ORIGINAL RESEARCH PAPER
Journal Section

Large Language Model-based Methodology for Data-Driven Health Prediction of Lithium-ion Batteries

Suleyman Tuncel[1] | Hasan Cinar[2,3] | Mehmet Gucyetmez[4] | Nuh Erdogan[5*]

[1]Electronics and Communication Engineering, Turkish Military Academy, National Defence University, Ankara, 06420, Türkiye.

[2]Department of Aeronautical Engineering, Necmettin Erbakan University, Konya, 42140, Türkiye.

[3]Faculty of Engineering and Applied Sciences, Centre for Autonomous and Cyber-Physical Systems, Cranfield University, Cranfield, UK

[4]Electrical and Electronics Engineering, Sivas Science and Technology University, Sivas, Turkey.

[5]School of Science and Technology, Nottingham Trent University, Nottingham, NG11 8NS, U.K.

Correspondence
Department of Engineering, School of Science and Technology, Nottingham Trent University, Nottingham, NG11 8NS, U.K.
Email: nuh.erdogan@ntu.ac.uk

Funding information
-

Accurate prediction of lithium-ion battery health is critical for the performance and safety of electric vertical takeoff and landing (eVTOL) vehicles. Traditional machine learning approaches require significant expertise in data preprocessing and model development, which limits their accessibility. This study introduces an innovative large language model (LLM)-based technique to automate the implementation and optimization of machine learning algorithms for battery state-of-health (SOH) forecasting. The proposed framework integrates ChatGPT into the complete machine learning pipeline, including data pre-processing, determining importance of characteristics, model recommendation and selection based on learning from reference studies, hyperparameter tuning, and performance evaluation. The LLM-driven approach involves iterative refinement of the model through structured prompts, ensuring continuous improvement and adaptation to the specific requirements of the SOH estimation. The study utilized a publicly available dataset of a lithium-ion battery used in the propulsion system of an eVTOL vehicle, which includes comprehensive flight missions and structured charge-discharge cycles. Three machine learning algorithms, i.e., Random Forest, XGBoost, and

CatBoost, were implemented and optimized using ChatGPT. The performance of the LLM-driven models was benchmarked against conventional methods, demonstrating a 52% reduction in Mean Absolute Percentage Error (MAPE) compared to traditional approaches. The findings highlight the potential of LLM-driven machine learning in enhancing battery health prediction, making advanced techniques more accessible to a broader audience. This study demonstrates that integrating ChatGPT into the machine learning workflow can significantly improve the accuracy and efficiency of SOH estimation for eVTOL applications.

KEYWORDS
battery health estimation, generative AI, large language model, eVTOL, machine learning, prompt engineering

1 | INTRODUCTION

Vertical Take-off and Landing (eVTOL) vehicles employed in urban air mobility operate with fully electric propulsion systems that utilize solely lithium-ion batteries or hybrid energy sources, thus eliminating the need for traditional fossil fuel engines [1, 2]. Among the various challenges facing the battery management system (BMS), two major concerns include the monitoring and estimation of essential battery performance parameters [3]. The BMS is essential for ensuring safe and reliable operations, extending operational life, and reducing overall costs [4, 5]. It is responsible for monitoring battery operating parameters such as voltage, current, and temperature, as well as managing battery degradation [6]. The state of health (SOH), which reflects long-term battery degradation, is a key performance indicator. Data-driven models, which are often preferred for estimating battery performance, can better manage complex nonlinear behaviors compared to electrochemical or equivalent circuit models, thereby providing improved efficiency [7]. Recent advances in battery performance estimation have introduced various techniques, including Kalman filters [6, 7] and hybrid support vector machines [4]. While these methods demonstrate improved accuracy, they typically require: *(i)* manual tuning of complex parameters (e.g., noise covariance matrices in Kalman filters), *(ii)* expert knowledge for feature engineering, and *(iii)* computationally intensive optimization processes. These limitations constrain their practical implementation, particularly in dynamic eVTOL applications where rapid, automated decision-making is crucial. In this regard, large language models (LLMs) have emerged as advanced machine learning (ML) models capable of understanding, generating, and interacting with human language [8].

Large language models (LLMs) have seen significant advancements in recent years. ChatGPT, developed by OpenAI [9], is one such implementation of the Generative Pre-trained Transformer (GPT) series of LLMs. Built on the principles of prompt engineering [10], ChatGPT leverages the capabilities of Generative Artificial Intelligence (AI) to simulate human-like interactions, understanding speech, and executing commands as instructed [11, 12]. Prompt engineering, a systematic approach involving conditions or rules, addresses the challenges faced by conventional AI in emulating human creativity, particularly within the emerging concept of generative AI [13]. By providing structured prompts or instructions, prompt engineering guides the AI model's output towards desired outcomes, thereby enhancing its ability to generate creative and contextually relevant content [14]. This iterative process allows AI to continuously refine and enhance its performance, enabling it to build upon previous levels of intelligence rather than starting from scratch when transferred to another system [15]. Recently, LLMs like ChatGPT have gained popularity across diverse applications, including solving mathematical equations [16], generating academic and literary content [17], debugging software [18], performing text classification [19], and automating code generation [20].

When integrated with ML techniques, LLM has the potential to provide solutions for engineering problems, thus enhancing applications and expanding the capabilities of electrical engineering systems. Bonadia et al. [21] investigate the potential of ChatGPT to generate distribution text networks for power flow studies, demonstrating that some user knowledge is required to effectively leverage ChatGPT in detecting and solving power distribution network problems. Huang et al. [22] used ChatGPT for fine-tuning pre-trained models by adopting the Knowledge Graph Completion approach to diagnose defects in the main electrical equipment of the power grid. In another study [23], a transformer-based model was used in wind power forecasting, showing that ChatGPT is effective in capturing complex temporal relationships in large-scale time series data. He et al. [24] evaluated the development of ChatGPT in robots, considering robot perceptions such as visual, auditory, and tactile, as well as intelligences such as linguistic, logical-mathematical, and spatial. Li et al. [25] utilized ChatGPT to solve several power engineering problems, including unit commitment and decentralized optimization of multi-vector energy systems. Zhang et al. [26] revealed the potential vulnerability of LLMs such as ChatGPT in smart grid applications. Recent studies demonstrate that ChatGPT can make ML techniques and tools easier and more efficient, making them accessible to individuals without a deep background in ML or programming. ChatGPT can save time and effort by developing and implementing ML algorithms, preprocessing data for model training and testing, and identifying and fixing errors in code. Additionally, it can create user-friendly interfaces and simplify complex processes, enabling a broader audience to apply ML to solve real-world problems without requiring specialist knowledge or coding skills.

Table 1 summarizes recent studies on the application of LLMs in battery monitoring and prognostics. primarily focus on monitoring, feature selection, and estimating key battery health indicators such as state of charge (SOC), SOH, and remaining useful life (RUL). However, relatively few studies provide a comprehensive approach to SOH estimation, particularly in the context of dynamic operational conditions. In electric propulsion systems, batteries experience highly variable demand loads due to fluctuating flight conditions, making accurate health monitoring crucial. Most existing research has focused on integrating SOC and SOH metrics into trajectory planning and energy management for health-aware electric aircraft [27]. These developments have highlighted the importance of accurate battery health assessments in enhancing aircraft performance and safety. To address this need, this study utilizes LLMs to improve the accuracy and robustness of SOH prediction for eVTOL applications.

This study proposes an innovative LLM-driven framework that automates the entire ML process from feature selection to hyperparameter optimization. Specifically, we evaluate the capability of ChatGPT to implement data-driven machine-learning algorithms for forecasting the state of health of Lithium-ion (Li-ion) batteries in eVTOL applications. Our methodology advances conventional approaches by demonstrating that LLMs can: *(i)* systematically explore parameter spaces through prompt-guided optimization, *(ii)* adaptively refine models based on performance feedback, and *(iii)* generate executable code without requiring deep programming expertise. Recent research has highlighted the importance of SOH prediction for optimizing eVTOL performance, with studies employing various algorithms including Multi-layer Perceptron, Support Vector Regression (SVR), Random Forest (RF), Gaussian Process Regression (GPR), Extreme Gradient Boosting (XGBoost), and CatBoost—to forecast key battery parameters such as SOH, RUL, and maximum operating temperature (MOT) [36, 37, 38, 39]. Building on these advancements, this study uses ChatGPT 4.0, guided by prompt engineering, to implement and optimize the Random Forest, XGBoost, and CatBoost algorithms for SOH forecasting. A feature importance analysis is performed to evaluate the effectiveness of ChatGPT in identify-

TABLE 1 Studies on the application of LLMs in battery monitoring and prognostics.

Ref.	Purpose	Highlights
[26]	RUL estimation	improving prediction accuracy according to common ML algorithms
[28]	Prognostics	combines the local knowledge method and large language model
[29]	SOC estimation	more accurate and robust estimates with a new soft prompt adapter
[30]	battery management	introduces the concept of Internet of Batteries in EVs
[31]	SOC estimation	a hybrid prompt-driven large language model
[32]	RUL estimation	a SHAP analysis based on large language model
[33]	SOC estimation	a prompt-driven fine-tuning method
[34]	SOH and RUL estimations	a transformer-based LLM framework
[35]	SOH estimation	innovative feature engineering technology

ing critical features. Using publicly available eVTOL data, the models are trained, tested, and compared with existing methods. Their precision is evaluated using mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) to assess forecasting accuracy. This approach not only demonstrates the potential of ChatGPT in automating ML workflows, but also provides a robust framework to improve battery health prediction in eVTOL applications. The proposed approach not only makes battery health monitoring more accessible and scalable but also maintains high accuracy, as demonstrated by our experimental results showing 52% improvement in MAPE compared to traditional methods. The main contribution of this study is the development of a structured, prompt-driven methodology that leverages LLMs (specifically ChatGPT) to automate the entire SOH estimation process. This approach goes beyond single-step code generation by enabling iterative, intelligent interaction with the LLM, thereby replicating and improving upon the analytical process typically performed by human experts. By providing the model with a reference study, prompting it to learn from previous methods and guiding it through the implementation, optimization, and evaluation of the model, our study demonstrates a reproducible workflow for SOH estimation that is both accessible and does not require prior coding expertise. In summary, ChatGPT 4.0 is utilized not only as a general tool for developing estimation and optimization frameworks but is specifically designed in this study to enhance battery SOH estimation for eVTOL systems. This is achieved through the application of domain-specific feature selection, guided hyperparameter tuning, and performance evaluation adapted to the dynamics of battery degradation.

The remainder of this paper is organized as follows. The methodology along with the eVTOL dataset used is presented in Section 2. The prompting mechanism with ChatGPT for the SOH prediction is also detailed. Experimental and comparison results are discussed in Section 3. Finally, conclusions are provided in Section 4. The remainder of this paper is structured as follows: Section 2 presents our innovative LLM-driven methodology, including (i) the eVTOL battery dataset characteristics, (ii) the structured prompt engineering framework for SOH prediction, and (iii) the implementation of ML algorithms through ChatGPT. Section 3 details the experimental results, including comparative performance analysis against conventional methods and quantitative assessment. Finally, Section 4 concludes with key findings, discusses practical implications for eVTOL battery management, and outlines future research directions for LLM-assisted battery prognostics.

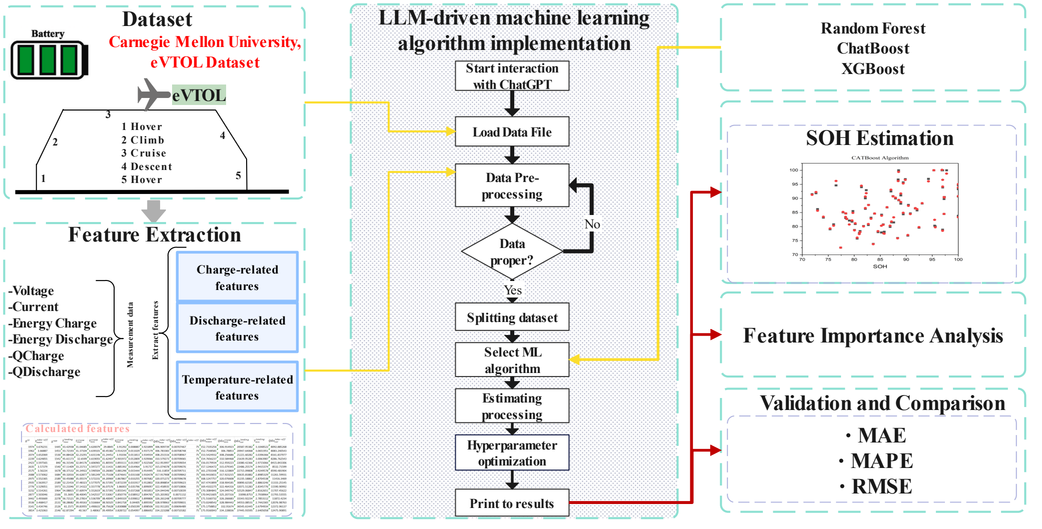

FIGURE 1 The proposed LLM-based methodology for Li-ion battery state-of-health prediction in eVTOL applications.

2 | METHODOLOGY

2.1 | Approach

Figure 1 illustrates the general framework of the proposed methodology, comprising four key steps: (i) collecting actual eVTOL operational data, (ii) extracting and validating relevant features, (iii) implementing and optimizing ML models using a LLM, i.e., ChatGPT 4.0 as the LLM, and (iv) validating the SOH estimation results. Rather than treating GPT as a coding assistant, ChatGPT is integrated into the full machine learning pipeline, including data preprocessing, model recommendation and selection, informed by reference study learning, hyperparameter tuning, and performance evaluation through LLM model implementation approach in this paper. Such comprehensive integration ensures that ChatGPT is not merely a tool for code generation but a transformative element that enhances the entire machine learning workflow, extending beyond simple code generation to active participation in the analytical process.

The process utilizing ChatGPT-initiates with a request for ChatGPT to list ML-based regression models suitable for SOH forecasting. From the listed models, the user selects Random Forest, XGBoost, and CatBoost for their interpretability, scalability, and robustness in managing noisy and high-dimensional data, as detailed in Section 2.4. Subsequently, a data set, comprising feature inputs and SOH capacity measurements is uploaded. ChatGPT 4.0 then preprocesses this dataset by eliminating irrelevant entries and validating the 21 pre-identified features from Mitici et al. [37]. Following preprocessing, the specified models are implemented, followed by hyperparameter optimization through methods such as Bayesian optimization to enhance prediction accuracy. Finally, the forecasting results are assessed with performance metrics such as MAE, MAPE, and RMSE. The interaction with ChatGPT is guided by structured through organized prompts, as illustrated in Figure 2, ensuring a systematic and reproducible workflow.

2.2 | Mathematical Formulation of the Proposed LLM-Guided Workflow

Consider the battery SOH dataset D comprising battery feature vectors and corresponding SOH values:

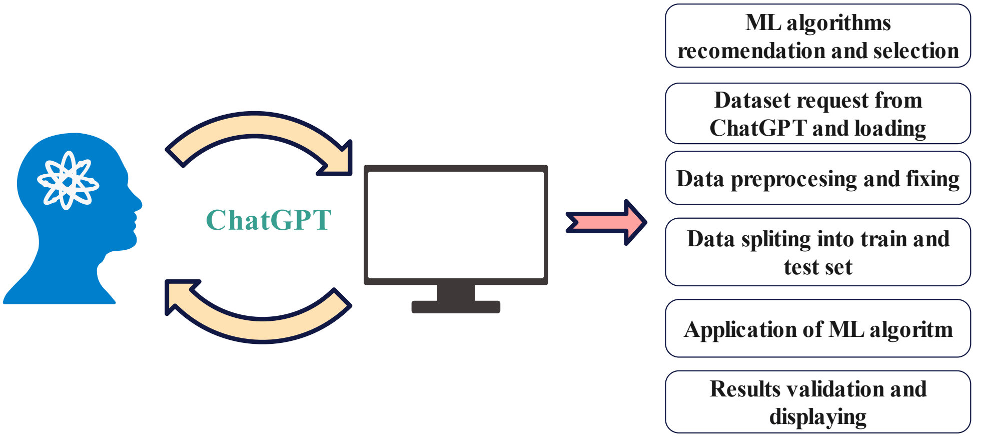

FIGURE 2 Workflow of the proposed LLM-based methodology for battery state-of-health prediction.

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}, \quad \mathbf{x}_i \in \mathbb{R}^d,\ y_i \in \mathbb{R}, \tag{1}$$

where each feature vector $\mathbf{x}_i$ specifically represents battery features-related measurements:

$$\mathbf{x}_i = [V_{\text{var}}^{\text{take-off}}, V_{\text{min}}^{\text{take-off}}, V_{\text{mean}}^{\text{take-off}}, \delta^{\text{CC}}, \delta^{\text{CV}}, T_{\text{max}}, Q_{\text{Charge}}, Q_{\text{Discharge}}, I_{\text{avg}}, E_{\text{Charge}}, \ldots]^T \tag{2}$$

and $\mathbf{y}_i$ represent battery health status-related indicator:

$$\mathbf{y}_i = [\mathcal{SOH}] \tag{3}$$

The critical battery-specific features explicitly utilized are defined as:

The LLM-driven method introduces a meta-function $g(\text{Prompt}_t)$, representing the structured prompting mechanism, which guides the LLM to output:

$$g(\text{Prompt}_t, \mathcal{R}_{\text{battery}}) \rightarrow \text{Code}_t = f_t(\mathcal{D}; \theta_t). \tag{4}$$

where $t \in \{1, 2, \ldots, T\}$ is the iteration step, and Code_t represents the ML implementation or modification produced at each prompt-response round. The LLM learns from both external reference material $\mathcal{R}_{\text{battery}}$ and the structure of $\mathcal{D}$ to iteratively improve the forecasting model.

Hyperparameter tuning is formulated as an automated optimization framework guided by ChatGPT's responses, through which the optimized model parameters (θ^*), specifically tailored for battery SOH prediction, are determined iteratively as follows:

$$\theta^* = \arg\min_{\theta\in\Theta} \mathcal{L}(f(\mathcal{D};\theta), y), \quad \text{where } \mathcal{L} \text{ is the prediction error (performance metrics such as } \mathcal{MAE}, \mathcal{RMSE}, \mathcal{MAPE}) \tag{5}$$

This loop continues until $\mathcal{L}$ is minimized within tolerance ϵ, and ChatGPT session exports the finalized model code Code* for offline execution.

2.3 | Dataset Description

This study employs a publicly available dataset of a Li-ion battery used in the propulsion system of an eVTOL vehicle, as detailed in [40, 41]. This dataset has gained significant attention in recent years as it is one of the few publicly available datasets in the literature for an eVTOL vehicle and considers relatively high discharge currents at the take-off and landing flight phases of the aircraft. The dataset comprises a comprehensive set of missions, including take-off, cruise, landing, resting 1, charging, and resting 2. During take-off and landing, the cells are discharged at high power for a short duration, while during cruise, they are discharged at low power for a longer period. The resting 1 continues until the cell temperature drops to 27°C or a minimum of 15 minutes has passed. The charging process includes a constant current (CC) phase, which continues until the voltage exceeds 4.2 V at 1C, followed by a constant voltage (CV) phase that continues until the current drops to C/30 at 4.2 V. Finally, during the resting 2, the cells remain in cooling until the temperature decreases to 35°C, and after 15 minutes, the battery is ready for the next mission. In this way, one full cycle of the cells is completed. VAH12, which has the longest operating duty, completes 2,347 cycles. Table 2 outlines the six variables measured within this dataset: Q Charge (the amount of charge supplied to the cell during charging), Q Discharge (the amount of charge extracted from the cell during discharging), Voltage, Current, Temperature, and Cycle number. These measurements are specific to the Sony-Murata 18650 VTC-6 Li-ion battery cells, which can provide energy up to 230 Wh/kg.

The dataset encompasses 22 distinct flight missions, each characterized by unique operational profiles. These missions include variations such as baseline operations, short and extended cruise lengths, power reduction during discharge, constant current charging with reduced current, constant voltage charging with reduced voltage, and different thermal chamber temperatures. To assess battery health, a capacity test is conducted every 50 cycles for each flight mission. As a result, each mission includes a capacity test corresponding to 1/50 of the total cycle count. Each capacity test is structured into several phases: Constant Current Charge, Constant Voltage Charge, 1st Resting Period, Take-off, Cruise, Landing, and 2nd Resting Period.

In Ref. [37], a feature importance analysis was performed to predict the battery SOH, identifying 21 features with importance scores exceeding 65%. These 21 features were also utilized in our study. Table 2.3 highlights the calculated values of the five most significant features before the battery SOH dropping to 85%. For this forecasting study, all 21 features were employed across the entire range of battery SOH levels, ensuring a comprehensive analysis of battery performance and degradation.

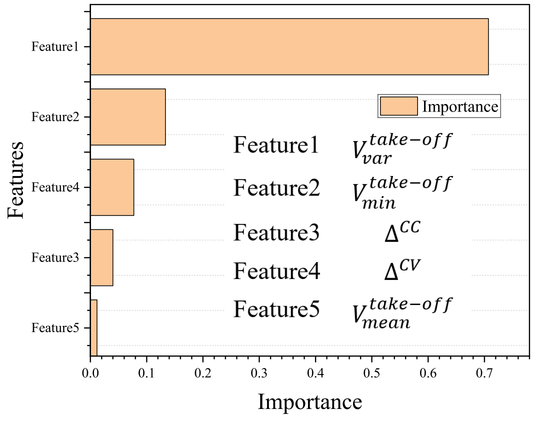

FIGURE 3 The top five features of greatest importance.

TABLE 2 The measured battery variables.

Inspected Battery Variables	Unites
Time	s
Voltage	V
Current	A
Energy Charge	Wh
Q_Charge	mAh
Energy Discharge	Wh
Q_Discharge	mAh
Temperature	Centigrade (°C)
Cycle number	-

2.4 | Preparing Dataset

The ML-based data-driven estimation method involves several critical stages, including data preprocessing, feature selection, model training and testing, and testing dataset processing [42]. Feature selection is particularly crucial, as it eliminates variables that are not strongly correlated with the target battery health parameters, such as SOH and RUL [43]. Selecting the most relevant features not only reduces preprocessing time but also enhances the overall performance of the ML algorithm. However, manually performing feature selection can be computationally intensive and requires significant expertise, which may limit its accessibility [42].

In this study, we introduce a methodology for SOH estimation that minimizes the need for software expertise by utilizing pre-identified features from Mitici et al. [37]. Using a LLM-driven approach, we validated the feature importance rankings from Mitici et al., ensuring consistency and reliability without manual intervention. Specifically,

TABLE 3 The top five most important features for predicting State of Health (SOH) values up to 85% using the Random Forest algorithm.

$V_{var}^{take-off}$	$V_{min}^{take-off}$	δ^{CC}	δ^{CV}	$V_{mean}^{take-off}$	SOH
0.002214	0.002214	2982.268	1973.506	3.6762	100
0.002083	3.6229935	2884.722	1962.386	3.6688	96.95
0.002058	3.6105096	2806.878	2074.074	3.6520	94.98
0.002086	3.5888891	2731.746	2239.308	3.6248	93.33
0.002151	3.5667963	2665.148	2422.816	3.5975	91.85
0.002267	3.5474994	2588.970	2634.578	3.5737	90.50
0.002305	3.5391505	2569.898	2575.476	3.5622	89.42
0.016118	3.5236735	2527.098	2687.868	3.5730	88.34
0.017841	3.4987845	2480.190	2975.434	3.5523	87.45
0.017221	3.512400	2475.032	2935.030	3.5639	86.83
0.019368	3.4688153	2397.414	2776.456	3.5290	85.26
0.021264	3.4485731	2387.870	3232.370	3.5142	85.04

Mitici et al. identified 21 features with over 65% importance out of 33 total features using the Random Forest algorithm. These features, categorized into temperature-, charge-, and discharge-related groups, were incorporated into the proposed LLM-based estimation method. The most influential features include the voltage variance during takeoff, the minimum takeoff voltage, and the constant current (CC) time, as illustrated in Figure 3 and Table 3. The remaining 18 features each contribute less than 10% to the relative importance. By using these pre-validated features, we streamlined the dataset preparation process, enabling efficient and accurate data-driven estimation.

2.5 | Integrated Machine Learning Algorithm

Various algorithms, including Support Vector Machine, Gaussian Process Regression, and Gradient Boosting methods, have been implemented to estimate battery health indicators such as SOH, RUL, and MOT on eVTOL battery datasets [41]. Among these, Random Forest (RF), Extreme Gradient Boosting (XGBoost), and CatBoost stand out as widely used methods for both classification and regression tasks. These algorithms were chosen over deep learning methods like Long Short-Term Memory (LSTM) networks and hybrid approaches due to their interpretability, scalability, and robustness in handling high-dimensional and noisy data. Deep learning methods, while powerful, often require large datasets and significant computational resources, and their "black-box" nature limits their interpretability in critical applications like eVTOL battery health monitoring. Hybrid approaches, though effective, can be complex to implement and tune. In contrast, RF, XGBoost, and CatBoost offer a balance of accuracy, efficiency, and ease of use, making them ideal for real-world battery health prediction tasks.

RF utilizes multiple decision trees to improve prediction accuracy. Its ensemble approach, known as "bagging," trains each tree on a random subset of the data, reducing overfitting and enhancing generalization [44]. RF is particularly effective in handling high-dimensional data and capturing complex interactions between features, making it a reliable choice for estimating battery health status. The flow chart of the RF algorithm is shown in Figure 4. Each decision tree consists of decision nodes that test input features, and leaf nodes, which provide output values. By averaging the predictions across all trees, RF produces robust and reliable estimates, even in the presence of noise.

XGBoost is a highly efficient and scalable ML algorithm that improves traditional gradient boosting by incorporating regularization of L1 and L2 to control overfitting [45, 46]. Its ability to handle missing data and provide ranking of features makes it particularly suitable for predicting battery health, where interpretability is critical [47]. XGBoost has been widely used to estimate SOH and RUL with high precision, using historical battery degradation data and feature engineering techniques to model complex non-linear patterns [48, 49]. The flow chart of the XGBoost algorithm is shown in Figure 5. CatBoost is a high-performance gradient boosting algorithm that excels in handling categorical features without extensive preprocessing [50]. Its advanced regularization and ordered boost techniques mitigate overfitting, while its native support for categorical variables eliminates the need for manual encoding [51]. CatBoost has been successfully applied to predict SOH and RUL in Li-ion batteries, demonstrating its ability to model non-linear degradation patterns and capture complex dependencies in battery aging data [52, 53]. The flow chart of the CatBoost algorithm is shown in Figure 6.

In summary, RF, XGBoost, and CatBoost were chosen for their interpretability, scalability, and robustness in handling the challenges of battery health prediction. These algorithms provide a practical and efficient alternative to deep learning and hybrid approaches, making them well-suited for real-world eVTOL applications.

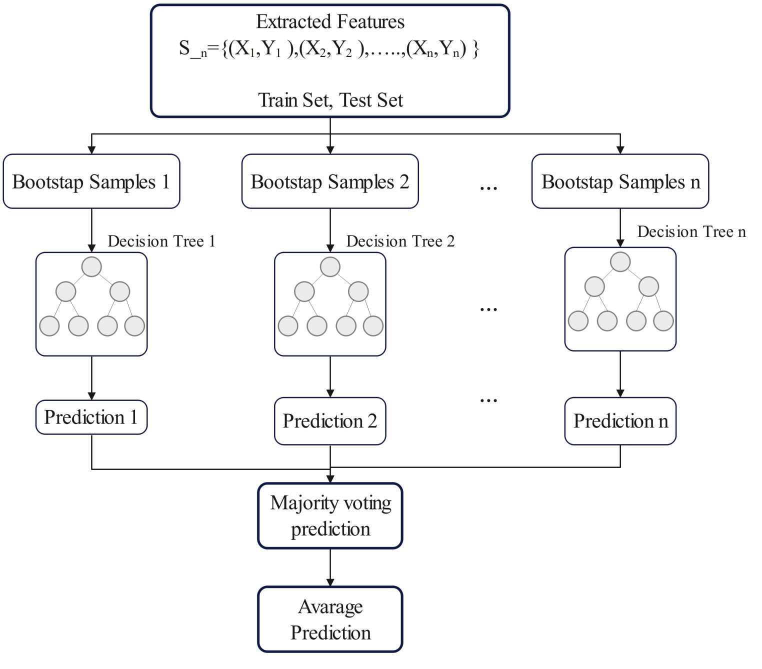

FIGURE 4 Main structure of Random Forest Algorithm (adapted from [44, 54, 55]).

2.6 | Performance Evaluation

This study implements several ML models, including RF, XGBoost, and CATBoost through assisted by ChatGPT 4.0 to forecast the SOH of Li-ion batteries used in eVTOL vehicles. The SOH of a battery is defined as the ratio between

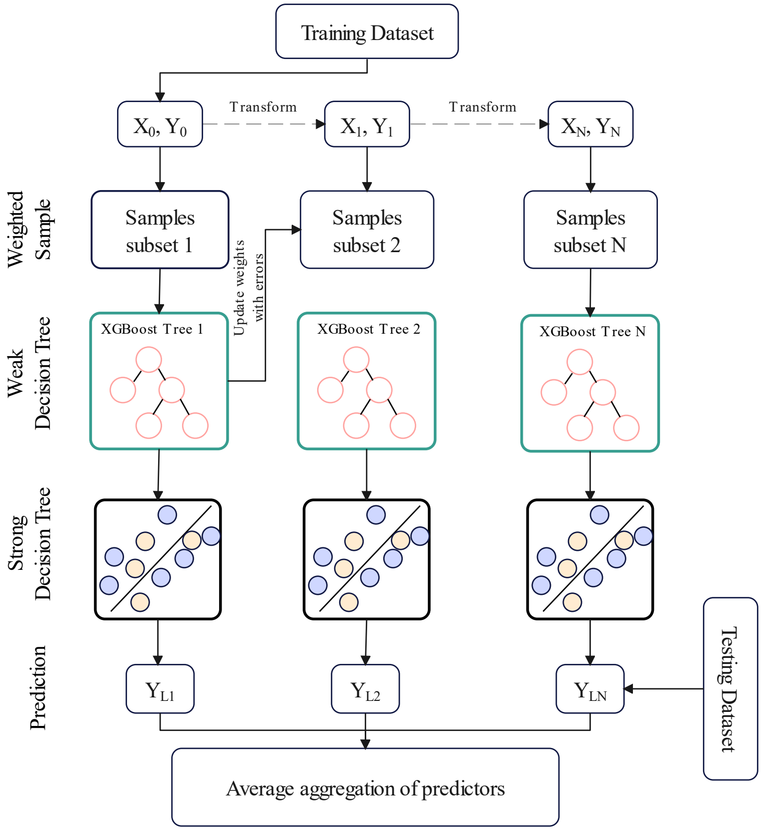

FIGURE 5 Main structure of XGBoost Algorithm (adapted from [56, 57]).

the charge capacity measured during a capacity test and the rated capacity of the battery, as given by

$$SOH^{m,c} = \frac{max_i(Qcharge_i^{m,c})}{max_i, (Qcharge_i^{m,0})} * 100\% \tag{6}$$

where $Qcharge_i^{(m,c)}$ is the maximum measured capacity during a^{th} capacity test c^{th} of mission profile m. $Qcharge_i^{(m,0)}$ is the maximum battery capacity measured during the first capacity test at $(c = 0)$ of mission profile m.

To test and validate the forecasting performance of the ML algorithms, three metrics, i.e., MAE, MAPE, and RMSE. They are defined for the estimated SOH of a battery under mission profile m, $1 \le m \le M$, as follows:

$$MAE_{SOH}^{m} = \frac{1}{c^m}\sum_{i=1}^{c^m}|SOH^{m,i} - SOH^{*m,i}|, \tag{7}$$

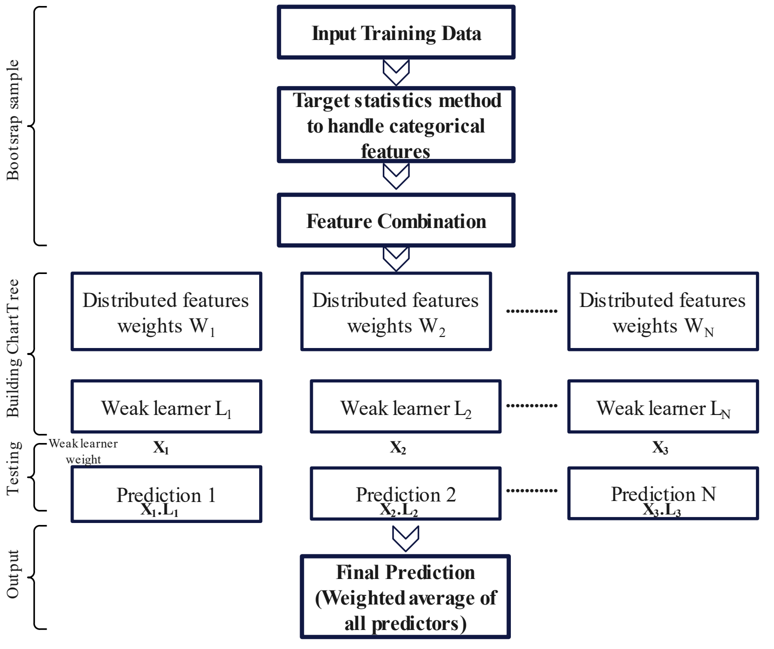

FIGURE 6 Main structure of CatBoost Algorithm (adapted from [58, 59]).

$$MAPE_{SOH}^{m} = \frac{max_i(Qcharge_i^{m,c})}{max_i,(Qcharge_i^{m,0})} * 100\% \tag{8}$$

where $SOH^{m,i}$ is the true battery SOH at capacity test i^{th} of mission profile m, $SOH^{*m,i}$ is the predicted SOH at capacity test c^{th} of mission profile m, $1 \leq \text{m} \leq M$. The overall performance of our SOH predictions across all M mission profiles is evaluated as follows:

$$MAE_{SOH}^{m} = \frac{1}{M}\sum_{J=1}^{c^m} MAE_{SOH}^{J}, \tag{9}$$

$$MAPE_{SOH}^{m} = \frac{1}{M}\sum_{J=1}^{c^m} MAPE_{SOH}^{J}. \tag{10}$$

3 | RESULTS AND DISCUSSION

Using the actual eVTOL dataset described in Section 2.1, ChatGPT-4.0 was employed to forecast the SOH of the battery through prompt engineering. Figure 7 illustrates the prompts used to interact with ChatGPT-4.0 for implementing

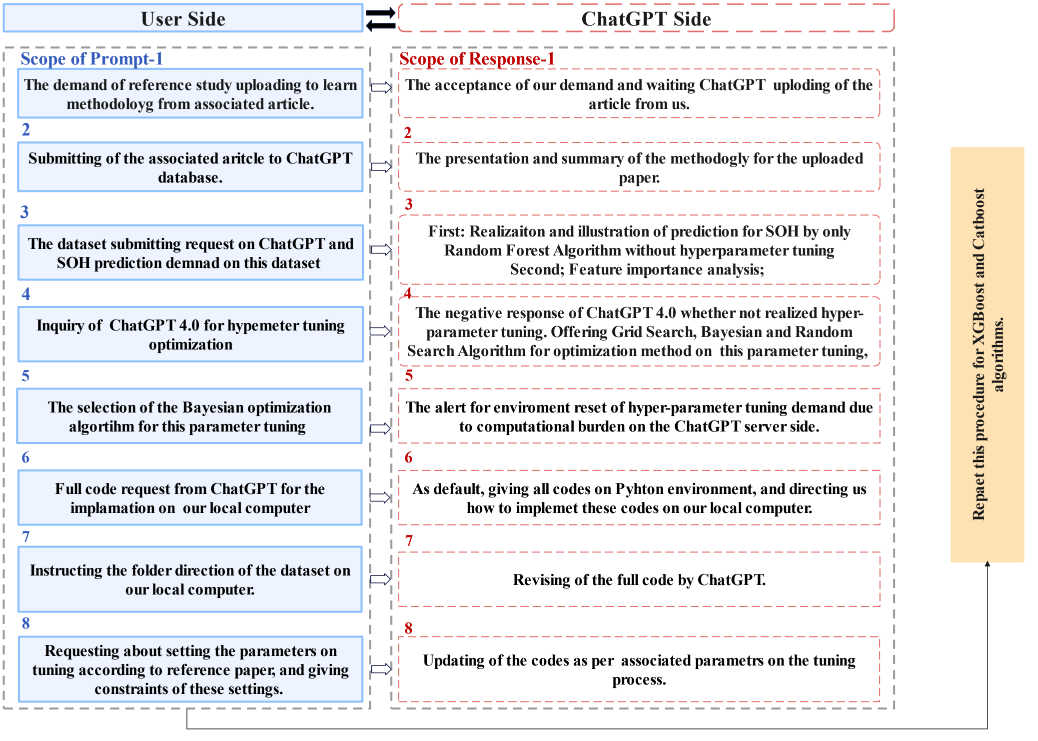

FIGURE 7 Prompts used to guide ChatGPT-4.0 in implementing a ML algorithm for SOH forecasting

ML algorithms for SOH forecasting.

The process began with instructing ChatGPT-4.0 to learn the ML-based methodology from the reference study in [37]. Upon receiving the instruction, ChatGPT-4.0 accepted the request and prepared for the upload of the reference study. After submitting the study, ChatGPT provided a concise summary of the methodology. Subsequently, the dataset was uploaded, and a request for SOH prediction was declared. Initially, ChatGPT performed predictions using the RF algorithm without hyperparameter optimization. It also calculated the feature importance values of the input data. Noticing the absence of hyperparameter tuning, a follow-up request was made to optimize the model. ChatGPT offered three optimization methods: Grid Search, Bayesian Optimization, and Random Search. Bayesian Optimization was selected to align with the reference study. Due to computational constraints on the ChatGPT server, the environment was reset, and the complete implementation code was requested for local execution. ChatGPT provided the Python code and detailed instructions for local implementation. After specifying the dataset's folder directory, ChatGPT revised the code accordingly. Finally, the parameters for tuning were set based on the reference study, with constraints provided in the prompt. ChatGPT delivered a fully functional code, enabling seamless execution. The same request-response loop was followed for implementing the XGBoost and CatBoost algorithms, as illustrated in Figure 7.

The dataset in Reference [41] is the only publicly available dataset that effectively simulates the dynamic power demand of an eVTOL vehicle. Various ML algorithms, such as Random Forest, XGBoost, Gaussian Process Regression (GPR), and Linear Support Vector Machine, have been applied for SOH estimation using this dataset, as reported in References [37] and [39]. The SOH estimation approach developed through ChatGPT 4.0 in this study is compared

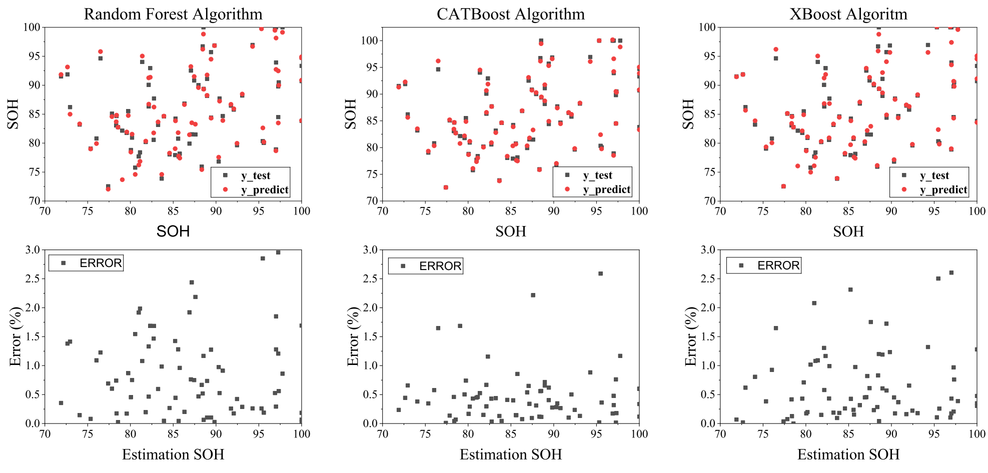

FIGURE 8 Comparison of SOH prediction results and error rates for the RF, XGBoost, and CatBoost algorithms.

with these previous approaches by employing performance metrics such as MAE, RMSE, and MAPE. The forecasting results, as reported in Table 4, demonstrate significant improvements achieved by the ML models enhanced by ChatGPT. These ChatGPT-driven models achieved lower MAE, MAPE, and RMSE than conventional models previously applied to the identical eVTOL dataset. Specifically, the RF model in [37] achieved an MAE of 1.33, an MAPE of 0.02, and an RMSE of 1.80, while the ChatGPT-driven RF model improved these metrics considerably to 0.8183, 0.0096, and 1.3463, respectively. Similarly, in Reference [39], which also utilized this dataset, the RF algorithm recorded RMSE and MAE values of 1.52 and 1.98, respectively, while the k-nearest neighbors (kNN) algorithm demonstrated better performance with scores of 1.4 and 1.16. Thus, the ChatGPT-based RF estimation method outperforms conventional RF estimations reported in prior research. This trend of improvement extends to XGBoost and Gaussian Process Regression methods as well, with the ChatGPT-driven versions surpassing their standalone counterparts. Particularly, the ChatGPT-driven CatBoost model exhibited superior performance, achieving an MAE of 0.47, an MAPE of 0.0054, and an RMSE of 0.74. These results highlight the potential of LLM-driven forecasting to significantly enhance predictive accuracy in eVTOL applications. The LLM-based prediction method outperforms traditional ML algorithms by enabling a more structured learning process and robust hyperparameter optimization, leading to enhanced model performance.

The results in Tables 5, 6, and 7 highlight the superior performance of ChatGPT 4.0-driven ML models in predicting battery SOH. The ChatGPT-enhanced XGBoost algorithm achieved the lowest average error rate of 0.0331%, followed closely by CatBoost with an error rate of 0.0246%, and the RF model with an error rate of 0.0353%. Among individual trials, the ChatGPT 4.0-driven XGBoost model achieved the most precise prediction, with a minimum error rate of 0.0009%. Similarly, CatBoost demonstrated consistently low error rates, with its best prediction deviating by only 0.0107% from the true SOH value. The RF model also performed well, maintaining errors below 0.05% across its top five trials. As shown in Figure 8, the maximum deviation of individual predictions from the actual SOH for all ChatGPT 4.0-driven models was below 3%.

While this study focused on RF, XGBoost, and CatBoost, ChatGPT-4.0 can be instructed to implement any ML algorithm, offering flexibility to explore the best-performing model for SOH forecasting. The success of ChatGPT-4.0 in implementing these algorithms highlights its versatility and potential for further improvements. ChatGPT-4.0

TABLE 4 Comparison of SOH prediction performance results between ChatGPT driven ML models and traditional ML models applied to the same eVTOL dataset.

Method	MAE	MAPE	RMSE
RF [37]	1.3300	0.02	1.80
XGBoost [37]	1.39	0.02	1.91
Gaussian Process regression [37]	1.4800	0.79	2.27
Support Vector Machine [37]	1.4800	0.02	2.20
RF [39]	1.52	-	1.98
kNN [39]	1.16	-	1.4
ChatGPT 4.0-driven RF	0.8183	0.0096	1.3463
ChatGPT 4.0-driven XGBoost	0.6417	0.0074	1.0461
ChatGPT 4.0-driven CATBoost	0.47	0.0054	0.74

TABLE 5 Five best SOH prediction results of the ChatGPT 4.0-driven RF Algorithm.

Trail number	True SOH	Estimated SOH	Error rates (%)
Best trail-I	82.7631	82.7444	0.0226
Best trail-II	96.8614	96.8353	0.0269
Best trail-III	94.7444	94.7124	0.0338
Best trail-IV	77.7245	77.7596	0.0452
Best trail-V	84.6192	84.6599	0.0480
Avarage of I-V	87.3425	87.3423	0.0353

delivers rapid and satisfactory results, even for computationally intensive tasks such as hyperparameter tuning and feature importance calculation. Its user-friendly interface eliminates the need for deep coding expertise, making advanced ML techniques accessible to a broader audience. Furthermore, ChatGPT-4.0 facilitates reproducible and adaptable offline analysis by providing complete Python code, enabling users to overcome computational limitations on the server. These advantages position ChatGPT-4.0 as a valuable tool for efficiently developing and implementing ML models for battery SOH prediction.

3.1 | Limitations

While this study demonstrates the effectiveness of ChatGPT-driven ML for battery SOH prediction, it has several limitations. First, the reliance on pre-identified features from Mitici et al. in [37] may limit the approach's generalizability to other datasets or battery types. Future work could explore ChatGPbased automated feature engineering to enhance adaptability. Second, the computational constraints of the ChatGPT server necessitated offline execution

TABLE 6 Five best SOH prediction results of the ChatGPT 4.0-driven XGBoost Algorithm.

Trail number	True SOH	Estimated SOH	Error rates (%)
Best trail-I	82.7631	82.7623	0.0009
Best trail-II	91.8746	91.8568	0.0193
Best trail-III	72.5449	72.5717	0.0369
Best trail-IV	89.3708	89.4061	0.0395
Best trail-V	91.5195	91.4564	0.0689
Avarage of I-V	85.6146	85.6106	0.0331

TABLE 7 Five best SOH prediction results of the ChatGPT 4.0-driven CatBoost Algorithm.

Trail number	True SOH	Estimated SOH	Error rates (%)
Best trail-I	72.5449	72.5371	0.0107
Best trail-II	84.4964	84.5106	0.0168
Best trail-III	100	100.02	0.0200
Best trail-IV	87.7068	87.6770	0.0340
Best trail-V	84.6753	84.7106	0.0417
Avarage of I-V	85.8847	85.8911	0.0246

for hyperparameter tuning, which may not be feasible for all users. Developing more efficient on-server optimization methods could address this challenge. Third, the study focused on three specific algorithms (RF, XGBoost, and CatBoost), while these were chosen for their interpretability and performance. However, exploring other algorithms or hybrid approaches could yield further improvements. Fourth, electric vehicle charging patterns often involve incomplete and irregular charging, which is a common scenario in real-world applications. It is important to note that the dataset used in this study consists of structured, complete charge and discharge cycles, typically observed under laboratory or mission-controlled environments. The application of the LLM-based workflow can be guided via prompt engineering to consider and handle irregular charging cycles. Future work can explore the integration of additional features and preprocessing steps to accommodate these complexities. For instance, the LLM can be instructed to identify and preprocess irregular charging patterns, ensuring that the model remains robust and accurate even when faced with incomplete or irregular charging data. Finally, the study was conducted on a single eVTOL dataset, and further validation on diverse datasets is needed to confirm the robustness of the proposed method. Addressing these limitations in future research will strengthen the applicability and impact of ChatGPT-driven ML in battery health forecasting.

4 | CONCLUSIONS

In this study, we introduced a new methodology, utilizing a LLM model for estimating the SOH of Li-ion batteries in eVTOL vehicles. By integrating ChatGPT into the full machine learning pipeline, our approach automates various tasks, including data preprocessing, feature importance determination, model recommendation and selection, hyper-parameter tuning, and performance evaluation. This comprehensive integration streamlines the machine learning workflow while enhancing the accuracy and efficiency of SOH estimation. The originality of the proposed method lies in the comprehensive integration of ChatGPT into every stage of the machine learning process. This holistic approach minimizes the need for manual intervention and expert knowledge, thereby providing a structured and systematic workflow. The LLM-driven approach involves iterative refinement of the model through structured prompts, allowing continuous improvement and adaptation to the specific requirements of SOH estimation.

The experimental validation was conducted using a publicly available dataset of a Li-ion battery used in the propulsion system of an eVTOL vehicle. Three machine learning algorithms—Random Forest, XGBoost, and CatBoost—were implemented and optimized using ChatGPT. The LLM-driven CatBoost model achieved MAE, MAPE, and RMSE values of 0.47, 0.0054, and 0.74, respectively, representing a significant improvement over traditional methods. Overall, when compared against conventional methods, our LLM-driven models demonstrated a 52% reduction in MAPE.

Future research can extend this approach to predict other battery health parameters, such as RUL and MOT, and explore the integration of additional ML algorithms or hybrid models. Automating feature engineering and optimizing hyperparameter tuning directly within the ChatGPT framework could further reduce error rates and improve model performance. Additionally, the application of the LLM-based workflow can be extended to handle irregular and incomplete charging patterns, which are common in real-world applications. The proposed framework can also be expanded to include other machine learning algorithms or hybrid models to further improve prediction accuracy and robustness. Additionally, ensemble learning could be employed by combining the best-performing machine learning algorithms to further improve the accuracy of SOH estimation within the prompt-based LLM framework.

AUTHOR CONTRIBUTIONS

Suleyman Tuncel: Data curation; formal analysis; methodology; software; validation; writing—review and editing.**Hasan Cinar:** conceptualisation; data curation; methodology; validation; writing—original draft; writing—review and editing. **Mehmet Gucyetmez:** Writing—original draft; validation; writing—review and editing. **Nuh Erdogan:** writing—review and editing; resources; supervision; validation;

ACKNOWLEDGEMENTS

The authors wish to acknowledge Carnegie Mellon University for publicly publishing a real-test dataset.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data openly available in a public repository at https://doi.org/10.1184/R1/14226830.v2

ORCID

Süleyman Tuncel, https://orcid.org/0000-0003-4565-2590
Hasan Cinar, https://orcid.org/0000-0001-8718-3767
Mehmet Gucyetmez, https://orcid.org/0000-0003-2191-8665
Nuh Erdogan, https://orcid.org/0000-0003-1621-2748

REFERENCES

[1] van Oosterom S, Mitici M. An environmentally-aware dynamic planning of electric vehicles for aircraft towing considering stochastic aircraft arrival and departure times. Transportation Research Part C: Emerging Technologies 2024;169:104857.

[2] Zewde L, Raptis IA. Conceptualizing UAM: Technologies and Methods for Safe and Efficient Urban Air Transportation. Green Energy and Intelligent Transportation 2025;p. 100265.

[3] Zhang C, Tu L, Yang Z, Du B, Zhou Z, Wu J, et al. A CMMOG-based lithium-battery SOH estimation method using multi-task learning framework. Journal of Energy Storage 2025;107:114884.

[4] Wang S, Wang C, Takyi-Aninakwa P, Jin S, Fernandez C, Huang Q. An improved parameter identification and radial basis correction-differential support vector machine strategies for state-of-charge estimation of urban-transportation-electric-vehicle lithium-ion batteries. Journal of Energy Storage 2024;80:110222.

[5] Zhang C, Zhao S, Yang Z, He Y. A multi-fault diagnosis method for lithium-ion battery pack using curvilinear Manhattan distance evaluation and voltage difference analysis. Journal of Energy Storage 2023;67:107575.

[6] Wang S, Dang Q, Gao Z, Li B, Fernandez C, Blaabjerg F. An innovative square root-untraced Kalman filtering strategy with full-parameter online identification for state of power evaluation of lithium-ion batteries. Journal of Energy Storage 2024;104:114555.

[7] Wang S, Zhang S, Wen S, Fernandez C. An accurate state-of-charge estimation of lithium-ion batteries based on improved particle swarm optimization-adaptive square root cubature kalman filter. Journal of power sources 2024;624:235594.

[8] Ding Q, Ding D, Wang Y, Guan C, Ding B. Unraveling the landscape of large language models: a systematic review and future perspectives. Journal of Electronic Business & Digital Economics 2023;(ahead-of-print).

[9] OpenAI, OpenAI; 2024. Accessed: 2024-06-06. `https://www.openai.com`.

[10] Korzynski P, Mazurek G, Krzypkowska P, Kurasinski A. Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. Entrepreneurial Business and Economics Review 2023;11:25–37.

[11] Nah FfH, Zheng R, Cai J, Siau K, Chen L. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research 2022;25:277–304.

[12] Meshram S, Naik N, VR M, More S T adn Kharche. Conversational AI: Chatbots. In: 2021 International Conference on Intelligent Technologies (CONIT); 2021. p. 9498508.

[13] Gu JG, Han Z, Beirami A, He B, Zhang G, Liao R, et al. A systematic survey of prompt engineering on vision-language foundation models. arXiv preprint 2023;2307:arXiv preprint.

[14] Sabit E. Prompt Engineering For ChatGPT: A Quick Guide To Techniques. Journal of Information Technology Case and Application Research 2023;.

[15] Ziegler D, Stiennon N, Wu j, B Brown T, Ranford A, Amodei D, et al. Fine-tuning language models from human preferences. arXiv preprint 2019;1909:08593.

[16] Wardat Y, A Tashtoush M, AlAli R, M Jarrah A. ChatGPT: A revolutionary tool for teaching and learning mathematics. Eurasia Journal of Mathematics, Science and Technology Education 2023;19.

[17] Tai AMY, Meyer M, Varidel M, Prodan A, Vogel M, Lorfino F, et al. Exploring the potential and limitations of ChatGPT for academic peer-reviewed writing: Addressing linguistic injustice and ethical concerns. Journal of Academic Language and Learning 2023;17:16–3.

[18] Haque MA, Li S. The potential use of ChatGPT for debugging and bug fixing. EAI Endorsed Transactions on AI and Robotics 2023;2.

[19] Alshami A, Elsayed M, Ali E, E E Eltoukhy A, Zayed T. Harnessing the Power of ChatGPT for Automating Systematic Review Process: Methodology, Case Study, Limitations, and Future Directions. Systems 2023;11:351.

[20] Jalil S, Rafi S, D LaToza T, Moran W K Lam. ChatGPT and Software Testing Education: Promises Perils. In: 2023 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW); 2023. .

[21] S Bonadia R, C L Trindade F, Freitas W, Venkatesh B. On the Potential of ChatGPT to Generate Distribution Systems for Load Flow Studies Using OpenDSS. IEEE Transactions on Power Systems 2023;38:5965–5968.

[22] Huang J, Quian J, Chen Y, Lin R, Weng Y, Lin G, et al. Improve Knowledge Graph Completion for Diagnosing Defects in Main Electrical Equipment. In: 19th International Conference; 2023. p. 738–748.

[23] Dai X, Liu GP, Hu W, Lei Z, Zhou H. Learning from ChatGPT: A Transformer-Based Model for Wind Power Forecasting. In: 2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC / ICPS Europe); 2023. p. 1–6.

[24] He H. Robotgpt: From chatgpt to robot intelligence. Authorea Preprints 2023;.

[25] Li R, Pu C, Fan F, Tao J, Xiang Y. Leveraging ChatGPT for Power System Programming Tasks. arXiv preprint arXiv:230511202 2023;.

[26] Zhang Z, Liu M, Sun M, Deng R, Cheng P, Niyato D, et al. Vulnerability of Machine Learning Approaches Applied in IoT-Based Smart Grid: A Review. IEEE Internet of Things Journal 2024;.

[27] Gao Q, Lei T, Yao W, Zhang X, Zhang X. A health-aware energy management strategy for fuel cell hybrid electric UAVs based on safe reinforcement learning. Energy 2023;283:129092.

[28] Wang H, Li YF, Xie M. Empowering ChatGPT-Like Large-Scale Language Models with Local Knowledge Base for Industrial Prognostics and Health Management. arXiv preprint arXiv:231214945 2023;.

[29] Bian C, Duan Z, Hao Y, Yang S, Feng J. Exploring large language model for generic and robust state-of-charge estimation of Li-ion batteries: A mixed prompt learning method. Energy 2024;p. 131856.

[30] Peng H, Liu C, Li H. Large Language Model Enabled Health Management for Internet of Batteries in Electric Vehicles. IEEE Internet of Things Journal 2024;.

[31] Bian C, Han X, Duan Z, Deng C, Yang S, Feng J. Hybrid prompt-driven large language model for robust state-of-charge estimation of multi-type li-ion batteries. IEEE Transactions on Transportation Electrification 2024;.

[32] Lee J, Rew J. Large Language Model-based SHAP Analysis for Interpretation of Remaining Useful Life Prediction of Lithium-ion Battery. Journal of Korea Society of Industrial Information Systems 2024;29(5):51–68.

[33] Qiu T, Hou L, Shang Y. Prompt-Driven Fine-Tuning of Large Language Model for Li-ion Battery State Estimation. In: 2024 8th CAA International Conference on Vehicular Control and Intelligence (CVCI) IEEE; 2024. p. 1–6.

[34] Yunusoglu A, Le D, Tiwari K, Isik M, Dikmen I. Battery State of Health Estimation Using LLM Framework. arXiv preprint arXiv:250118123 2025;.

[35] Zhang Z, Zhu Y, Zhang Q, Cui N, Shang Y. Multi-cycle charging information guided state of health estimation for lithium-ion batteries based on pre-trained large language model. Energy 2024;313:133993.

[36] Wang L, Jiang S, Mao Y, Li Z, Zhang Y, Li M. Lithium-ion battery state of health estimation method based on variational quantum algorithm optimized stacking strategy. Energy Reports 2024;11:2877–2891.

[37] Mitici M, Hennink B, Pavel M, Dong J. Prognostics for Lithium-ion batteries for electric Vertical Take-off and Landing aircraft using data-driven machine learning. Energy and AI 2023;12:100233.

[38] Clarke MA, Alonso JJ. Forecasting the Operational Lifetime of Battery-Powered Electric Aircraft. Journal of Aircraft 2023;60(1):47–55.

[39] Granado L, Ben-Marzouk M, Saenz ES, Boukal Y, Jugé S. Machine learning predictions of lithium-ion battery state-of-health for eVTOL applications. Journal of Power Sources 2022;548:232051.

[40] Bills A, Sripad S, Fredericks WL, Guttenberg M, Charles D, Frank E, et al. Universal battery performance and degradation model for electric aircraft. arXiv preprint arXiv:200801527 2020;.

[41] Bills A, Sripad S, Fredericks L, Guttenberg M, Charles D, Frank E, et al. A battery dataset for electric vertical takeoff and landing aircraft. Scientific Data 2023;10(1):344.

[42] Hu X, Che Y, Lin X, Onori S. Battery health prediction using fusion-based feature selection and machine learning. IEEE Transactions on Transportation Electrification 2020;7(2):382–398.

[43] Rauf H, Khalid M, Arshad N. A novel smart feature selection strategy of lithium-ion battery degradation modelling for electric vehicles based on modern machine learning algorithms. Journal of Energy Storage 2023;68:107577.

[44] Jafari S, Byun YC. Optimizing battery RUL prediction of lithium-ion batteries based on Harris hawk optimization approach using random forest and LightGBM. IEEE Access 2023;.

[45] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM; 2016. p. 785–794.

[46] Chen T, Introduction to XGBoost; 2017. Retrieved from `https://xgboost.readthedocs.io`.

[47] Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics 2001;29(5):1189–1232.

[48] Liu K, Li Y, Zhang C, Zhang Z. Battery health state estimation based on XGBoost algorithm and grid search optimization. Energy 2020;191:116514.

[49] Li Z, Liu H, Wu B, Zhang F. An improved XGBoost model for state-of-health estimation of lithium-ion batteries. Journal of Energy Storage 2021;42:103040.

[50] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. arXiv preprint arXiv:181011363 2018;.

[51] Hancock JT, Khoshgoftaar TM. CatBoost for big data: An interdisciplinary review. In: 2020 International Conference on Big Data (Big Data) IEEE; 2020. p. 3124–3133.

[52] Zhao W, Wang M, Li X. Battery State-of-Health Estimation Using CatBoost and Feature Selection. IEEE Transactions on Industrial Electronics 2023;70(5):4521–4532.

[53] Liu K, Wang Y, Zhang C. Battery remaining useful life prediction based on CatBoost and Bayesian optimization. Energy Reports 2021;7:5120–5132.

[54] Liu K, Hu X, Zhou H, Tong L, Widanage WD, Marco J. Feature analyses and modeling of lithium-ion battery manufacturing based on random forest classification. IEEE/ASME Transactions on Mechatronics 2021;26(6):2944–2955.

[55] Li Y, Zou C, Berecibar M, Nanini-Maury E, Chan JCW, Van den Bossche P, et al. Random forest regression for online capacity estimation of lithium-ion batteries. Applied energy 2018;232:197–210.

[56] Ma M, Zhao G, He B, Li Q, Dong H, Wang S, et al. XGBoost-based method for flash flood risk assessment. Journal of Hydrology 2021;598:126382.

[57] Ali ZH, Burhan AM. Hybrid machine learning approach for construction cost estimation: An evaluation of extreme gradient boosting model. Asian Journal of Civil Engineering 2023;24(7):2427–2442.

[58] Pandey M, Karbasi M, Jamei M, Malik A, Pu JH. A comprehensive experimental and computational investigation on estimation of scour depth at bridge abutment: emerging ensemble intelligent systems. Water Resources Management 2023;37(9):3745–3767.

[59] Sapkota SC, Saha P, Das S, Meesaraganda LP. Prediction of the compressive strength of normal concrete using ensemble machine learning approach. Asian Journal of Civil Engineering 2024;25(1):583–596.