Proceedings of the 4th International Conference on Functional Materials and Civil Engineering
DOI: 10.54254/2755-2721/2026.36922

Residual Value Prediction of New Energy Vehicles Based on Machine Learning Algorithms

Peiguo Zheng[1,*], Hongyu Zhu[2]

[1]*School of Automotive Engineering, Hubei Vocational College of Transportation, Wuhan, China*
[2] *School of Aeronautics, Hubei Vocational College of Transportation, Wuhan, China*
**Corresponding Author. Email: 642754436@qq.com*

Abstract. Under the impetus of global energy transition and the promotion of the dual carbon goals, there is an urgent need for quantitative decision-making in the prediction of the residual value of new energy vehicles. Traditional assessment methods and conventional machine learning models have problems such as insufficient feature mining, limited prediction accuracy, and high risk of overfitting. This paper proposes the FVIM-XGBoost regression algorithm, which integrates the four-vector intelligent meta-heuristic optimization algorithm and the extreme gradient boosting model. Through the feature value optimization module, it screens and reconstructs features, filters out redundant noise, combines the regularization advantages of XGBoost and the gradient optimization advantages, and improves the model's fitting accuracy and generalization ability. The experiment is based on multi-source collected and standardized processed data of new energy vehicles. FVIM is used to optimize the key hyperparameters of XGBoost, and an efficient and robust prediction scheme is constructed. Compared with multiple mainstream machine learning models, this algorithm performs better in regression prediction tasks and can provide precise and reliable technical support for the assessment of the residual value of new energy vehicles, pricing decisions, etc.

Keywords: New energy vehicles, Residual value prediction, FVIM, XGBoost

1. Introduction

Under the background of the global energy structure transformation and the continuous advancement of dual-carbon goals, the new energy automobile industry has become the core pillar of green development in the transportation field [1]. With the rapid increase of market ownership, the demand for data-driven decision-making in vehicle residual value prediction, pricing decision-making, consumption guidance and other links is increasingly urgent [2]. The value evaluation of traditional new energy vehicles mostly relies on empirical judgment and simple statistical models, which makes it difficult to effectively capture the complex correlation of multi-dimensional characteristics such as battery performance, model configuration, market sales, etc. Under the trend of data scale expansion and scenario complexity, the prediction accuracy and generalization ability are gradually difficult to meet the practical application needs [3]. At the same time, the related data of new energy vehicles presents the characteristics of multi-source heterogeneity and strong

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.36922

nonlinear correlation. Traditional methods can't fully tap the potential laws between features, resulting in deviations between the evaluation results and the market reality. Building efficient and accurate prediction models has become an important topic of common concern between industry and academic fields [4].

With powerful feature learning and fitting capabilities, machine learning technology provides a reliable technical path for the quantitative prediction task of new energy vehicles. All kinds of regression algorithms can learn high-dimensional features automatically and mine the mapping relationship between features and target variables, which can effectively solve the nonlinear problems that traditional models can't deal with. From basic linear regression to tree model and ensemble model, machine learning methods have gradually improved the prediction accuracy of new energy vehicle related indicators, providing quantitative support for pricing strategy, residual value management, industry analysis and other scenarios. However, the existing algorithms still have shortcomings in feature contribution distinction, redundant information filtering and model fitting efficiency. Some models are sensitive to outliers, and problems such as overfitting or slow convergence are prone to occur under complex real data, which restricts the further improvement of prediction effect [5].

In order to solve the limitations of traditional machine learning models in the prediction task of new energy vehicles, this paper proposes the FVIM-XGBoost regression algorithm. This method combines the feature importance optimization mechanism and the extreme gradient boosting model, and constructs a feature value optimization module to screen the input variables and reconstruct their weights, so as to improve the contribution of key features to the prediction results and reduce the interference of redundant information and noisy data. On this basis, the regularization strategy and gradient optimization advantages of XGBoost model are combined to improve the fitting accuracy and generalization performance of the model on nonlinear data, effectively alleviate the overfitting problem and speed up the convergence of the model. The FVIM-XGBoost algorithm takes into account both feature optimization and model strengthening, which can more accurately adapt to the distribution characteristics of new energy vehicle data, provide a more efficient and robust solution for the quantitative forecasting task of new energy vehicles, and also provide a more reliable scientific basis for related decisions.

2. Data sources

The residual value prediction data set of new energy vehicles constructed in this paper is obtained through the joint collection of public vehicle parameter database, industry statistical report and real transaction data. The original data covers multi-dimensional characteristic information such as vehicle brand attributes, model categories, battery capacity, cruising range, fast charging power, new car price, motor power, driving form, body grade, charging configuration, battery warranty life and annual sales volume. In the data processing stage, the cleaning and regularization are completed by eliminating outliers, numerical range constraints and normal distribution fitting, and at the same time, the discrete features are reasonably distributed and set, finally forming 1627 standardized samples. First, a correlation analysis was performed on each variable, and a correlation heat map was plotted, as shown in Figure 1.

DOI: 10.54254/2755-2721/2026.36922

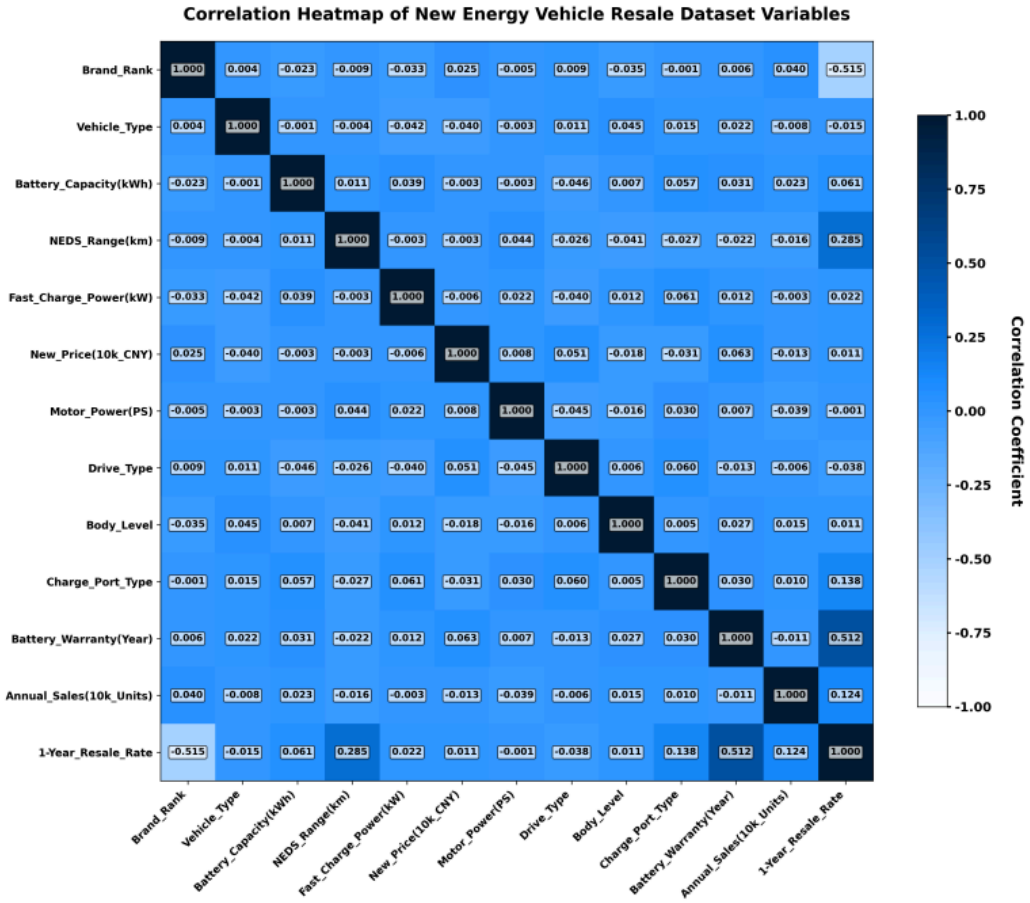

Figure 1. The correlation heat map

Fiddle graph distribution plot of output variables, as shown in Figure 2.

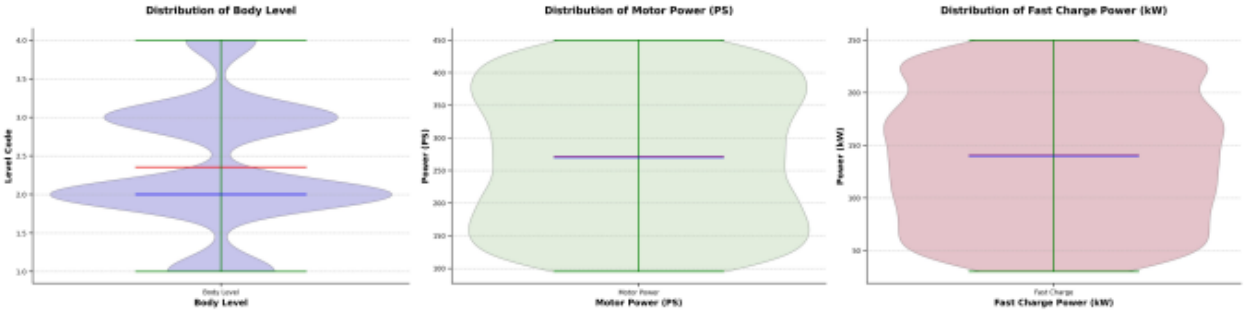

Figure 2. The violin plots

3. Method

3.1. FVIM

FVIM algorithm, which is called four-vector intelligent meta-heuristic algorithm, is a new global optimization algorithm proposed in 2024. The core of the algorithm is to guide the particle motion through four optimal individuals in the population, so as to balance the global exploration and local development capabilities [6]. In the initialization stage, the algorithm will randomly generate

DOI: 10.54254/2755-2721/2026.36922

particle populations, set the upper and lower boundaries of the search space, the number of iterations and the number of particles, and initialize the fitness-related parameters at the same time to adapt to two kinds of optimization problems: minimization and maximization. The iteration stage is the core of the algorithm. In each iteration, the fitness values of all particles are calculated firstly, and the four particles with the best fitness in the current population are selected. The average positions of these four particles are calculated as new reference vectors. When the particle position is updated, it will combine its own historical optimum, global optimum and the average position information of the four optimal particles, and adjust the position through specific motion equations, which not only expands the search range to avoid falling into the local optimum, but also accurately converges to the global optimum [7]. The network structure of FVIM is shown in Figure 3.

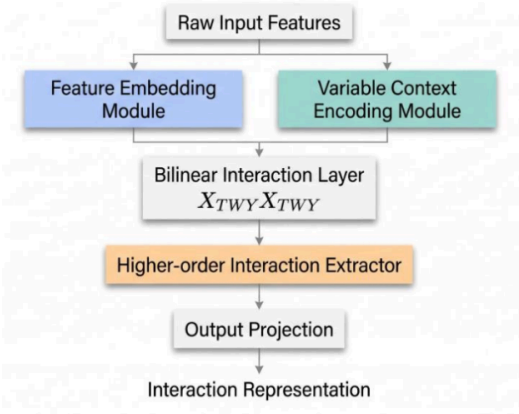

Figure 3. The network structure of FVIM

3.2. XGBoost

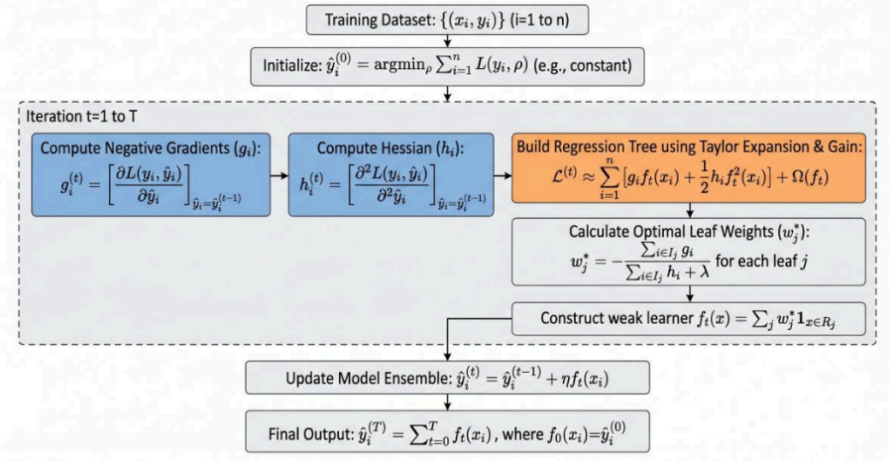

Figure 4. The algorithm flow chart of XGBoost

XGBoost is an extreme gradient boosting algorithm, which belongs to the optimized and upgraded version of the gradient boosting tree framework. Its core is to serially build multiple decision trees and integrate them to achieve high-precision prediction. The algorithm aims at minimizing the loss function. The objective function is composed of a loss term and a regular term. The loss term measures the error between the predicted value and the true value, while the regular term prevents the model from overfitting by controlling the number of leaf nodes and the sum of

DOI: 10.54254/2755-2721/2026.36922

squares of leaf weights of the tree [8]. Starting from a single base tree, each iteration trains a new tree. The new tree specifically fits the prediction residuals of all trees in the previous order, and determines the optimization direction by the negative gradient direction of the loss function. Different from the traditional GBDT, XGBoost performs second-order Taylor expansion on the objective function, and at the same time uses the first-order derivative and second-order derivative information to make the model converge faster and optimize more accurately. The tree is constructed using CART tree, and the weighted histogram algorithm is used to quickly find the best split point of features, which improves the training efficiency. The algorithm flow chart of XGBoost is shown in Figure 4.

3.3. FVIM-XGBoost

FVIM-XGBoost regression algorithm is a hybrid algorithm that combines FVIM optimization algorithm with XGBoost regression model. The core is to optimize the key hyperparameters of XGBoost with FVIM to improve the accuracy and stability of regression prediction [9]. The performance of XGBoost regression model is highly dependent on hyperparameters such as learning rate, maximum depth of tree, minimum weight of leaf nodes, regularization parameters, etc. Manual parameter adjustment is inefficient and it is difficult to find the optimal combination. The algorithm takes the prediction error (such as mean square error) of XGBoost regression model as the fitness function of FVIM, and transforms the hyperparameter optimization into a global optimization problem. In the training process, the FVIM particle population is initialized first, and each particle corresponds to a set of XGBoost hyperparameter combinations. Then iteratively execute, train the XGBoost regression model with each set of hyperparameters, calculate the prediction error as the particle fitness value, screen out four optimal particles and calculate the average position, and guide the particles to update the hyperparameter combination. After multiple rounds of iteration, FVIM continuously explored the hyperparameter space and found the optimal hyperparameter combination that minimized the regression error of XGBoost. Finally, the XGBoost model trained by this combination completed the regression prediction, which effectively solved the XGBoost hyperparameter tuning problem and improved the generalization ability and prediction effect of the regression model [10]. The network structure of FVIM-XGBoost is shown in Figure 5.

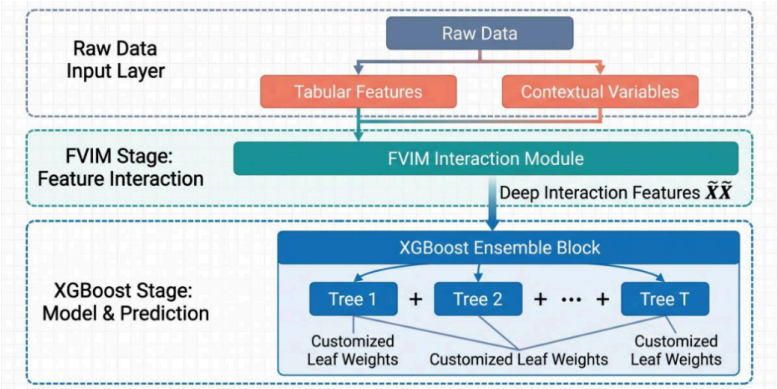

Figure 5. The network structure of FVIM-XGBoost

4. Result

This project uses FVIM algorithm to optimize three core parameters of XGBoost model. The specific values are set to the maximum number of iterations between 10 and 200, the maximum

DOI: 10.54254/2755-2721/2026.36922

depth of the tree between 10 and 20, the learning rate between 0.01 and 1, the number of search agents of the algorithm is set to 5, and the maximum number of optimization iterations is set to 15. At the same time, the basic regression objective function of XGBoost model is set to linear regression, the minimum child node weight is set to 1, the sample sampling ratio is set to 0.9, and the column sampling ratio is set to 1.

Comparative experiments were performed using a variety of machine learning algorithms, and the results are shown in Table 1.

Table 1. The results of the comparative experiment

Model	MSE	RMSE	MAE	MAPE	R^2
ExtraTrees	2.873	1.695	1.242	1.403	0.657
Decision tree	3.654	1.911	1.314	1.493	0.595
GBDT	2.999	1.732	1.25	1.419	0.652
CatBoost	2.913	1.707	1.23	1.391	0.671
Random forest	2.976	1.725	1.256	1.421	0.668
BP neural network	3.841	1.96	1.573	1.776	0.602
XGBoost	2.628	1.621	1.201	1.36	0.676
AdaBoost	2.992	1.73	1.172	1.325	0.691
LightGBM	3.176	1.782	1.292	1.467	0.675
Linear regression	3.333	1.826	1.479	1.664	0.618
FVIM-XGBoost	2.45	1.565	1.13	1.277	0.72

Plot the metrics comparison histogram of the FVIM-XGBoost test set, as shown in Figure 6.

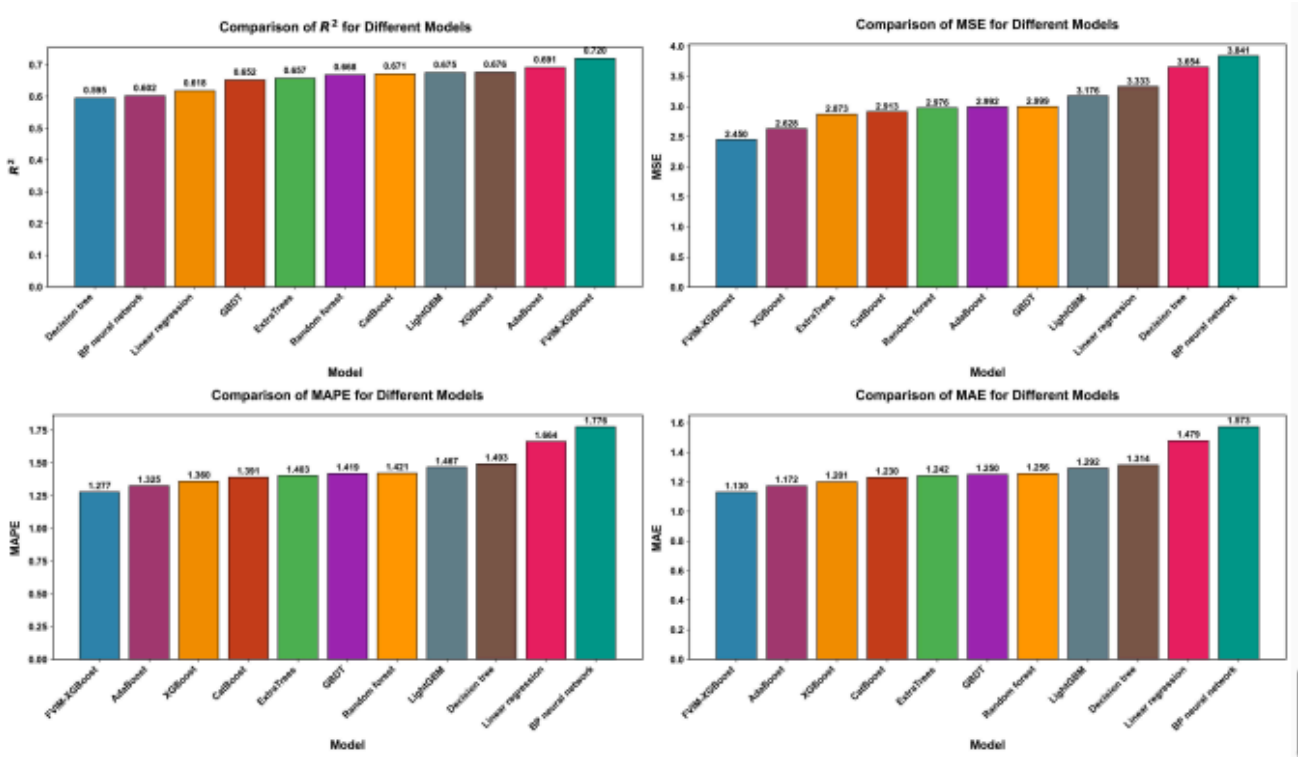

Figure 6. The comparison results of each index of each FVIM-XGBoost test set

DOI: 10.54254/2755-2721/2026.36922

From the comparison results of various regression evaluation indicators, it can be clearly seen that the FVIM-XGBoost algorithm proposed in this paper has the best comprehensive performance among all the models participating in the comparison, and its MSE, RMSE, MAE and MAPE are all the lowest values, and R^2 reaches the highest value. Compared with the traditional single models such as linear regression, decision tree and BP neural network, the error index is significantly reduced and the fitting goodness is greatly improved. Compared with the classical ensemble learning algorithms such as ExtraTrees, GBDT, CatBoost, Random Forest, XGBoost, AdaBoost and LightGBM, FVIM-XGBoost also achieves comprehensive surpassing in various error control and model fitting effects, which fully verifies the effectiveness of the improved algorithm in improving the accuracy and stability of regression prediction, and can better capture data feature patterns and reduce prediction bias.

5. Conclusion

Aiming at the characteristics of multi-source heterogeneity and strong nonlinear correlation of new energy vehicle data, as well as the limitation of traditional models in feature processing and hyperparameter tuning, this paper proposes the FVIMXGBoost regression prediction method. This method relies on the global optimization ability of FVIM to realize the automatic optimization of XGBoost core hyperparameters, and at the same time complete the feature screening and weight reconstruction, effectively reduce the interference of redundant information, alleviate overfitting and speed up the model convergence. The experimental results show that FVIMXGBoost is lower than the comparative model in terms of mean square error, root mean square error, mean absolute error, mean absolute percentage error, etc. The determination coefficient reaches the highest level, and the prediction accuracy and stability are better than linear regression, decision tree, BP neural network and many classical ensemble learning algorithms. The model can accurately adapt to the data distribution characteristics of new energy vehicles, efficiently mine the complex mapping relationship between multi-dimensional features and residual values, and significantly improve the quantitative prediction effect. The research results can provide scientific basis for residual value management, market pricing, consumption guidance and industrial analysis of new energy vehicles, promote the intelligent upgrading of data-driven decision-making in the aftermarket of new energy vehicles, and also provide technical ideas for similar nonlinear high-dimensional data regression forecasting tasks.

References

[1] Alshboul, O., et al. (2022). Deep and machine learning approaches for forecasting the residual value of heavy construction equipment: A management decision support model. *Engineering, Construction and Architectural Management, 29*(10), 4153–4176.

[2] Miao, X., & Li, B. (2024). Research on the development of new energy vehicles based on multiple machine learning models and the ARIMA algorithm. In *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*. IEEE.

[3] Liu, B., et al. (2023). Carbon emission reduction prediction of new energy vehicles in China based on GRA-BiLSTM model. *Atmospheric Pollution Research, 14*(9), 101865.

[4] Li, Y., et al. (2025). Machine learning for new energy vehicle sales prediction and carbon emission impact. *Journal of Computing and Electronic Information Management, 16*(3), 48–53.

[5] Refaai, M. R. A., et al. (2022). Battery energy forecasting in electric vehicle using deep residual neural network. *International Journal of Photoenergy, 2022*(1), 5959443.

[6] Pan, Y., Fang, W., & Zhang, W. (2023). Development of an energy consumption prediction model for battery electric vehicles in real-world driving: A combined approach of short-trip segment division and deep learning.

DOI: 10.54254/2755-2721/2026.36922

Journal of Cleaner Production, 400, 136742.

[7] Liu, Z., et al. (2022). Durability estimation and short-term voltage degradation forecasting of vehicle PEMFC system: Development and evaluation of machine learning models. *Applied Energy, 326*, 119975.

[8] Naresh, V. S., Rao, G. V. R., & Prabhakar, D. V. N. (2024). Predictive machine learning in optimizing the performance of electric vehicle batteries: Techniques, challenges, and solutions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14*(5), e1539.

[9] Shah, A., et al. (2022). State of charge, remaining useful life and knee point estimation based on artificial intelligence and machine learning in lithium-ion EV batteries: A comprehensive review. *Renewable Energy Focus, 42*, 146–164.

[10] Li, D., et al. (2022). Battery thermal runaway fault prognosis in electric vehicles based on abnormal heat generation and deep learning algorithms. *IEEE Transactions on Power Electronics, 37*(7), 8513–8525.