This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.23919/CCC64809.2025.11178919,
Proceedings of the 44th Chinese Control Conference
July 28-30, 2025, Chongqing, China

Energy Consumption Prediction and Feature Contribution Analysis of Unmanned Mining Trucks Based on XGBoost and SHAP

GuoWei Cao[1], DaXin Chen[2], Hua Zhao[3], MingWei Zeng[4], Tao Chen[5]

1. State Key Laboratory of Engines, Tianjin University, Tianjin 300350
E-mail: cgw_123456@tju.edu.cn

2. State Key Laboratory of Engines, Tianjin University, Tianjin 300350
E-mail: chendaxin@tju.edu.cn

3. Brunel University of London, London 999020
E-mail: mesthhz@163.com

4. State Key Laboratory of Engines, Tianjin University, Tianjin 300350
E-mail: zmw_1234@tju.edu.cn

5. State Key Laboratory of Engines, Tianjin University, Tianjin 300350
E-mail: tao.chen@tju.edu.cn

Abstract: With the rapid development of mining automation, unmanned mining trucks are increasingly used in mine transportation. Accurate energy consumption prediction is crucial for optimizing energy management and control strategies. This study uses actual operational data of unmanned mining trucks and employs an XGBoost-based prediction model for energy forecasting, with SHAP used to interpret the model and quantify the contribution of each feature. Operating conditions are classified into unloaded downhill and fully loaded upslope conditions, with data analyzed by speed intervals. XGBoost models are constructed for each condition. SHAP analysis reveals that battery current and generator current significantly impact the model under unloaded downhill conditions, while battery current dominates in most speed ranges for fully loaded upslope conditions. At low speeds, generator speed has a strong influence. SHAP dependence plots show a linear relationship between battery current and energy consumption. Feature selection is performed by removing features with minimal contributions, simplifying the model and improving efficiency. The optimized model maintains predictive accuracy while reducing complexity. The results show that the XGBoost and SHAP-based model effectively predicts energy consumption, providing a basis for energy-saving optimization in smart mining operations.

Key Words: Unmanned Mining Trucks, Energy Consumption Prediction, XGBoost, SHAP, Feature Selection

1 Introduction

With the development of mining automation and intelligence, the application of unmanned mining trucks in mine transportation is becoming increasingly widespread. Compared to traditional manually operated mining trucks, unmanned mining trucks can improve transportation efficiency, reduce labor costs, and minimize safety accidents caused by human errors. Energy consumption is a critical factor in enhancing operational efficiency in mining areas. Mining transport operations typically occur in complex terrains and harsh environments, with vehicle energy consumption influenced by various factors such as speed, slope, load, and power system parameters. Therefore, constructing an accurate energy consumption prediction model is essential for optimizing the energy management of unmanned mining trucks, enhancing endurance, and reducing operating costs.

Machine learning methods are widely used in vehicle energy consumption modeling and prediction due to their advantages, including efficient handling of complex relationships, automated feature learning, adaptability to large-scale data, and strong generalization capabilities. In recent years, these methods have been increasingly applied in the field of vehicle energy consumption prediction. Jerome H. Friedman [1] made pioneering contributions to the gradient boosting algorithm, detailing how the performance of decision trees can be enhanced through additive models and optimal gradient search, laying the theoretical foundation for algorithms such as XGBoost. Tianqi Chen et al. [2] proposed the XGBoost algorithm, describing its optimization strategies based on Gradient Boosted Decision Trees (GBDT), including regularization, block-wise computation, and cache optimization, making XGBoost an efficient and powerful machine learning tool. Ke, Guolin et al. [3] proposed LightGBM, a GBDT variant similar to XGBoost but with higher computational efficiency, improving training speed on large-scale data through histogram binning and leaf-wise growth strategies. These studies on algorithms have laid a solid theoretical foundation for applying machine learning methods to energy consumption prediction.

However, relying solely on the prediction results of machine learning models is insufficient. The key issue lies in how to interpret the model's predictions and quantify the contribution of each feature to energy consumption. Lundberg et al. [4] introduced the SHAP method, which provides a unified and fair way to calculate feature importance based on Shapley values. SHAP calculates the marginal contribution of each feature, offering a consistent

---

[*] This work is supported by National Key Research and Development Program of China 2024YFE0107700 and Guangxi Science and Technology Major Program 2023AA11008.

and fair approach to measuring feature importance. SHAP analysis can identify the most significant variables impacting energy consumption and further optimize feature selection, thereby improving model efficiency and prediction performance.

Typical research topics related to hybrid and electric vehicles include but are not limited to battery size [5], charging schedules [6], driving range, route prediction [7, 8], speed planning and control [9], road condition analysis [10], etc. Accurate and reliable energy consumption models are fundamental for addressing these challenges. In the field of vehicle energy consumption prediction, Qingbo Zhu et al. [11, 12] compared the performance of machine learning methods such as XGBoost, random forest, and neural networks in electric vehicle energy consumption prediction, exploring the impact of feature engineering on prediction accuracy. Their findings concluded that XGBoost, with its gradient boosting strategy, improved model fitting compared to other methods, and even performed well with small sample data. Sugam Pokharel et al. [13] researched how to use SHAP to quantify the influence of different driving behaviors and surrounding environmental factors on vehicle energy consumption, and optimized feature sets to enhance prediction performance. However, their study did not discuss the relationship between the vehicle's kinematic parameters, such as speed and generator-related parameters, and energy consumption.

Although many studies have applied different machine learning methods to predict vehicle energy consumption, few have investigated the complex relationships between various features and energy consumption. This study utilizes actual operational data from unmanned mining trucks on a fixed route in a mining area. First, the data is categorized based on the vehicle’s operation cycles, and XGBoost energy consumption prediction models are constructed for different speed intervals. Then, the SHAP method is applied to analyze the contribution of different features to energy consumption at various speed intervals, and the model is optimized based on SHAP values. The results not only improve the accuracy of energy consumption prediction for unmanned mining trucks but also provide data support for energy management and control strategies for mining trucks, offering valuable references for energy-saving optimization in smart mining operations.

2 Data and Methods

2.1 Dataset Description

The data used in this study is derived from the actual operational data of unmanned mining trucks on a fixed route in a mining area. This data is recorded in real-time by the vehicle’s Electronic Control Unit (ECU) and Battery Management System (BMS), covering key power parameters, electrical parameters, and energy consumption data during vehicle operation. The data is sampled at a frequency of 1 Hz, meaning that sensor data is recorded once per second, ensuring high temporal resolution and enabling the dynamic changes in the vehicle's operational status to be fully captured.

The data is stored in Excel file format and includes time-series information collected by multiple sensors, covering core parameters such as vehicle speed, generator and motor current, voltage, power, speed, torque, and the battery’s State of Charge (SOC). The relevant parameters and their units are shown in Table 1, which includes various variables closely related to energy consumption prediction during vehicle operation. This data provides the foundation for subsequent feature selection, model training, and interpretation.

Table 1: Parameters Used for Energy Consumption Prediction

Feature Name	Unit	Category
Vehicle Speed	Kph	Kinematic Parameters
Generator Speed	rpm	Generator Parameters
Generator Torque	Nm	
Generator Voltage	V	
Generator Current	A	
Generator Power	kW	
Battery Current	A	Battery Parameters
Battery Voltage	V	
Battery SOC	%	

2.2 Data Preprocessing

After obtaining the operational data of the unmanned mining truck, it is essential to properly classify and select features to construct an accurate energy consumption prediction model. Given that the mining truck operates cyclically on a fixed route, its speed profile exhibits strong periodicity over time. To determine the operational cycle, this study employs the Autocorrelation Function (ACF) to analyze the speed time series data. ACF measures the similarity of a time series at different time lags by computing the autocorrelation coefficient, which is mathematically expressed as follows:

$$\mathrm{ACF}(k)=\frac{\sum_{t=1}^{N-k}\ (x_t-\bar{x})(x_{t+k}-\bar{x})}{\sum_{t=1}^{N}\ (x_t-\bar{x})^2} \tag{2}$$

where x_t represents the speed at time step t, $\bar{x}$ is the mean speed, N is the total length of the time series, and k is the lag step. By computing ACF, the significant periodicity in the speed profile can be identified, allowing for the segmentation of complete operational cycles.

Once the operational cycle is segmented, further analysis of the speed variation pattern within each cycle is conducted. By integrating the mining truck's operational workflow and examining the relationships and regularities across different cycles, four distinct operating conditions are identified within each cycle: unloaded downhill, loading stop, loaded upslope, and unloading stop. The figure below illustrates the speed profile of a representative operational cycle from the route data.

After dividing the operation cycle, the speed variation pattern within each cycle was further analyzed. Based on the mining truck's operational process and by examining the relationships and regularities between different operation cycles, four main operating conditions within a single cycle were identified: unloaded downhill, loading stop, loaded upslope, and unloading stop. Fig. 1 below shows the speed curve of a specific operation cycle from the road profile data.

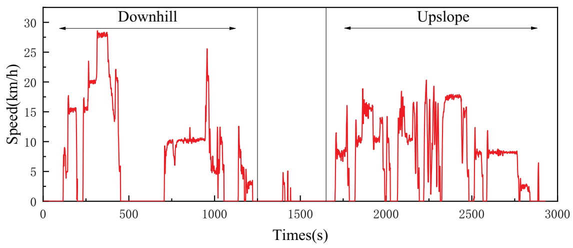

Fig. 1: Speed Curve and Operating Condition Classification for a Single Cycle

2.3 XGBoost Model Construction

XGBoost (eXtreme Gradient Boosting) is a machine learning algorithm based on Gradient Boosting, known for its computational efficiency, strong generalization ability, and resistance to overfitting. XGBoost improves traditional Gradient Boosted Decision Trees (GBDT) by incorporating optimizations like second-order gradient optimization, parallel computation, L1 and L2 regularization, and automatic handling of missing values. It is widely used for tasks such as regression, classification, and ranking. In this study, the XGBoost regression model is applied to predict the energy consumption of unmanned mining trucks. The model predicts real-time energy consumption using features like speed, generator power, torque, battery current, and voltage, aiding in optimizing energy consumption management. The model adjusts the decision tree structure to better fit the data and reduce errors. Additionally, SHAP (SHapley Additive exPlanations) is used to interpret the model and identify the impact of different features on energy consumption, providing support for optimization efforts in unmanned mining trucks.

2.4 SHAP Model Interpretation

SHAP (Shapley Additive Explanation) is a game-theoretic interpretability framework that quantifies feature contributions to machine learning predictions via Shapley values. Rooted in cooperative game theory, SHAP computes the marginal impact of each feature across all possible subsets, ensuring fairness by accounting for interaction effects. Unlike traditional methods that evaluate features in isolation, SHAP aggregates subset-specific contributions to assign unbiased importance scores. Mathematically, the SHAP value for a given feature i is calculated as follows:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (1)$$

In this equation, ϕ_i represents the SHAP value of feature i; N is the set of all features; S is a subset of features excluding i; $f(S)$ denotes the model's prediction when trained using only the feature subset S; and $f(S \cup \{i\}) - f(S)$ represents the change in the model's prediction when feature iii is added. The summation in the formula iterates over all possible feature subsets. By computing the marginal contribution of feature iii across different subsets, SHAP evaluates the average impact of that feature on the final prediction.

By summing over all possible feature subsets, this formulation ensures that SHAP values accurately reflect each feature's average contribution under different conditions, eliminating biases related to feature order or grouping. This makes SHAP a powerful and reliable model interpretability tool, particularly suited for high-dimensional data and complex models. In this study, SHAP values are employed to interpret the predictions of the XGBoost model, helping to identify key factors influencing energy consumption in unmanned mining trucks. The insights obtained from SHAP analysis provide valuable guidance for energy optimization strategies.

3 Results and Discussion

3.1 Impact of Different Speed Intervals on Feature Contribution to Model Prediction

In the unloaded downhill condition, the vehicle primarily relies on inertia for coasting and utilizes the energy recovery system for regenerative braking, resulting in distinct energy consumption characteristics compared to other conditions. To analyze the energy consumption patterns more accurately in this condition, this study further divides it into three speed ranges: 0-10 km/h, 10-20 km/h, and above 20 km/h. Separate XGBoost energy consumption prediction models are constructed for each speed range to improve model specificity and prediction accuracy.

Since the vehicle's power demand, electrical system load, and energy recovery efficiency vary across different speed ranges, the key features influencing energy consumption may also differ. Therefore, this study applies the SHAP method to interpret the models for each speed range, calculating SHAP values to quantify the real contribution of each feature to the model's predictions and explain the prediction outcomes. The results are presented in Figs. 2-4.

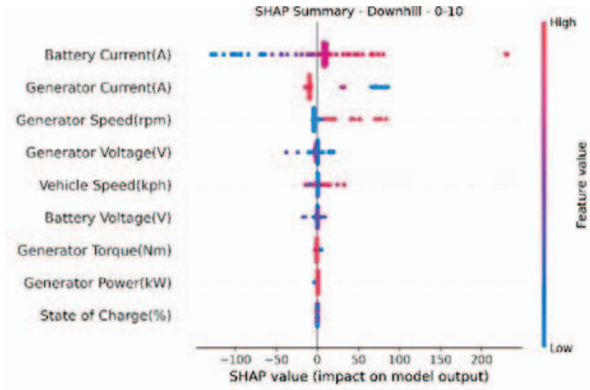

Fig. 2: SHAP Summary Plot for Energy Consumption Prediction (Downhill 0-10 km/h)

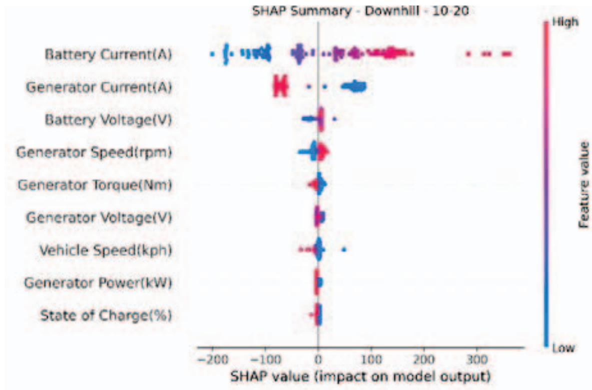

Fig. 3: SHAP Summary Plot for Energy Consumption Prediction (Downhill 10-20 km/h)

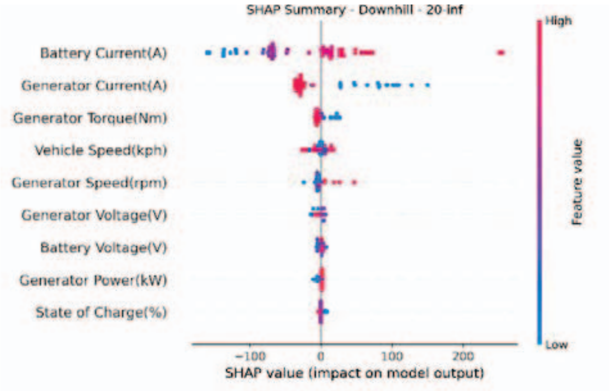

Fig. 4: SHAP Summary Plot for Energy Consumption Prediction (Downhill >20 km/h)

Under unloaded downhill conditions, SHAP value analysis of the XGBoost energy consumption prediction model reveals that battery current and generator current are the dominant features across all speed ranges (Figs. 2-4). In the 0-10 km/h range, generator speed exhibits comparable SHAP values (0-100) to generator current, highlighting its significance in this regime, while other features show negligible influence. This aligns with the energy recovery mechanism during downhill motion, where battery current dominates due to regenerative power generation.

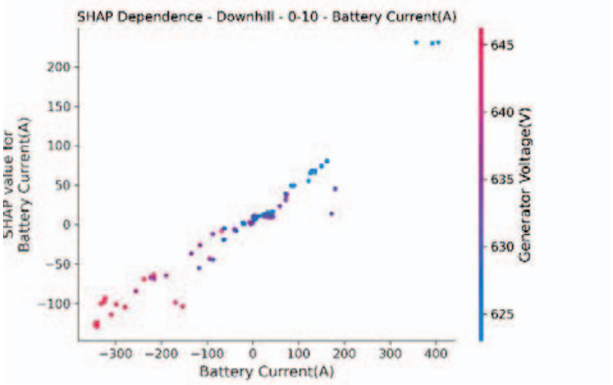

Fig. 5: SHAP Dependence Plot for Battery Current (Downhill 0-10 km/h)

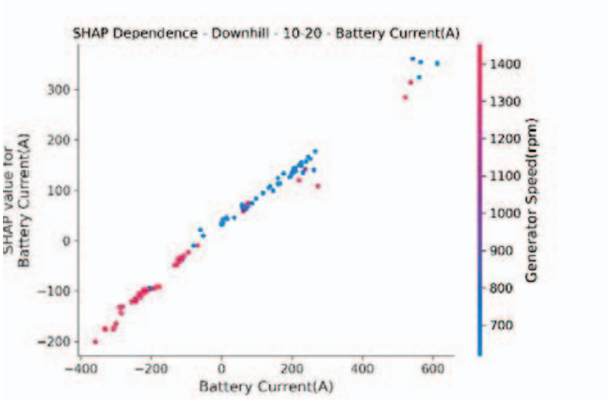

Fig. 6: SHAP Dependence Plot for Battery Current (Downhill 10-20 km/h)

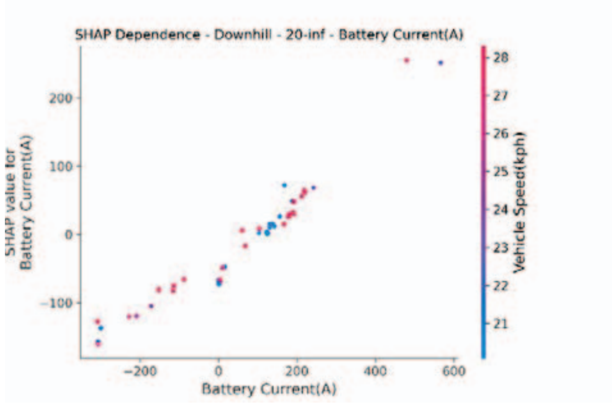

Fig. 7: SHAP Dependence Plot for Battery Current (Downhill >20 km/h)

Figs. 5-7 demonstrate a quasi-linear relationship between battery current and SHAP values in the XGBoost model across three speed segments. Specifically, negative battery currents correspond to negative SHAP values, whereas increasing currents elevate SHAP outputs. In the 0-10 km/h segment, higher generator voltage (color shift from blue to red) reduces SHAP values; similar trends are observed for generator speed in the 10-20 km/h segment, highlighting their negative correlations with predictive contributions under varying current levels.

Under upslope conditions, energy consumption diverges from downhill scenarios due to nonlinear load-resistance dynamics across speed intervals (0-5, 5-10, 10-15, >15 km/h). XGBoost models for each interval reveal that motor power, battery state, and generator state dominate energy consumption. SHAP analysis quantifies feature contributions, emphasizing motor-driven demand as the primary factor.

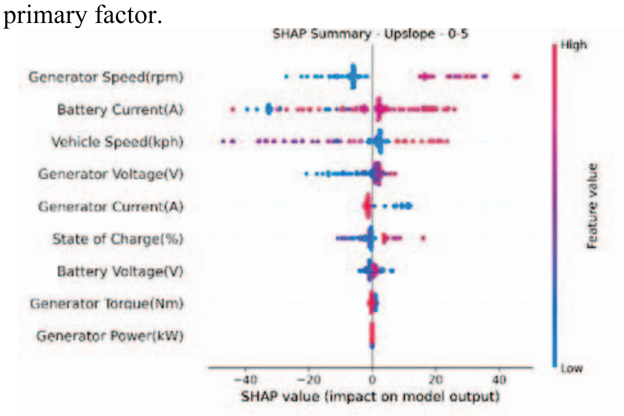

Fig. 8: SHAP Summary Plot for Energy Consumption Prediction (Upslope 0-5 km/h)

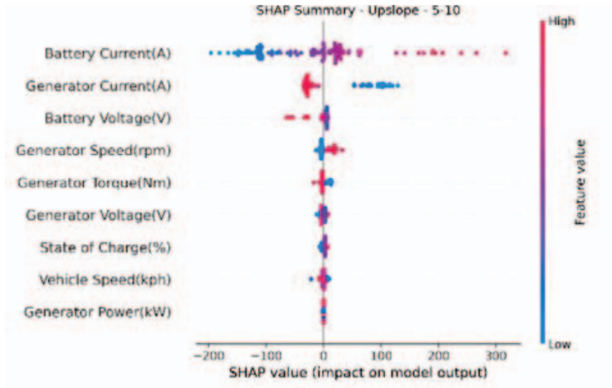

Fig. 9: SHAP Summary Plot for Energy Consumption Prediction (Upslope 5-10 km/h)

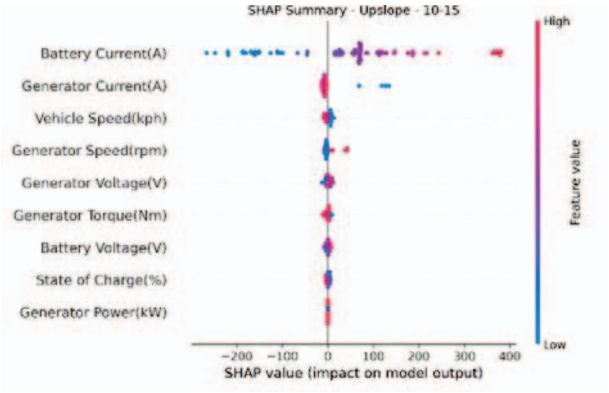

Fig. 10: SHAP Summary Plot for Energy Consumption Prediction (Upslope 10-15 km/h)

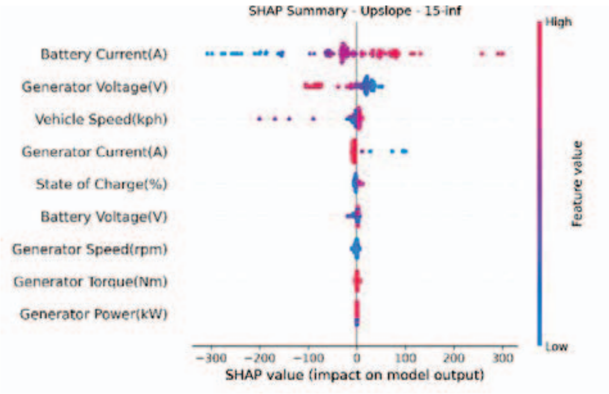

Fig. 11: SHAP Summary Plot for Energy Consumption Prediction (Upslope >15 km/h)

In the 0-5 km/h speed segment model, the feature with the greatest impact on the model is the generator speed, while in other speed segments, the feature with the greatest impact is the battery current. Through the feature contribution analysis in Figs. 8-11 , as well as the analysis of the unloaded downhill condition, it is evident that the battery current has a significant influence on the energy consumption prediction model in all conditions, with its SHAP value reaching a maximum value of over 300. This indicates that paying attention to the relevant state parameters of the battery is crucial for the energy consumption prediction and energy management of unmanned mining trucks.

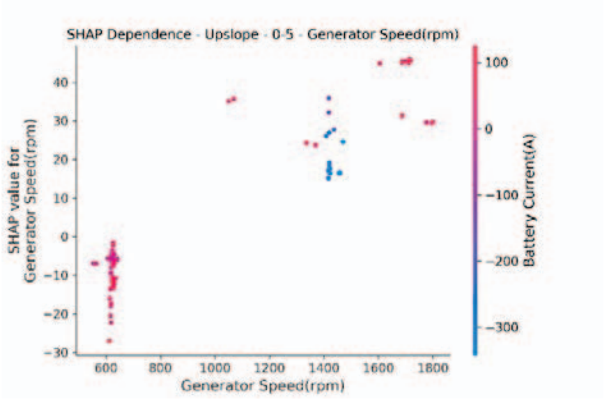

Fig. 12: SHAP Dependence Plot for Generator Speed (Upslope 0-5 km/h)

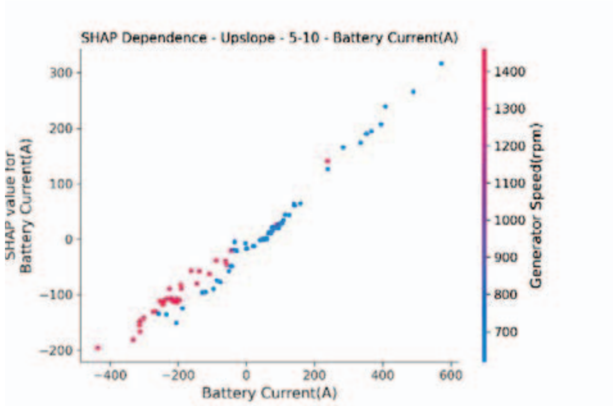

Fig. 13: SHAP Dependence Plot for Battery Current (Upslope 5-10 km/h)

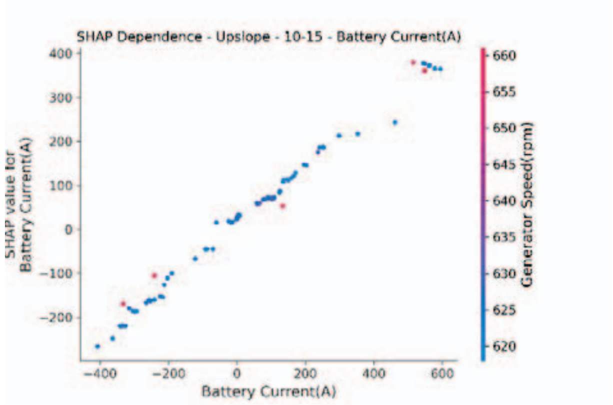

Fig. 14: SHAP Dependence Plot for Battery Current (Upslope 10-15 km/h)

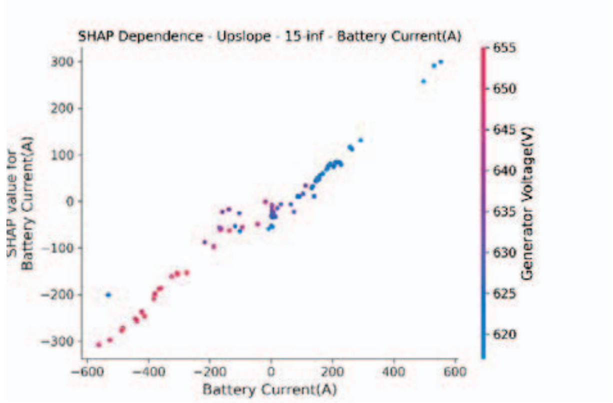

Fig. 15: SHAP Dependence Plot for Battery Current (Upslope >15 km/h)

Figs. 12-15 show the relationship between the generator speed and battery current size and the SHAP values in the XGBoost models of the four speed segments under fully loaded downhill conditions. In the 0-5 km/h speed segment model, there is no strong correlation between the generator speed and the SHAP value. However, in the other speed segments, there is a clear relationship between the battery current size and the SHAP value. It is evident that the battery current size and the SHAP value exhibit a strong linear relationship, which is similar to the conclusion drawn under the unloaded downhill condition.

3.2 Optimized Model

SHAP-based feature contribution analysis enabled optimized feature selection by excluding features with near-zero SHAP values (indicating negligible predictive impact or noise). Retraining models using only the top three influential features per speed segment achieved comparable accuracy to the original models (Tables 2-3), validating the redundancy of low-contribution features.

Table 2: Comparison of RMSE Values Before and After Model Optimization for Different Speed Segments under Downhill Condition

Speed Segment	0-10km/h	10-20km/h	>20km/h
RMSE of Optimized Model (kW)	10.6622	9.2874	7.6924
RMSE of Optimized Model (kW)	10.5832	9.3468	7.3847

Table 3: Comparison of RMSE Values Before and After Model Optimization for Different Speed Segments under Upslope Condition

Speed Segment	0-5km/h	5-10km/h	10-15km/h	>15km/h
RMSE of Optimized Model (kW)	11.2874	8.7623	10.2586	11.4461
RMSE of Optimized Model (kW)	10.6652	8.8567	11.3259	11.7531

From the data in the table, it can be seen that for the unloaded downhill condition, the RMSE of the prediction models in the 0-10 km/h and greater than 20 km/h speed segments decreased. This indicates that while the model was simplified, its prediction accuracy improved. For the 10-20 km/h speed segment, the RMSE increased by 0.0594, which suggests that the simplification of the model was achieved while maintaining its accuracy. For the fully loaded upslope condition, the RMSE of the prediction model in the 0-5 km/h speed segment decreased by 0.6222, achieving both model simplification and improved accuracy. For the other speed segments, the RMSE increased by 0.0994, 1.0673, and 0.3070, respectively. These increases are within an acceptable range.

4 Conclusion

This study uses the XGBoost model to predict the energy consumption of unmanned mining trucks and calculates the SHAP values of each feature to reflect their actual contributions to the prediction model, thereby optimizing and simplifying the model. The study first classifies the operating conditions of the autonomous mining trucks, then constructs energy consumption prediction models for different speed segments under each operating condition, and analyzes the SHAP values of each feature. The findings reveal that under unloaded downhill conditions, the influence of features on the model varies across different speed segments, but the two features with the greatest contribution to the model are battery current and generator current, both of which have SHAP maximum values exceeding 200. For fully loaded upslope conditions, the feature with the greatest influence on the model in the 0-5 km/h speed segment is the generator speed, while in other speed segments, the most influential feature is the battery current. By selecting the top three most influential features, the energy consumption prediction model was retrained, and the RMSE of the simplified model was compared to that of the original model. The results show that after simplification, all models maintained their original accuracy, and for the unload downhill condition in the 0-10 km/h speed segment and the fully loaded upslope condition in the 0-5 km/h speed segment, the accuracy of the simplified models improved while maintaining model simplification.

References

[1] Jerome H. Friedman.Greedy Function Approximation: A Gradient Boosting Machine[J].The Annals of Statistics,2001,Vol.29(5): 1189-1232

[2] Tianqi Chen;Carlos Guestrin.XGBoost: A Scalable Tree Boosting System[A].KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C],2016

[3] Ke, Guolin;Meng, Qi;Finley, Thomas;Wang, Taifeng;Chen, Wei;Ma, Weidong;Ye, Qiwei;Liu, Tie-Yan.LightGBM: A Highly Efficient Gradient Boosting Decision Tree[A].NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems[C],2017

[4] Lundberg, Scott M.;Lee, Su-In.A Unified Approach to Interpreting Model Predictions[A].NIPS[C],2017

[5] X.Hu, S. J. Moura, N. Murgovski, B. Egardt, and D. Cao, "Integrated optimization of battery sizing, charging, and power management in plug in hybrid electric vehicles," IEEE Trans. Control Syst. Technol., vol. 24, no. 3, pp. 1036–1043, May 2016.

[6] Y.Xu, S. Çolak, E. C. Kara, S. J. Moura, and M. C. González, "Planning for electric vehicle needs by coupling charging profiles with urban mobility," Nature Energy, vol. 3, no. 6, pp. 484–493, 2018.

[7] R.Baek et al., "Battery-aware operation range estimation for terrestrial and aerial electric vehicles," IEEE Trans. Veh. Technol., vol. 68, no. 6, pp. 5471–5482, Jun. 2019.

[8] R. Basso, B. Kulcsár, and I. Sanchez-Diaz, "Electric vehicle routing problem with machine learning for energy prediction," Transp. Res. B, Methodol., vol. 145, pp. 24–55, Mar. 2021.

[9] Y.Zhang, X. Qu, and L. Tong, "Optimal eco-driving control of autonomous and electric trucks in adaptation to highway topography: Energy minimization and battery life extension," IEEE Trans. Transport. Electrific., vol. 8, no. 2, pp. 2149–2163, Jun. 2022.

[10] Y. Correa, P. M. Muñoz, and C. R. Rodriguez, "A comparative energy and environmental analysis of a diesel, hybrid, hydrogen and electric urban bus," Energy, vol. 187, Nov. 2019, Art. no. 115906.

[11] Qingbo Zhu;Yicun Huang;Chih Feng Lee;Peng Liu;Jin Zhang;Torsten Wik.Predicting Electric Vehicle Energy Consumption From Field Data Using Machine Learning[J].IEEE Transactions on Transportation Electrification,2025,Vol.11(1): 2120-2132

[12] Chunling Liu.Energy Consumption Prediction and Optimization of Electric Vehicles Based on RLS and Improved SOA[J].IEEE Access,2024,Vol.12: 38180-38191

[13] Sugam Pokharel;Pradip Sah and Deepak Ganta.Improved Prediction of Total Energy Consumption and Feature Analysis in Electric Vehicles Using Machine Learning and Shapley Additive Explanations Method[J].World Electric Vehicle Journal,2021,Vol.12(3): 94