with dynamic changes. It involves four steps: Mixing, where previous state estimates are combined based on mode transition probabilities and model likelihoods; Filtering, where each filter model is updated with current measurements; Combination, where state estimates from each model are combined based on model probabilities; and Mode Probability Calculation, indicating the confidence of each filter model. The paper introduces two strategies for applying IMM with dual filters, dual-KF-IMM and dual-SIF-IMM, comparing their performance and analyzing mode probabilities to understand battery degradation modes on NASA dataset.

A Feedforward Neural Network (FNN) with PCA for SOH estimation in lithium-ion batteries [25]. The work also introduces a synthetic feature, SOCxI, which is the product of SOC and current (I). This feature captures the relationship between current magnitude and state of charge, which can affect battery degradation and capacity loss. The work compares the proposed algorithm with other machine learning models like ELM, SVM, and LSTM and shows that it outperforms them in terms of accuracy.

The work in [37] proposes a method for estimating the SOH of lithium-ion batteries using dynamic discharge conditions over a wide temperature range. The method extracts three aging features (discharge voltage integration, discharge time, and net discharge energy) and two operating condition features (mean current and discharge capacity ratio) from the voltage and current data of the batteries. The method fuses these features to obtain three health indicators that are highly correlated with battery capacity and are not affected by temperature or operating conditions. The method uses a GPR model to establish a mapping between the health indicators and battery capacity, achieving high accuracy and low computational complexity. The method is validated using experimental data from two types of lithium-ion batteries (LFP and NMC) under various dynamic operating conditions and temperatures. The results demonstrate the effectiveness and universality of the proposed method.

In [38] using an equivalent circuit model with fixed parameters, the integral voltage error is extracted as an aging feature. The average current is extracted as an operating condition feature. The two features are fused to obtain a fused feature that is input into a back propagation neural network (BPNN) to estimate SOH. The method achieves accurate and generalized SOH estimation under various dynamic operating conditions, with mean absolute errors around 1%. The method also eliminates the dependence on state of charge accuracy and has low computational requirements.

In [39] the authors propose an indirect SOH estimation method for online EV lithium-ion batteries based on arctangent function adaptive genetic algorithm combined with back propagation neural network (ATAGA-BP).The authors use constant current drop time (CCDT), constant current drop capacity (CCDC) and maximum constant current drop rate (MCCDR) in the constant voltage charging stage as health indicators to evaluate battery SOH and indirectly quantify the degradation process of lithium-ion batteries. The work optimizes and validates the health indicators using PauTa criterion and Pearson Product-Moment Correlation Coefficient (PPMCC), and establish the relationship between health indicators and available battery capacity using the ATAGA-BP algorithm.

B. BATTERY HEALTH PREDICTION USING DEEP LEARNING

In [7], the study introduces a deep learning-based approach for real-world electric vehicles to predict the SOH of battery systems. The method utilizes long short-term memory (LSTM) networks and Pearson correlation analysis to identify relevant features and model the battery degradation process. Through the use of Pearson correlation analysis, the study identifies four highly correlated features with the SOH: constant current charging time, constant voltage charging time, voltage change rate from 300s to 1000s, and voltage per cycle at 200s intervals. The proposed method claims to achieve high accuracy and low prediction error in SOH estimation by leveraging actual battery data from electric buses.

The authors in [13] develop a predictive model using LSTM neural networks to estimate the RUL of lithium-ion batteries. The model uses a Variational Mode Decomposition (VMD) technique to decompose the battery voltage signal into IMFs and selects the most relevant IMFs as input features. The model also uses a particle swarm optimization (PSO) algorithm to optimize the hyperparameters of the LSTM network. The model achieves high accuracy with a RMSE of 4.32 cycles and a Mean Absolute Error (MAE) of 3.21 cycles. The drawbacks are interpretability challenges, limited discussion on external factors and dependency on maintenance data.

The work in [19] integrates Convolutional Autoencoders (CAEs) and Bi-directional Long Short-Term Memory (BiLSTM) neural networks.CAE is used to extract key fused timing features like voltage, current, and temperature directly from raw data.The extracted features are then used as input for the BiLSTM, which includes dropout technology and a fully connected layer to map to the battery's SOH. The method includes data interpolation and augmentation to enhance the training process.An end-to-end model is proposed that combines the CAE and BiLSTM for effective SOH prediction, demonstrating high accuracy and robustness in experiments.

A deep learning method that uses a Deep Neural Network (DNN) and Convolutional Neural Network (CNN) to estimate cell-level capacity in lithium-ion batteries based on voltage, current, and state of charge is proposed in [20]. The approach involves a multi-physics battery model that integrates electrical, thermal, and aging models to generate

TABLE 1. Machine learning based methods.

Reference	Algorithm	Results	Dataset	Advantage	Disadvantage
[8]	AR + RVM	MAE = 0.0042, RMSE = 0.0056	NASA Li-ion battery aging dataset	Uses error compensation and adaptive learning to refine estimates	Does not provide uncertainty representation or confidence intervals
[9]	Two-Step Noise Reduction + Domain-Specific Features + Stacking Ensemble Learning	MAPE= 0.28% Root mean squared percent error (RMSPE) of 0.55%	CALCE battery dataset	Addresses noise reduction, feature engineering, and predictive modeling in a unified framework	Requires significant computational resources and careful model evaluation
[10]	EMD + PF	One-step prediction: MAPE 0.3931% to 0.7818%, RMSE 0.6537% to 1.1985% Long-term prediction: MAPE 0.6366% to 1.8317%, RMSE 0.7157% to 1.4207% Multi-step prediction: MAPE 0.5309% to 1.5259%, RMSE 0.6917% to 1.6625%	NASA Li-ion battery aging dataset	Uses EMD to capture trends and fluctuations in battery data and PF to provide uncertainty representation	Does not consider the influence of operating conditions or battery aging mechanisms
[11]	VMD + DBO-SVR	MAPE= 0.3906% RMSE= 0.4771%	NASA Li-ion battery aging dataset	Uses VMD to reduce noise interference and DBO-SVR to optimize model parameters	Requires significant computational resources and may not be feasible for online and real-time prediction
[12]	Optimized random forest regression	RMSE ranges from 0.0039 to 0.225	NASA Li-ion battery aging dataset	Uses PCA to reduce dimensionality and a synthetic feature to capture the relationship between current and SOC	Depends on data quality and specific datasets and faces a trade-off between model size and accuracy
[14]	DTW-XGBoost	RMSE=0.674 MAE= 0.843	Real-world electric	Clustering-based method for multi-source data fusion efficiently reduces data dimensionality and extracts representative features from real-time battery data during charging	Does not consider the influence of operating conditions or battery aging mechanisms
[15]	M-SVM	RMSE=0.244 MAE=0.17 R2=0.999	NASA Li-ion battery aging dataset	Uses a modified SVM to improve the performance of the base learner	Does not provide uncertainty representation or confidence intervals

TABLE 1. ***(Continued.)*** **Machine learning based methods.**

[16]	Catboost + Interval capacity	MAPE= 2.74% RMSE= 1.12%	Real-world electric taxi	Uses Catboost to handle categorical features and ordered boosting to reduce overfitting	The battery aging evaluation strategy and future research plan are preliminary and require additional validation and implementation
[17]	PSO-ELM	RMSE (%) = 0.0849 MAPE (%) = 0.0710 Mean Error (%) = 0.1786	Real-world electric vehicles	Uses PSO to optimize ELM parameters and enhance prediction precision	Assuming uniform battery pack capacity, which may not reflect practical variations arising from cell imbalances or degradation, potentially impacting SOH estimation accuracy and health feature extraction
[18]	Dual filters + IMM	RMSE: Dual-KF-IMM: 0.1758 Dual-SIF-IMM: 0.2198	NASA Li-ion battery aging dataset	Uses dual filters to handle nonlinearities and IMM to enhance state estimates for dynamic changes	Work assumes that the battery system can be modeled by a linear state-space model, which may not capture the nonlinear dynamics and uncertainties of the battery
[25]	FNN + PCA	RMSPE = 1.03% MAPE = 0.68%	NASA Li-ion battery aging dataset	Uses PCA to reduce dimensionality and a synthetic feature to capture the relationship between current and SOC	Does not consider the influence of operating conditions or battery aging mechanisms
[37]	GPR	LiFePO4 battery at a single temperature: MAE = 0.3578% RMSE = 0.4125% NMC batteries over a wide temperature range: MAE = 0.4452% RMSE = 0.4971%	LiFePO4 and lithium nickel manganese cobalt oxide (NMC) batteries with different dynamic operating conditions and temperatures	Extracted effective and universal health indicators by fusing aging features and operating condition features	Required sufficient number of known dynamic operating conditions to determine the relationship between aging features and operating condition features
[38]	BPNN	For the Dynamic Stress Test: RMSE = 0.91307% , MAE = 0.71768% For the Federal Urban Driving Schedule: RMSE = 1.3828% , MAE = 0.91755% For the Urban Dynamometer Driving Schedule: RMSE = 1.204% , MAE = 0.82136%	A123 company	With low computational complexity, it's ideal for real-time use in electric vehicles. The method combines battery aging and operating features, employing a simple neural network for SOH estimation.	Unable to handle the uncertainty and noise in the measurement data, which could affect the extraction and fusion of the features.

TABLE 1. ***(Continued.)*** **Machine learning based methods.**

		For the Beijing Dynamic Stress Test: RMSE = 1.5307%, MAE = 1.1967%			
[39]	ATAGA-BP	RMSE = 4.9%, MAE = 3.7%	NASA	The work uses measurable parameters of the actual charging process of EV as HI, which can indirectly quantify the degradation process of lithium-ion batteries and avoid complicated and expensive online capacity and internal resistance measurement.	The work only considers the constant voltage charging stage as HI, which may not fully reflect the degradation characteristics of lithium-ion batteries under different working conditions.

input and output data for machine learning algorithms. The machine learning process involves training and testing the DNN and CNN models using data from the multi-physics model, with performance evaluated using MAE and mean squared error (MSE) metrics. The results show that the DNN model has the lowest error and the best stability among the compared capacity estimation methods.

In [21] raw battery data is processed using a 1D Convolutional Neural Network (1DCNN) to extract local and global features. These features serve as inputs for subsequent models. The study combines LSTM and transformer models. LSTM excels in sequence modeling by using a smaller window of several consecutive cycles, the model does not need to store or process large data sequences. LSTM captures local dependencies and short-term patterns. Transformer captures global dependencies and long-term context. The hybrid model combines these two architectures to improve prediction accuracy.

The Deep Learning Feed-Forward Neural Network (DL-FFNN) [22] employs a FFNN model featuring two hidden layers, each comprising five neurons, to train a dataset derived from actual driving trips of a BMW i3 EV. The input parameters include voltage, current, battery temperature, and ambient temperature, while the output is the SOC of the battery. The optimization of weights and biases in the FFNN model is accomplished through a metaheuristic algorithm inspired by the mating process in organisms. This algorithm, termed EMA, categorizes candidate solutions into males and females, selects the best performers for mating based on variance, and generates new offspring by combining their features with a random factor. EMA also considers the effect of environmental factors such as predators on the survival of the offspring.

The approach in [23] utilizes equal voltage interval discharge time, Incremental Capacity (IC), and Differential Thermal Voltammetry (DTV) for feature extraction. The deep learning model is structured around a Bi-LSTM with an incorporated Attention Mechanism (AM) to emphasize crucial features. Bi-LSTM facilitates capturing context from both past and future states, enhancing sequence prediction. The attention mechanism focuses on significant portions of the input sequence during predictions, assigning weights based on importance to selectively attend to informative elements.

A data-driven prediction model, Nonlinear Autoregressive with External Input (NARXRNN) recurrent neural network, that focuses on forecasting the long-term health and remaining lifespan of lithium-ion batteries in electric vehicles [24]. The NARXRNN model uses battery voltage, current, and temperature inputs to predict the SOH and RUL of lithium-ion batteries. The model comprises three layers with specific activation functions and employs a Backpropagation Through Time (BPTT) algorithm for training, minimizing MSE between predicted and actual values. Validation on NASA and CALCE datasets, encompassing diverse lithium-ion batteries and operating conditions, demonstrates the model's accuracy for long horizons (up to 100 cycles). Notably, the NARXRNN outperforms alternative methods such as ARIMA, ELM, and LSTM, showcasing robustness to noise and generalization across varying battery types and conditions.

In [40] a method for predicting the SOH of lithium-ion batteries using Gated Recurrent Unit (GRU) neural networks and Hidden Markov Model (HMM) with considering uncertainty quantification is proposed. The work decomposes the battery capacity into the global downward trend and the local fluctuations by using Empirical Mode Decomposition (EMD), and trains a GRU network to fit the global trend and an HMM to fit the local fluctuations. The work conducts numerical experiments on two public datasets with fixed and random working conditions, and compares the proposed method with other existing methods in terms of accuracy and reliability. The result demonstrates that the proposed method can capture the long-term and dynamic characteristics of battery degradation, and outperforms other methods on the SOH prediction of lithium-ion batteries.

The developed model [41], named CNN-CBAM-LSTM, combines convolution neural network, convolutional block attention module, and long short-term memory neural network to achieve reliable SOH estimation for lithium-ion batteries. The model uses partial charging voltage fragments as input to automatically extract features related to battery degradation from the raw data. The developed method is validated using two battery degradation datasets, Oxford and NASA, and the results show that the model can achieve high accuracy estimation even in the case of short charging process. The SOH estimation can be achieved by fusing the personality features of the target battery and the degradation features of the source batteries through transfer learning with fine-tuning strategy, so that only a small amount of the target battery data is required.

Table 3 provides a comparative analysis of different Machine Learning (ML) and Deep Learning (DL) approaches based on their effectiveness in various contexts, highlighting their relative advantages and limitations.

C. BATTERY HEALTH MANAGEMENT USING DIGITAL TWIN DATA-DRIVEN TECHNOLOGY

Lithium-ion battery performance and longevity are optimized through the use of digital twin technology in a variety of applications, including smart grids and electric cars. With the use of this technology, a physical system may be virtually represented to mimic its behavior and offer decision-making insights.

Within the context of the Digital Twin architecture, the study [26] integrates an Extended Kalman Filter (EKF) and an Extreme Gradient Boost (XGBoost) model to provide full battery status estimates, monitoring, and prediction. The XGBoost model estimates the incremental SOH and SOC of the battery, drawing insights from historical data. The EKF plays a crucial role in refining the estimation by correcting errors and providing a robust representation of uncertainties associated with the battery's state.

In [27] an accurate method for predicting the SOH of electric vehicle battery packs, emphasizing real-world data integration and temperature correction for practical implementation is proposed. The methodology employs a double exponential function and PSO for SOH prediction, incorporating model migration to address limited data scenarios. Additionally, ambient temperature correction is introduced to enhance the accuracy of battery performance modeling. The strengths of the work lie in its utilization of real-world electric vehicle data, contributing to the practicality of SOH prediction. The proposed modeling approach is robust, offering a flexible solution for situations with limited data. However, limitations include the absence of specific details on the accuracy achieved by the method and a lack of consideration for potential computational complexity or scalability issues, particularly when applying the models to large datasets or real-time applications.

Reference [28] introduces a digital twin model structured in three phases: data acquisition, data processing, and regression modeling. During the data acquisition stage, a voltage sensor measures the battery pack voltage, transmitting the data to the digital twin server through wireless communications. In the data processing stage, the voltage data is segregated into normal and charging subsets, eliminating and organizing abnormal data. Finally, in the regression model stage, a straightforward linear regression model is employed to forecast the maximum and minimum cell voltage using the pack voltage. Additionally, a multi-linear regression model is utilized to predict other variables based on the pack voltage and the forecasted cell voltages.

It is suggested to use a digital twin structure [29] for managing the batteries in electric vehicles. This methodology makes use of an on-vehicle system to assess the SOC and a cloud-based model to anticipate the battery's SOH. The framework uses a data-driven methodology to analyze time-frequency pictures of the voltage and current signals from the battery in order to forecast SOH. To avoid catastrophic forgetting and overfitting, a selection of unimportant parameters is used while updating the model incrementally with fresh data from the vehicle. SOC estimation is performed through a coulomb counting method, integrating current over time, and is executed on the vehicle to provide real-time feedback for driving range calculations. The proposed framework is implemented on Microsoft Azure, incorporating a deep learning model and a Raspberry-Pi board as the battery management system.

Three components make up the framework in [30] for determining the SOH of lithium-ion batteries. The first part is the synchronization of variable cycling data, which is achieved using energy discrepancy-aware temporal warping to align data from several cycles. The second part includes a time-attention SOH estimating model that regresses on the Maximum Available Capacity (MAC) and captures the temporal importance of different sampling times using a LSTM with a time-attention mechanism. The third part focuses on real-time SOH estimation through future data reconstruction, utilizing data matching and reconstruction to supplement unknown future data and estimate the MAC at any point during the ongoing cycle.

In [31], a method is proposed for continuous learning in predicting the SOH of electric vehicle batteries using data from a digital twin. To preserve the data distribution, a memory buffer is created from the base data using a random sampling technique. The base model is then fine-tuned by

TABLE 2. Deep learning based methods.

Reference	Algorithm	Results	Dataset	Advantage	Disadvantage
[7]	LSTM + Pearson correlation	RMSE < 0.02 for driving, < 0.05 for charging	Real-world electric buses	Selects relevant features and models battery degradation process	Does not consider external factors such as temperature and discharge rate
[13]	VMD + PSO + LSTM	For the HPPC test at room temperature, RMSE and MAE for LiNCA battery: 0.508%, 0.086%; for LiNMC battery: 0.382%, 0.076%. For DST drive cycle at varying temperatures, RMSE and MAE for LiNCA battery: 0.312%-0.685%, 0.193%-0.424%	NASA Li-ion battery aging dataset	Uses VMD to select relevant features and PSO to optimize LSTM hyperparameters	Faces interpretability challenges and does not discuss external factors or maintenance data
[19]	CAEs + BiLSTM	Oxford dataset: MAE = 0.038% RMSE = 0.074% NASA dataset: MAE = 1.14% RMSE = 3.76%	NASA & Oxford	Feature extraction and temporal modeling	Requires complete charging data
[20]	DNN	MAE = 0.2253 MSE = 0.0875	Multi-physics model	High accuracy and stability	Requires voltage, current, and SOC inputs
[21]	LSTM	MSE= 0.43 RMSE= 0.66 MAE= 0.53 MAPE= 0.58 Maximum Absolute Error (MAXE)= 1.65	MIT & NASA	It does not rely on past SOH values, but only on ageing features extracted from current, voltage and temperature curves	The approach requires a feature selection process to reduce the dimensionality and redundancy of the input features, which can be time-consuming and dependent on the dataset
[22]	DL-FFNN + EMA	RMSE=4.702 MAE= 3.458 STD.DEV=4.694	NASA	Analysis of 70 BMW i3 EV driving trips bolsters the practical relevance and validity of the proposed approach	Impact of different input features, such as elevation, speed, throttle, or traffic conditions, on the SOC estimation accuracy could have been considered
[23]	Bi-LSTM + AM	RMSE = 0.648% MAE = 0.543%	NASA and Oxford	Improved temporal modeling using the GRU neural network to learn temporal information from health feature time series data	More experimentation required to assess the state of health estimation under varying charging protocols, environmental conditions, and battery types

TABLE 2. ***(Continued.)*** **Deep learning based methods.**

[24]	NARXRNN	RMSE= 0.46% to 1.82% MAE= 0.38% to 1.60%	NASA and CALCE	Long horizon forecasting and generalization	The battery degradation data is available and sufficient for the prognostic models, which may not be the case for different battery chemistries, operating conditions, and applications
[40]	GRU + HMM	Oxford dataset MAE = 0.36% RMSE = 0.42% NASA dataset MAE = 1.45% RMSE = 1.76% Three-step-ahead prediction: Oxford dataset MAE = 0.33% RMSE = 0.41% NASA dataset MAE = 2.27% RMSE = 2.65%	NASA and Oxford	Usage of EMD to decompose the battery capacity into global trend and local fluctuations, and trains GRU and HMM to fit them respectively	Estimating the SOH for lithium-ion batteries using randomly obtained partial charging data could be considered
[41]	CNN-CBAM-LSTM	For the Oxford dataset: - MAE: 0.72% for B0005, 0.88% for B0029, and 2.1% for B0055 - RMSE: 0.61% for B0005, 0.64% for B0029, and 1.62% for B0055 For the NASA dataset: - MAE: 0.81% for B0005, 0.94% for B0029, and 2.18% for B0055 - RMSE: 0.65% for B0005, 0.69% for B0029, and 1.62% for B0055	NASA and Oxford	The model uses partial charging voltage fragments as input to automatically extract features related to battery degradation from the raw data, which reduces the need for manual feature engineering	The work does not provide a detailed analysis of the impact of different sampling frequencies and voltage windows on the performance of the developed model, which may limit its generalizability to other datasets and working conditions

training it with a mini-batch of data, which includes one sample from the new data and k-1 samples from the memory buffer. This approach helps prevent overfitting and catastrophic forgetting. The fine-tuned model is registered in the cloud and utilized for making predictions. Additionally, the new data sample is incorporated into the base data to generate the subsequent reservoir sample.

A method for enhanced condition monitoring of lithium-ion batteries in EVs using digital twin technology is presented in [32]. This approach enables real-time

TABLE 3. Comparative analysis of machine learning and deep learning approaches.

Context	ML Approach	DL Approach	Effectiveness	Comments
Noise Reduction & Feature Engineering	Two-Step Noise Reduction + Domain-Specific Features + Stacking Ensemble Learning	Convolutional Autoencoders (CAEs) + Bi-directional Long Short-Term Memory (BiLSTM)	**ML:** Effective in reducing noise and improving feature representation. **DL:** Captures complex temporal dependencies.	**ML:** Provides interpretable features. **DL:** Requires substantial data and computational resources.
Trend & Fluctuation Analysis	Empirical Mode Decomposition (EMD) + Particle Filter (PF)	Gated Recurrent Unit (GRU) + Hidden Markov Model (HMM)	**ML:** Captures trends and fluctuations. **DL:** Models long-term dependencies.	**ML:** Limited by assumptions of EMD. **DL:** Requires large labeled datasets.
Real-Time SOH Estimation	Autoregressive (AR) + Relevance Vector Machine (RVM)	Deep Neural Network (DNN) + Convolutional Neural Network (CNN)	**ML:** Efficient for online prediction. **DL:** Learns complex patterns.	**ML:** Lacks uncertainty representation. **DL:** Prone to overfitting without regularization.
Feature Extraction from Charging Data	Modified Support Vector Machine (M-SVM)	CNN-CBAM-LSTM	**ML:** Extracts relevant features. **DL:** Adapts to hierarchical features.	**ML:** Sensitive to kernel choice. **DL:** Requires large labeled data and model tuning.
Uncertainty Representation	Catboost + Interval Capacity	LSTM + Pearson Correlation	**ML:** Provides uncertainty intervals. **DL:** Captures temporal correlations.	**ML:** Limited to specific models. **DL:** Computationally expensive.
Optimization & Parameter Tuning	PSO-ELM (Particle Swarm Optimization-Extreme Learning Machine)	VMD + PSO + LSTM	**ML:** Optimizes model parameters. **DL:** Handles noise and uncertainty.	**ML:** May converge to local optima. **DL:** Requires careful hyperparameter tuning.
Generalization Across Conditions	GPR (Gaussian Process Regression)	NARXRNN (Nonlinear Autoregressive with External Input Recurrent Neural Network)	**ML:** Generalizes well. **DL:** Adapts to varying conditions.	**ML:** Scalable to large datasets. **DL:** Complex architecture may overfit small datasets.

monitoring signals without requiring additional hardware circuits or sensor calibration, due to its seamless integration with the embedded BMS. The method, consists of two main parts physical modeling and digital modeling, uses offline modeling in the physical modeling stage with a LSTM algorithm to guarantee accurate SOC predictions with different learning rates (LRs) and optimization by three different kinds of optimizers. A digital twin is used in the digital modelling component to forecast and monitor battery behavior in real time. Interestingly, the method uses a Time-Series Generative Adversarial Network (TS-GAN) to produce fake data, which makes the monitoring procedure better.

III. PUBLIC DATASETS

In the realm of battery research and innovation, access to comprehensive and diverse datasets is pivotal for advancing scientific understanding and fostering breakthroughs in energy storage technologies. This section presents a selection of public data collections that have played a key role in shaping the discourse around lithium-ion batteries.

The 'Oxford Battery Degradation Dataset 1' [48] comprises long-term battery ageing tests involving eight Kokam (SLPB533459H4) 740 mAh lithium-ion pouch cells. The data were collected in a thermal chamber at 40°C, subjecting the cells to a constant-current-constant-voltage charging profile followed by a drive cycle discharging profile derived from the

TABLE 4. Digital twin technology based methods.

Reference	Algorithm	Results	Dataset	Advantage	Disadvantage
[26]	XGBoost + EKF	RMSE= 0.009 MAE= 0.007	NASA	Real-time predictions of SOH and SOC	Assumption of constant battery parameters
[27]	Double exponential function + PSO	Relative error < 1.5%	EV battery pack data	Real-world data integration and temperature correction	Lack of accuracy details and scalability issues
[28]	Simple linear regression + Multi-linear regression	R=0.94	Battery pack voltage data	Simple and fast	Low accuracy and limited applicability
[29]	Deep learning model + Coulomb counting	MSE=0.023	NASA	Low computational and memory requirements of the on-vehicle system	Does not consider the impact of driving patterns, environmental factors, and user preferences on the battery performance and degradation
[30]	LSTM + Time-attention	RMSE= 0.64% to 1.66% R^2= 95.03% to 99.41%	MIT & Li-ion battery data	Variable cycling data synchronization and real-time estimation	Needs partially discharged cycling data and consumes substantial memory
[31]	Incremental learning	MSE=0.022	NASA	Adapts to different operating conditions and avoid overfitting and catastrophic forgetting	Extensive experimentation needed to verify the generalizability and robustness of the proposed method
[32]	LSTM + TS-GAN	MAE= 0.888 RMSE= 0.912	Multi-variate time series	Incorporating TS-GAN for the generation of synthetic data aims to improve the accuracy of SOC prediction and address the issue of limited availability of real-time data	More experiments needed to validate the impact of data augmentation on the SOC prediction results

urban Artemis profile, with characterization measurements taken every 100 cycles.

The dataset originates from a custom-built battery prognostics testbed at the NASA Ames Prognostics Center of Excellence (PCoE) [49]. This dataset contains data from lithium-ion batteries that underwent three operational profiles (charge, discharge, and Electrochemical Impedance

Spectroscopy) across different temperatures. The dataset includes repeated charge and discharge cycles, simulating accelerated battery aging until meeting the end-of-life criteria of a 30% fade in rated capacity. The testbed is equipped with commercially available Li-ion 18650-sized rechargeable batteries, a programmable 4-channel DC electronic load and power supply, voltmeter, ammeter, thermocouple sensors, custom EIS equipment, an environmental chamber, PXI chassis-based DAQ, and MATLAB-based experiment control, operating at an approximate data acquisition rate of 10Hz.

The dataset by CALCE [50] includes experimental lithium-ion battery test data for applications like state estimation, life prediction, degradation modeling, and reliability analysis. It covers various battery types (cylindrical, pouch, prismatic) and chemistries (LCO, LFP, NMC). The tests involve OCV tests, dynamic driving profiles, impedance measurements, and temperature cycling. The A123 cell (LFP) underwent SOC estimation via low-current and incremental OCV tests at 0°C, 25°C, and 45°C, favoring the latter. The CS2 cell (LCO) and CX2 cell (NMC) contributed to prognostics and health management with different discharge profiles, aiding in state estimation, remaining useful life prediction, and reliability analysis. The PLN cell (NCA) was examined for capacity and impedance degradation under varying SOC levels and temperatures, contributing to battery degradation modeling.

The dataset [51] presents experimental results for four Lithium-ion batteries undergoing charge, discharge, and impedance operations at room temperature. In charge cycles, batteries were charged in constant current (CC) mode at 1.5A until 4.2V, followed by constant voltage (CV) mode until the current dropped to 20ma. Discharge cycles maintained a Constant Current (CC) of 2A until specific voltage thresholds were reached. Impedance measurements utilized Electrochemical Impedance Spectroscopy (EIS) frequency sweeps from 0.1Hz to 5khz. Operation cycles involve specific measurements. For charge, track 'Voltage,' 'Current,' and 'Temperature' alongside 'Current_charge' and 'Voltage_charge.' Dischargcycles incorporate 'Capacity' until a specific voltage. Impedance cycles encompass 'Sense_current,' 'Battery_current,' 'Current_ratio,' 'Battery_impedance,' 'Rectified_impedance,' 'Re,' and 'Rct,' giving a comprehensive overview of the system's behavior.

The dataset [52] comprises 124 lithium-ion phosphate/graphite cells (A123 Systems, APR18650M1A) cycled to failure under fast-charging conditions. The cells, with a nominal capacity of 1.1 Ah and a voltage of 3.3 V, were tested in a forced convection temperature chamber at 30°C using a 48-channel Arbin LBT potentiostat. The fast-charging policy involves one or two steps denoted as 'C1(Q1)-C2,' where C1 and C2 represent the first and second constant-current steps, and Q1 is the state-of-charge (SOC) at which the currents switch. The dataset is organized into three batches, each representing around 48 cells, with unique irregularities detailed for each batch.

IV. FUTURE DIRECTIONS

In this section, we explore the potential future directions for forecasting the condition of battery health in electric vehicles, drawing insights from the existing literature and referring to Table 1, Table 2, and Table 4.

- Developing hybrid models that combine data-driven and physics-based methods for more accurate and interpretable SOH estimation and RUL prediction, taking into account the complex electrochemical and thermal dynamics of batteries. Hybrid models can combine the advantages of both methods, and overcome their limitations, by integrating data-driven and physics-based sub-models, or by using different types of sub-models, such as neural networks, equivalent circuit models, and transfer learning. Hybrid models can improve the accuracy, interpretability, and generalization of SOH estimation and RUL prediction, by learning from data and physics, and by adapting to different scenarios and factors.
- SOH prediction is challenging due to the complex and nonlinear dynamics of battery degradation, as well as the uncertainties and noise in battery data. Therefore, novel deep learning techniques, such as graph neural networks, transformers, and attention mechanisms, have been explored to improve the accuracy and robustness of SOH prediction models. These techniques can capture the spatiotemporal dependencies and nonlinear relationships in battery data, and enhance the generalization and scalability of SOH prediction models. For example, graph neural networks can model the interactions among different battery cells or modules, transformers can encode the sequential and contextual information of battery data, and attention mechanisms can focus on the important features or time steps for SOH prediction.
- SOH prediction model is necessary to provide confidence intervals and reliability measures for the predictions, and to handle the uncertainties and noise in battery data. Uncertainty quantification can also improve the robustness and safety of battery systems, as it can help identify the worst-case scenarios and avoid overconfidence or under confidence in the predictions. Some methods for uncertainty quantification include Bayesian inference, Monte Carlo simulation, or interval analysis, which can capture both aleatoric and epistemic uncertainties in SOH prediction.
- SOH prediction is challenging due to the influence of various external factors, such as environmental conditions, driving patterns, charging strategies, and battery aging mechanisms, which may affect the performance and degradation of batteries in different ways. Therefore, evaluating the impact of external factors on SOH prediction performance is necessary to understand the sources of uncertainties and variations in battery data, and to identify the most relevant and sensitive factors for SOH prediction. Moreover, developing adaptive models that can adjust to different scenarios and optimize battery management is essential to improve the accuracy and robustness of SOH prediction, and to enhance the efficiency and safety of battery systems.
- SOH prediction is challenging due to the diversity and variability of batteries and applications, such as smart grids, renewable energy systems, and microgrids, which may have different types, sizes, configurations, operating conditions, and degradation patterns of batteries. Therefore, applying transfer learning and domain adaptation techniques to enable SOH prediction for different types of batteries and applications is necessary to improve the accuracy and robustness of SOH prediction models, and to reduce the data and computational requirements. Transfer learning and

domain adaptation techniques can leverage the knowledge and information from existing or related tasks, and adapt them to new or target tasks, by minimizing the distribution discrepancy or maximizing the feature similarity between the source and target domains.

- Accurate prediction of SOH is crucial for effective battery management, aiding in estimating remaining useful life and capacity. Battery aging is intricate, influenced by factors like chemistry, operating conditions, and environment. Diverse batteries exhibit varied aging modes, such as power fade or capacity fade. Addressing this diversity requires extensive data and knowledge of the battery system and degradation mechanisms. In lithium-ion batteries, common aging mechanisms include structural disordering, phase transition, metal dissolution, and electrolyte decomposition, each impacting the ability of electrodes to store and release lithium ions, causing capacity loss, and affecting charge transfer and mass transport processes.
- Applying explainable artificial intelligence (XAI) techniques to enhance the interpretability and transparency of SOH prediction models, and to provide explanations and justifications for the predictions, such as feature importance, decision rules, and causal relationships.
- Exploring the trade-off between accuracy and complexity of SOH prediction models, and developing efficient and scalable models that can achieve high accuracy with low computational cost and memory requirement, such as sparse models, and compressed models.

V. CONCLUSION

This paper has presented a comprehensive survey of data-driven models for predicting battery state of health in electric vehicles. The paper has reviewed the main methods, challenges, and future directions of data-driven techniques, such as machine learning, deep learning, and digital twin technology, for battery health estimation and management. The survey has also provided a comparative analysis of the performance, advantages, and disadvantages of different models and algorithms, using various datasets and metrics. The paper has highlighted the importance of data quality, availability, and security, as well as model validation, verification, and generalization, for developing reliable and robust machine learning models for battery SOH prediction. The survey foster additional research and innovation in this domain by offering a thorough overview of the current advancements and the challenges they entail.

REFERENCES

[1] Inc42. (2018). *Electric Vehicles Overview: The Evolution of EVs In India*. Accessed: Dec. 2, 2023. [Online]. Available: https://inc42.com/features/electric-vehicles-overview-indiae-evs/

[2] Int. Energy Agency. (2021). *Trends and Developments in Electric Vehicle Markets*. Accessed: Dec. 2, 2023. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2021/trends-and-developments-in-electric-vehicle-markets

[3] Int. Energy Agency. (2023). *Electric Vehicles*. Accessed: Dec. 2, 2023. [Online]. Available: https://www.iea.org/energy-system/transport/electric-vehicles

[4] S. Khaleghi Rahimian and Y. Tang, "A practical data-driven battery state-of-health estimation for electric vehicles," *IEEE Trans. Ind. Electron.*, vol. 70, no. 2, pp. 1973–1982, Feb. 2023.

[5] Y. Zhi, H. Wang, and L. Wang, "A state of health estimation method for electric vehicle Li-ion batteries using GA-PSO-SVR," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2167–2182, Jun. 2022.

[6] J. Wu, L. Fang, G. Dong, and M. Lin, "State of health estimation for lithium-ion batteries in real-world electric vehicles," *Sci. China Technol. Sci.*, vol. 66, no. 1, pp. 47–56, Jan. 2023.

[7] H. Yang, J. Hong, F. Liang, and X. Xu, "Machine learning-based state of health prediction for battery systems in real-world electric vehicles," *J. Energy Storage*, vol. 66, Aug. 2023, Art. no. 107426.

[8] H. Feng and H. Yan, "State of health estimation of large-cycle lithium-ion batteries based on error compensation of autoregressive model," *J. Energy Storage*, vol. 52, Aug. 2022, Art. no. 104869.

[9] J. Zhao, H. Ling, J. Liu, J. Wang, A. F. Burke, and Y. Lian, "Machine learning for predicting battery capacity for electric vehicles," *ETransportation*, vol. 15, Jan. 2023, Art. no. 100214.

[10] L. Cai, J. Lin, and X. Liao, "A data-driven method for state of health prediction of lithium-ion batteries in a unified framework," *J. Energy Storage*, vol. 51, Jul. 2022, Art. no. 104371.

[11] C. Wu, J. Fu, X. Huang, X. Xu, and J. Meng, "Lithium-ion battery health state prediction based on VMD and DBO-SVR," *Energies*, vol. 16, no. 10, p. 3993, May 2023.

[12] B. Zhang and G. Ren, "Li-ion battery state of charge prediction for electric vehicles based on improved regularized extreme learning machine," *World Electr. Vehicle J.*, vol. 14, no. 8, p. 202, Jul. 2023.

[13] M. S. Hossain Lipu, M. A. Hannan, A. Hussain, S. Ansari, S. A. Rahman, M. H. M. Saad, and K. M. Muttaqi, "Real-time state of charge estimation of lithium-ion batteries using optimized random forest regression algorithm," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 1, pp. 639–648, Jan. 2023.

[14] N. Yan, Y.-B. Yao, Z.-D. Jia, L. Liu, C.-T. Dai, Z.-G. Li, Z.-H. Zhang, W. Li, L. Wang, and P.-F. Wang, "Online battery health diagnosis for electric vehicles based on DTW-XGBoost," *Energy Rep.*, vol. 8, pp. 121–128, Nov. 2022.

[15] R. Swarnkar, R. Harikrishnan, P. Thakur, and G. Singh, "Electric vehicle lithium-ion battery ageing analysis under dynamic condition: A machine learning approach," *SAIEE Afr. Res. J.*, vol. 114, no. 1, pp. 4–13, Mar. 2023.

[16] R. Li, J. Hong, H. Zhang, and X. Chen, "Data-driven battery state of health estimation based on interval capacity for real-world electric vehicles," *Energy*, vol. 257, Oct. 2022, Art. no. 124771.

[17] X. Zhao, J. Hu, G. Hu, and H. Qiu, "A state of health estimation framework based on real-world electric vehicles operating data," *J. Energy Storage*, vol. 63, Jul. 2023, Art. no. 107031.

[18] R. Bustos, S. A. Gadsden, M. Al-Shabi, and S. Mahmud, "Lithium-ion battery health estimation using an adaptive dual interacting model algorithm for electric vehicles," *Appl. Sci.*, vol. 13, no. 2, p. 1132, Jan. 2023.

[19] C. Zhu, M. Gao, Z. He, H. Wu, C. Sun, Z. Zhang, and Z. Bao, "State of health prediction for Li-ion batteries with end-to-end deep learning," *J. Energy Storage*, vol. 65, Aug. 2023, Art. no. 107218.

[20] J. Hemdani, L. Degaa, N. Rizoug, and A. Chaari, "State of health prediction of lithium-ion battery using machine learning algorithms," in *Proc. 9th Int. Conf. Control, Decis. Inf. Technol. (CoDIT)*, Jul. 2023, pp. 2729–2733.

[21] I. Jorge, T. Mesbahi, A. Samet, and R. Boné, "Time series feature extraction for lithium-ion batteries state-of-health prediction," *J. Energy Storage*, vol. 59, Mar. 2023, Art. no. 106436.

[22] M. H. Sulaiman, Z. Mustaffa, N. F. Zakaria, and M. M. Saari, "Using the evolutionary mating algorithm for optimizing deep learning parameters for battery state of charge estimation of electric vehicle," *Energy*, vol. 279, Sep. 2023, Art. no. 128094.

[23] H. Zhao, Z. Chen, X. Shu, J. Shen, Z. Lei, and Y. Zhang, "State of health estimation for lithium-ion batteries based on hybrid attention and deep learning," *Rel. Eng. Syst. Saf.*, vol. 232, Apr. 2023, Art. no. 109066.

[24] S. Bamati and H. Chaoui, "Lithium-ion batteries long horizon health prognostic using machine learning," *IEEE Trans. Energy Convers.*, vol. 37, no. 2, pp. 1176–1186, Jun. 2022.

[25] J. Chen, P. Kollmeyer, F. Chiang, and A. Emadi, "Lithium-ion battery state-of-health estimation via histogram data, principal component analysis, and machine learning," in *Proc. IEEE Transp. Electrific. Conf. Expo. (ITEC)*, Jun. 2023, pp. 1–6.

[26] S. Jafari and Y.-C. Byun, ''Prediction of the battery state using the digital twin framework based on the battery management system,'' *IEEE Access*, vol. 10, pp. 124685–124696, 2022.
[27] X. Li, T. Wang, C. Wu, J. Tian, and Y. Tian, ''Battery pack state of health prediction based on the electric vehicle management platform data,'' *World Electr. Vehicle J.*, vol. 12, no. 4, p. 204, Oct. 2021.
[28] H. Li, M. Bin Kaleem, I.-J. Chiu, D. Gao, and J. Peng, ''A digital twin model for the battery management systems of electric vehicles,'' in *Proc. IEEE 23rd Int. Conf. High Perform. Comput. Commun.*, Hainan, China, Dec. 2021, pp. 1100–1107.
[29] N. D. K. M. Eaty and P. Bagade, ''Electric vehicle battery management using digital twin,'' in *Proc. IEEE Int. Conf. Omni-Layer Intell. Syst. (COINS)*, Aug. 2022, pp. 1–5.
[30] Y. Qin, A. Arunan, and C. Yuen, ''Digital twin for real-time Li-ion battery state of health estimation with partially discharged cycling data,'' *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 7247–7257, May 2023.
[31] N. D. K. M. Eaty and P. Bagade, ''Digital twin for electric vehicle battery management with incremental learning,'' *Expert Syst. Appl.*, vol. 229, Nov. 2023, Art. no. 120444.
[32] M. Pooyandeh and I. Sohn, ''Smart lithium-ion battery monitoring in electric vehicles: An AI-empowered digital twin approach,'' *Mathematics*, vol. 11, no. 23, p. 4865, Dec. 2023.
[33] Y. Zhang, X. Feng, M. Zhao, and R. Xiong, ''In-situ battery life prognostics amid mixed operation conditions using physics-driven machine learning,'' *J. Power Sources*, vol. 577, Sep. 2023, Art. no. 233246.
[34] J. Yang, C. Du, W. Liu, T. Wang, L. Yan, Y. Gao, X. Cheng, P. Zuo, Y. Ma, G. Yin, and J. Xie, ''State-of-health estimation for satellite batteries based on the actual operating parameters—Health indicator extraction from the discharge curves and state estimation,'' *J. Energy Storage*, vol. 31, Oct. 2020, Art. no. 101490.
[35] Y. Liu, L. Wang, D. Li, and K. Wang, ''State-of-health estimation of lithium-ion batteries based on electrochemical impedance spectroscopy: A review,'' *Protection Control Mod. Power Syst.*, vol. 8, no. 1, pp. 1–17, Dec. 2023.
[36] X. Lai, Y. Yao, X. Tang, Y. Zheng, Y. Zhou, Y. Sun, and F. Gao, ''Voltage profile reconstruction and state of health estimation for lithium-ion batteries under dynamic working conditions,'' *Energy*, vol. 282, Nov. 2023, Art. no. 128971.
[37] S. Liu, Y. Nie, A. Tang, J. Li, Q. Yu, and C. Wang, ''Online health prognosis for lithium-ion batteries under dynamic discharge conditions over wide temperature range,'' *ETransportation*, vol. 18, Oct. 2023, Art. no. 100296.
[38] Q. Yu, Y. Nie, S. Liu, J. Li, and A. Tang, ''State of health estimation method for lithium-ion batteries based on multiple dynamic operating conditions,'' *J. Power Sources*, vol. 582, Oct. 2023, Art. no. 233541.
[39] N. Li, F. He, W. Ma, R. Wang, L. Jiang, and X. Zhang, ''An indirect state-of-health estimation method based on improved genetic and back propagation for online lithium-ion battery used in electric vehicles,'' *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 12682–12690, Dec. 2022.
[40] M. Lin, Y. You, W. Wang, and J. Wu, ''Battery health prognosis with gated recurrent unit neural networks and hidden Markov model considering uncertainty quantification,'' *Rel. Eng. Syst. Saf.*, vol. 230, Feb. 2023, Art. no. 108978.
[41] A. Tang, Y. Jiang, Q. Yu, and Z. Zhang, ''A hybrid neural network model with attention mechanism for state of health estimation of lithium-ion batteries,'' *J. Energy Storage*, vol. 68, Sep. 2023, Art. no. 107734.
[42] M. Chen, G. Ma, W. Liu, N. Zeng, and X. Luo, ''An overview of data-driven battery health estimation technology for battery management system,'' *Neurocomputing*, vol. 532, pp. 152–169, May 2023.
[43] Y. Li, K. Liu, A. M. Foley, A. Zülke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo, and H. E. Hoster, ''Data-driven health estimation and life-time prediction of lithium-ion batteries: A review,'' *Renew. Sustain. Energy Rev.*, vol. 113, Oct. 2019, Art. no. 109254.
[44] R. Swarnkar, H. Ramachandran, S. H. M. Ali, and R. Jabbar, ''A systematic literature review of state of health and state of charge estimation methods for batteries used in electric vehicle applications,'' *World Electr. Vehicle J.*, vol. 14, no. 9, p. 247, Sep. 2023.
[45] M. Elmahallawy, T. Elfouly, A. Alouani, and A. M. Massoud, ''A comprehensive review of lithium-ion batteries modeling, and state of health and remaining useful lifetime prediction,'' *IEEE Access*, vol. 10, pp. 119040–119070, 2022.
[46] E. Kim, M. Kim, J. Kim, J. Kim, J.-H. Park, K.-T. Kim, J.-H. Park, T. Kim, and K. Min, ''Data-driven methods for predicting the state of health, state of charge, and remaining useful life of Li-ion batteries: A comprehensive review,'' *Int. J. Precis. Eng. Manuf.*, vol. 24, no. 7, pp. 1281–1304, Jul. 2023.
[47] T. Oji, Y. Zhou, S. Ci, F. Kang, X. Chen, and X. Liu, ''Data-driven methods for battery SOH estimation: Survey and a critical analysis,'' *IEEE Access*, vol. 9, pp. 126903–126916, 2021.
[48] (2023). *Ora.ox.ac.uk*. [Online]. Available: https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac
[49] (2023). *Data.nasa.gov*. [Online]. Available: https://data.nasa.gov/browse?q=battery&sortBy=relevance
[50] (2023). *Calce.Umd.Edu*. [Online]. Available: https://calce.umd.edu/battery-data
[51] C. Fei. (2022). *Lithium-Ion Battery Data Set*. Accessed: Dec. 30, 2023. [Online]. Available: https://ieee-dataport.org/documents/lithium-ion-battery-data-set
[52] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, and R. D. Braatz, ''Data-driven prediction of battery cycle life before capacity degradation,'' *Nature Energy*, vol. 4, no. 5, pp. 383–391, Mar. 2019.

A. PRAVIN RENOLD received the Ph.D. degree from Anna University, Chennai. He is currently an Assistant Professor (Sr.G) with the School of Computer Science and Engineering, Vellore Institute of Technology Chennai. His current research interests include the IoT, edge intelligence, electric vehicles, and sustainable computing.

NEERAJ SINGH KATHAYAT was born in Jasari, Khatima, Uttarakhand, India, in 2002. He received the Bachelor of Computer Applications (B.C.A.) degree from Graphic Era Hill University, Dehradun. He is currently pursuing the Master of Computer Applications (M.C.A.) degree with Vellore Institute of Technology. His academic interests revolve around machine learning and electric vehicle (EV) battery health detection. His current research interest includes leveraging machine learning techniques to enhance the detection and monitoring of EV battery health.