Received 31 December 2023; revised 3 June 2024 and 15 July 2024; accepted 16 July 2024. Date of publication 22 July 2024; date of current version 31 July 2024.
Digital Object Identifier 10.1109/OJITS.2024.3430843

A Novel Federated & Ensembled Learning-Based Battery State-of-Health Estimation for Connected Electric Vehicles

PRAVEEN ABBARAJU (Member, IEEE), AND SUBRATA KUMAR KUNDU (Member, IEEE)
Advanced Technology Development Department, Hitachi Astemo Americas, Inc., Farmington Hills, MI 48335, USA
CORRESPONDING AUTHOR: P. ABBARAJU (e-mail: praveen.abbaraju.oj@hitachiastemo.com)

ABSTRACT Electric vehicles (EV) are gaining wide traction and popularity despite the operational range and charging time limitations. Therefore, to ensure the reliability of EVs for realizing improved customer satisfaction, it is necessary to monitor and track its battery condition. This paper introduces a novel federated & ensembled learning (FEL) algorithm for precise estimation of battery State of Health (SoH). FEL algorithm leverages real-world data from diverse stakeholders and geographical factors like traffic and weather data. A Long-Short Term Memory (LSTM) model has been implemented as a base-model for SoH estimation, continuously updating for each trip as an edge scenario using data-centric federated learning strategy. A stacked ensemble learning algorithm is employed to combine data from heterogenous data sources for retraining the base-model. The effectiveness of the proposed FEL algorithm has been evaluated using NASA battery dataset, showing significant improvement in SoH estimations with a mean average error of 3.24% after 30 iterations. Comparative analysis, including LSTM model with and without ensembled stakeholder data, reveals up to 75% accuracy improvement. The proposed model-agnostic FEL algorithm shows its effectiveness in precise SoH estimation through efficient data sharing among stakeholders and could bring significant benefits for realizing data-centric intelligent solutions for connected EVs.

INDEX TERMS Data-centric AI, federated learning, state of health (SoH), connected vehicles.

I. INTRODUCTION

ELECTRIC VEHICLE (EV) adoption is widely advocated as a sustainable mode of transportation that is cleaner and greener, marching us towards the carbon neutral society [1]. The adoption of EVs is outpacing the alternative options such as hydrogen-powered vehicles due to several factors, including manufacturing simplicity and maintenance processes, as well as the affordability to materials for energy sources. One of the fundamental components of EVs is their batteries, which serves as a finite energy source. However, these batteries undergo degradation over multiple charging and discharging cycles before reaching their end of life [2]. Moreover, external factors such as operating temperature and humidity can significantly impact the battery health and its ability to retain maximum capacity [3]. Therefore, precise monitoring of battery health, while considering these influential factors, is significant for ensuring longevity of battery and optimal operational efficiency.

The review of this article was arranged by Associate Editor Goncalo Correia.

The State of Health (SoH) of a battery is a metric to determine its degradation throughout its operational lifespan [2], [4]. Precise estimation and tracking of battery SoH in real-time is critically important for various automotive applications. Moreover, accurate estimation of battery SoH improves overall battery state estimations including State of Charge (SoC) and State of Power (SoP). Although the accurate estimation and tracking of battery SoH significantly helps to realize improved operational efficiency, the inherent non-linearity of battery model coupled with unaccounted external factors poses significant challenge with accurately estimating the battery SoH.

The estimation of SoH relies on three primary parameters: battery capacity, impedance, and internal resistance. As the battery ages, these parameters undergo changes and impact the operational lifetime. According to the EV norms, an EV battery capacity may degrade up to 20%, while internal resistance can surge up to 160%, which signifies its end-of-life (EoL) [5]. In this paper, the SoH is defined as the percentage ratio of available capacity to the rated capacity. When the battery SoH drops below 80%, the battery has reached its EoL. Therefore, monitoring real-time conditions influencing these parameters are essential for accurate tracking of battery SoH. Traditional experimental-based SoH estimation methods provide precise tracking of battery parameters but are primarily designed in controlled laboratory settings [6], [7]. However, data collection from an EV to track its battery state considering such controlled laboratory experimental setup is inefficient and expensive, and therefore not suitable for real-world tracking. Alternatively, model-based and machine learning-based methods are comparatively better suited for EV battery SoH estimation [3]. However, the accuracy of health indicators that reflect battery SoH are affected by the selection of model inputs. Therefore, the key features such as voltage, current, and temperature from both charging and discharge data are employed for improved accuracy in battery SoH estimation [8].

Machine learning-based methods combines aspects of both model-based & experimental methods. Numerous reviews articles extensively evaluate the advantages and disadvantages of different SoH estimation models [6], [9], [10], [11]. However, the efficacy of such models is contingent upon quality, diversity and quantity of the training data. This paper introduces a novel federated & ensemble learning based battery SoH estimation algorithm for realizing improve accuracy leveraging data from multiple stakeholders. By integrating real-time data from stakeholders, this method ensures acquisition of high-quality, diversity, and quantity of data. Therefore, the contributions of this paper are as follows:

- Novel federated & ensembled learning algorithm for SoH estimation leveraging data from multiple stakeholders in real-time which are based on discharging and charging the battery during the course of its usage.
- Identify and verify the significance of stakeholders based on the provided data and improved accuracy of estimations.

The proposed battery SoH estimation, based on federated & ensembled learning algorithm, captures the essential characteristics of EV batteries. This approach enables accurate estimation or prediction of battery aging under diverse operating conditions. Section II provides a detailed report on the literature corresponding to SoH estimations. The proposed federated & ensembled learning algorithm is described in Section III. Section IV provides a real-world example of EV operation that formulates the problem statement. The results generated based on NASA battery datasets are evaluated and compared in Section V. Section VI shows the effectiveness of the proposed algorithm, concludes the paper, and provides the future directions.

II. RELATED WORKS

In the literature, battery SoH estimation methods are typically categorized into two primary groups: model-based methods and data-driven methods. Model-based methods have been primarily explored using electrochemical models, empirical/ semi-empirical models, and equivalent circuit models. Electrochemical model-based methods rely on the description of battery internal electrochemical processes using first-principles equations, to accurately calculate the SoH [9]. However, their real-time applicability is limited by high computational overheads. Several methods have been proposed within the battery equivalent circuit model-based methods, employing electric models such as RC equivalent circuit model in [10], fractional order equivalent circuit model in [11], and impedance spectrum growth model in [12]. These methods use data derived from the battery equivalent circuit models to estimate SoH parameters through filtering algorithms [13]. Such methods could be employed for real-time computations due to the low computational cost. On the other hand, empirical and semi-empirical models are favored for their simplicity and computation efficiency in real-time applications. Data on the loss of capacity and increasing internal resistance with respect to time or battery cycles are fitted using particle filter [24], and particle swarm optimization (PSO) algorithm [17], etc. However, such data is often subjected to noise, impacting the accuracy, and robustness of the estimation models. Additionally, the fitted models may not be generalizable to all battery types and requires individualized attention, increasing the cost associated with data collection and resources.

In recent years, data-driven methods based on artificial intelligence (AI) are receiving more attention due their popularity in providing accurate estimations with fewer real-time computation requirements. Battery operational metrics such as voltage, current, and temperature serve as inputs for machine learning algorithms targeted for monitoring SoH and provide prognostic estimates [15], [16]. The accuracy of these SoH estimates is dependent on the quality of corresponding health indicators (HIs) data [14], [20]. Such HIs data can be acquired through direct extraction methods based on measured variables or indirect extraction methods based on calculated variables.

The literature presents numerous machine learning algorithms targeted for monitoring the EV battery SoH. Early research was focused on energy management strategies for fuel cell hybrid electric vehicles, employing a support vector regression model to estimate battery Remaining Useful Life (RuL) [14]. In the case of Lead Acid batteries, SoH estimation was conducted by combining EIS (Electrochemical Impedance Spectroscopy) and fuzzy logic data analysis [20]. For Lithium-Ion batteries, gaussian process algorithm was proposed for SoH estimation using the training data from WLTC (worldwide light duty driving test cycle) profiles [21], [22]. However, Neural Networks (NN) algorithms have gained popularity among machine learning techniques. Back Propagation Neural Network (BPNN) was introduced using the battery internal parameters and interval capacity [17], [18]. Additionally, a single-layer feedforward NN was developed in [19], with increased operational speed and estimation accuracy compared to BPNN. Another variant of NN, based on long-short-term memory (LSTM) demonstrated superior accuracy in tracking battery SoH [23]. Compared to the data-driven models, the LSTM NN algorithm demonstrated lower average root mean square, 0.0216, for SoH estimations and conjunct error, 0.0831, for RuL predictions. Further, a particle filter-based algorithm was proposed for SoH estimations [24]. It's important to note that the accuracy of these machine learning algorithms heavily relies on the availability of battery operational data under diverse conditions. However, Procuring real-world operational data remains a challenge due to heterogeneity in operational conditions and data sources.

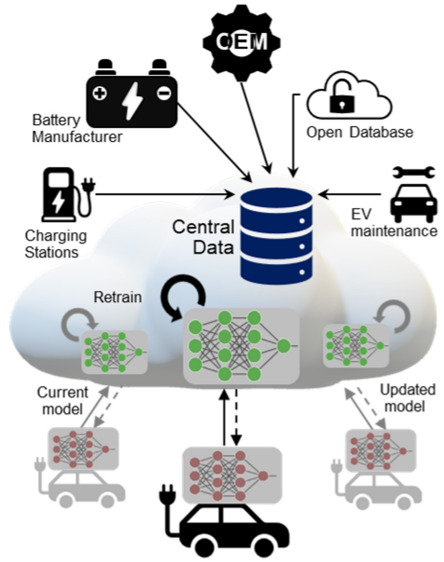

FIGURE 1. Data-Centric Federated Learning architecture to update ML models using the data from multiple stakeholders.

III. BATTERY SOH ESTIMATION ALGORITHM

The EV ecosystem benefits from end-to-end information sharing among major players within mobility, energy, infrastructure, and policymakers [25], [26]. Among these stakeholders, which directly or indirectly influence battery operations, are battery manufacturers, charging stations, weather & traffic, etc. Battery manufacturers are significant stakeholders who provide initial estimation models. Other stakeholders contributing to battery degradation have the option to share charging and discharging datasets or locally-trained Machine Learning (ML) models. However, leveraging these datasets and different ML models poses challenges due to the heterogeneity in data and models. To address this, a novel Federated & Ensembled Learning (FEL) algorithm is introduced to effectively combine heterogeneous data and local ML models, thereby generating diverse and large quantities of data. This data is subsequently used by EVs to retrain their SoH estimation models tailored to their operations, as depicted in Fig. 1.

The proposed FEL algorithm is designed to be model-agnostic, ensuring data quantity and diversity to refine the SoH estimation model and improve tracking accuracy. Since the SoH estimation can be characterized as a time-series processing problem, an LSTM based deep neural network is selected as the base model to evaluate the effectiveness of the proposed FEL algorithm [3], [15].

A. ENSEMBLED LEARNING ALGORITHM

Stakeholders have the option to contribute either local datasets or locally-trained ML models. Directly shared datasets undergo a conversion process to derive charging and discharging profiles from battery current, voltage, and temperature curves. Whereas, sharing locally-trained ML models provides learnings derived from large datasets of each stakeholder. However, to effectively leverage these learnings from the heterogeneous pool of ML models, a meta learning technique called ensembled learning algorithm is employed to generate estimations. The resultant data from estimations of ensembled models is combined with the directly shared datasets to form a new pool of data. This new dataset is then utilized for retraining SoH estimation models.

Stakeholders within the EV ecosystem, including charging stations, EV maintenance companies, and fleet operators, possess extensive historical datasets of charge and discharge profiles. However, sharing these large datasets is challenging due to increased communication overheads and concerns regarding security and proprietary information. To overcome these challenges, the preferred approach is to share locally trained ML model derived from these extensive datasets.

A stacked ensembled learning algorithm is proposed to combine the predictions from all the ML models, as outlined in Algorithm 1. The heterogeneous ML models, obtained from various stakeholders, serve as the base-level models, denoted as $L_1, \ldots, L_k$. Predictions from these base-level models are generated using the ideal changing and discharging profile data as inputs, which is supplied by the battery manufacturer or OEMs. Subsequently, these predictions are stacked to form the training data for the meta-level model.

A new dataset is constructed as the union of base-level model predictions and converted data from direct datasets, which is also illustrated in Fig. 2. The LSTM-based SoH estimation model is then retrained as the meta-model using this updated dataset. The retrained SoH estimation model is shared back with the EV in a data-centric federated fashion. This integration of federated learning and ensemble learning algorithms are presented in the next subsection.

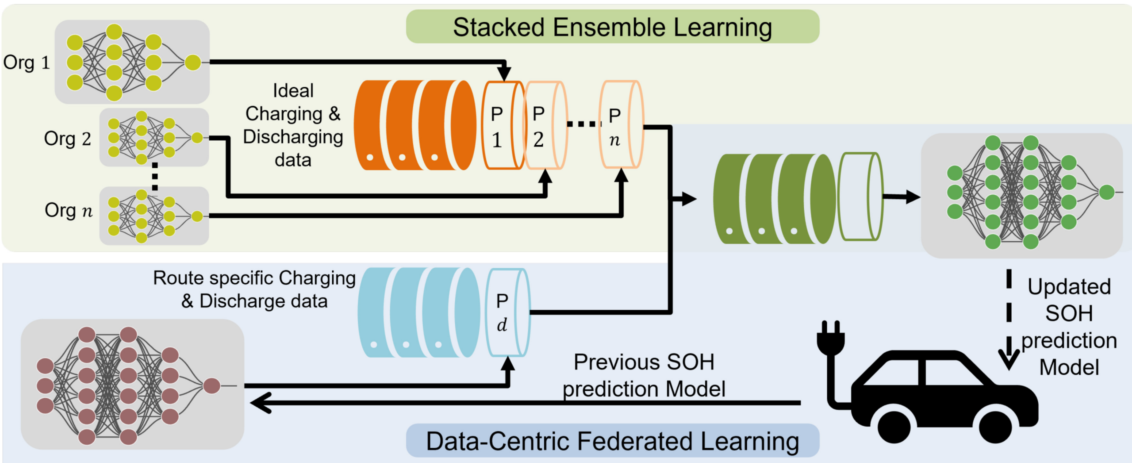

FIGURE 2. Federated & Ensembled Learning model architecture.

Algorithm 1 Stacked Ensemble Learning Algorithm

Input: Ideal charge & discharge dataset, $D = \{x_i, c_i\}_{i=1}^{n}$.
Route specific data, D_R.
Base-level ML models, $L_1, \ldots, L_k$.
Previously trained SoH estimation (LSTM) model, L^0.
Output: Ensembled model L^*
BEGIN:
Step 1: Estimates of directly shared dataset using L^0
$D_R^* = L^0(D_R)$
Step 2: Base-level models estimate using ideal dataset
for $i = 1$ to n **do**
  $D_b = \{x_i', c_i\}$, where $x_i' = \{L_1(x_i), \ldots, L_k(x_i)\}$
end for
Step 3: construct new data set of estimations D^*
$D^* = D_b \cup D_R^*$
Step 4: retrain SoH model (LSTM), L^0 as meta-model
Retrain L^0 based on D^* as L^*
return L^*

Algorithm 2 Algorithm for Data-Centric Federated Learning Technique for Updating SoH Estimation Model

Edge EV executes:
Initialize Ω
Select new trip and update Ω with trip info
B = TripUpdate(Ω) from updated trip.
END
TripUpdate(Ω)**:**
$B \leftarrow$ (local model from a new trip)
$\omega \leftarrow$ (current LSTM model)
Initialization (Q =>Query, *evID* =>EV id, *tripID* =>trip id)
Initialize: Q, EVID, TripID
if (Q ==Query, *EVID* == *evID*) **then**
  update $\omega \rightarrow$ (Algo 1) $\rightarrow$ (Step 3)
  if (*tripID* $\leq$ match the value with new trip) **then**
    $B \leftarrow$ (request to update the LSTM model)
  else
    $B \leftarrow$ (query the previous LSTM model)
    (Send message that update is not available)
  end if
  Ask for next query
end if
return B

B. FEDERATED & ENSEMBLED LEARNING ALGORITHM

Federated Learning (FL) is a decentralized approach for training ML models by leveraging data from edge devices [27]. FL can be broadly classified into Model-Centric & Data-Centric approaches [28]. Model-Centric FL is a common technique, where a central model is refined through federated retraining across distributed data centers or devices [29]. Conversely, Data-Centric FL is an emerging technique, where the data is a primary asset and hosted in central server for retraining of edge ML models [30]. In this paper, a data-centric FL algorithm is chosen for retraining the SoH estimation model using the newly ensembled dataset. The Data-Centric FL illustrated in Fig. 1 is used by each EV to retrain its SoH estimation model using ensembled data from diverse stakeholders.

The architecture of the FEL algorithm is shown in Fig. 2, combining principles of data-centric federated learning with stacked ensembled learning algorithms. Algorithm 2 provides detailed steps involved in the data-centric federated learning approach, in which an LSTM model is iteratively retrained as a meta-model using centralized data ensembled from diverse stakeholders. The EV, acting as an edge device, initiates requests to update the SoH estimation model based on the intended trip. The trip details such as the starting and ending locations, current location, and current SoH estimation model are transmitted to the server following a data-centric federated approach. Leveraging the data collected through the stacked ensembled learning algorithm,

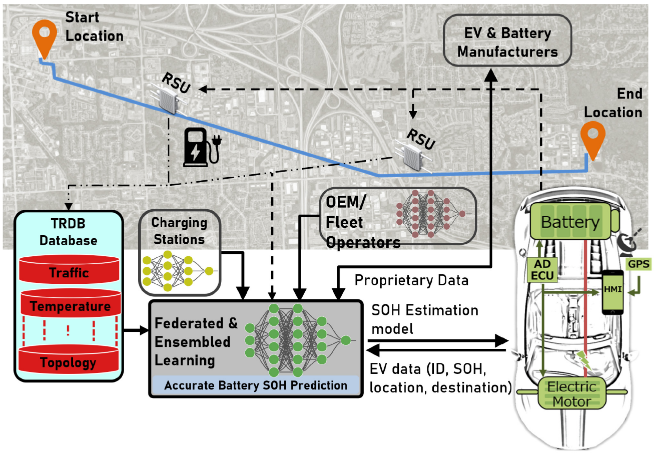

FIGURE 3. **Use-Case scenario: Daily commute route and the data flow among multiple stakeholders.**

the SoH estimation model undergoes meta-model retraining and is shared back with the EV.

Each trip undertaken by the EV is regarded as an edge scenario and the data-centric federated learning algorithm is applied to update the SoH estimation model. The datasets used for retraining are constructed based on the stakeholders involved in each of the trips. Leveraging this unique combination of Federated & Ensembled Learning algorithms, the curation of data is of high quality, quantity, and diverse.

IV. PROBLEM SCENARIO

This section presents a practical example showcasing the effectiveness of the proposed FEL algorithm in conducting meta-analysis on data sourced from diverse stakeholders and providing accurate battery SoH estimates. Consider an EV navigating between origin and destination points (refer to Fig. 3), the FEL algorithm updates the SoH estimation model by ensembling charging and discharging data from relevant stakeholders involved in the trip.

At the start, the EV shares its identification data, along with the current SoH, location, and destination with the service provider. Based on this information, the service provider plans a trip route and executes FEL algorithm to update SoH estimation model provided by the OEM/ fleet operator. Two road-side units (RSUs), one charging station (CS) along the planned route, and one wall charger in home have been considered in this study. The CS stakeholder contributes data when the SoC of the EV battery drops below 20% and requires charging. The RSUs share traffic information, which is then converted into acceleration-deceleration cycles of traffic flow. Moreover, vehicle traffic data within each road topology, sourced from the TRDB (Traffic & Routing Database), is utilized to generate vehicle drive cycles [31]. Weather data such as temperatures are sourced from the CDC's publicly available records. Essential supplementary data, such as battery configuration, initial SoH estimation model parameters, optimal charging and discharging profiles, are provided by battery manufacturers or OEMs.

Ensuring secure and reliable data sharing among stakeholders is significant, while also encouraging their participation. Recent developments in blockchain-based standards are facilitating multi-stakeholders participation in data sharing, thereby ensuring security and integrity, provided stakeholders derive benefits from the applications [32], [33], [34].

V. RESULTS

In this section, the proposed FEL algorithm is evaluated within the context of the problem scenario as described in Section IV. The quality and quantity of data is ensured through the acquisitions from diverse stakeholders and facilitate data exchange using blockchain-based standards. However, the challenge with diverse data araises from the lack of requirements from each stakeholder. Publicly available data such as traffic, temperature, and topology data are readily accessible. This data is converted to vehicle drive cycles that are used to derive the corresponding discharge and charge profiles based on EV & battery configurations. NASA dataset is used to fabricate the data for FEL model training and verification as detailed in Section V-A.

Charging profile data from CS presents a challenge due to the multitude of CS operators processing extensive datasets. In such scenarios, CS operators can share locally trained ML models using historical charging cycle data derived from the battery current, voltage, and temperature curves. However, the heterogenity of the ML model is addressed by employing a weighted stacking ensembled learning algorithm. The optimal charge and discharge profiles provided by the OEM/ battery manufacturers form the inputs of the shared ML models and their SoH predictions stacked. Section V-B details the two different ML models at CS and home charger. The data with charging and discharging profiles along with the predicted SoH values is combined with the dataset from Section V-A to form the training data to retrain the meta-model.

Validations are conducted in MATLAB linked to a private blockchain network. The deep learning toolbox in MATLAB is employed to train all the models, while data is being transmitted between the service provider and various stakeholders. Python scripts are used to represent different stakeholders and simulate data sharing. A private blockchain network was established in local PC using Hyperledger fabric and facilitate data sharing based on Mobility Open Blockchain Initiative (MOBI) standards [33], [35].

A. STAKEHOLDER DATASET

One full cycle of EV battery is defined by 1 full discharging cycle and 1 full charging cycle. The vehicle drive cycles derived from the traffic and topology data are converted to their corresponding discharge and regenerative charge cycles. This data is further segmented at 10 minute intervals to represent the data received from multiple RSUs. This data is

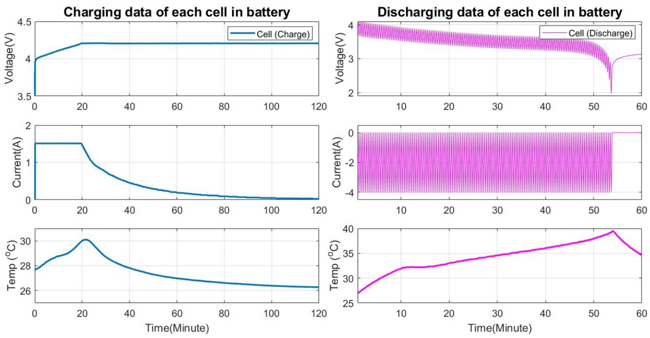

FIGURE 4. Sample dataset of a full cycle of charging and discharging profile data for each cell in battery from NASA dataset.

fabricated from battery data sets that are provided by NASA Prognostics Center of Excellence Data Repository [36].

The NASA battery datasets consist of lithium-ion batteries data that run through three different operational profiles, *i.e.,* charging, discharging, and rest period, at room temperature. Their experiments consisted of applying repeated charging and discharging cycles to commercially available 18650 lithium-ion cells for achieving accelerated aging. Batteries were charged by constant current constant voltage (CCCV) principle; charging at constant current of $1.5A$ until the voltage reaches the cell upper voltage limit of $4.2V$, then applying constant voltage until the current drops to $20mA$. Discharging is done at the constant current of $2A$ until the cell voltage falls to $2.4V$ $2.7V$, $2.5V$, and $2.2V$ for batteries #4, #5, #6, & #7, respectively. The experiments were performed until the batteries lose 30% of the rated capacity, *i.e.,* $1.4Ah$. Additional electrochemical impedance data are also provided in this dataset, but not used in this work. Fig. 4, shows sample dataset of charging and discharging data of each cell in a battery.

B. ML MODELS PARAMETERS

The SoH estimation is based on the LSTM model, that is retained for every trip as the edge scenario. A common LSTM unit that is composed of a cell, an input gate, an output gate, and a forget gate is adopted. The cell remembers values over arbitrary time intervals and three gates regulate the flow of information in and out of the cells. LSTM defines an internal memory cell state to store long-term information. The memory cell state interacts with the previous output and the following input to determine which elements of internal state vector should be updated, maintained, or erased.

Two different ML models were used for the two charging stations as defined in the use-case scenario. A commonly used feed-forward Neural Networks (FNN) model is used for one charging station. For another charging station, a convolution neural networks (CNN) model, which is a well-known deep neural networks that uses convolution operation in at least one of its layers instead of general matrix multiplication, is used [15]. The model structures and hyper-parameters are summarized in Table 1.

TABLE 1. The structures of learning models for each stakeholder. FC stands for fully connected.

Model	Stakeholder	Model structure
FNN	Charging station 1	Input-Hidden(10 neurons)-Output
CNN	Charging station 2	Input-Conv1-Conv2-FC-Output
LSTM	Service provider (SoH estimation)	Sequence length : 5 Input dimensions : 10(single)

TABLE 2. ML model training performance.

Model	RSME	MAE	MAPE
FNN	0.0736	0.0655	4.71%
CNN	0.0701	0.0623	4.002%
LSTM	0.0288	0.0210	1.377%

Evaluations were carried out to estimate the accuracy of the trained models using Mean Absolute Percentage Error (MAPE) as a representative error index. In addition, the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) were also computed for evaluations. Table 2 lists the capacity estimation performance of each individual models for both charging stations and the base model for SoH estimation.

The trained models were utilized in the FEL architecture without any further optimization of their parameters. This approach was taken to ensure fairness and clarity in the subsequent model comparisons in the next subsection.

C. SOH ESTIMATION RESULTS

The FEL algorithm is implemented using the set of ML models and the battery datasets as explained above in sections. A pretrained LSTM model was used as the baseline for comparison with the updated LSTM model with FEL algorithm. The simulations were conducted on three differently aged batteries at beginning-of-life (BoL) (Battery 1 with 100% SoH), moderately used (Battery 2 & 3 with 90% SoH), and EoL (Battery 4 with 80% SoH). The SoH estimations were determined over 30 cycles of full discharge and charge of the EV battery.

The SoH estimation results shown in Fig. 5, compares the estimation of baseline model with the FEL algorithm estimations. The FEL algorithm estimations start at the true battery SoH values, unlike the baseline estimations that always start with maximum SoH. This is due to the availability of real-world data that steers the battery degradation. Additionally, the FEL estimations showed better estimated towards the EoL, where the SoH estimations are more accurately tracked, in comparison to the baseline estimates.

Analyzing battery degradation is challenging without ground truth data. Therefore, the battery capacity degradation trends were compared for baseline LSTM model with FEL algorithm, in Table 3. The negative slope is due to the degrading nature of batteries. However, the lower absolute value of the degradation trend for FEL algorithm shows the closeness to the linear trends in theoretical models

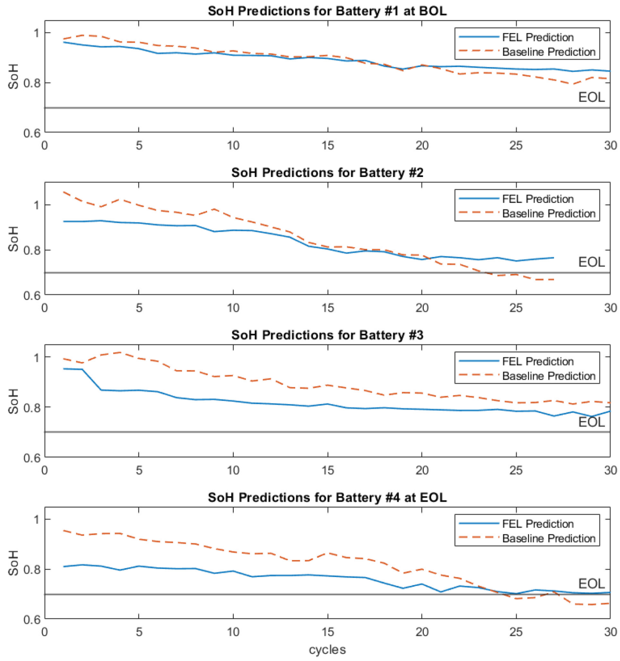

FIGURE 5. Degradation of battery SoH and comparison of trends in FEL estimations and Baseline model estimations.

TABLE 3. Degradation trends of batteries compared for Baseline model and FEL algorithm.

Degradation trends	LSTM (Baseline)	FEL algorithm
Battery 1	-0.015381	-0.008235
Battery 2	-0.006256	-0.003778
Battery 3	-0.006807	-0.004254
Battery 4	-0.010595	-0.004370

of batteries. These results support the claims of improved accuracy of the proposed FEL algorithm compared to the baseline model.

D. SOH ESTIMATION COMPARISON

To further measure the accuracy of the SoH estimations with respect to the real-world conditions, additional simulations were conducted. The moderately used batteries 2 & 3, were chosen for evaluation as the observed trends from Fig. 6, provide good variations and with distinguishable start and end SoH values. The evaluations were conducted using the different discharge profiles from the NASA battery dataset, which also provides battery capacity at the end of each cycle. The ground truth SoH values were generated from the available battery capacity data.

As shown in Fig. 6, the real-world data bring the FEL estimations closer to the ground truth values in comparison to the baseline estimations. The significance of FEL algorithm is evident over higher battery cycles, where the estimations have much higher tracking accuracy. The estimation accuracy is evaluated using the RMSE and MAPE of the estimations from ground truth data and are shown in Table 4. The FEL

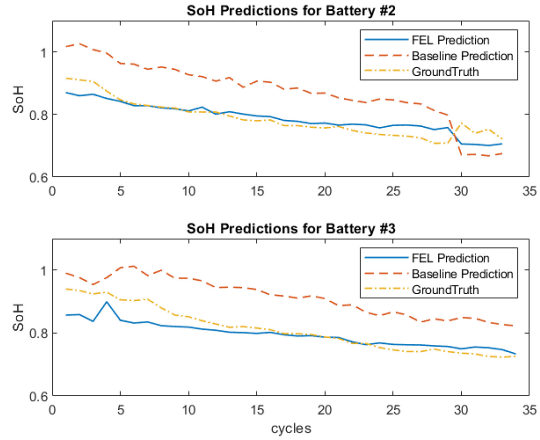

FIGURE 6. Comparison of SoH estimations of baseline model and FEL algorithm with ground truth data.

TABLE 4. Comparison of SoH estimations performance of baseline model and FEL algorithm for batteries 2 & 3.

		LSTM (Baseline)	FEL algorithm
RMSE	Battery 2	-0.01538	-0.00823
RMSE	Battery 3	-0.00625	-0.00377
MAPE (%)	Battery 2	-0.00680	-0.00425
MAPE (%)	Battery 3	-0.01059	-0.00437

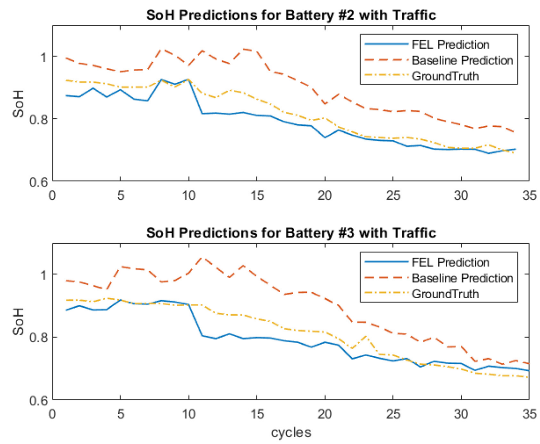

FIGURE 7. Comparison of SoH estimations of baseline model and FEL algorithm with ground truth data with high traffic.

algorithm estimations have significantly lower RMSE and MAPE, in comparison to the baseline model.

E. SOH ESTIMATIONS COMPARISON WITH TRAFFIC DATA

To evaluate the significance of stakeholder data, additional simulation results were generated for real-world conditions with higher traffic situations, where larger variations in charge & discharge profiles can be observed. Sudden accelerations in certain conditions can seriously damage the battery and degrade much faster. Therefore, such scenario data from stakeholders, is used for further evaluations. Fig. 7, plots the FEL algorithm results with traffic data and baseline

TABLE 5. Comparison of SoH estimations performance of baseline model and FEL algorithm for batteries 2 & 3 in high traffic scenario.

		LSTM (Baseline)	FEL algorithm
RMSE	Battery 2	0.0935	0.0379
RMSE	Battery 3	0.1061	0.0374
MAPE (%)	Battery 2	10.59	3.56
MAPE (%)	Battery 3	12.92	3.24

model estimates without traffic data, in comparison to ground truth data. The FEL algorithm has comparatively bettery tacking of battery SoH, over the baseline model.

The RMSE and MAPE of the estimations from ground truth data are shown in Table 5. Despite the higher traffic conditions, the FEL algorithm estimations have significantly lower RMSE and MAPE, in comparison to the baseline estimations. The reduced error in accurately tracking battery SoH highlights the significance of the ensembled data from various stakeholders.

The simulation results presented in this section provide evidence to support the claim that the proposed FEL algorithm achieves higher accuracy for SoH estimations. Higher accuracy was achieved with basic algorithms within FEL algorithm, such as weighted staking EL, and mildly tuned ML models. The proposed FEL algorithm facilitates the quantity and diversity of data to enable higher accuracy, while staying model agnostic.

VI. CONCLUSION

The proposed federated & ensembled learning (FEL) algorithm has demonstrated its efficacy in accurately estimating the battery SoH, empowering EV operators to ensure highest SoH levels at the end of each trip. This accuracy is achieved from leveraging real-time attributes such as traffic, temperature, topology data sourced from various stakeholders. However, a key challenge lies in facilitating data sharing among these stakeholders while addressing proprietary data concerns. While publicly available data such as traffic and temperature pose no acquisition challenges, data from stakeholders such as charging stations, battery manufacturers remain inaccessible. To address these challenges, this paper introduces a novel meta-learning technique fused with federated learning to conduct meta-analysis of extensive datasets. Heterogeneous ML models from stakeholders such as charging stations are combined using the stacked ensemble learning approach. Further, incorporating additional data from RSUs, such as discharge profiles, enhances the retraining process of base model (LSTM model). The updated LSTM model exhibits improved accuracy in SoH estimations compared to the previously trained base model, highlighting the efficacy of the proposed FEL algorithm. Therefore, FEL algorithm ensures the highest quality, quantity and diversity in data that is employed for retraining.

While the study primarily focused on developing and evaluating FEL algorithm for SoH estimation, the insights regarding ML models and emerging technologies require further exploration. The FEL algorithm emphasizes the significance of leveraging both charging and discharging data for SoH estimations. Discharging data offers valuable insights into the battery's performance under different operating conditions, such as fluctuating load demands and discharge rates, which significantly impacts battery health. Additionally, the charging strategies, such as CCCV & MSCC (Multi-Stage Constant Current), have significant influence on battery health. Although not addressed in this paper, further research is needed to understand their impact on SoH estimation. Furthermore, by providing insights into battery degradation based on chosen routes and charging stations, the FEL algorithm enhances transparency in battery usage, addressing range anxiety concerns for both individual EV owners and fleet operators.

ACKNOWLEDGMENT

This study was conducted when the authors worked at the Automotive Products Research Laboratory of R&D Division, Hitachi America, Ltd., Michigan, USA.

REFERENCES

[1] T. Tsukiji, N. Zhang, Q. Jiang, B. Y. He, and J. Ma, "A multifaceted equity metric system for transportation electrification," *IEEE Open J. Intell. Transp. Syst.*, vol. 4, pp. 690–707, 2023, doi: 10.1109/OJITS.2023.3311689.

[2] X. Hu, F. Feng, K. Liu, L. Zhang, J. Xie, and B. Liu, "State estimation for advanced battery management: Key challenges and future trends," *Renew. Sustain. Energy Rev.*, vol. 114, Oct. 2019, Art. no. 109334.

[3] 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, doi: 10.1109/ACCESS.2022.3221137.

[4] X. Wang, X. Wei, and H. Dai, "Estimation of state of health of lithium-ion batteries based on charge transfer resistance considering different temperature and state of charge," *J. Energy Storage*, vol. 21, pp. 618–631, Feb. 2019.

[5] M. Berecibar, I. Gandiaga, I. Villarreal, N. Omar, J. Van Mierlo, and P. Van den Bossche "Critical review of state of health estimation methods of Li-ion batteries for real applications," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 572–587, Apr. 2016.

[6] R. Xiong, L. Li, and J. Tian, "Towards a smarter battery management system: A critical review on battery state of health monitoring methods," *J. Power Sources*, vol. 405, pp. 18–29, Nov. 2018.

[7] S. M. Rezvanizaniani, Z. Liu, Y. Chen, and J. Lee, "Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility," *J. Power Sources*, vol. 256, pp. 110–124, Jun. 2014.

[8] Y. Li, D.-I. Stroe, Y. Cheng, H. Sheng, X. Sui, and R. Teodorescu, "On the feature selection for battery state of health estimation based on charging-discharging profiles," *J. Energy Storage*, vol. 33, Jan. 2021, Art. no. 102122.

[9] H. Meng and Y. F. Li, "A review on prognostics and health management (PHM) methods of lithium-ion batteries," *Renew. Sustain. Energy Rev.*, vol. 116, Dec. 2019, Art. no. 109405.

[10] Y. Li et al., "Data-driven health estimation and lifetime prediction of lithium-ion batteries: A review," *Renew. Sustain. Energy Rev.*, vol. 113, Oct. 2019, Art. no. 109254.

[11] A. Farmann, W. Waag, A. Marongiu, and D. U. Sauer, "Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles," *J. Power Sources*, vol. 281, pp. 114–130, May 2015.

[12] R. Xiong, Y. Zhang, J. Wang, H. He, S. Peng, and M. Pecht, "Lithi-ion battery health prognosis based on a real battery management system used in electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4110–4121, May 2019, doi: 10.1109/TVT.2018.2864688.

[13] L. Cai, J. Meng, D.-I. Stroe, G. Luo, and R. Teodorescu, "An evolutionary framework for lithium-ion battery state of health estimation," *J. Power Sources*, vol. 412, pp. 615–622, Feb. 2019.
[14] A. Widodo, M.-C. Shim, W. Caesarendra, and B.-S. Yang, "Intelligent prognostics for battery health monitoring based on sample entropy," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11763–11769, 2011.
[15] Y. Choi, S. Ryu, K. Park, and H. Kim, "Machine learning-based lithium-ion battery capacity estimation exploiting multi-channel charging profiles," *IEEE Access*, vol. 7, pp. 75143–75152, 2019.
[16] M. Yue, S. Jemei, R. Gouriveau, and N. Zerhouni, "Developing a health-conscious energy management strategy based on prognostics for a battery/fuel cell hybrid electric vehicle," in *Proc. IEEE Veh. Power Propuls. Conf. (VPPC)*, Chicago, IL, USA, 2018, pp. 27–30.
[17] M. Wu, Y. Zhong, J. Wu, Y. Wang, and L. Wang, "State of health estimation of the lithium-ion power battery based on the principal component analysis-particle swarm optimization-back propagation neural network," *Energy*, vol. 283, Nov. 2023, Art. no. 129061.
[18] Q. Yang, J. Xu, X. Li, D. Xu, and B. Cao, "State-of-health estimation of lithium-ion battery based on fractional impedance model and interval capacity," *Int. J. Electr. Power Energy Syst.*, vol. 119, Jul. 2020, Art. no. 105883.
[19] H. Pan, Z. Lü, H. Wang, H. Wei, and L. Chen, "Novel battery state-of-health online estimation method using multiple health indicators and an extreme learning machine," *Energy*, vol. 160, pp. 466–477, Oct. 2018.
[20] P. Singh, S. Kaneria, J. Broadhead, X. Wang, and J. Burdick, "Fuzzy logic estimation of SOH of 125Ah VRLA batteries," in *Proc. 26th Annu. Int. Telecommun. Energy Conf.*, Chicago, IL, USA, 2004, pp. 19–23.
[21] S. Khaleghi, Y. Firouz, J. van Mierlo, and P. van den Bossche, "Developing a real-time data-driven battery health diagnosis method, using time and frequency domain condition indicators," *Appl. Energy*, vol. 255, Dec. 2019, Art. no. 113813.
[22] F. Li, Y. Min, Y. Zhang, Y. Zhang, H. Zuo, and F. Bai, "State-of-health estimation method for fast-charging lithium-ion batteries based on stacking ensemble sparse Gaussian process regression," *Rel. Eng. Syst. Saf.*, vol. 242, Feb. 2024, Art. no. 109787. [Online]. Available: https://doi.org/10.1016/j.ress.2023.109787
[23] P. Li et al., "State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network," *J. Power Sources*, vol. 459, May 2020, Art. no. 228069.
[24] X. Tang, K. Liu, X. Wang, B. Liu, F. Gao, and W. Widanage, "Real-time aging trajectory prediction using a base model-oriented gradient-correction particle filter for Lithium-ion batteries," *J. Power Sources*, vol. 440, Nov. 2019, Art. no. 227118.
[25] F. Bouali et al., "5G for vehicular use cases: Analysis of technical requirements, value propositions and outlook," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 73–96, 2021, doi: 10.1109/OJITS.2021.3072220.
[26] A. Kose, H. Lee, C. H. Foh, and M. Dianati, "Beam-based mobility management in 5G millimetre wave V2X communications: A survey and outlook," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 347–363, 2021, doi: 10.1109/OJITS.2021.3112533.
[27] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. Stat.*, 2016, pp. 1273–1282.
[28] P. Kairouz et al., "*Advances and Open Problems in Federated Learning*. Norwell, MA, USA: Now Publ., 2021.
[29] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," 2019, *arXiv:1902.04885*.
[30] N. Majcherczyk, N. Srishankar, and C. Pinciroli, "Flow-FL: Data-driven federated learning for spatio-temporal predictions in multi-robot systems," 2020, *arXiv:2010.08595*.
[31] J. A. Laval and L. Leclercq, "A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic," *Philos. Trans. Royal Soc.*, vol. 368, no. 1928, pp. 4519–4541, 2010.
[32] R. A. Mallah, D. López, and B. Farooq, "Cyber-security risk assessment framework for blockchains in smart mobility," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 294–311, 2021, doi: 10.1109/OJITS.2021.3106863.
[33] R. Rajbhandari, "Battery state of health (SOH) business white paper," MOBI, Honolulu, HI, USA, White paper MOBI SOH0001, 2022.
[34] R. Rajbhandari, B. Kahya, and M. Shi, "Electric vehicle battery state of health (SOH) value chain map," MOBI, Honolulu, HI, USA, White paper MOBI SOH0003, 2023.
[35] T. Vo. "MOBI (mobility open blockchain initiative) consortium." Sep. 2023. [Online]. Available: https://dlt.mobi/
[36] B. Saha and K. Goebel (NASA AMES Progn. Data Repos., NASA AMES Res. Center, Moffett Field, CA, USA). *Battery Data Set, NASA AMES Prognostics Data Repository*. 2024. [Online]. Available: http://ti.arc.nasa.gov/project/prognostic-data-repository

PRAVEEN ABBARAJU (Member, IEEE) received the M.S. degree in robotics and automation from the Darmstadt University of Applied Sciences, Darmstadt, Germany, in 2014 and the Ph.D. degree in robotics and autonomous systems from Purdue University, West Lafayette, IN, USA, in 2021.

He is currently the part of Advanced Technology Development Team at Hitachi Astemo Americas, Inc., focusing on Automated Driving and the Connected Mobility Ecosystem. Previously, he was a Researcher with the Automotive Products Research Laboratory, Research and Development Division, Hitachi America, Ltd. His research interests include intelligent systems and control, generative AI, and robotics.

SUBRATA KUMAR KUNDU (Member, IEEE) received the M.E. and Ph.D. degrees in robotics and intelligent systems from Saga University, Saga, Japan.

He is currently working as a Director with Hitachi Astemo Americas, Inc. He leads an Advanced Technology Development Team focused on Advanced Driver Assistance Systems, Automated Driving, and Connected Mobility Ecosystem. He was a Senior Manager and the Head of Automotive Products Research Laboratory, Research and Development Division, Hitachi America, Ltd. Prior to working at Hitachi, he worked as a Postdoctoral Fellow with Toyota Technological Institute, Nagoya, and as a Project Assistant Professor with the University of Tokyo, Japan. His research interests include connected automated driving, robotics, and intelligent systems.