International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

ADVANCED BATTERY HEALTH PREDICTION IN ELECTRIC VEHICLES USING OPERATIONAL PARAMETER ANALYSIS

DR. G. RAMU[1], DHARA ANKITHA[2], ANDOLE SREEJA[3], ERGADHINDLA MAHESHWARI[4], ALLE SHYAM SUNDAR[5], CHAMAKURI YASHWANTH[6]

PROFESSOR[1], UG SCHOLAR[2,3,4,5&6]

DEPARTMENT OF CSE, NARSIMHA REDDY ENGINEERING COLLEGE (UGC- AUTONOMOUS) MAISAMMAGUDA (V), KOMPALLY, SECUNDERABAD, TELANGANA-500100

PROBLEM STATEMENT

Electric vehicles (EVs) are becoming increasingly important for sustainable transportation, but **battery degradation remains a major challenge** affecting vehicle performance, reliability, and lifespan. The health of an EV battery gradually decreases due to various operational factors such as **charging patterns, temperature variations, driving behavior, depth of discharge, and load conditions**.

Traditional battery monitoring systems mainly rely on simple estimation techniques that do not fully consider the complex relationship between these operational parameters and battery degradation. As a result, it becomes difficult to **accurately predict battery health and remaining useful life**, which can lead to unexpected battery failures, higher maintenance costs, and reduced vehicle efficiency. Therefore, there is a need for an **advanced predictive framework** that can analyze operational data and accurately estimate battery health to improve the reliability and performance of electric vehicles.

OBJECTIVES

1. To analyze the **impact of operational parameters** such as temperature, charging cycles, driving patterns, and discharge rates on EV battery degradation.
2. To develop a **data-driven predictive model** for estimating battery health and degradation levels.
3. To apply **machine learning or deep learning techniques** to improve the accuracy of battery health prediction.
4. To monitor and evaluate **battery performance over time** using operational data.
5. To provide an **early warning system** for battery degradation to reduce maintenance costs and improve EV reliability.
6. To support **efficient battery management systems (BMS)** for better energy utilization and longer battery lifespan.

ABSTRACT

Electric vehicles (EVs) are playing a vital role in the transition toward sustainable and eco-friendly transportation systems. However, battery degradation significantly affects the performance, efficiency, and lifespan of EV batteries. Accurate prediction of battery health is essential for improving vehicle reliability and optimizing battery management strategies. This study proposes an **advanced battery health prediction framework for electric vehicles based on operational parameter analysis**. The proposed system analyzes various operational factors such as **temperature conditions, charging and discharging cycles, driving behavior, and load variations** to understand their impact on battery degradation. Using data-driven techniques and machine learning algorithms, the system predicts the battery's state of health and identifies patterns associated with degradation over time. By integrating intelligent predictive models, the framework enables **early detection of battery performance decline**, allowing timely maintenance and improved battery management. The proposed approach enhances the accuracy of battery degradation prediction and contributes to the development of **more reliable and efficient electric vehicle systems**. This framework ultimately helps extend battery life, reduce operational costs, and support the advancement of intelligent transportation technologies.

1.INTRODUCTION

A S ENERGY shortage, climate change, and pollutan development of the world automotive industry, the development of electric vehicles (EVs) has been gradually regarded as essential to facilitating the low-carbon transformation of the global automotive industry and fulfilling a carbon-neutral vision in the transportation sector [1]. As a vital component of EVs, the battery pack significantly affects the overall vehicle performance and safety [2]. However, due to the inherent nature of the electrochemical systems, the battery inevitably deteriorates over time during usage or storage, as evidenced by a decrease in capacity and an increase in internal resistance [3]. Thus, accurately assessing battery aging degree and predicting further degradation trends, especially under highly dynamic and stochastic operating conditions, are crucial for service life evaluation, predictive maintenance, and second-life utilization. Existing approaches for battery health prediction can be categorized into model-based, data-driven, and hybrid [4]. Model-based methods describe the battery dynamic and degradation behavior through mathematical or physics-based models. Empirical and semi-empirical models are two widely used mathematical models that fit the battery capacity curve using some influencing factors, such as temperature, rate, and state of charge (SOC) [5]. In contrast, electrochemical and equivalent circuit models (ECMs), commonly employed physical models, utilize circuitry elements to simulate the reaction process inside the battery [6]. Although model-based methods have achieved remarkable progress in battery degradation prediction, some obstacles restrict their applications on inservice vehicles. First, battery degradation results from diverse, interlaced, and nonlinear degradation mechanisms. No single physics-based model can describe all the mechanisms comprehensively, nor can an effective approach quantify them separately [7]. Second, the modeling process of both mathematical and physics-based methods necessitates complete battery tests and involves lots of model parameters. Nevertheless, comprehensive testing is not feasible for in-service battery packs, and a significant number of the model parameters cannot be observed or calibrated using the existing onboard sensing technologies. In addition, the primary limitation of model-based approaches is that they are developed under well-controlled conditions, which results in a lack of extrapolation ability when applied to complex working conditions. In contrast to the physics-based methods, data-driven approaches focus more on the data under investigation to map the relationship between features and predictions, bypassing the complex mechanisms and propagation. Essential steps in data-driven modeling encompass extracting battery health-relevant features and constructing prediction models for specific applications. From a feature extraction perspective, well-extracted features with strong correlations to battery health are crucial to model accuracy. Therefore, various approaches have been proposed to extract features from the charging and discharging process. Among them, incremental capacity (IC) [8] and differential voltage (DV) analysis [9] have been extensively explored for battery health prognostics. However, low current rates and constant ambient temperatures are key premises for effectively extracting IC/DV features, which impedes its application in actual scenarios with large current rates. To fulfill the health prognostics demand under the high current rate scenarios, the peak/valley voltage and regional capacity features extracted from the voltage curve of the multistep fast charging process are verified to be strongly correlated with battery health in our previous research [10]. It should be noted that the effectiveness of the mentioned features is highly dependent on the operating conditions. Because of this, several statistical characteristics of the capacity sequence are explored to have strong correlations with battery health [11]. In addition, benefiting from the preceding network layers of deep learning algorithms, the original highdimensional time-series data can be directly used as the input [12], [13]. Although using the statistical characteristics of the capacity sequence and the raw time-series data as model input shows excellent performance in both low and high current rate scenarios, they have high requirements for data quality and sensor sampling frequency. The primary objective of hybrid methods is to enhance prediction accuracy by integrating signal processing, model-based approaches, and various data-driven methods [14], [15]. Despite their advantages, these methods also inherit certain limitations from their component models, such as diminished generalizability to unfamiliar systems and a strong dependence on highquality data. It is also worth noting that the aforementioned methods derive features from standard sensors, such as those measuring current, voltage, and temperature. However, with advancements in sensor technology, new types of sensing information, such as electrochemical impedance spectroscopy (EIS) [16], mechanical stress [17], and acoustic waves [18], offer alternative perspectives on battery health, complementing traditional electrical and thermal measurements. These technologies are recognized as essential components of the next generation of battery management. Once the features are extracted and selected, the intrinsic relationships between features and battery health indicators can be built by data-driven methods. Considering the nonlinear relationships between features and capacity or lifetime, nonlinear mapping algorithms such as support vector machines (SVMs) [19], relevance vector machines (RVMs) [20], and Gaussian process regression (GPR) [21] have become popular. However, their prediction accuracy heavily depends on the kernel functions, and it is hard to select one kernel function that suits various degradation patterns. In contrast, the neural network (NN)-based approach is more flexible and feasible, including the long short-term memory network (LSTM) [22], [23], deep convolutional NN (DCNN) [24], and their variants. However, it has been proved that external stress factors such as loading profiles and temperature significantly affect the rate of side reactions, thus leading to different degradation patterns between batteries. To enhance the model generalization ability in various operational scenarios, the transfer learning [25], [26] and model migration [27] strategies are utilized for fine-tuning and domain adaptation

2.LITERATURE REVIEW

The rapid adoption of electric vehicles (EVs) has increased the need for reliable battery monitoring systems to ensure safety, efficiency, and long battery life. Lithium-ion batteries are the primary energy storage systems in EVs, but their performance gradually degrades due to operational and environmental factors. Accurate prediction of **Battery State of Health (SOH)** using operational parameters such as voltage, current, temperature, and charge–discharge cycles has become an important research area. This literature review summarizes recent developments in battery health prediction techniques.

Importance of Battery Health Prediction in EVs

Battery State of Health (SOH) represents the remaining capacity of a battery relative to its original capacity and is a key parameter in battery management systems (BMS). Accurate SOH prediction helps determine battery lifetime, prevent failures, and optimize charging strategies. However, predicting battery degradation is challenging because it is influenced by multiple nonlinear factors such as temperature, charging rates, and depth of discharge.

Traditional battery monitoring approaches rely on electrochemical models or direct measurements, but these methods often require complex modeling or expensive sensors. Therefore, recent research focuses on **data-driven methods that analyze operational parameters collected during battery operation**.

Traditional Methods for Battery Health Estimation

Early research on battery health prediction mainly relied on **physics-based and experimental methods**. These approaches model battery degradation through electrochemical reactions and aging mechanisms.

Physics-based models provide detailed insights into battery behavior but require extensive knowledge of battery chemistry and are computationally expensive. Direct measurement techniques, such as capacity testing, are simple but cannot be applied continuously during real vehicle operation.

Although these methods offer good interpretability, they are not suitable for real-time EV applications because of their high complexity and limited scalability.

Machine Learning Approaches for Battery Health Prediction

Recent studies have explored **machine learning (ML) techniques** for predicting battery SOH using operational parameters. ML algorithms can analyze large datasets from battery management systems and capture complex relationships between battery degradation and operating conditions.

Researchers have used algorithms such as:

- **Random Forest**
- **Support Vector Machines (SVM)**
- **Artificial Neural Networks (ANN)**
- **Gradient Boosting**

For example, Random Forest models have been used to analyze battery parameters such as voltage, current, temperature, and charging cycles to predict SOH with improved accuracy and reliability. These models can handle complex datasets and identify the most influential parameters affecting battery degradation.

Machine learning methods have demonstrated strong predictive capabilities, making them suitable for real-time EV battery monitoring systems.

Deep Learning–Based Battery Health Prediction

With the growth of large battery datasets, **deep learning techniques** have become increasingly popular. Deep learning models can automatically extract relevant features from battery operational data and provide high-accuracy predictions.

Recent research has proposed hybrid deep learning frameworks combining multiple architectures such as:

- Convolutional Neural Networks (CNN)
- Long Short-Term Memory (LSTM)
- Temporal Convolutional Networks (TCN)
- Attention mechanisms

These models analyze features derived from battery charge-discharge cycles, including differential voltage and incremental capacity curves. Hybrid architectures have shown high prediction accuracy, achieving significant improvements compared with traditional machine learning approaches.

Similarly, advanced models such as **DCRNN combined with SVM-RFE feature selection** have demonstrated low prediction errors and high reliability in SOH estimation using public battery datasets.

Hybrid and Explainable AI Models

Recent studies emphasize **hybrid models** that integrate machine learning, deep learning, and explainable AI techniques. For instance, hybrid frameworks combining CNNs, GRUs, transformers, and attention mechanisms can capture both spatial and temporal patterns in battery data.

Such models have shown exceptional performance, achieving very low prediction errors and strong generalization across different battery datasets. Additionally, explainable AI techniques such as SHAP analysis help identify key operational parameters that influence battery degradation.

Hybrid approaches are considered promising because they combine the interpretability of physics-based models with the predictive power of data-driven methods.

Key Operational Parameters Used in Battery Health Prediction

Operational parameter analysis plays a critical role in battery health prediction models. The most commonly used parameters include:

- Voltage and current profiles
- Battery temperature
- State of Charge (SOC)
- Charge/discharge cycles
- Depth of discharge
- Charging rate and driving patterns

These parameters provide valuable information about battery degradation patterns and help improve prediction accuracy in machine learning models.

Research Challenges and Future Directions

Despite significant progress, several challenges remain in EV battery health prediction:

1. **Limited real-world datasets** for training accurate prediction models.
2. **Variability in operating conditions** such as temperature and driving behavior.
3. **Difficulty in integrating models into real-time Battery Management Systems (BMS).**
4. **Generalization issues** across different battery chemistries and manufacturers.

Future research is expected to focus on:

- Hybrid physics-informed machine learning models
- Transfer learning for cross-battery prediction
- Real-time edge-based battery health monitoring
- Integration of AI techniques into EV battery management systems.

Conclusion:

Advanced battery health prediction techniques using operational parameter analysis have evolved from traditional electrochemical models to sophisticated machine learning and deep learning approaches. Data-driven models provide high accuracy and adaptability, making them suitable for real-time EV applications. Hybrid AI models that integrate multiple techniques and operational parameters are expected to play a crucial role in improving the reliability and efficiency of future electric vehicle battery management systems.

3.SYSTEM ANALYSIS

EXISTING SYSTEM

Existing approaches for battery health prediction can be categorized into model-based, data-driven and hybrid [4]. Model-based methods describe the battery dynamic and degradation behaviour through mathematical or physics-based models. Empirical and semi-empirical models are two widely used mathematical models that fit the battery capacity curve using some influencing factors, such as temperature, rate, and state of charge (SOC) [5]. In contrast, Electrochemical and Equivalent Circuit models (ECM), commonly employed physical models, utilize circuitry elements to simulate the reaction process inside the battery [6]. Although model-based methods have achieved remarkable progress in battery degradation prediction, some obstacles restrict their applications on in-service vehicles.

- First, battery degradation results from diverse, interlaced and nonlinear degradation mechanisms. No single physics-based model can describe all the mechanisms comprehensively, nor can an effective approach quantify them separately [7].
- Second, the modelling process of both mathematical and physics-based methods necessitates complete battery tests and involves lots of model parameters. Nevertheless, comprehensive testing is not feasible for in-service battery packs, and a significant number of the model parameters cannot be observed or calibrated using the existing onboard sensing technologies.

Additionally, the primary limitation of model-based approaches is that they are developed under well controlled conditions, which results in a lack of extrapolation ability when applied to complex working conditions.

DISADVANTAGES

- The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets for Battery Degradation Prediction.
- Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.
- Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

PROPOSED SYSTEM

In the proposed system, a feasible framework that specifically constructs a feature pool by recognizing the primary charging pattern of vehicles in operation, and then integrates transfer learning with the neural network for predicting battery health across various vehicle communities. The main contributions are summarized as follows:

1. Proposing a cluster-based approach for identifying charging patterns, which provides the orientation for subsequent feature extraction and enhances the utilization and processing efficiency of field data.
2. Introducing a multi-level feature selection strategy to eliminate irrelevant and redundant features, achieving comprehensive and effective feature pool construction.
3. Integrating LSTM with transfer learning to realize the historical capacity curve reconstruction and degradation prediction for vehicles across different scenarios.
4. The impact of training and adaptation dataset scale on model accuracy and efficiency is systematically analyzed in real-world scenarios.

ADVANTAGES

In the proposed system, a primary charging pattern identification approach is proposed based on the systematical analysis in . Subsequently, feature engineering involving feature extraction and multi-level screening is performed for effective feature pool construction. Thirdly, neural networks are proposed to reconstruct historical capacity curves and predict further degradation of vehicles in TD. Additionally, the transfer learning technique is utilized to improve the accuracy of the models.

IMPLEMENTATION

MODULES

SERVICE PROVIDER

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Upload Datasets Train & Test Datasets, View Trained and Tested Datasets Accuracy in Bar Chart, View Trained and Tested Datasets Accuracy Results, View Prediction Of Battery Degradation Details, View All Uploaded Datasets Download Predicted Data Sets, View All Remote Users.

VIEW AND AUTHORIZE USERS

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

REMOTE USER

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT BATTERY DEGRADATION STATUS, VIEW YOUR PROFILE.

ALGORITHMS

DECISION TREE CLASSIFIERS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class
Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

GRADIENT BOOSTING

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

K-NEAREST NEIGHBORS (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

LOGISTIC REGRESSION CLASSIFIERS

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

NAÏVE BAYES

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).

While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique.

Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.

An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance.

Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space. SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

CONCLUSION Accurate health prognostics are crucial for effective status management and maintenance decisions in lithium-ion batteries. However, existing health prediction methods require advanced knowledge of loading profiles to extract various features and are primarily developed under constant operating conditions, posing significant challenges for practical application. In this study, we propose a comprehensive health prognostic framework for in-service EVs. This framework begins with the introduction of a charging pattern identification method to recognize the primary charging patterns of EVs, followed by the targeted extraction of both instantaneous and statistical features. Then, a multi-level feature selection strategy is employed to construct a robust feature pool by eliminating redundant and irrelevant features. Building on this foundation, two multi-layer neural network models are designed to reconstruct historical capacity trajectories and predict degradation. To improve the generalization ability of these models in unknown scenarios, the pre-trained models are refined using limited data samples collected from the offline maintenance processes of vehicles in the TD. The proposed framework is

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

evaluated under various scenarios, considering different combinations of training sets, sampling intervals, and training / predicting steps. Results demonstrate that the refined capacity estimation model is superior in reconstructing historical capacity trajectories, even with a smaller training set and limited data samples. Consequently, the prediction model refined using the reconstructed capacity trajectory shows optimal accuracy when performing the health prediction task in TD, with mean MAPE and RMSE within 1.30% and 3.05 Ah (2.03%), respectively. This represents an absolute error reduction of over 3.63% and a maximum relative error reduction of over 73.63% compared to traditional neural network based methods. Moreover, the proposed framework achieves an excellent trade-off between accuracy and efficiency, with the time required for model refining being significantly lower than that for initial modeling, underscoring its practical applicability. The relevant results for vehicles charged with the CC-CV pattern also validate the superiority and feasibility of the proposed framework. Under the guidance of the proposed framework, and with the extensive data resources and powerful computing capabilities of cloud platforms, the seamless integration of online and offline data could be facilitated, thereby achieving refined vehicle management. Despite these advancements, several challenges persist. These include bridging the data barrier between online and offline systems, ensuring the framework's adaptability to diverse EV models and battery chemistries, and addressing concerns related to data quality and security. We remain optimistic that through continued research and development, these issues can be effectively addressed, paving the way for the successful large-scale implementation of the framework within the EV industry.

REFERENCES

[1] H. He, F. Sun, Z. Wang, C. Lin, C. Zhang, R. Xiong, J. Deng, X. Zhu, P. Xie, S. Zhang, Z. Wei, W. Cao, and L. Zhai, "China's battery electric vehicles lead the world: achievements in technology system architecture and technological breakthroughs," Green Energy and Intelligent Transportation, vol. 1, no. 1, p. 100020, 2022.

[2] Z. Mao, X. Gu, J. Li, K. Liu, T. Wang, and Y. Shang, "An applicable minor short-circuit fault diagnosis method for automotive lithium-ion
batteries based on extremum sample entropy," IEEE Transactions on Power Electronics, vol. 39, no. 4, pp. 4636–4644, 2024.

[3] Y. Li, K. L. Liu, A. M. Foley, A. Zulke, M. Berecibar, E. Nanini-Maury, J. Van Mierlo, and H. E. Hoster, "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review," Renewable & Sustainable Energy Reviews, vol. 113, 2019, 109254.

[4] X. Hu, L. Xu, X. Lin, and M. Pecht, "Battery lifetime prognostics," Joule, vol. 4, no. 2, pp. 310–346, 2020.

[5] I. Baghdadi, O. Briat, J.-Y. Del´etage, P. Gyan, and J.-M. Vinassa, "Lithium battery aging model based on dakin's degradation approach,"
Journal of Power Sources, vol. 325, pp. 273–285, 2016.

[6] C. Lyu, Q. Z. Lai, T. F. Ge, H. H. Yu, L. X. Wang, and N. Ma, "A leadacid battery's remaining useful life prediction by using electrochemical
model in the particle filtering, framework," Energy, vol. 120, pp. 975–984, 2017.

[7] D. Zhang, Z. Wang, P. Liu, C. She, Q. Wang, L. Zhou, and Z. Qin, "A multi-step fast charging-based battery capacity estimation framework of real-world electric vehicles," Energy, vol. 294, p. 130773, 2024.

[8] L. Zhou, Z. Zhang, P. Liu, Y. Zhao, D. Cui, and Z. Wang, "Data-driven battery state-of-health estimation and prediction using ic based features and coupled model," Journal of Energy Storage, vol. 72, p. 108413, 2023.

[9] I. Bloom, J. Christophersen, and K. Gering, "Differential voltage analyses of high-power lithium-ion cells 2. applications," Journal of Power Sources, vol. 139, no. 1-2, pp. 304–313, 2005.

[10] D. Zhang, Z. Wang, L. Peng, Z. Qin, Q. Wang, C. She, and P. Bauer, "Multi-step fast charging based state of health estimation of lithium-ion batteries," IEEE Transactions on Transportation Electrification, pp. 1–1, 2023.

[11] 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. Bazan,
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, 2019.

[12] J. Tian, R. Xiong, W. Shen, J. Lu, and X.-G. Yang, "Deep neural network battery charging curve prediction using 30 points collected in 10 min," Joule, vol. 5, no. 6, pp. 1521–1534, 2021.

[13] C. Zhang, L. Luo, Z. Yang, B. Du, Z. Zhou, J. Wu, and L. Chen, "Flexible method for estimating the state of health of lithium-ion batteries using partial charging segments," Energy, vol. 295, p. 131009, 2024.

[14] J. Meng, M. Yue, and D. Diallo, "A degradation empirical-model-free battery end-of-life prediction framework based on gaussian process
regression and kalman filter," IEEE Transactions on Transportation Electrification, vol. 9, no. 4, pp. 4898–4908, 2023.

[15] J. Meng, T. Azib, and M. Yue, "Early-stage end-of-life prediction of lithium-ion battery using empirical mode decomposition and particle

filter," Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, vol. 237, no. 5, pp. 1090–1099, 2023.