Franklin Open 11 (2025) 100293

Contents lists available at ScienceDirect

Franklin Open

journal homepage: www.elsevier.com/locate/fraope

Electric vehicle battery state of charge estimation using metaheuristic-optimized CatBoost algorithms

Mohd Herwan Sulaiman [a,d,*], Zuriani Mustaffa [b], Ahmad Salihin Samsudin [c], Amir Izzani Mohamed [a], Mohd Mawardi Saari [a]

[a] Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), 26600 Pekan, Pahang, Malaysia
[b] Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), 26600 Pekan, Pahang, Malaysia
[c] Ionic Materials Team, Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), 26300 Gambang, Pahang, Malaysia
[d] Center for Advanced Industrial Technology (AIT), Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA), 26600 Pekan, Pahang, Malaysia

A R T I C L E I N F O

Keywords:
Battery state of charge
CatBoost algorithm
Machine learning
Metaheuristic algorithms

A B S T R A C T

State of Charge (SoC) estimation plays a crucial role in battery management systems for electric vehicles, directly impacting their operational efficiency and reliability. This study presents a hybrid approach combining the CatBoost algorithm with metaheuristic optimization techniques to enhance SoC estimation accuracy and robustness. The methodology was validated using an extensive dataset collected from 72 real-world driving trips of a BMW i3 (60 Ah), comprising 1053,910 instances of battery and vehicle operation metrics. A comprehensive data preprocessing pipeline was implemented, including missing value treatment, outlier removal, and feature normalization using Min-Max scaling. Three distinct metaheuristic algorithms were investigated: Barnacles Mating Optimizer (BMO), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Whale Optimization Algorithm (WOA), each integrated with CatBoost to optimize critical parameters including learning rate, tree depth, regularization, and bagging temperature. Experimental results demonstrate that the BMO—CatBoost approach achieved superior performance with best-case metrics of RMSE = 6.1031, MAE = 4.1303, and R^2 = 0.8211, outperforming both PSO—CatBoost, GA-CatBoost, and WOA-CatBoost implementations. The framework's effectiveness was validated through rigorous testing, establishing its potential for real-world electric vehicle applications. This research contributes to the advancement of battery management technology, offering promising implications for electric vehicle energy management and broader energy storage applications.

1. Introduction

Battery Electric Vehicles (EVs) are essential for sustainable transportation, reducing emissions, lowering fuel costs, and promoting energy efficiency for cleaner futures. Accurate estimation of battery State of Charge (SoC) is critical for optimizing battery performance and ensuring reliable operation in various applications, from electric vehicles to renewable energy systems [1,2]. Precise SoC estimation allows for better energy management, preventing overcharging and deep discharging, which can degrade battery health and shorten lifespan [3]. It also supports accurate range predictions, enhancing user confidence and system efficiency. Advanced SoC estimation techniques help avoid costly downtimes and safety issues by providing real-time data for effective monitoring and maintenance. Thus, accurate SoC estimation is essential for maximizing battery reliability, safety, and overall performance in energy storage solutions [4,5].

Battery SoC estimation plays a crucial role in modern battery management systems. Industrial applications typically categorize SoC estimation methods into five main groups: Coulomb counting methods, Voltage methods, Kalman Filter-based methods, machine learning approaches, and hybrid methods [6]. These methods have evolved significantly over the past decade, offering distinct advantages for specific applications while addressing various technological challenges in battery monitoring and management [7]. Traditional methods, including voltage-based and Coulomb counting approaches, remain prevalent due to their simplicity and cost-effectiveness. Voltage-based methods estimate SoC through direct voltage measurements, while Coulomb counting tracks battery charge through current integration [8–10]. However, these methods often face accuracy limitations due to their sensitivity to environmental factors, particularly temperature

* Corresponding author.
E-mail address: herwan@umpsa.edu.my (M.H. Sulaiman).

https://doi.org/10.1016/j.fraope.2025.100293
Received 20 January 2025; Received in revised form 14 May 2025; Accepted 1 June 2025
Available online 4 June 2025

variations, and degradation in estimation quality as batteries age [11, 12].

Kalman filter-based methods are more sophisticated, employing mathematical models to adapt to changing conditions and battery characteristics [13]. These methods have demonstrated superior accuracy in various applications, particularly in electric vehicles where precise SoC estimation is critical [14–17]. Nevertheless, they present implementation challenges, requiring complex battery models and significant computational resources to maintain real-time performance [18,19]. Machine learning approaches have gained significant attention due to their ability to capture complex, nonlinear relationships in battery behavior [20–22]. These methods excel in adaptability, learning from historical data to improve estimation accuracy over time [19]. In particular, recent studies have explored the application of deep neural networks [23,24], convolutional neural networks [25], long short-term memory (LSTM) architectures [25–27], and ensemble learning algorithms such as XGBoost and CatBoost [28,29] for enhanced SoC prediction accuracy. Such advancements demonstrate improved robustness against aging effects and variations in driving conditions. Recent advancements in deep learning architectures have further enhanced their capabilities in handling varying operating conditions and battery aging effects [30,31]. However, these approaches necessitate substantial training data and computational resources, potentially limiting their application in resource-constrained systems [32].

Combining multiple estimation techniques, hybrid methods represent the latest evolution in SoC estimation [2,12,33–36]. These approaches typically integrate the advantages of different methods while mitigating their individual limitations. For instance, combining machine learning with model-based approaches can enhance robustness while maintaining computational efficiency [37]. Furthermore, metaheuristic optimization algorithms have recently gained traction for hyperparameter tuning of machine learning models in battery applications, as they help automate and refine the learning process [38,39]. In this context, recent studies on intelligent active cell balancing in electric vehicle battery management have shown the potential for machine learning to optimize performance and enhance SoC accuracy, pointing to the growing importance of advanced algorithms in the Battery Management System (BMS) field [40]. Additionally, the application of machine learning for active balancing has provided insights into improving the overall performance of EV battery management [41]. Although hybrid methods often achieve superior accuracy and reliability, they require careful integration of different techniques and increased computational resources [12].

Building upon these developments and addressing existing gaps, this study proposes a novel hybrid approach that combines the CatBoost algorithm with metaheuristic optimization techniques, specifically the Barnacles Mating Optimizer (BMO), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Whale Optimization Algorithm (WOA). The proposed methodology leverages CatBoost's superior learning capabilities for handling complex, nonlinear relationships in battery behavior, while employing metaheuristic optimization to fine-tune the model's hyperparameters. This integration aims to overcome the limitations of traditional methods by optimizing model performance through systematic parameter tuning, ultimately enhancing the accuracy and reliability of SoC estimation in real-world driving conditions. The approach was validated using extensive real-world data collected from BMW i3 electric vehicle operations, demonstrating its practical applicability in actual driving scenarios.

The main contributions of this paper are listed as follows:

Novel Hybrid Approach:

- Development of a unique hybrid methodology combining the CatBoost algorithm with metaheuristic optimization for SoC estimation
- Integration of advanced machine learning with optimization techniques to enhance battery monitoring accuracy

Real-World Data Implementation:

- Successful validation using extensive real-world data (72 driving trips from BMW i3)
- Processing and refinement of over 1 million data instances (1053,910), demonstrating practical applicability

Parameter Optimization Framework:

- Implementation of metaheuristic optimization for fine-tuning critical CatBoost parameters
- Optimization of multiple parameters including learning rate, depth, regularization, and bagging temperature

Practical Applications:

- Direct applicability to electric vehicle energy management systems
- Potential extensibility to broader battery management applications
- Contribution to the advancement of battery monitoring technology

The paper is organized as follows: Section 2 presents a concise development of the SoC estimation based on the CatBoost algorithm, followed by the methodology proposed in this paper in Section 3. Section 4 details the selected metaheuristic algorithm as an optimizer of the parameters of CatBoost, and Section 5 discusses the results obtained. Finally, Section 6 provides the concluding remarks.

2. Battery state of charge estimation based on CatBoost algorithm

The CatBoost algorithm [42], short for Categorical Boosting, is a gradient boosting method particularly effective for handling categorical data and managing overfitting, making it well-suited for complex regression tasks like battery SoC estimation. CatBoost is an implementation of gradient-boosting decision trees (GBDT), where predictions are made by constructing an ensemble of weak learners, typically decision trees, sequentially. Each tree is trained to correct the errors of the preceding one, refining the model's accuracy iteratively. The fundamental goal of gradient boosting is to minimize the loss function $L(y, \hat{y})$, where y is the actual SoC value, and $\hat{y}$ is the predicted SoC. In each iteration t, a new tree $f_t(x)$ is added to minimize the loss, and the model is updated as:

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta \cdot f_t(x). \tag{1}$$

where η is the learning rate, which controls the contribution of each new tree, ensuring a balance between speed and accuracy. This sequential learning approach allows CatBoost to capture non-linear relationships in the battery data, which is essential for predicting SoC accurately, as battery dynamics often exhibit complex, non-linear patterns.

A notable feature of CatBoost is its unique handling of categorical features through its "ordered boosting" mechanism, which reduces prediction bias and avoids overfitting, particularly important in SoC estimation, where accurate and generalizable predictions are critical. CatBoost estimates the residuals by taking derivatives of the loss function with respect to the predictions, allowing it to update each tree based on the gradient of the error. Mathematically, for a given loss function L, the gradient $g_i^{(t)}$ for a sample i at iteration t is computed as:

$$g_i^{(t)} = -\frac{\partial L(y_i, \hat{y}^{(t-1)})}{\partial \hat{y}^{(t-1)}}. \tag{2}$$

These gradients are then used to fit the new tree $f_t(x)$, effectively capturing and correcting errors in SoC estimation. Furthermore, CatBoost incorporates L2 regularization, which penalizes large coefficients and reduces the risk of overfitting. This is particularly useful for SoC estimation, where consistent performance across varying battery conditions is desired. Overall, the algorithm's robust handling of complex data patterns, categorical feature encoding, and regularization capabilities make it an effective tool for modelling the highly variable and non-linear characteristics of battery SoC.

In this paper, key parameters of the CatBoost algorithm—such as learning rate, depth, regularization, subsample, features of each level, bagging temperature, and random strength—are optimized using advanced metaheuristic algorithms. These optimization methods, particularly metaheuristic approaches, are employed to enhance the CatBoost model's performance for battery SoC estimation. The learning rate controls the contribution of each tree in the ensemble, while depth determines the complexity of individual trees. Regularization prevents overfitting by penalizing large model coefficients, and the subsample decides the fraction of data used to build each tree, thus introducing randomness for better generalization.

Other parameters like features of each level influence feature sampling at each level of tree construction, while bagging temperature and random strength introduce controlled randomness to diversify trees in the ensemble. By utilizing metaheuristic optimization techniques, this study aims to systematically search for optimal parameter values, balancing model accuracy and generalization. This parameter tuning through metaheuristic algorithms enhances the robustness of the CatBoost model for SoC estimation, enabling it to capture the complex, nonlinear relationships in battery data more effectively.

In the proposed optimization process, each metaheuristic algorithm is applied to tune the CatBoost parameters within predefined bounds. Each parameter was assigned a bounded search range based on prior knowledge and preliminary experiments—for instance, the learning rate was set within [0.01–0.1] and tree depth between [3–10]. The optimization process used Mean Square Error (MSE) as the objective function to evaluate each parameter combination. The final optimized values and their corresponding performance metrics, viz. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Standard Deviation (STD), and Coefficient of Determination (R^2) for all approaches are summarized in Table 4. Further methodological details of the optimization setup, including the fitness function, algorithmic flow, and evaluation metrics, are elaborated in Section 3.

3. Methodology

This section outlines the methodology used for metaheuristic-CatBoost to estimate the SoC of lithium-ion batteries. It includes a description of the dataset, dataset analysis, data normalization, metaheuristic algorithms, and performance evaluation criteria. Fig. 1 visualizes the metaheuristic-CatBoost for the SoC estimation framework.

The approach hybridizes the CatBoost algorithm with a metaheuristic optimizer to estimate the SoC in lithium-ion batteries. The dataset consists of real-world data collected from 72 trips, containing critical battery parameters such as voltage, current, temperature, and other operational metrics essential for accurate SoC predictions. For training and testing purposes, the dataset was divided into a 70 % training set and a 30 % testing set. The split was performed randomly to ensure the model was not biased by any particular trip or time period within the dataset.

The training data was used to train the CatBoost model, enabling it to capture the intricate relationships between input features and SoC. The testing data, which the model had not seen during training, was used to evaluate the model's performance and generalization capability. To ensure randomization and avoid potential issues like temporal correlation or overfitting, the data was randomly shuffled before the split. This shuffling process ensures that each subset contains a representative sample of the data from all 72 trips, which mitigates any sequential dependencies that might affect model performance. Metaheuristic algorithms were then applied to fine-tune the CatBoost model's hyperparameters iteratively. The optimization process continues until a termination criterion (such as a predefined number of iterations or convergence to a stable solution) is met, producing the final optimized model.

After parameter optimization, the model's performance was assessed on the testing set, with performance metrics such as RMSE, MAE, STD, and R^2 used to evaluate the quality of SoC predictions. By combining CatBoost's robust learning capabilities with metaheuristic optimization, this hybrid approach aims to deliver more precise and reliable SoC estimations than traditional models. Each component is discussed in detail in the following subsections.

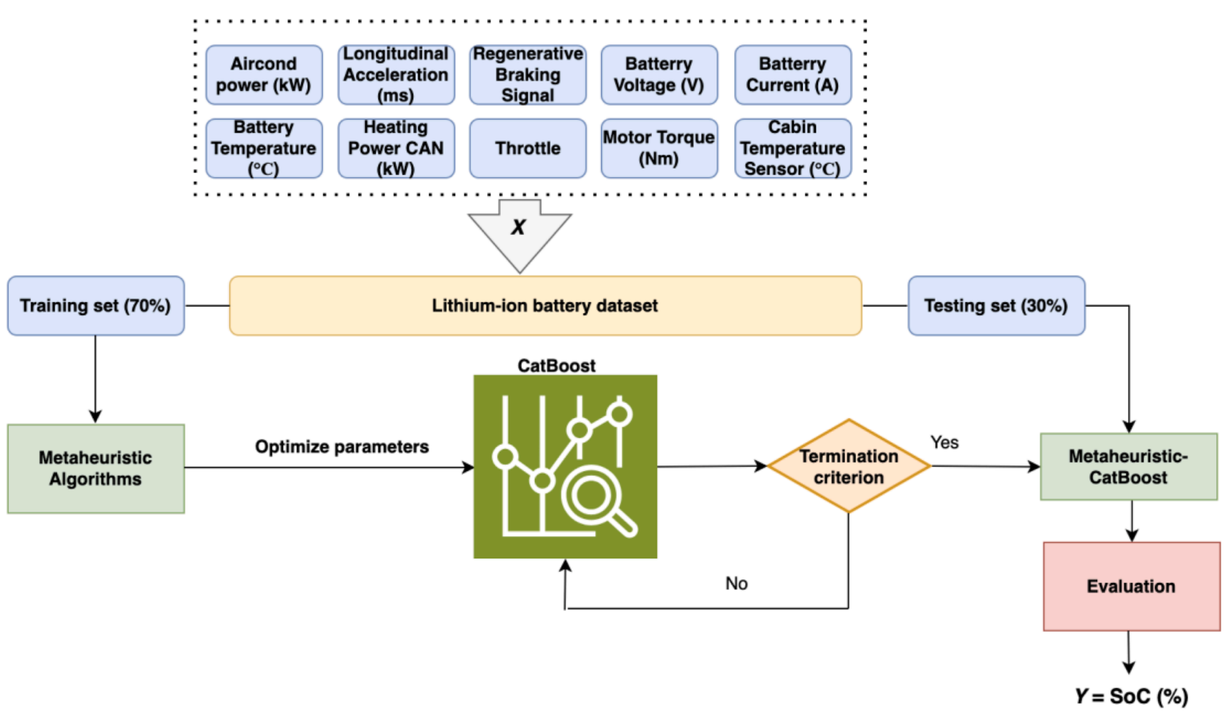

Fig. 1. Metaheuristic-CatBoost for SoC Estimation Framework.

3.1. Dataset description

Data quality is known to significantly impact machine learning model performance, making careful data collection and processing essential. In this study, data from 72 real-world driving trips of a BMW i3 electric car with a battery capacity of 60 Ah, are used to validate a detailed vehicle model, accessible in [43,44]. The estimation model incorporates 10 input variables: air conditioning power (AC Power) in kW, longitudinal acceleration (LA) in m/s^2, regenerative braking signal (RBS), battery voltage (V_batt) in volts, battery current (I_batt) in amps, battery temperature (T_batt) in °C, heating power CAN in kW, throttle position (TP), motor torque (T_motor) in Nm, and cabin temperature (T_cabin). The output is the State of Charge (SoC) represented as a percentage. Compared to previous studies utilizing deep learning [2,20,38] or random forest models [45], the present work incorporates an expanded set of ten real-world input features related to driving dynamics, thermal behavior, and auxiliary systems. These features, particularly AC power, heating power, cabin temperature, and motor torque have not been simultaneously considered in earlier SoC estimation studies, enabling a more comprehensive modeling of electric vehicle energy behavior. The initial dataset included missing values and irregular entries, which were addressed through a structured preprocessing process. Missing values, either denoted as '*NaN*' or symbolized with double dashes ('–') were handled using linear interpolation, particularly when they occurred sporadically within continuous time-series records. This approach was employed to maintain temporal consistency without introducing artificial bias.

With respect to outliers, no explicit statistical outlier removal was performed. This decision was guided by the inherent robustness of the CatBoost algorithm, which is designed to tolerate irregularities through its gradient boosting framework. Retaining such variations allows the model to generalize better under real-world conditions and avoids the risk of excluding meaningful edge-case patterns. After preprocessing, including removal of clearly invalid entries and interpolation of sparse missing data, the final cleaned dataset contained 1053,910 valid instances, used for model training and testing. Table 1 presents a sample of these input parameters.

3.2. Data normalization and evaluation

To maintain consistent scaling and boost model performance, Min-Max normalization was applied to the dataset. This approach transforms data to a predefined range, usually [0, 1], which supports numerical stability and enhances optimization algorithm convergence [46,47]. Min-Max normalization is selected for its straightforward yet effective approach to rescaling features uniformly. By compressing values within a fixed interval, this method prevents any large values from overpowering the learning process and ensures that the model's parameters update consistently. This normalization technique is especially beneficial when data features differ in scale or units, promoting more stable and efficient training. The Min-Max normalization formula can be expressed as:

$$x' = \frac{x - min(x)}{max(x) - min(x)}, \tag{3}$$

where x is the original value, min(x) is the minimum value of the feature, max(x) is the maximum value of the feature, and x' is the normalized value.

For evaluation purposes, the performances of metaheuristic-CatBoost models were evaluated subject to the following metrics:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{4}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{5}$$

$$\text{STD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}((y_i - \hat{y}_i) - \bar{e})^2}, \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{7}$$

where $n = 1, 2, \ldots, n$; y_i = actual values; $\hat{y}_i$ = predicted values while n is the number of training, validation, or test data.

The RMSE and MAE quantify average prediction error magnitude, with RMSE placing greater weight on larger errors. The STD captures the dispersion of the prediction error around its mean, providing insight into model stability and robustness. The R^2 score indicates the proportion of variance in the actual SoC that is captured by the model, offering a measure of goodness-of-fit. The use of these metrics ensures a holistic evaluation of model performance across multiple perspectives: accuracy, reliability, and error variability.

4. Metaheuristic algorithms

This section offers a concise overview of the chosen benchmarking techniques, including four hybrid CatBoost algorithms combined with metaheuristic optimization algorithms: namely BMO, PSO, GA and WOA.

4.1. Barnacles mating algorithm

BMO is a metaheuristic optimization algorithm inspired by the unique reproductive behaviors of barnacles [48]. This algorithm models the mating strategies and environmental adaptations of barnacles, simulating a population of candidate solutions that interact and "mate" based on their fitness levels. By mimicking genetic recombination and mutation, BMO guides the search for optimal solutions. Leveraging the natural behaviors of barnacles, BMO effectively explores and exploits the solution space, making it a versatile tool for solving complex optimization problems. This approach has been successfully applied to various tasks, including engineering design and scheduling, demonstrating its potential to address challenges where traditional methods

Table 1
Input-output sample.

Instance #	AC Power (kW)	LA (m/s^2)	RBS	V_batt (V)	I_batt (A)	T_batt (°C)	heating power CAN (kW)	TP	T_motor (Nm)	T_cabin	SoC (%)
15	0.4	−0.03	0	391.4	−2.2	21	0	0	0	24.5	80.3
16	0.4	0	0	391.4	−2.21	21	0	0	0	24.5	80.3
17	0.4	−0.01	0	391.4	−2.26	21	0	0	0	24.5	80.3
18	0.4	−0.03	0	391.4	−2.3	21	0	0	0	24.5	80.28
19	0.4	−0.03	0	391.4	−2.3	21	0	0	0	24.5	80.23
20	0.4	−0.01	0	391.4	−2.3	21	0	0	0	24.5	80.2
21	0.4	−0.01	0	391.4	−2.3	21	0	0	0	24.5	80.2
22	0.4	−0.03	0	391.4	−2.31	21	0	0	0	24.5	80.2
23	0.4	−0.01	0	391.4	−2.36	21	0	0	0.38	24.5	80.2
24	0.4	−0.01	0	391.4	−2.37	21	0	0	0.12	24.5	80.2

may fall short.

In BMO, new offspring are generated through a fertilization process involving neighboring solutions. Barnacles, known for their long penises adapted to their sedentary lifestyle and changing tides, serve as a model for this process. The selection of barnacle parents for creating new offspring is random, with the length of the barnacle's penis, denoted as *pl*, acting as a tuning parameter. The exploitation process involves generating new offspring inspired by the Hardy-Weinberg principle, while exploration is guided by the sperm cast situation. BMO is particularly effective at maintaining population diversity and avoiding local optima in complex search spaces. This makes it well-suited for non-convex and highly nonlinear optimization tasks such as SoC prediction, where multiple interacting parameters exist. However, BMO may require careful parameter tuning and can be computationally intensive in large-scale datasets.

4.2. Particle swarm optimization

PSO [49] is a robust optimization technique that mimics the social behavior of swarms to solve complex problems. Each particle in the swarm represents a potential solution and adjusts its position based on its own experience and the experiences of neighboring particles. This method leverages both social and cognitive behaviors to explore and exploit the solution space effectively. PSO is favored for its rapid convergence and minimal algorithm parameters [50]. Critical tuning parameters include the number of particles, which affects the algorithm's exploration and convergence behavior; the cognitive coefficient ($c1$), which influences a particle's attraction to its own best-known position; and the social coefficient ($c2$), which determines the impact of the best-known position found by the swarm. The inertia weight (w) controls the influence of a particle's previous velocity on its current velocity, balancing exploration and exploitation, while velocity limits set boundaries on particle movement to maintain effective search dynamics. These parameters are vital for guiding the swarm towards optimal solutions and ensuring an efficient search process. PSO offers rapid convergence and relatively low computational complexity, making it highly efficient for problems requiring fast approximation, such as real-time SoC estimation. Its simplicity and ease of implementation are major strengths. However, PSO is more prone to premature convergence, especially in high-dimensional or multimodal problems, unless diversity-enhancing mechanisms are introduced.

4.3. Genetic algorithm

GA represents a pioneering advancement in metaheuristic optimization, drawing its fundamental principles from biological evolution and natural selection processes. The foundation of GA was established by John Holland in the 1970s [51], with significant enhancements contributed by David E. Goldberg during the 1980s [52], leading to its widespread adoption across diverse domains of application. The evolutionary framework of GA operates through several key mechanisms, including population dynamics and selection, genetic operators, and an iterative evolution process.

Population dynamics and selection in GA implement the "survival of the fittest" principle through strategic selection methods. Various techniques, such as tournament selection and roulette wheel selection are utilized to select individuals based on fitness metrics for subsequent generations. Genetic operators, including crossover (recombination) and mutation, play crucial roles in this process. Crossover mimics biological genetic recombination by combining genetic information from parent solutions to create new offspring, facilitating the exploration of the solution space. Mutation introduces random variations in genetic information, maintaining population diversity and enabling the exploration of unexplored solution spaces through random modifications based on predetermined mutation rates.

The iterative evolution process in GA involves selection favoring higher fitness individuals, crossover combining beneficial traits, and mutation introducing necessary variability. This process continues until optimal or near-optimal solutions are reached. The theoretical foundations established by Holland's seminal work highlighted the critical role of population dynamics and genetic operators. Goldberg's subsequent contributions expanded these concepts, providing both theoretical frameworks and practical implementation guidelines. Their combined work established GA as a robust optimization methodology, particularly effective in solving complex, multi-dimensional problems. This evolutionary approach continues to influence modern optimization strategies, with applications ranging from engineering design to artificial intelligence, demonstrating the enduring impact of Holland and Goldberg's pioneering work in the field of evolutionary computation. GA excels in exploring wide and complex solution landscapes through its robust global search capability. This is advantageous for SoC optimization, where the input space may contain nonlinearity, noise, and uncertainty. Nonetheless, GA can be slower to converge than PSO and is sensitive to parameter settings like crossover and mutation rates. Additionally, without diversity control, GA may stagnate in later generations.

4.4. Whale optimization algorithm

The WOA is a population-based metaheuristic optimization method inspired by the social and hunting behavior of humpback whales, particularly their bubble-net feeding strategy [53]. Developed by Mirjalili and Lewis in 2016, WOA simulates the cooperative foraging mechanism of whales to search for optimal solutions in a high-dimensional space. The algorithm utilizes three primary operators: encircling prey, bubble-net attacking (exploitation phase), and search for prey (exploration phase). These operators are governed by mathematical models that allow WOA to dynamically switch between exploration and exploitation based on adaptive control parameters. This mechanism enables WOA to avoid local optima and achieve robust convergence towards global solutions.

WOA starts with a randomly initialized population and updates candidate solutions iteratively by modeling the whales' spiral movements and random encircling behavior. During the bubble-net feeding method, whales either follow a shrinking encircling mechanism or perform a spiral update to simulate the helix-shaped movement towards the prey. The transition between these strategies is probabilistic, promoting a balance between intensification and diversification. Due to its simple structure and minimal control parameters, WOA is computationally efficient and has been applied successfully in various engineering optimization domains, including feature selection, parameter tuning, and energy systems. However, its performance can be sensitive to problem dimensionality, and fine-tuning may be required for optimal effectiveness in specific applications such as battery SoC estimation.

5. Results and discussion

Table 2 outlines the parameter settings used for four metaheuristic algorithms: BMO, PSO, GA, and WOA in optimizing the CatBoost model for battery SoC estimation. All algorithms were set to a population size of 10 and a maximum of 10 iterations, with five simulation runs, ensuring consistency in experimental conditions. The BMO was

Table 2
Parameter settings used for all algorithms.

Algorithm	Parameter setting
# All algorithms	Population size, *pop size* =10, maximum iteration =10, Simulation runs =5
BMO	$pl = 9$
GA	*Crossover rate* =0.7; *Mutation rate* =0.1
PSO	$w = 1$, w_{damp}=0.5, c_1= 1.5 c_2= 1.5
WOA	Vector a linearly decreased from 2 to 0 throughout iteration.

configured with a parameter *pl* set to 9, which influences the mating process within the solution space, encouraging a diverse search to avoid premature convergence. GA parameters included a crossover rate of 0.9 and a mutation rate of 0.1, which promote exploration by allowing information exchange between solutions while avoiding local optima. PSO was tuned with an inertia weight (w) of 1, a damping factor (w_{damp}) of 0.5, and acceleration coefficients (c_1 and c_2) set to 1.5. These settings enabled PSO to balance exploration and exploitation by adjusting particles' movement towards optimal solutions. In addition, the WOA was also included, which does not require extensive parameter tuning. Its main control parameter, the coefficient vector a, is linearly decreased from 2 to 0 over the course of the iterations, guiding the search balance between exploration and exploitation.

This study is distinct in its integration of MATLAB and Python, where the metaheuristic algorithms are implemented in MATLAB and linked with CatBoost in Python. Such a hybrid platform approach leverages MATLAB's robust optimization capabilities alongside Python's advanced machine learning framework. This configuration facilitates an efficient search for CatBoost's optimal parameters, thus enhancing model accuracy in SoC estimation while exploring the novel potential of multi-platform optimization in the field.

Table 3 presents the performance metrics for battery SoC estimation using hybridized CatBoost models optimized with BMO, PSO, GA, and WOA. For each metric: RMSE, MAE, STD DEV, and R^2, the BMO—CatBoost model demonstrates superior performance overall, achieving the lowest average RMSE (6.2247) compared to PSO—CatBoost (6.5490), GA-CatBoost (6.5039), and WOA-CatBoost (6.6362). BMO—CatBoost also achieves the lowest worst-case RMSE (6.4492), suggesting it maintains greater consistency across trials. This consistency is further reflected in MAE, where BMO—CatBoost also shows the lowest average value (4.1744) and the best-case value (4.1303), indicating that BMO effectively reduces prediction errors and enhances the accuracy of SoC estimation. Interestingly, WOA-CatBoost shows promising performance, outperforming both PSO and GA in best-case MAE (4.1563), though its average MAE (4.4287) remains higher than BMO's. This suggests that while WOA can occasionally achieve highly accurate results, its stability over repeated runs is comparatively lower.

The implicit findings from Table 3 suggest that BMO not only provides more accurate estimates but also achieves a narrower range in terms of error variability when compared to PSO, GA, and WOA. Regarding standard deviation, BMO—CatBoost achieves a lower average STD DEV (5.8151), reflecting a stable performance across different runs, which is especially important in applications like SoC estimation that require reliable and repeatable results. PSO—CatBoost and GA-CatBoost have higher variability, with GA showing the highest worst-case STD DEV (6.2726). This variability might suggest that while PSO and GA are effective for certain types of optimization tasks, BMO's unique mating-inspired mechanisms are more robust for the type of parameter tuning needed in CatBoost. Although WOA-CatBoost does not outperform BMO, it demonstrates a slightly lower best-case STD DEV (5.8149) than PSO and GA, indicating potential in achieving consistent results under ideal conditions. However, its higher average and worst-case STD DEV values suggest room for improvement in robustness.

Table 3
Performance of SoC Estimation using BMO, PSO, and GA-CatBoost Algorithms.

Metrics		BMO—CatBoost	PSO—CatBoost	GA-CatBoost	WOA-CatBoost
RMSE	Best	6.1031	6.3784	6.2992	6.2283
	Worst	6.4492	6.7073	6.7893	7.0383
	Average	6.2247	6.5490	6.5039	6.6362
MAE	Best	4.1303	4.2184	4.2772	4.1563
	Worst	4.3105	4.5116	4.5771	4.6659
	Average	4.1744	4.3900	4.3832	4.4287
STD	Best	5.7183	5.9760	5.8337	5.8149
	Worst	5.9833	6.1627	6.2726	6.4859
	Average	5.8151	6.0751	6.0164	6.1522
R^2	Best	0.8211	0.8017	0.8111	0.8175
	Worst	0.7919	0.7734	0.7652	0.7381
	Average	0.8127	0.7849	0.7894	0.7768

A key insight from the R^2 values across algorithms highlights the explanatory power of each hybrid model. BMO—CatBoost achieves the highest average R^2 (0.8127) and best-case R^2 (0.8211), indicating a stronger fit of the model to the data compared to PSO—CatBoost, GA-CatBoost, and WOA-CatBoost. Although PSO—CatBoost and GA-CatBoost perform reasonably well, they fall slightly behind BMO—CatBoost in capturing the underlying patterns of the data. WOA-CatBoost, while not surpassing BMO, achieves a best-case R^2 of 0.8175, which is higher than both PSO and GA, highlighting its capacity to capture underlying data patterns effectively in certain runs. Its average R^2 (0.7768), however, remains the lowest among all compared models, emphasizing variability in its predictive accuracy. These results reveal that BMO—CatBoost not only achieves lower errors but also exhibits higher reliability and better interpretability, positioning it as a potentially more effective hybrid model for SoC estimation tasks in electric vehicles. This advantage may stem from BMO's unique approach to balancing exploration and exploitation, allowing it to achieve a more precise search for optimal hyperparameters within the CatBoost framework.

Table 4 provides a detailed breakdown of the optimized CatBoost hyperparameters using the BMO, PSO, GA, and WOA metaheuristic algorithms. For the learning rate, BMO—CatBoost achieved an optimal value of 0.08 within the specified range (0.01–0.1), while PSO—CatBoost found a slightly lower rate at 0.05. GA-CatBoost, on the other hand, selected a learning rate just outside the upper bound, at 0.104, which suggests a more aggressive training approach. Interestingly, the WOA-CatBoost model selected a learning rate of 0.081, very close to BMO—CatBoost's optimal value, suggesting convergence toward this rate as potentially ideal for the dataset. The similar learning rates between BMO and WOA likely contributed to their comparable performance metrics, with both outperforming PSO and GA variants.

The regularization and subsample parameters show significant variance across the algorithms, highlighting different optimization approaches. For regularization, both BMO and GA settled on a value of 5, which balances model complexity and prevents overfitting. PSO set this parameter to the upper limit of 10, indicating a higher penalization of model weights. Notable is WOA-CatBoost's selection of a much lower regularization value of 2, suggesting a less restrictive approach that still delivered strong performance with an R^2 of 0.8175, second only to BMO—CatBoost. This lower regularization in WOA may have allowed greater model expressiveness while maintaining effective generalization. In terms of subsample, BMO—CatBoost selected the minimum value of 0.1, likely favoring a conservative approach to subsampling data for model training. WOA-CatBoost chose a similarly conservative value of 0.19, contrasting with PSO—CatBoost's higher subsample rate of 0.72 and GA-CatBoost's moderate rate of 0.50. The superior performance of both BMO and WOA models suggests that more conservative subsampling strategies may be advantageous for this particular dataset.

Examining additional parameters like features of each level, bagging temperature, and random strength reveals further nuances. The features of each level value chosen are similar across algorithms, with BMO, PSO, and WOA selecting high values (0.83, 0.82, and 0.81, respectively), which emphasizes feature diversity within trees. GA-CatBoost's choice of 0.74 indicates a slightly lower feature diversity per level, possibly influencing its lower R^2 value (0.8111) compared to BMO—CatBoost (0.8211) and WOA-CatBoost (0.8175). For bagging temperature, BMO—CatBoost set this parameter near the upper bound at 0.97, allowing for higher sample variability across iterations, while PSO chose 0, and GA set it at 0.91. WOA-CatBoost selected a moderate value of 0.50, striking a balance between exploration and exploitation. Lastly, random strength, which controls randomness in feature selection, is lowest in BMO—CatBoost at 1, suggesting minimal randomization,

Table 4
Results of optimized values based on metaheuristic-CatBoost algorithms.

Parameters	Min-Max	BMO—CatBoost	PSO—CatBoost	GA-CatBoost	WOA-CatBoost
Learning rate	[0.01–0.1]	0.08	0.05	0.104	0.081
Depth	[3–10]	5	5	5	5
Regularization	[1–10]	5	10	5	2
Subsample	[0.1–1]	0.1	0.72	0.50	0.19
Feature of each level	[0.1–1]	0.83	0.82	0.74	0.81
Bagging temperature	[0–1]	0.97	0	0.91	0.50
Random strength	[1–20]	1	5	6	9
RMSE		6.1031	6.3784	6.2992	6.2283
MAE		4.1303	4.2184	4.2772	4.1563
STD		5.7183	5.976	5.8337	5.8149
R^2		0.8211	0.8017	0.8111	0.8175

potentially enhancing precision. PSO, GA, and WOA opted for progressively higher values (5, 6, and 9, respectively), with WOA's higher random strength possibly contributing to its robust performance despite using less restrictive regularization. Overall, BMO—CatBoost's parameter selections delivered superior performance across RMSE, MAE, and R^2, followed closely by WOA-CatBoost, which achieved the second-best performance in all metrics (RMSE: 6.2283, MAE: 4.1563, R^2: 0.8175). This suggests that while BMO optimization remains the top performer, the WOA-based hybrid approach offers a compelling alternative that outperforms both PSO and GA variants.

The performances of all metaheuristic-CatBoost algorithms for SoC estimation are visualized in Figs. 2 to 5, covering approximately 350,000 instances. Fig. 2 shows the results of BMO—CatBoost, while Figs. 3, 4, and 5 show the results of PSO—CatBoost, GA-CatBoost and WOA-CatBoost for the SoC estimation task, respectively. In the top plot of all figures, the actual SoC values are shown in blue, while the SoC values predicted by the BMO—CatBoost, PSO—CatBoost, and GA-CatBoost models are shown in red. All models appear to track the overall trend of the actual SoC reasonably well, capturing the major fluctuations over time. However, noticeable discrepancies between the predicted and actual SoC values can be observed at certain intervals, particularly between instances 1.7×10^5 and 2.0×10^5. These discrepancies likely arise from rapid changes in operating conditions, such as abrupt shifts in current or temperature, which introduce nonlinear behavior that is challenging to capture accurately.

The bottom plots of Figs. 2 to 5 show the normalized error between the actual and predicted SoC values. The error fluctuates between −0.4 and 0.1, suggesting that the models' predictions generally vary within this range compared to the actual SoC. Among all methods, BMO—CatBoost demonstrates the most stable error distribution and lowest error margins, indicating superior generalization and robustness under varying conditions. These results demonstrate that the hybrid BMO—CatBoost approach can provide a reasonably accurate estimation of the battery SoC, although the observed discrepancies highlight the need for further refinement. Moreover, these discrepancies are recognized as inherent limitations of the current modeling framework, which are explicitly addressed in this paper. Specifically, improvements could be pursued by incorporating enhanced input features such as time-

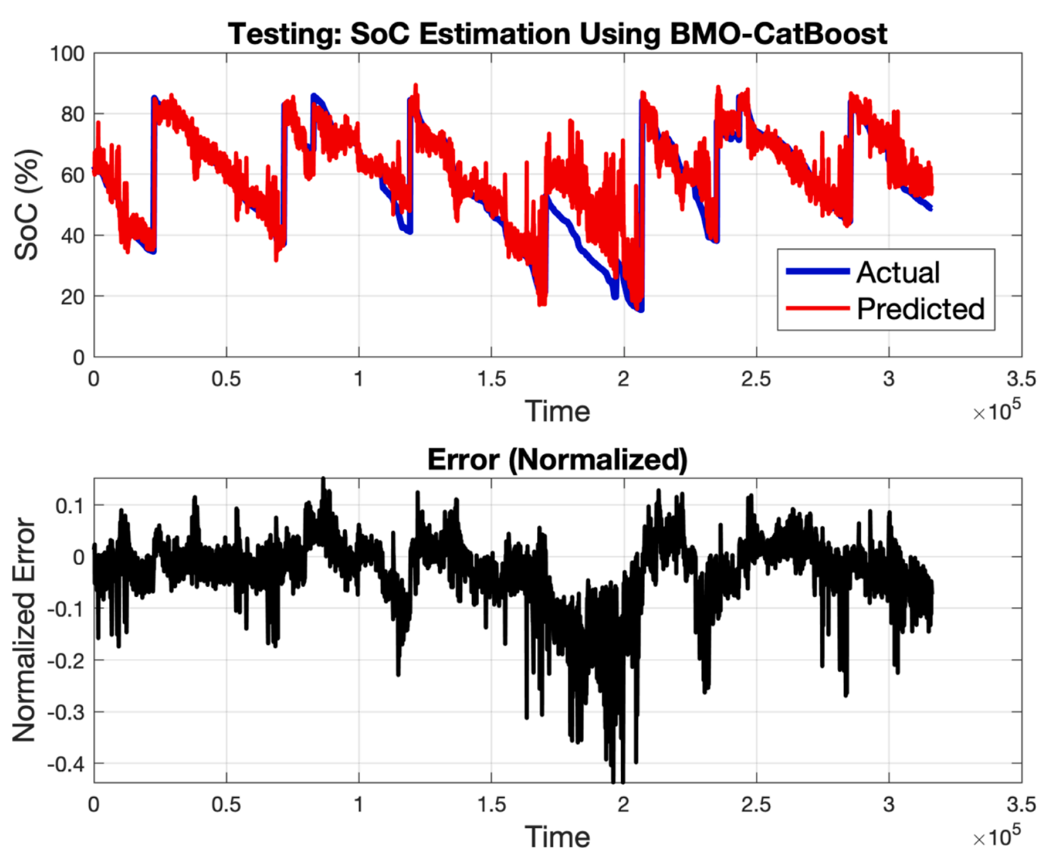

Fig. 2. Soc Estimation using BMO—CatBoost.

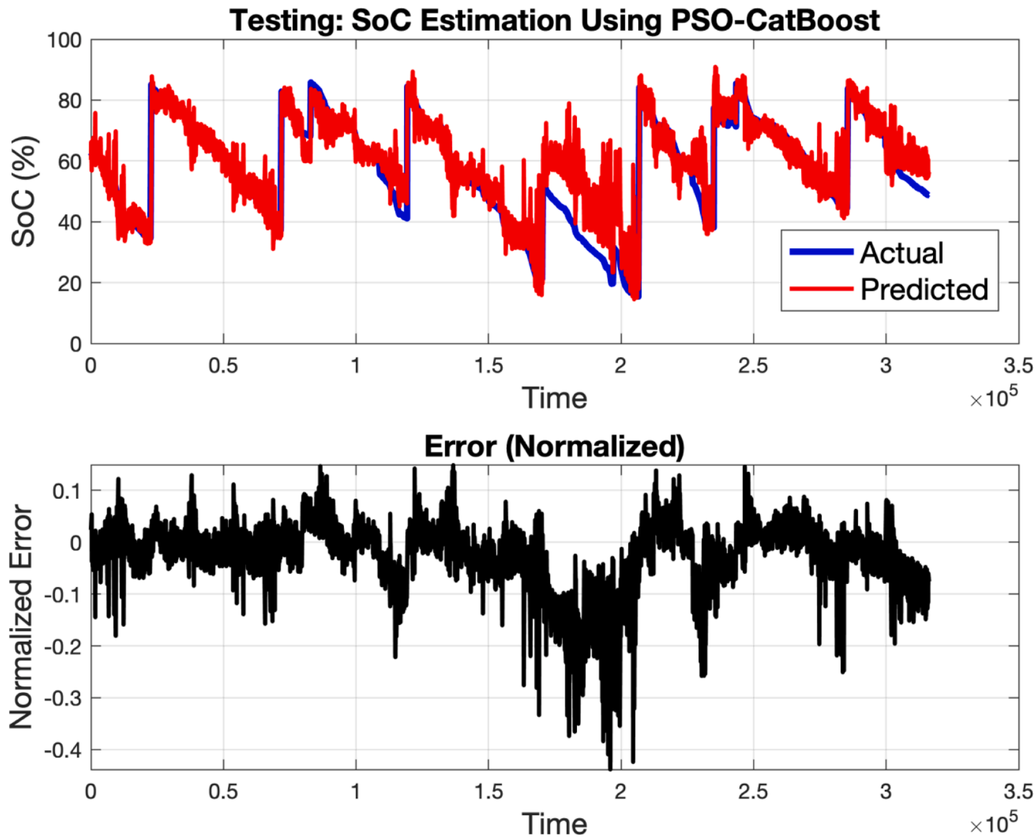

Fig. 3. Soc Estimation using PSO—CatBoost.

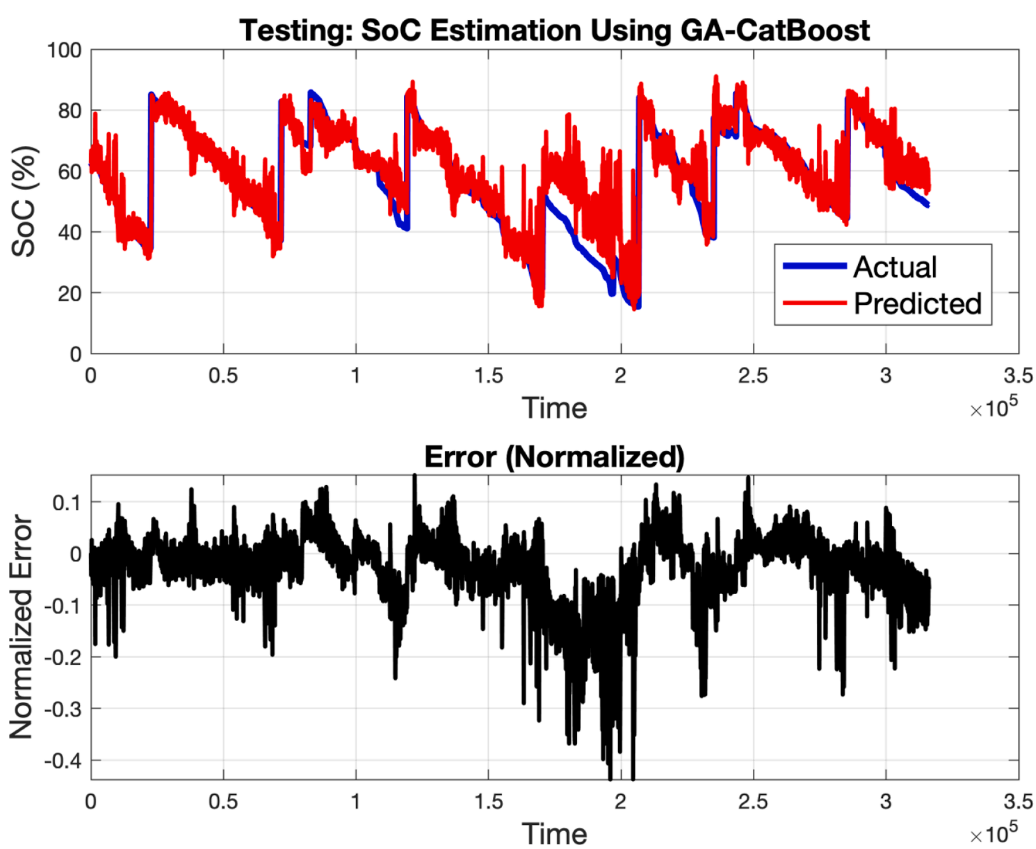

Fig. 4. Soc Estimation using GA-CatBoost.

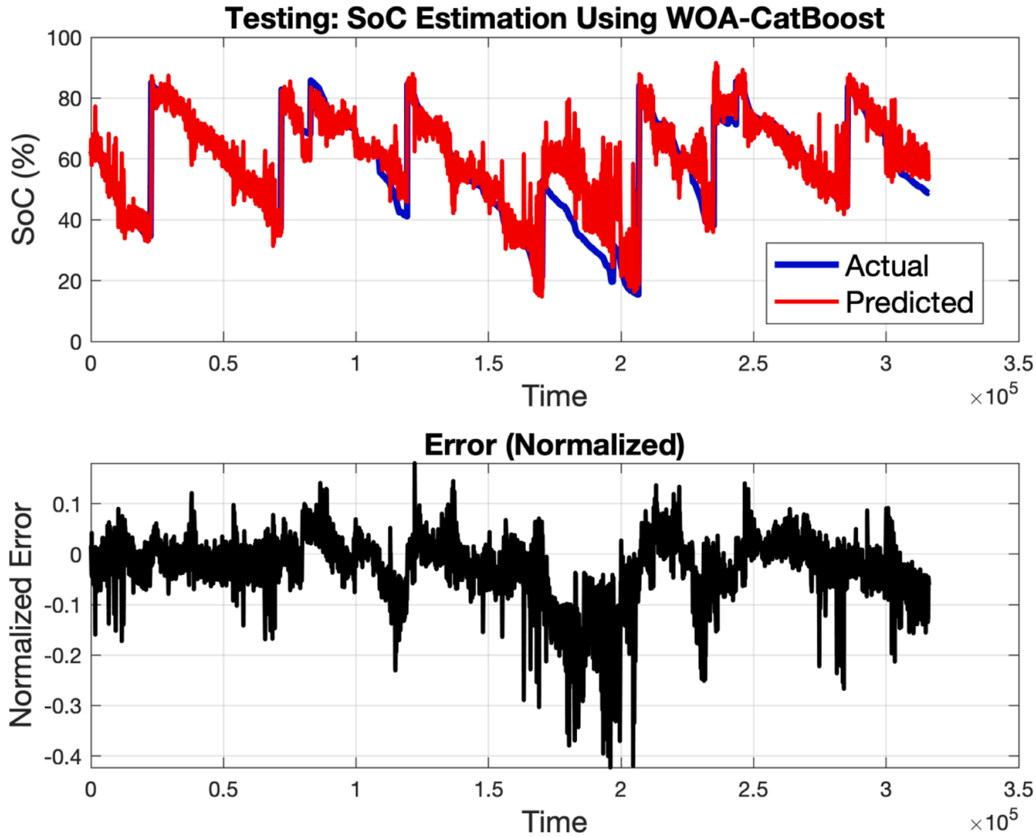

Fig. 5. Soc Estimation using WOA-CatBoost.

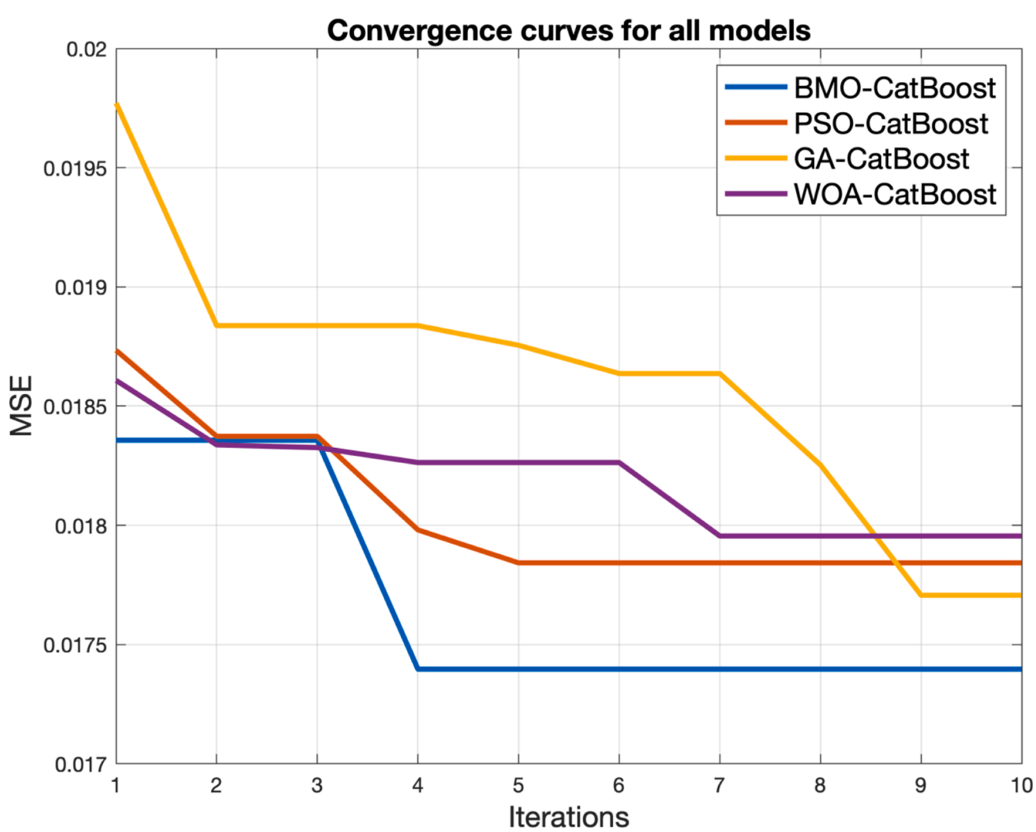

Fig. 6. Convergence curves of BMO—CatBoost, PSO—CatBoost, GA-CatBoost, and WOA-CatBoost.

series-based derivatives or dynamic load profile indicators to better capture temporal dependencies and evolving system dynamics. The adoption of advanced regularization techniques or ensemble learning strategies may further improve model adaptability under transient conditions by reducing overfitting and increasing robustness to outlier behaviors. Additionally, integrating external environmental parameters such as ambient temperature and humidity may aid in modeling complex nonlinear interactions that affect battery performance, thereby enabling more precise SoC estimations under diverse operating scenarios. These directions are expected to significantly contribute to the development of a more resilient and accurate predictive framework.

Fig. 6 illustrates the convergence behavior of the BMO, PSO, GA, and WOA algorithms when hybridized with CatBoost for the training phase of State of Charge (SoC) estimation. The convergence curves reveal that BMO consistently outperforms the other algorithms, achieving the lowest final objective function value of 0.017396969. This suggests that BMO is particularly effective at minimizing the objective function, which is a critical aspect in enhancing the accuracy of SoC estimation. BMO's superior performance in terms of convergence indicates its strong ability to explore and exploit the solution space efficiently, reaching optimal or near-optimal solutions with fewer iterations compared to PSO, GA, and WOA. WOA-CatBoost demonstrates interesting convergence characteristics, starting with a relatively low MSE value (approximately 0.0185) and exhibiting a steady, incremental improvement throughout the iterations. Unlike BMO's sharp drop around iteration 4, WOA shows a more gradual optimization path with a notable improvement at iteration 7, ultimately settling at a final objective function value of approximately 0.0179. This positions WOA-CatBoost as the third-best performer, behind BMO and GA but ahead of PSO in terms of final convergence value.

Although GA shows initially poor performance with the highest starting MSE values (around 0.0197), it demonstrates remarkable improvement over time with a final objective function value of 0.01770694, eventually surpassing both PSO and WOA by the final iterations. This dramatic improvement highlights GA's strong exploitation capabilities in later iterations, despite its slower initial convergence. In contrast, PSO demonstrates a faster initial convergence rate compared to GA, yet it fails to achieve the same level of objective function minimization as BMO and GA by the end of the training phase, settling at approximately 0.0178. A particularly noteworthy observation is the convergence pattern between iterations 2–3, where BMO, PSO, and WOA all demonstrate nearly identical objective function values before diverging significantly. BMO shows a steep improvement at iteration 4, whereas WOA maintains a more consistent, gradual improvement trajectory throughout the optimization process. This suggests that while BMO excels at finding breakthrough improvements in specific iterations, WOA offers more predictable, stable convergence behavior that might be advantageous in certain applications where consistency is valued. These differences highlight the unique strengths of each algorithm; while PSO excels in rapid early convergence, BMO achieves both the fastest significant improvement and the best final result, GA demonstrates powerful late-stage optimization despite poor initial performance, and WOA offers stable, consistent improvement throughout the process. Overall, BMO's convergence profile confirms its advantage as the most effective metaheuristic algorithm for optimizing CatBoost in the SoC estimation task, while WOA presents a compelling alternative with its steady convergence characteristics.

The results and analysis provided offer several key insights into the effectiveness and limitations of hybridized CatBoost models optimized with different metaheuristic algorithms for battery SoC estimation. Overall, the BMO—CatBoost model consistently outperforms the PSO—CatBoost, GA-CatBoost, and WOA-CatBoost models across various performance metrics, such as RMSE, MAE, STD DEV, and R^2. This suggests that the BMO algorithm is highly effective at fine-tuning the CatBoost parameters to reduce prediction errors and maintain stability across trials. This level of stability and precision is essential for SoC estimation, as it ensures the model can reliably track the battery's performance over time, making it suitable for real-world applications in electric vehicles. However, while the BMO—CatBoost model demonstrates superior performance, the noticeable discrepancies between predicted and actual SoC values at specific time intervals, particularly between instances 1.7×10^5 and 2.0×10^5. These discrepancies suggest challenges in modeling dynamic transitions in battery behavior, which may result from unaccounted external factors such as temperature fluctuations, rapid load changes, or battery aging effects.

Despite its effectiveness, the hybrid BMO—CatBoost model still faces challenges that warrant further investigation. The deviations observed in the error plots suggest that there may be underlying dynamics in battery behavior that the current model configuration fails to capture fully. Future work could explore incorporating additional environmental or operational data, such as temperature variations, charge/discharge rates, and battery age, to enhance the model's ability to generalize across a broader range of conditions. Moreover, while BMO's approach to parameter tuning demonstrates promise, it may benefit from combining with adaptive optimization techniques or multi-objective frameworks to further balance between minimizing errors and enhancing model robustness. Integrating domain-specific knowledge or applying regularization techniques could also help address the occasional overfitting issues that might arise from BMO's parameter tuning.

Finally, while BMO—CatBoost shows clear advantages, the computational demands of using metaheuristic algorithms like BMO, PSO, GA, and WOA in the training phase present practical challenges. Metaheuristic optimization often requires significant computational resources, especially when handling large datasets or complex parameter spaces. This limitation could impact the scalability of these models for large-scale applications or real-time systems. Therefore, one potential area for improvement is developing more computationally efficient optimization strategies, such as lightweight versions of BMO or hybrid models that utilize faster optimization algorithms during initial training phases, followed by fine-tuning with BMO. Additionally, examining the trade-offs between computational efficiency and predictive accuracy will be essential for advancing SoC estimation models and making them more feasible for widespread deployment in battery management systems.

In comparison with prior studies, which primarily employ standard machine learning methods such as Support Vector Machines (SVM) [54] and Random Forests (RF) for SoC estimation [45], the proposed BMO—CatBoost hybrid model demonstrates enhanced predictive accuracy and improved generalization performance. Traditional models often lack built-in mechanisms for handling categorical features and may suffer from overfitting when trained on large, high-dimensional datasets. The presented results indicate that CatBoost when effectively optimized using metaheuristic algorithms, can outperform these conventional approaches by achieving lower error metrics (RMSE and MAE) and more stable performance across repeated trials. Additionally, while many previous works rely on limited or simulated datasets, this study utilizes over one million real-world driving instances from an actual electric vehicle, thereby strengthening the model's practical relevance and applicability to real-time battery management systems.

Moreover, while this study focuses on enhancing CatBoost using metaheuristic algorithms, we acknowledge that alternative machine learning methods such as RF, XGBoost, and neural networks have also shown considerable promise in SoC estimation tasks. A comparative investigation involving these algorithms could offer further insights into model performance and suitability under various conditions. However, such an analysis is beyond the scope of this study and is proposed as part of our future research directions.

6. Conclusion

A hybrid approach for lithium-ion battery SoC estimation by combining the CatBoost algorithm with metaheuristic optimization, aimed at enhancing predictive accuracy and robustness, has been presented in this study. Data from 72 real-world driving trips of a BMW i3 electric car with a 60 Ah battery capacity were used, encompassing a comprehensive set of input features, including battery and vehicle operation metrics. Systematic data cleaning and processing addressed missing values and outliers, resulting in a refined dataset of 1053,910 instances. Each input feature was normalized using Min-Max scaling, ensuring consistent scaling and numerical stability during model training. The CatBoost model was trained on this prepared dataset to capture complex relationships between features and SoC, with metaheuristic algorithms employed to optimize key parameters, including learning rate, depth, regularization, subsample, features by each level, bagging temperature, and random strength. This hybrid approach not only achieved a high level of predictive accuracy on the test dataset but also demonstrated enhanced generalization capabilities, providing a more robust and accurate SoC estimation compared to standalone methods.

Nonetheless, several limitations should be acknowledged. Firstly, the model's complexity, particularly with the integration of metaheuristic optimization, increases the computational burden during training, which may limit real-time applicability. Secondly, although the dataset is extensive and based on real-world trips, it is constrained to a single vehicle model (BMW i3) and specific usage patterns, which may affect the generalizability of the findings to other battery chemistries, vehicle types, or driving behaviors. Moreover, the model's performance could potentially vary under conditions or input distributions not represented in the training data.

From a practical standpoint, the proposed hybrid model offers a data-driven solution that enhances the accuracy and reliability of SoC estimation in battery management systems. This can lead to improved range predictions, better charging control, and longer battery lifespan in electric vehicles. Moreover, the model's ability to generalize across varied driving profiles makes it suitable for deployment in diverse real-world scenarios, including fleet management and energy storage applications. The robustness of the optimized CatBoost model also supports its integration into embedded BMS hardware, where predictive precision is crucial for safety and operational efficiency.

Future research could explore the application of additional feature engineering techniques or incorporate environmental variables, such as humidity and altitude, to further improve model performance. Additionally, extending this approach to other battery types or integrating real-time learning mechanisms may expand its utility across broader battery management applications. By leveraging the powerful learning capabilities of CatBoost with the stabilization benefits of metaheuristic optimization, this approach contributes a valuable method to the field of battery SoC estimation, with promising implications for electric vehicle energy management and other energy storage applications.

Data availability declaration

Data will be made available on request.

Ethics statement

Ethical approval is not applicable to this article. This article does not contain any studies with human or animal subjects. There are no human subjects in this article and informed consent is not applicable.

Author statement

Mohd Herwan Sulaiman: Conceptualization; Methodology; Formal analysis; Writing original draft

Zuriani Mustaffa: Data curation; Validation; writing – review and editing

Ahmad Salihin Samsudin: Resources; Project administration; writing – review and editing

Amir Izzani Mohamed: Investigation; writing – review and editing

Mohd Mawardi Saari: Validation; writing – review and editing

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to improve language and readability, with caution. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

CRediT authorship contribution statement

Mohd Herwan Sulaiman: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Zuriani Mustaffa:** Writing – review & editing, Validation, Data curation. **Ahmad Salihin Samsudin:** Writing – review & editing, Resources, Project administration. **Amir Izzani Mohamed:** Writing – review & editing, Investigation. **Mohd Mawardi Saari:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Ministry of Higher Education Malaysia (MOHE) under the Fundamental Research Grant Scheme (FRGS/1/2024/ICT02/UMP/02/2) and Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) under Distinguished Research Grant (# RDU223003).

References

[1] N. Khan, C.A. Ooi, A. Alturki, M. Amir, Shreasth, T. Alharbi, A critical review of battery cell balancing techniques, optimal design, converter topologies, and performance evaluation for optimizing storage system in electric vehicles, Energy Reports 11 (2024) 4999–5032, https://doi.org/10.1016/j.egyr.2024.04.041, 2024/06/01/.

[2] Z. Mustaffa, M.H. Sulaiman, Enhancing battery state of charge estimation through hybrid integration of barnacles mating optimizer with deep learning, Franklin Open 5 (2023) 100053, https://doi.org/10.1016/j.fraope.2023.100053, 2023/12/01/.

[3] M. Maurya, S. Gawade, N. Zope, A study of different machine learning algorithms for state of charge estimation in lithium-ion battery pack, Energy Storage 6 (4) (2024) e658, https://doi.org/10.1002/est2.658.

[4] O. Demirci, S. Taskin, E. Schaltz, B.Acar Demirci, Review of battery state estimation methods for electric vehicles - Part I: SOC estimation, Journal of Energy Storage 87 (2024) 111435, https://doi.org/10.1016/j.est.2024.111435, 2024/05/15/.

[5] M.R. Harikrishnan, P. Kanakasabapathy, Modeling and performance analysis of a lithium-ion battery pack with an electric vehicle power-train for different drive cycles and highway conditions, Energy Storage 6 (2) (2024) e584, https://doi.org/10.1002/est2.584.

[6] P. Dini, A. Colicelli, S. Saponara, Review on Modeling and SOC/SOH Estimation of Batteries for Automotive Applications, Batteries 10 (1) (2024) 34 [Online]. Available: https://www.mdpi.com/2313-0105/10/1/34.

[7] M. Amer, J. Masri, A. Dababat, U. Sajjad, K. Hamid, Electric vehicles: Battery technologies, charging standards, AI communications, challenges, and future directions, Energy Conversion and Management: X 24 (2024) 100751, https://doi.org/10.1016/j.ecmx.2024.100751, 2024/10/01/.

[8] F. Mohammadi, Lithium-ion battery State-of-Charge estimation based on an improved Coulomb-Counting algorithm and uncertainty evaluation, Journal of Energy Storage 48 (2022) 104061, https://doi.org/10.1016/j.est.2022.104061, 2022/04/01/.

[9] S. Zhang, X. Guo, X. Dou, X. Zhang, A data-driven coulomb counting method for state of charge calibration and estimation of lithium-ion battery, Sustainable Energy Technologies and Assessments 40 (2020) 100752, https://doi.org/10.1016/j.seta.2020.100752, 2020/08/01/.

[10] S. Zhang, X. Guo, X. Dou, X. Zhang, A rapid online calculation method for state of health of lithium-ion battery based on coulomb counting method and differential

voltage analysis, Journal of Power Sources 479 (2020) 228740, https://doi.org/10.1016/j.jpowsour.2020.228740, 2020/12/15/.
[11] Y. Liu, C. Liu, Y. Liu, F. Sun, J. Qiao, T. Xu, Review on degradation mechanism and health state estimation methods of lithium-ion batteries, Journal of Traffic and Transportation Engineering (English Edition) 10 (4) (2023) 578–610, https://doi.org/10.1016/j.jtte.2023.06.001, 2023/08/01/.
[12] H. Sorouri, A. Oshnoei, Y. Che, R. Teodorescu, A comprehensive review of hybrid battery state of charge estimation: Exploring physics-aware AI-based approaches, Journal of Energy Storage 100 (2024) 113604, https://doi.org/10.1016/j.est.2024.113604, 2024/10/20/.
[13] R.P. Priya, S. R, R. Sakile, State of charge estimation of lithium-ion battery based on extended Kalman filter and unscented Kalman filter techniques, Energy Storage 5 (3) (2023) e408, https://doi.org/10.1002/est2.408.
[14] H. Li, et al., A cubature Kalman filter for online state-of-charge estimation of lithium-ion battery using a gas-liquid dynamic model, Journal of Energy Storage 53 (2022) 105141, https://doi.org/10.1016/j.est.2022.105141, 2022/09/01/.
[15] Y. Jin, C. Su, S. Luo, Improved Algorithm Based on AEKF for State of Charge Estimation of Lithium-ion Battery, International Journal of Automotive Technology 23 (4) (2022) 1003–1011, https://doi.org/10.1007/s12239-022-0087-x, 2022/08/01.
[16] G. Li, C. Liu, E. Wang, L. Wang, State of Charge Estimation for Lithium-Ion Battery Based on Improved Cubature Kalman Filter Algorithm, Automotive Innovation 4 (2) (2021) 189–200, https://doi.org/10.1007/s42154-021-00134-4, 2021/05/01.
[17] C. Ge, Y. Zheng, Y. Yu, State of charge estimation of lithium-ion battery based on improved forgetting factor recursive least squares-extended Kalman filter joint algorithm, Journal of Energy Storage 55 (2022) 105474, https://doi.org/10.1016/j.est.2022.105474, 2022/11/15/.
[18] P. Mei, et al., Battery state estimation methods and management system under vehicle–cloud collaboration: a Survey, Renewable and Sustainable Energy Reviews 206 (2024) 114857, https://doi.org/10.1016/j.rser.2024.114857, 2024/12/01/.
[19] C. Yang, Z. Ye, X. Xiong, C. Su, M. Xie, Improving electric vehicles sustainability: Accurate forecasting of lithium-ion battery health using machine learning models, Journal of Energy Storage 103 (2024) 114280, https://doi.org/10.1016/j.est.2024.114280, 2024/12/01/.
[20] M.H. Sulaiman, Z. Mustaffa, S. Razali, M.R. Daud, Advancing battery state of charge estimation in electric vehicles through deep learning: a comprehensive study using real-world driving data, Cleaner Energy Systems 8 (2024) 100131, https://doi.org/10.1016/j.cles.2024.100131, 2024/08/01/.
[21] X. Yang, J. Hu, G. Hu, X. Guo, Battery state of charge estimation using temporal convolutional network based on electric vehicles operating data, Journal of Energy Storage 55 (2022) 105820, https://doi.org/10.1016/j.est.2022.105820, 2022/11/30/.
[22] J. Tian, C. Chen, W. Shen, F. Sun, R. Xiong, Deep Learning Framework for Lithium-ion Battery State of Charge Estimation: Recent Advances and Future Perspectives, Energy Storage Materials 61 (2023) 102883, https://doi.org/10.1016/j.ensm.2023.102883, 2023/08/01/.
[23] B.P. Adedeji, G. Kabir, A feedforward deep neural network for predicting the state-of-charge of lithium-ion battery in electric vehicles, Decision Analytics Journal 8 (2023) 100255, https://doi.org/10.1016/j.dajour.2023.100255, 2023/09/01/.
[24] A. Shokry, M.A.E. Qassime, A. Espuña, E. Moulines, Health-aware optimal charging of lithium-ion batteries using deep-neural networks-based explicit constrained model predictive control, Computers & Chemical Engineering 199 (2025) 109096, https://doi.org/10.1016/j.compchemeng.2025.109096, 2025/08/01/.
[25] C. Li, et al., State of charge estimation of lithium-ion battery based on state of temperature estimation using weight clustered-convolutional neural network-long short-term memory, Green Energy and Intelligent Transportation 4 (1) (2025) 100226, https://doi.org/10.1016/j.geits.2024.100226, 2025/02/01/.
[26] J. Tao, S. Wang, W. Cao, C. Fernandez, F. Blaabjerg, L. Cheng, An innovative multitask learning - Long short-term memory neural network for the online anti-aging state of charge estimation of lithium-ion batteries adaptive to varying temperature and current conditions, Energy 314 (2025) 134272, https://doi.org/10.1016/j.energy.2024.134272, 2025/01/01/.
[27] Y. Zou, S. Wang, W. Cao, N. Hai, C. Fernandez, Enhanced transformer encoder long short-term memory hybrid neural network for multiple temperature state of charge estimation of lithium-ion batteries, Journal of Power Sources 632 (2025) 236411, https://doi.org/10.1016/j.jpowsour.2025.236411, 2025/03/15/.
[28] C. Lin, et al., Multi-model ensemble learning for battery state-of-health estimation: Recent advances and perspectives, Journal of Energy Chemistry 100 (2025) 739–759, https://doi.org/10.1016/j.jechem.2024.09.021, 2025/01/01/.
[29] X. Sun, J. Fu, Many-objective optimization of BEV design parameters based on gradient boosting decision tree models and the NSGA-III algorithm considering the ambient temperature, Energy 288 (2024) 129840, https://doi.org/10.1016/j.energy.2023.129840, 2024/02/01/.
[30] U. Khan, S. Kirmani, Y. Rafat, M.U. Rehman, M.S. Alam, Improved deep learning based state of charge estimation of lithium ion battery for electrified transportation, Journal of Energy Storage 91 (2024) 111877, https://doi.org/10.1016/j.est.2024.111877, 2024/06/30/.
[31] M.S. Reza, et al., Towards enhanced remaining useful life prediction of lithium-ion batteries with uncertainty using optimized deep learning algorithm, Journal of Energy Storage 98 (2024) 113056, https://doi.org/10.1016/j.est.2024.113056, 2024/09/15/.
[32] M. Pietrołaj, M. Blok, Resource constrained neural network training, Scientific Reports 14 (1) (2024) 2421, https://doi.org/10.1038/s41598-024-52356-1, 2024/01/29.
[33] Z. Liu, Z. Zhao, Y. Qiu, B. Jing, C. Yang, H. Wu, Enhanced state of charge estimation for Li-ion batteries through adaptive maximum correntropy Kalman filter with open circuit voltage correction, Energy 283 (2023) 128738, https://doi.org/10.1016/j.energy.2023.128738, 2023/11/15/.
[34] S.A.A. Shah, et al., A hybrid deep learning framework integrating bidirectional sliding windows, TCN, and external attention for accurate state-of-charge estimation in lithium-ion batteries, Journal of Power Sources 622 (2024) 235312, https://doi.org/10.1016/j.jpowsour.2024.235312, 2024/12/01/.
[35] M.H. Zafar, et al., Hybrid deep learning model for efficient state of charge estimation of Li-ion batteries in electric vehicles, Energy 282 (2023) 128317, https://doi.org/10.1016/j.energy.2023.128317, 2023/11/01/.
[36] Z. Sherkatghanad, A. Ghazanfari, V. Makarenkov, A self-attention-based CNN-Bi-LSTM model for accurate state-of-charge estimation of lithium-ion batteries, Journal of Energy Storage 88 (2024) 111524, https://doi.org/10.1016/j.est.2024.111524, 2024/05/30/.
[37] M.J. Salehpour, M.J. Hossain, Leveraging machine learning for efficient EV integration as mobile battery energy storage systems: Exploring strategic frameworks and incentives, Journal of Energy Storage 92 (2024) 112151, https://doi.org/10.1016/j.est.2024.112151, 2024/07/01/.
[38] Z. Mustaffa, M.H. Sulaiman, J. Isuwa, State of Charge Estimation of Lithium-ion Batteries in an Electric Vehicle using Hybrid Metaheuristic - Deep Neural Networks Models, Energy Storage and Saving (2025), https://doi.org/10.1016/j.enss.2025.01.002, 2025/02/16/.
[39] N. Khosravi, M. Dowlatabadi, M.B. Abdelghany, M. Tostado-Véliz, F. Jurado, Enhancing battery management for HEVs and EVs: a hybrid approach for parameter identification and voltage estimation in lithium-ion battery models, Applied Energy 356 (2024) 122364, https://doi.org/10.1016/j.apenergy.2023.122364, 2024/02/15/.
[40] R. PVLN, V. Lavanya, D. Manasa, B. S, S. BP, Battery Models and Estimation Techniques for Energy Storage Systems in Residential Buildings, Journal of Modern Technology 1 (1) (September 2024) 47–58, https://doi.org/10.71426/jmt.v1.i1.pp47-58, 2024;1(1):47~58.
[41] S. Arandhakar, N. Jayaram, V.B.M. Krishna, A comprehensive analysis and future prospects on battery energy storage systems for electric vehicle applications, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 46 (1) (2024) 13003–13030, https://doi.org/10.1080/15567036.2024.2401118, 2024/12/31.
[42] A. Veronika Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," p. doi: 10.48550/arXiv.1810.11363.
[43] G. Wadhwa. "Predicting car battery heating data." https://www.kaggle.com/code/gitanjali1425/predicting-car-battery-heating-data (accessed.
[44] M.S.J.B.D. Trifonov. Battery and Heating Data in Real Driving Cycles, doi: 10.21227/6jr9-5235.
[45] M.H. Sulaiman, Z. Mustaffa, State of Charge Estimation for Electric Vehicles Using Random Forest, Green Energy and Intelligent Transportation (2024) 100177, https://doi.org/10.1016/j.geits.2024.100177, 2024/01/19/.
[46] Y.S. Kim, M.K. Kim, N. Fu, J. Liu, J. Wang, J. Srebric, Investigating the Impact of Data Normalization Methods on Predicting Electricity Consumption in a Building Using different Artificial Neural Network Models, Sustainable Cities and Society (2024) 105570, https://doi.org/10.1016/j.scs.2024.105570, 2024/06/03/.
[47] L.A. Shalabi and Z. Shaaban, "Normalization as a Preprocessing Engine for Data Mining and the Approach of Preference Matrix," in 2006 International Conference on Dependability of Computer Systems, 25-27 May 2006 2006, pp. 207–214, doi: 10.1109/DEPCOS-RELCOMEX.2006.38.
[48] M.H. Sulaiman, Z. Mustaffa, M.M. Saari, H. Daniyal, Barnacles Mating Optimizer: a new bio-inspired algorithm for solving engineering optimization problems, Engineering Applications of Artificial Intelligence 87 (2020) 103330, https://doi.org/10.1016/j.engappai.2019.103330, 2020/01/01/.
[49] J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proceedings of ICNN'95 - International Conference on Neural Networks, 27 Nov.1 Dec. 1995 1995, vol. 4, pp. 1942–1948 vol.4, doi: 10.1109/ICNN.1995.488968.
[50] J. Xin, L. ShangKe, X. YanLi, W. Ye, Prediction model for cost data of a power transmission and transformation project based on Pearson correlation coefficient–IPSO–ELM, Clean Energy 5 (4) (2021) 756–764, https://doi.org/10.1093/ce/zkad054.
[51] J.H. Holland, Genetic Algorithms, Scientific American 267 (1) (1992) 66–73 [Online]. Available: http://www.jstor.org/stable/24939139.
[52] D.E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning," 1989.
[53] S. Mirjalili, A. Lewis, The Whale Optimization Algorithm, Advances in Engineering Software 95 (2016) 51–67, https://doi.org/10.1016/j.advengsoft.2016.01.008, 5//.
[54] M. Zheng, X. Luo, Joint estimation of State of Charge (SOC) and State of Health (SOH) for lithium ion batteries using Support Vector Machine (SVM), Convolutional Neural Network (CNN) and Long Sort Term Memory Network (LSTM) models, International Journal of Electrochemical Science 19 (9) (2024) 100747, https://doi.org/10.1016/j.ijoes.2024.100747, 2024/09/01/.