https://doi.org/10.62836/iaet.v4i1.319

Global Science Publishing

IAET

Innovations in Applied Engineering and Technology

https://ojs.sgsci.org/journals/iaet

Article

An Ensemble Model Based on Fusion of Multiple Machine Learning Algorithms for Remaining Useful Life Prediction of Lithium Battery in Electric Vehicles

Weidong Huang [1], Tong Zhou [2], Jiahuai Ma [3] and Xiaoyang Chen [4,*]

[1] *Upower Energy Technology (Guang Zhou) Co., Ltd., Guangzhou 510000, China*

[2] *Air China Cargo Co., Ltd., Beijing 101318, China*

[3] *Herbert Wertheim College, University of Florida, Gainesville, FL 32608, USA*

[4] *The Ohio State University, Columbus, OH 43210, USA*

Abstract: As the demand for renewable energy solutions increases, electric vehicles (EVs) have become a critical component of sustainable transportation. Lithium-ion batteries, the core of EVs, determine vehicle performance and efficiency. Accurate Remaining Useful Life (RUL) prediction of these batteries is essential for effective battery management, reducing unexpected failures, and supporting sustainability goals. Traditional RUL prediction methods often fail to capture the complex degradation processes of batteries. To address these limitations, we propose a novel machine learning framework combining Artificial Neural Networks (ANN) for feature extraction and ensemble modeling with Random Forest (RF), K-Nearest Neighbors (KNN), and Gradient Boosting Decision Tree (GBDT). The final ensemble fusion further refines predictions by leveraging the complementary strengths of the models. Experimental results demonstrate that the proposed framework significantly improves prediction accuracy, with the final ensemble model achieving an R^2 of 0.92 and reducing MAE and RMSE to 74.52 and 110.07, respectively. While promising, the framework faces challenges in real-world generalization and computational efficiency, highlighting the need for further research.

Keywords: ensemble model; machine learning; electric vehicle; renewable energy; remaining useful life prediction

1. Introduction

The growing urgency to transition from conventional fossil fuels to sustainable energy sources has become a defining challenge of the 21st century. As global concerns about climate change and environmental degradation intensify, renewable energy solutions are increasingly regarded as vital pathways toward reducing carbon footprints and achieving long-term ecological balance [1–3]. Among the myriad of sustainable alternatives, electric vehicles (EVs) have emerged as a cornerstone of the green energy revolution. By replacing internal combustion engines with electric motors powered by lithium-ion batteries [4,5], EVs promise to not only cut greenhouse gas emissions but also provide an effective solution for a cleaner, more energy-efficient future.

At the heart of every EV lies its battery system, a critical component that determines the vehicle's performance, efficiency, and lifespan [6,7]. Accurate prediction of the battery's Remaining Useful Life (RUL) has thus gained significant attention in both academic and industrial domain [8–10]. RUL prediction refers to the estimation of the

Received: 13 February 2025; Accepted: 27 February 2025.

* Corresponding: Xiaoyang Chen (chen.13688@buckeyemail.osu.edu)

remaining time or cycles before a battery reaches a state where it can no longer perform its intended function effectively [11,12]. This capability is vital for ensuring optimal battery management, preventing unexpected failures, and enhancing the economic viability of EVs by maximizing battery utilization and minimizing maintenance costs. Moreover, precise RUL prediction facilitates the development of more efficient recycling and disposal strategies, contributing further to the sustainability goals associated with EV adoption.

Traditional methods for RUL prediction, including physics-based models and simple statistical approaches [13–16], often fall short in capturing the complex, nonlinear degradation behaviors of lithium-ion batteries. These conventional techniques typically rely on simplified assumptions and static parameters [17–19], which limit their adaptability to real-world operating conditions and diverse battery chemistries. For instance, physics-based models require extensive domain knowledge and computational resources, making them impractical for widespread deployment. On the other hand, statistical models struggle to generalize across different datasets and fail to provide the accuracy required for modern EV applications. As a result, there is a pressing need for more robust and versatile methodologies capable of handling the intricacies of battery degradation.

In recent years, the rapid advancement of Artificial Intelligence (AI) [20–23] has opened new horizons for addressing the limitations of traditional RUL prediction methods. AI techniques, particularly those rooted in machine learning and deep learning [24–26], excel at identifying hidden patterns and extracting meaningful insights from large, high-dimensional datasets. By leveraging sensor data from batteries, these methods can learn the underlying relationships between various features and the degradation process without relying on explicit physical models. Furthermore, AI-driven approaches offer the flexibility to adapt to different battery types and usage scenarios, making them highly suitable for the dynamic and diverse landscape of EVs.

To address the challenges in RUL prediction, we propose an innovative framework shown in Figure 1 that integrates multiple machine learning algorithms through a unified ensemble model. Our methodology begins with the extraction of high-level representations from raw battery sensor data using an Artificial Neural Network (ANN) [27, 28]. The ANN serves as a feature extractor, capturing intricate relationships within the data and providing a robust foundation for subsequent analysis. These high-level representations are then utilized by three complementary machine learning models: Random Forest (RF), K-Nearest Neighbors (KNN), and Gradient Boosting Decision Tree (GBDT). Each model predicts the RUL independently, leveraging its unique strengths to contribute diverse perspectives to the overall estimation process. The final step in our framework involves fusing the outputs of the RF, KNN, and GBDT models using a Random Forest-based ensemble. This model fusion strategy combines the individual predictions into a single, cohesive output by assigning optimal weights to each component model based on its performance. By integrating these predictions, the ensemble model capitalizes on the complementary advantages of the underlying algorithms, enhancing accuracy and robustness while mitigating individual model biases.

2. Literature Review

Remaining Useful Life Prediction of Lithium-Ion Batteries

Accurate prediction of the RUL of lithium-ion batteries is crucial for the reliability and efficiency of electric vehicles (EVs). Traditional approaches, such as physics-based models, simulate the internal chemical and physical processes of battery degradation. While these models offer valuable insights, they often require extensive domain knowledge and are computationally intensive, limiting their practical application [29]. In contrast, data-driven methods, particularly those utilizing machine learning (ML) techniques, have gained prominence due to their ability to model complex, nonlinear relationships without explicit physical assumptions. Support Vector Regression (SVR) and Gaussian Process Regression (GPR) are among the commonly used ML methods for RUL prediction. SVR constructs a hyperplane in a high-dimensional space to perform regression, while GPR provides a probabilistic approach, offering not only predictions but also uncertainty estimates [30]. Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have further enhanced RUL prediction capabilities. CNNs are adept at extracting hierarchical features

from data, making them suitable for capturing spatial dependencies, whereas RNNs, including Long Short-Term Memory (LSTM) networks, are proficient in modeling temporal sequences, effectively capturing the temporal dynamics of battery degradation [31]. Despite the advancements brought by ML and deep learning [32-34], challenges remain in feature extraction and model fusion. Integrating multiple models can leverage the strengths of each, leading to improved prediction performance. However, such integration requires careful consideration to avoid increased complexity and overfitting.

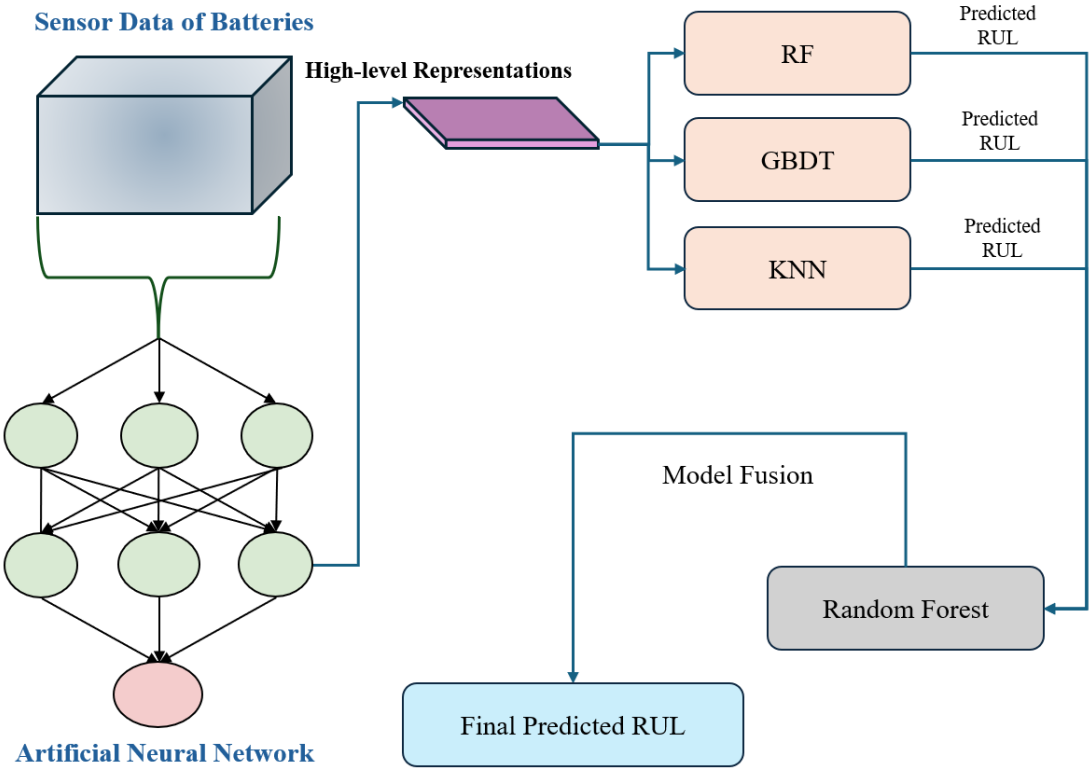

Figure 1. The workflow of the proposed approach.

3. Method

3.1. Dataset Preparation

The dataset utilized in this study is sourced from the publicly available repository provided by AWS, containing comprehensive sensor data from lithium-ion batteries. This dataset consists of 100,377 samples with 10 key features, including internal resistance (IR), charge capacity (QC), discharge capacity (QD), average temperature (Tavg), minimum temperature (Tmin), charge time, and additional metadata such as cycle number, cycle life, battery name, and battery index. The RUL) of the batteries was calculated as the difference between the total cycle life and the current cycle number. Features such as IR, QC, QD, Tavg, Tmin, and charge time were selected as inputs to predict the RUL. The data was divided into training, validation, and test sets, with 70% used for training, 20% of the training set for validation, and the remaining 30% for testing. This dataset enables a detailed analysis of battery degradation behavior, offering a robust foundation for developing machine learning models for RUL prediction. The sensor data, characterized by diverse conditions and extensive cycles, ensures sufficient complexity to validate advanced predictive frameworks. The distribution of computed RUL is provided in Figure 2.

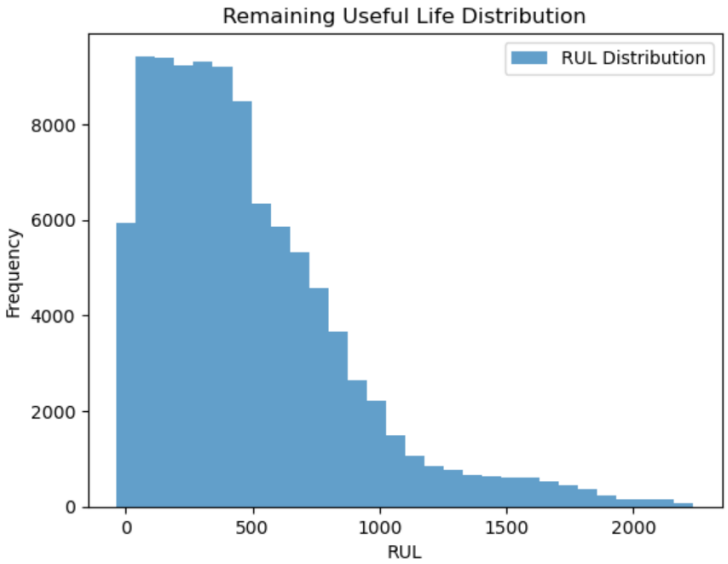

Figure 2. The RUL distribution of the dataset.

3.2. The Introduction of Used Machine Learning Models

3.2.1. ANN Model

Artificial Neural Networks (ANNs) [35–37] are computational models inspired by the structure and function of biological neural networks in the human brain. They consist of interconnected layers of artificial neurons that process and transmit information in a hierarchical manner. ANNs are widely used in various domains for tasks such as classification, regression, and pattern recognition due to their ability to model complex, nonlinear relationships within data. Their power lies in their adaptability, as they can learn intricate patterns directly from raw data through a process of optimization and backpropagation. In this study, we utilized an ANN architecture tailored for the regression task of predicting the RUL of lithium-ion batteries. The designed ANN model consists of three layers: an input layer, two hidden layers, and an output layer. The input layer takes sensor data from lithium-ion batteries, including features such as internal resistance, charge capacity, discharge capacity, average temperature, and charge time, which are crucial indicators of battery health. These features were preprocessed to ensure normalization and scaled to enhance the learning process.

The first hidden layer in the ANN architecture contains 64 neurons, employing a Rectified Linear Unit (ReLU) activation function [38–40]. This activation function introduces nonlinearity into the model, enabling it to learn and represent complex patterns in the data effectively. The second hidden layer contains 32 neurons, also utilizing the ReLU activation function. The gradual reduction in the number of neurons across layers allows the network to extract higher-level features while reducing the dimensionality of the data. The output layer consists of a single neuron with a linear activation function. This setup ensures that the model outputs a continuous numerical value, which is essential for the regression task of predicting RUL. The chosen architecture balances model complexity and computational efficiency, making it suitable for the dataset and task at hand.

The ANN model is trained using a supervised learning approach, where the network iteratively adjusts its weights and biases to minimize the error between predicted and actual RUL values. The ANN model was trained using the Adam optimizer [41,42] with a learning rate of 0.1 and the Mean Absolute Error (MAE) as the loss function, suitable for regression tasks. The training process ran for 50 epochs with a batch size of 32, balancing computational efficiency and gradient stability to ensure effective learning for the RUL prediction task.

3.2.2. K-Nearest Neighbors

K-Nearest Neighbors (KNN) [43,44] is a simple yet powerful non-parametric machine learning algorithm widely used for classification and regression tasks. In the context of regression, KNN predicts the target value of a data point by averaging the target values of its **k** nearest neighbors in the feature space. The proximity of these

neighbors is typically determined using a distance metric such as Euclidean distance, Manhattan distance, or Minkowski distance. The simplicity of KNN lies in its instance-based learning approach—it does not rely on any explicit training process but instead memorizes the training data and performs computations during prediction.

3.2.3. Gradient Boosted Decision Trees

Gradient Boosting Decision Tree (GBDT) [45, 46] is a highly effective ensemble learning algorithm that builds predictive models by sequentially combining weak learners, typically decision trees. The fundamental idea behind GBDT is to iteratively minimize a specified loss function by fitting new decision trees to the residual errors of the previous iterations. This gradient-based approach allows the model to correct its mistakes incrementally, resulting in a robust and accurate predictor.

3.2.4. Random Forest

Random Forest (RF) [47,48] is a widely used ensemble learning method that constructs multiple decision trees during training and aggregates their predictions to produce a final output. Each tree in the forest is trained on a random subset of the data and features, introducing diversity and reducing the risk of overfitting. For regression tasks, RF outputs the average of the predictions from individual trees, ensuring robustness against noisy data and overfitting.

3.3. Multi-Model Fusion Strategy for RUL Prediction

To further improve the performance of RUL prediction, we propose an innovative framework that integrates multiple machine learning algorithms through a unified ensemble model. The ANN mentioned before serves as a feature extractor, capturing intricate relationships within the data and providing a robust foundation for subsequent analysis. These high-level representations are then utilized by RF, KNN, and GBDT. Each model predicts the RUL independently, leveraging its unique strengths to contribute diverse perspectives to the overall estimation process. Then, the outputs of the RF, KNN, and GBDT models were fused using a Random Forest-based ensemble. This model fusion strategy combines the individual predictions into a single, cohesive output by assigning optimal weights to each component model based on its performance. By integrating these predictions, the ensemble model capitalizes on the complementary advantages of the underlying algorithms, enhancing accuracy and robustness while mitigating individual model biases.

4. Results and Discussion

4.1. The Performance of Different Machine Learning Models

The performance of the ANN model in predicting the RUL of lithium-ion batteries in EV is depicted in Figure 3. The results highlight several challenges associated with the standalone use of the ANN model for this task, particularly concerning overfitting and the reliability of predictions on the test set. In the training curve (Figure 3a), the training loss demonstrates a gradual decline over 50 epochs, indicative of the model effectively learning patterns from the training data. However, the validation loss exhibits significant fluctuations and fails to converge, suggesting that the model struggles to generalize to unseen data. This discrepancy between training and validation losses is a classic sign of overfitting, where the model becomes excessively tailored to the training data while losing the ability to perform well on new data. The test set predictions, as shown in Figure 3b, further underline the limitations of the ANN model in its current form. While many predictions align reasonably well with the true RUL values, several outliers demonstrate extreme deviations. For instance, the model occasionally outputs predicted RUL values in the range of tens of thousands, which are far beyond the realistic bounds of the dataset. Such anomalies not only degrade the model's overall performance but also highlight its instability when exposed to variations in the input data. These observations emphasize the necessity of integrating the ANN model with other complementary machine learning techniques to enhance the robustness and reliability of the predictions [48–52]. Although the ANN excels at extracting high-level representations from

the input features, its predictions alone may lack the consistency required for practical applications in battery health management.

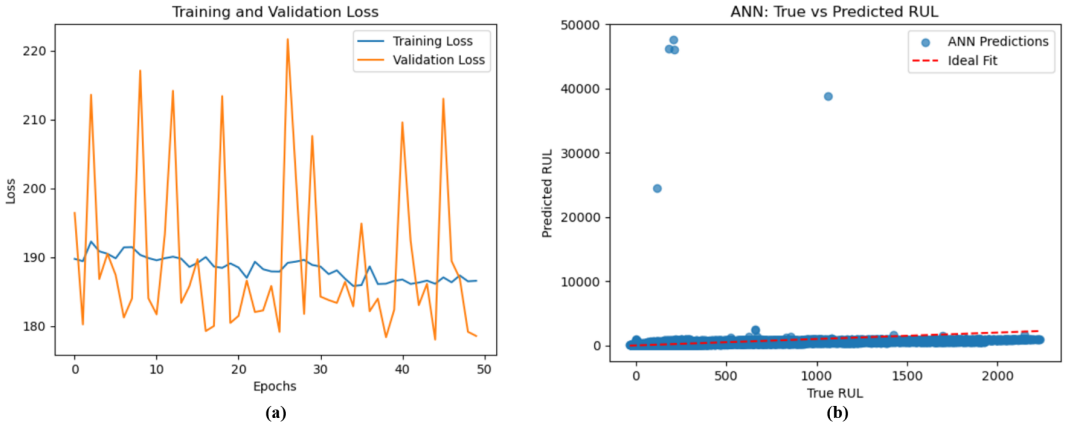

Figure 3. The (**a**) training curve and (**b**) prediction performance of the ANN.

The results presented in Table 1, Figures 4 and 5 provide a comprehensive analysis of the performance of different models used for RUL prediction. The models under consideration include ANN+RF, ANN+KNN, ANN+GBDT, and the final ensemble ANN+RF+KNN+GBDT. Each of these models builds upon the high-level features extracted by the ANN, enhancing predictive performance through complementary machine learning techniques.

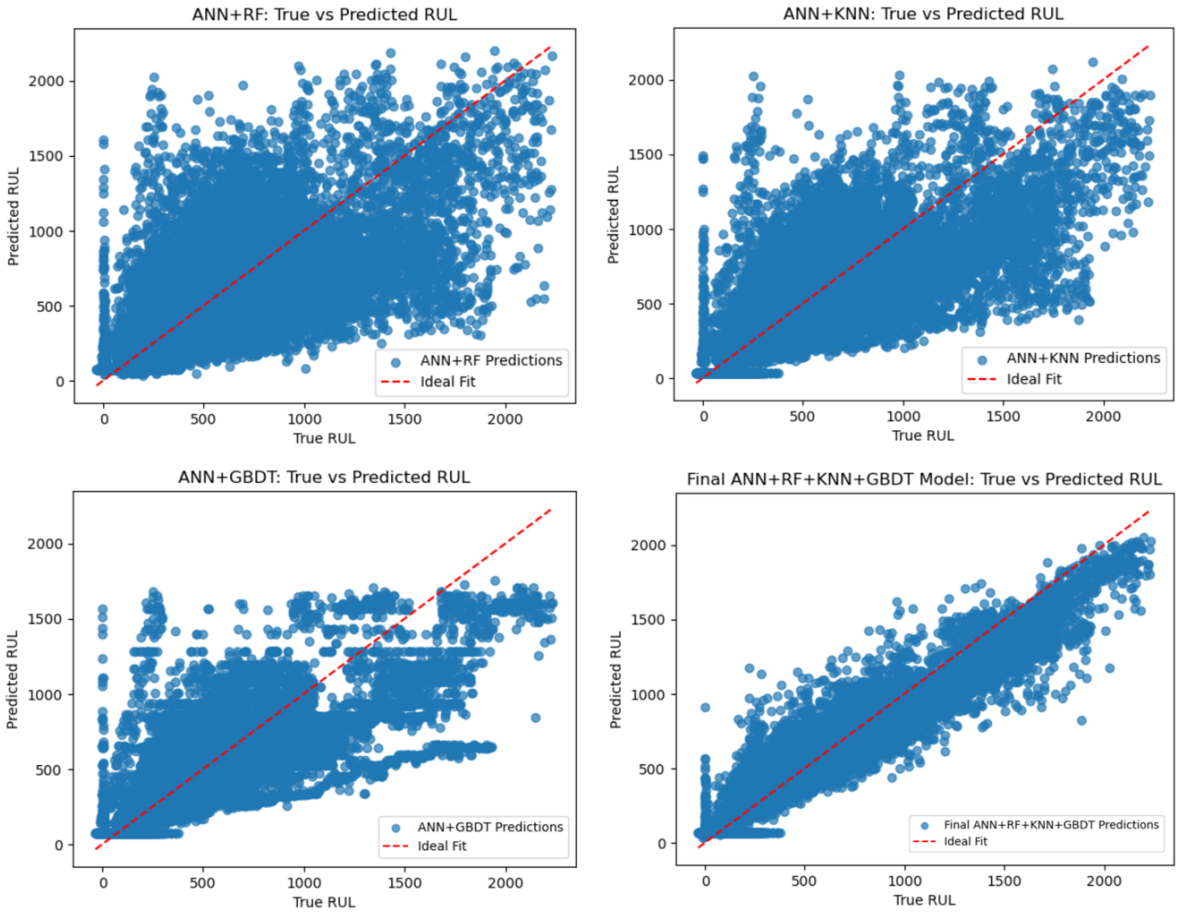

Figure 4. The performance of different machine learning models combined with the ANN.

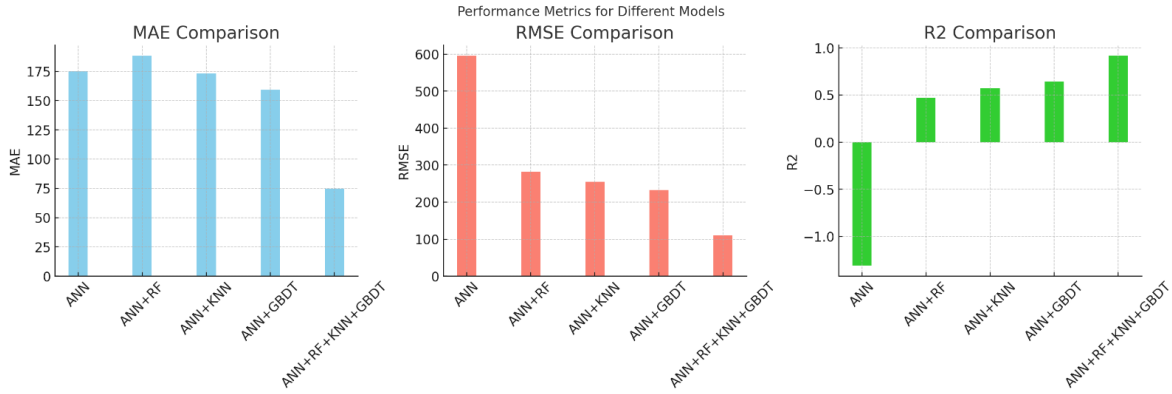

Figure 5. The visualization of performance comparison based on different machine learning models.

Table 1. The performance of different models in the testing dataset evaluated by various metrics.

Model Name	MAE	RMSE	R^2
ANN	175.28	595.26	−1.31
ANN + RF	188.15	282.54	0.47
ANN + KNN	173.22	254.87	0.57
ANN + GBDT	159.26	232.47	0.64
ANN + RF + KNN + GBDT	**74.52**	**110.07**	**0.92**

In the first stage, the ANN model acts as a feature extractor, capturing intricate relationships in the input sensor data. The extracted features are then utilized by three distinct models: RF, KNN, and GBDT. Each of these models leverages its unique strengths to improve RUL prediction: (1) ANN + RF**:** The Random Forest model effectively utilizes the extracted features, benefiting from its ability to handle high-dimensional data and identify important features. This approach achieves moderate improvements over the standalone ANN model, as indicated by a reduction in RMSE from 595.26 to 282.54. (2) ANN + KNN**:** The K-Nearest Neighbors model performs slightly better, with MAE and RMSE values of 173.22 and 254.87, respectively. KNN's ability to capture local patterns in the extracted features enables it to outperform RF in this setup. However, KNN's performance is limited by its sensitivity to the choice of hyperparameters, such as the number of neighbors. (3) ANN + GBDT: Gradient Boosting Decision Tree achieves the best performance among the individual models, with MAE and RMSE values of 159.26 and 232.47, respectively. GBDT's ability to model complex, nonlinear interactions between features extracted by the ANN contributes to its superior performance. Moreover, GBDT demonstrates better robustness against overfitting compared to RF and KNN.

The final ensemble model integrates the outputs of RF, KNN, and GBDT using another Random Forest model for computing their weights of combination. This fusion strategy further significantly enhances predictive accuracy by leveraging the complementary strengths of the three models. The resulting MAE is reduced to 74.52, and RMSE is dramatically lowered to 110.07, while the R^2 score increases to 0.92. This improvement highlights the effectiveness of combining diverse perspectives in the prediction process. The ensemble approach mitigates individual model biases and overcomes their respective weaknesses. For example, RF's robustness, KNN's ability to capture local patterns, and GBDT's capacity to model complex interactions complement each other.

4.2. The Influence of n_estimators on the Final Ensembled Model Performance

Since the RF model was employed to determine the optimal weight combination of predictions from ANN + RF, ANN + KNN, and ANN + GBDT models, the investigation related to the setting of its hyperparameters e.g., n_estimators is crucial. Experiments were conducted with varying values of n_estimators ranging from 50 to 500 shown in Figure 6. The left plot illustrates the changes in MAE and RMSE as n_estimators increases. Both metrics show a decreasing trend as the number of estimators grows. This indicates that increasing the number of

decision trees enhances the model's ability to generalize and reduces prediction errors. The RMSE, in particular, sees a significant decline initially, emphasizing the RF model's capacity to handle residual errors effectively. The right plot demonstrates the evolution of the R^2 score with increasing n_estimators. The R^2 value improves consistently, plateauing near 0.92 at around 300 estimators.

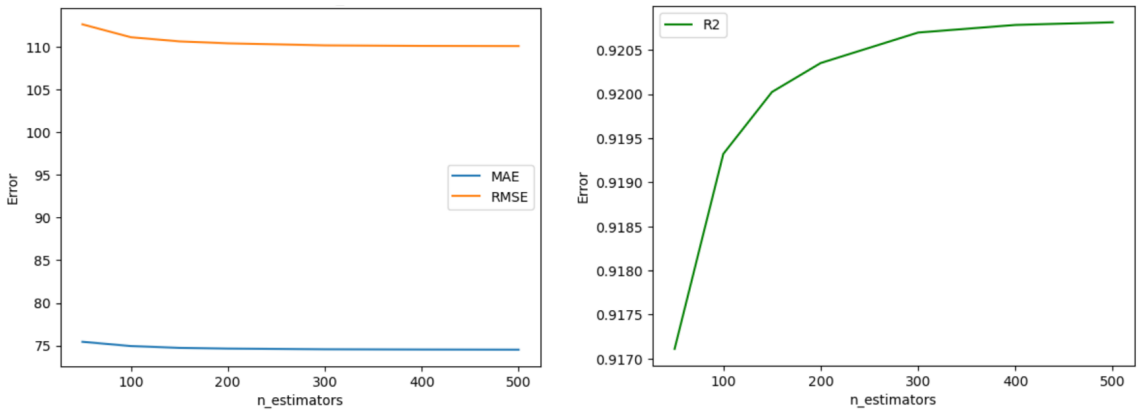

Figure 6. The influence of the number of estimators on the final ensembled model's performance.

4.3. Feature Importance Analysis Based on the Random Forest

To better understand feature importance from the perspective of machine learning models, we trained a standalone Random Forest model on the dataset from scratch. This process aims to evaluate the relative contributions of each feature to the prediction of RUL. The results, illustrated in Figure 7, provide insights into how different features influence the model's performance.

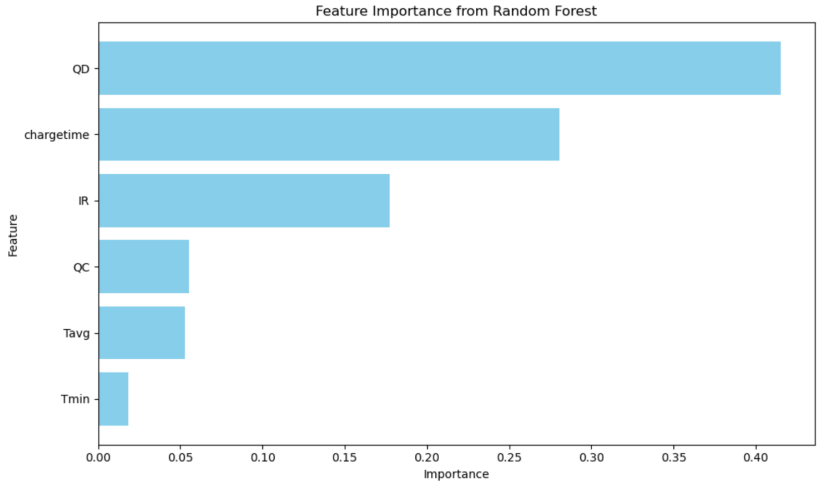

Figure 7. The influence of different features based on the random forest.

From the figure, it is evident that discharge capacity (QD) is the most critical feature, with the highest importance score. This indicates that the RUL of a battery is strongly correlated with its discharge behavior, as it reflects the battery's ability to deliver power over time. The second most important feature is charge time, which provides indirect information about the battery's efficiency and charging patterns, further emphasizing its relevance to degradation analysis. The internal resistance (IR) ranks third, showcasing its significance in monitoring battery health. Higher resistance is often associated with aging and reduced performance, making it a valuable predictor for RUL. In contrast, features such as charge capacity (QC), average temperature (Tavg), and minimum temperature (Tmin) exhibit lower importance scores. While these features contribute to the prediction, their influence is relatively minor compared to QD, charge time, and IR. The results are consistent with general battery science, but they also highlight the limitations of data-driven models like Random Forest. The model interprets feature importance based on statistical patterns in the dataset, which may not fully capture all domain-

specific nuances or edge cases. For instance, temperature effects, though less prominent in this analysis, are critical in specific real-world scenarios like extreme operating environments or fast charging.

4.4. Discussion

While the proposed framework demonstrates significant improvements in RUL prediction, several limitations should be addressed in future research: (1) The study relies on a specific dataset with sensor data from lithium-ion batteries under controlled conditions. Real-world environments may introduce additional variability, such as extreme temperatures or irregular charging behaviors, which are not accounted for in the current model. This limits the generalizability of the findings across diverse battery applications. (2) The ensemble framework, while effective, introduces increased complexity by combining multiple machine learning models. This complexity may hinder interpretability, making it challenging to understand how each component contributes to the final prediction. Developing more transparent ensemble techniques or explainable AI methods could address this issue. (3) Training and fine-tuning multiple models, including Random Forest, KNN, and GBDT, are computationally intensive, especially with large datasets. This could limit the practical application of the framework in scenarios requiring real-time or on-device RUL predictions. (4) Although the framework leverages ANN-extracted features, the absence of domain-specific feature engineering may have overlooked critical interactions between variables, such as temperature effects under specific conditions.

5. Conclusion

This study presents an innovative framework for predicting the RUL of lithium-ion batteries using an ensemble machine learning approach. By combining ANN-based feature extraction with RF, KNN, and GBDT models, the proposed method demonstrates enhanced predictive accuracy and robustness. The final ensemble model effectively integrates the strengths of individual algorithms, achieving significant improvements in RUL estimation. Furthermore, feature importance analysis highlights the relevance of discharge capacity and charge time as key indicators of battery health. Despite these advancements, challenges remain, including computational demands, limited generalizability to real-world conditions, and the need for more domain-specific feature engineering. Future work should focus on addressing these limitations to ensure practical applicability and scalability of the framework in diverse battery management scenarios.

Funding

Not applicable.

Author Contributions

Conceptualization, W.H.; writing—original draft preparation, writing—review and editing, W.H., T.Z., J.M. and X.C.; All of the authors read and agreed to the published the final manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The authors declare no conflict of interest.

References

1 Lei J, Nisar A. Investigating the Influence of Green Technology Innovations on Energy Consumption and Corporate Value: Empirical Evidence from Chemical Industries of China. *Innovations in Applied Engineering and Technology* 2023; **2**(**1**): 1–6.

2 Lei J. Efficient Strategies on Supply Chain Network Optimization for Industrial Carbon Emission Reduction. *arXiv* 2024, arXiv:2404.16863.

3 Lei J. Green Supply Chain Management Optimization Based on Chemical Industrial Clusters. *arXiv* 2024, arXiv:2406.00478.

4 Sanguesa JA, Torres-Sanz V, Garrido P, *et al*. A Review on Electric Vehicles: Technologies and Challenges. *Smart Cities* 2021; **4**(**1**): 372–404.

5 Hannan MA, Azidin FA, Mohamed A. Hybrid Electric Vehicles and Their Challenges: A Review. *Renewable and Sustainable Energy Reviews* 2014; **29**: 135–50.

6 Dai W. Safety Evaluation of Traffic System with Historical Data Based on Markov Process and Deep-Reinforcement Learning. *Journal of Computational Methods in Engineering Applications* 2021; **1**(**1**): 1–4.

7 Dai W. Design of Traffic Improvement Plan for Line 1 Baijiahu Station of Nanjing Metro. *Innovations in Applied Engineering and Technology* 2023; **2**(**1**): 10.

8 Ren L, Zhao L, Hong S, *et al*. Remaining Useful Life Prediction for Lithium-Ion Battery: A Deep Learning Approach. *IEEE Access* 2018; **6**: 50587–50598.

9 Liu K, Shang Y, Ouyang Q, *et al*. A Data-Driven Approach with Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-Ion Battery. *IEEE Transactions on Industrial Electronics* 2020; **68**(**4**): 3170–3180.

10 Liao L, Köttig F. Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction. *IEEE Transactions on Reliability* 2014; **63**(**1**): 191–207.

11 Liu Y, Wang J. AI-Driven Health Advice: Evaluating the Potential of Large Language Models as Health Assistants. *Journal of Computational Methods in Engineering Applications* 2023; **3**(**1**): 1–7. https://doi.org/10.62836/jcmea.v3i1.030106.

12 Tang CY, Li C. Examining the Factors of Corporate Frauds in Chinese A-share Listed Enterprises. *OAJRC Social Science* 2023; **4**(**3**): 63–77.

13 Tang Y, Li C. Exploring the Factors of Supply Chain Concentration in Chinese A-Share Listed Enterprises. *Journal of Computational Methods in Engineering Applications* 2023; **3**(**1**): 1–17. https://doi.org/10.62836/jcmea.v3i1.030105.

14 Moayedi H, Zhao Y, Foong LK, *et al*. Slime Mold and Four Other Nature-Inspired Optimization Algorithms in Analyzing the Concrete Compressive Strength. *Smart Structures and Systems, An International Journal* 2024; **33**(**1**): 65–91.

15 Boden-Albala B, Ding X, Ryan N, *et al*. Anti-Asian Racism Related Stigma, Racial Discrimination, and Protective Factors Against Stigma: A Repeated Cross-Sectional Survey among University Students during the COVID-19 Pandemic. *Frontiers in Public Health* 2023; **11**: 958932.

16 Chen X, Wang M, Zhang H. Machine Learning-Based Fault Prediction and Diagnosis of Brushless Motors. *Engineering Advances* 2024; **4**(**3**): 130–142.

17 Chen X, Zhang H. Performance Enhancement of AlGaN-Based Deep Ultraviolet Light-Emitting Diodes with AlxGa1-xN Linear Descending Layers. *Innovations in Applied Engineering and Technology* 2023; **2**(**1**): 1–10.

18 Wenjun D, Fatahizadeh M, Touchaei HG, *et al*. Application of Six Neural Network-Based Solutions on Bearing Capacity of Shallow Footing on Double-Layer Soils. *Steel and Composite Structures* 2023; **49**(**2**): 231–244.

19 Ding X, Meng Y, Xiang L, *et al*. Stroke Recurrence Prediction Using Machine Learning and Segmented Neural Network Risk Factor Aggregation. *Discover Public Health* 2024; **21**(**1**): 119.

20 Ma J, Chen X. Fingerprint Image Generation Based on Attention-Based Deep Generative Adversarial Networks and Its Application in Deep Siamese Matching Model Security Validation. *Journal of*

Computational Methods in Engineering Applications 2024; **4**: 1–13.

21 Zhang H, Zhu D, Gan Y, *et al*. End-to-End Learning-Based Study on the Mamba-ECANet Model for Data Security Intrusion Detection. *Journal of Information, Technology and Policy* 2024; **2**(**1**): 1–17.

22 Zhang G, Zhou T. Finite Element Model Calibration with Surrogate Model-Based Bayesian Updating: A Case Study of Motor FEM Model. *Innovations in Applied Engineering and Technology* 2024; **3**(**1**): 1–13.

23 Zhang G, Zhou T, Cai Y. CORAL-Based Domain Adaptation Algorithm for Improving the Applicability of Machine Learning Models in Detecting Motor Bearing Failures. *Journal of Computational Methods in Engineering Applications* 2023; **3**(**1**): 1–17. https://doi.org/10.62836/jcmea.v3i1.030108.

24 Wang Z, Zhao Y, Song C, *et al*. A New Interpretation on Structural Reliability Updating with Adaptive Batch Sampling-Based Subset Simulation. *Structural and Multidisciplinary Optimization* 2024; **67**(**1**): 7.

25 Ye X, Luo K, Wang H, *et al*. An Advanced AI-Based Lightweight Two-Stage Underwater Structural Damage Detection Model. *Advanced Engineering Informatics* 2024; **62**: 102553.

26 Hao Y, Chen Z, Jin J, *et al*. Joint Operation Planning of Drivers and Trucks for Semi-Autonomous Truck Platooning. *Transportmetrica A: Transport Science* 2023: 1–37.

27 Hao Y, Chen Z, Sun X, *et al*. Planning of Truck Platooning for Road-Network Capacitated Vehicle Routing Problem. *Transportation Research Part E: Logistics and Transportation Review* 2025; **194**: 103898.

28 Wang S, Jin S, Deng D, *et al*. A Critical Review of online Battery Remaining Useful Lifetime Prediction Methods. *Frontiers in Mechanical Engineering* 2021; **7**: 719718.

29 Qiu S, Zhang B, Lv Y, *et al*. A Lithium-Ion Battery Remaining Useful Life Prediction Model Based on CEEMDAN Data Preprocessing and HSSA-LSTM-TCN. *World Electric Vehicle Journal* 2024; **15**(**5**): 177.

30 Suh S, Mittal DA, Bello H, *et al*. Remaining Useful Life Prediction of Lithium-Ion Batteries Using Spatio-Temporal Multimodal Attention Networks. *Heliyon* 2024; **10**(**16**): e36236.

31 Gan Y, Zhu D. The Research on Intelligent News Advertisement Recommendation Algorithm Based on Prompt Learning in End-to-End Large Language Model Architecture. *Innovations in Applied Engineering and Technology* 2024; **3**(**1**): 1–19.

32 Gan Y, Ma J, Xu K. Enhanced E-Commerce Sales Forecasting Using EEMD-Integrated LSTM Deep Learning Model. *Journal of Computational Methods in Engineering Applications* 2023; **3**(**1**): 1–11.

33 Gan Y, Chen X. The Research on End-to-end Stock Recommendation Algorithm Based on Time-frequency Consistency. *Advances in Computer and Communication* 2024; **5**(**4**): 243–259.

34 Agatonovic-Kustrin S, Beresford R. Basic Concepts of Artificial Neural Network (ANN) Modeling and Its Application in Pharmaceutical Research. *Journal of Pharmaceutical and Biomedical Analysis* 2000; **22**(**5**): 717–727.

35 Zupan J. Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them. *Acta Chimica Slovenica* 1994; **41**(**3**): 327.

36 Ghritlahre HK, Prasad RK. Application of ANN Technique to Predict the Performance of Solar Collector Systems—A review. *Renewable and Sustainable Energy Reviews* 2018; **84**: 75–88.

37 Hara K, Saito D, Shouno H. Analysis of Function of Rectified Linear Unit Used in Deep Learning. In Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 12–17 July 2015; pp. 1–8.

38 Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th international Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.

39 Arora R, Basu A, Mianjy P, *et al*. Understanding Deep Neural Networks with Rectified Linear Units. *arXiv* 2016, arXiv:1611.01491.

40 Zhang Z. Improved adam Optimizer for Deep Neural Networks. In Proceedings of the 2018 IEEE/ACM 26th international symposium on quality of service (IWQoS), Banff, AB, Canada, 4–6 June 2018; pp. 1–2.

41 Bock S, Goppold J, Weiß M. An Improvement of the Convergence Proof of the ADAM-Optimizer. *arXiv* 2018, arXiv:1804.10587.

42 Peterson LE. K-Nearest Neighbor. *Scholarpedia* 2009; **4**(**2**): 1883.

43 Steinbach M, Tan PN. kNN: k-Nearest Neighbors. In *The Top Ten Algorithms in Data Mining*; Chapman and

Hall/CRC: London, UK, 2009; pp. 165–176.

44 Ke G, Meng Q, Finley T, *et al*. Lightgbm: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems* 2017; **30**.

45 Rao H, Shi X, Rodrigue AK, *et al*. Feature Selection Based on Artificial Bee Colony and Gradient Boosting Decision Tree. *Applied Soft Computing* 2019; **74**: 634–642.

46 Breiman L. Random Forests. *Machine Learning* 2001; **45**: 5–32.

47 Oshiro TM, Perez PS, Baranauskas JA. How Many Trees in a Random Forest?. In Proceedings of the Machine Learning and Data Mining in Pattern Recognition: 8th International Conference—MLDM 2012, Berlin, Germany, 13–20 July 2012; pp. 154–168.

48 Luo Z, Yan H, Pan X. Optimizing Transformer Models for Resource-Constrained Environments: A Study on Model Compression Techniques. *Journal of Computational Methods in Engineering Applications* 2023; **3**(**1**): 1–12. https://doi.org/10.62836/jcmea.v3i1.030107.

49 Yan H, Shao D. Enhancing Transformer Training Efficiency with Dynamic Dropout. *arXiv* 2024, arXiv: 2411.03236. https://doi.org/10.48550/arXiv.2411.03236.

50 Cui W, Zhang J, Li Z, *et al*. Evaluating and Improving Generation Consistency of Large Language Models via A Divide-Conquer-Reasoning Approach. Available online: https://openreview.net/forum?id=wk77w7DG1N (accessed on 15 February 2025).

51 Cui W, Zhang J, Li Z, *et al.* Phaseevo: Towards Unified in-Context Prompt Optimization for Large Language Models. *arXiv* 2024, arXiv:2402.11347.

52 Zhang J, Cui W, Huang Y, *et al*. Synthetic Knowledge Ingestion: Towards Knowledge Refinement and Injection for Enhancing Large Language Models. *arXiv* 2024, arXiv:2410.09629. https://doi.org/10.48550/arXiv.2410.09629.

© The Author(s) 2025. Published by Global Science Publishing (*GSP*).

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.