iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XI **Month of publication:** November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75181

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Machine Learning Approaches for Battery Management: A Hybrid Stacked Model for Predicting RUL and SoH

Mr. Mudathir Mowia Mohmoud Elhag
Intelligent Manufacturing and Control Engineering Zhejiang University of Science and Technology

Abstract: This study presents a comparative evaluation of four predictive models—Support Vector Regression (SVR), XGBoost, Random Forest, and a Hybrid Stacked Ensemble model—designed to estimate the Remaining Useful Life (RUL) and State of Health (SoH) of lithium-ion batteries using data from a distributed Battery Management System (BMS). The models were assessed based on multiple metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), training time, and inference time, with a focus on real-time deployment feasibility. The results show that while Random Forest is the most robust base model, the Hybrid Stacked model, which integrates predictions from SVR, XGBoost, and Random Forest via a Random Forest meta-learner, delivers superior performance. The Hybrid model achieved a 53% reduction in MAE and a 55% reduction in RMSE compared to the best base model. Visual and statistical analysis further supports the Hybrid model's accuracy, stability, and applicability to real-time battery health management. The findings suggest that ensemble methods, particularly stacking, offer substantial improvements in predictive reliability and generalization.
Keywords: Remaining Useful Life (RUL), State of Health (SoH), Lithium-ion Batteries, Machine Learning, Hybrid Stacked Ensemble

I. INTRODUCTION

As the demand for sustainable energy solutions grows, accurately predicting the Remaining Useful Life (RUL) and State of Health (SoH) of lithium-ion batteries has become a critical task. The Battery Management System (BMS) is integral to monitoring battery performance, but the raw data from such systems is often complex, noisy, and requires thorough preprocessing before it can be used effectively for predictive modeling. This article presents a comprehensive evaluation of several machine learning models for predicting battery RUL and SoH, specifically comparing Support Vector Regression (SVR), XGBoost, and Random Forest, with a focus on a Hybrid Stacked Ensemble Model. Through systematic preprocessing, including outlier removal, feature normalization, and SMOTE-based balancing, we aim to establish the effectiveness of these models in real-world applications [1]. The evaluation involves quantitative metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), along with qualitative analysis of prediction accuracy and computational efficiency. The results suggest that ensemble methods, particularly the proposed hybrid stacked model, significantly outperform traditional models in both accuracy and stability.

II. DATASET STRUCTURE AND PREPROCESSING

A. Graph Structure of the Battery Management System (BMS) Dataset

The dataset that is known as the Battery Management System (BMS) is made up of a number of different variables that indicate the performance parameters of the battery cells and modules. The (Figure 1) that is displayed is a correlation matrix for a number of different factors, some of which are cell voltage, temperature, current, and power, amongst others. cell_voltage_v is a variable that represents the voltage of the cells in the battery. the cell_temperature_c variable is a representation of the temperature of the battery cells expressed in degrees Celsius. Module_current_a is a variable that displays the current, measured in amperes, that is flowing through the battery modules. The module_power_kw variable stands for the power, measured in kilowatts, that the battery modules possess. The term "converter_command_pct" refers to the proportion of command signals that are received by the converter in the BMS. The soc_pct variable represents the proportion of the battery that is in a state of charge. The term "soh_pct" refers to the percentage of the battery that is in a condition of health. The anomaly score is represented as a percentage for the purpose of system health diagnostics via the anomaly_score_pct variable. latency_ms is a variable that indicates the amount of time that has passed in milliseconds within the system's data processing or transmission [2].

The following variables are shown to have relationships with one another in the correlation matrix: It has been found that certain variables, such as cell_temperature_c and soc_pct, exhibit strong negative correlations, which means that when one variable grows, the other variable tends to drop. There is a high positive correlation between the module_current_a and the module_power_kw, which indicates that they move together closely. Latency_ms has poor correlations with other parameters, suggesting it's less dependent on them.

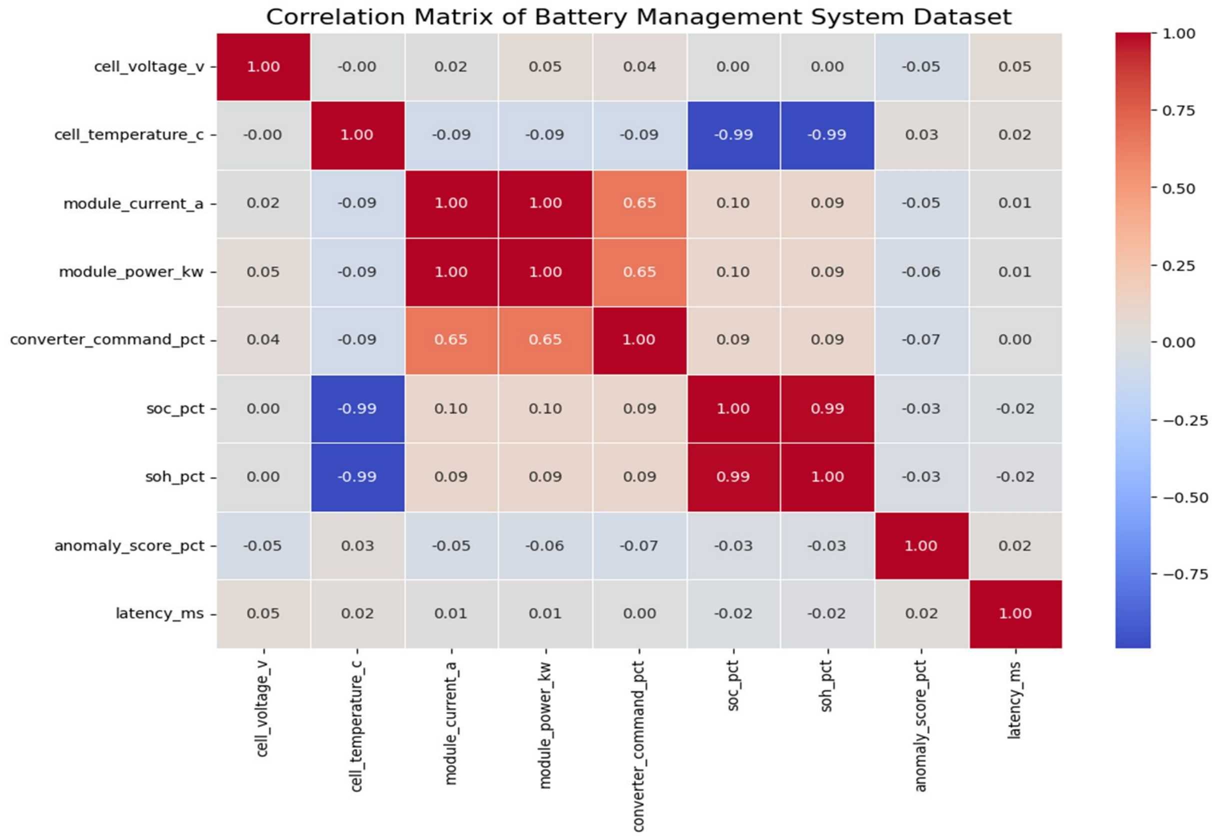


Figure 1: Correlation Matrix of Battery Management System Dataset

B. *Date Preprocessing*

Data preparation is a crucial first step for turning raw Battery Management System (BMS) data into a form suitable for efficient analysis and model training. Raw data often contain missing values, duplicates, outliers and noisy or uninformative patterns that can undermine model performance if left unaddressed. In this study, data preprocessing was carried out to clean and structure the dataset before modeling [3]. Missing data—caused by sensor errors or incomplete readings—were handled either by removing rows or columns with excessive gaps or by applying simple imputation methods (such as mean filling), to reduce bias and avoid instability during training. Duplicate records were identified and removed to prevent overfitting and to ensure that each observation contributes unique information. Outliers, which can distort statistical relationships and mislead the learning process, were detected using methods such as the Interquartile Range (IQR) and either removed or adjusted so that extreme, non-representative values do not dominate the model. In parallel, feature engineering was applied to transform raw signals into more informative representations: rolling averages of key variables (e.g., temperature, voltage, current) were computed to smooth short-term fluctuations and highlight longer-term trends, while rates of change were derived to capture dynamic shifts in battery behaviour that may signal degradation or impending failure. Because many machine learning algorithms are sensitive to feature scale, standardization was used to rescale all numerical variables to zero mean and unit variance, ensuring that variables with larger numeric ranges do not dominate the learning process. The dataset also exhibited class imbalance in scenarios such as failure detection, where failure events are rare but critical; therefore, the Synthetic Minority Over-sampling Technique (SMOTE) was employed as shown in diagram (figure 2) to generate synthetic samples of the minority class, improving the model's ability to recognize rare but important patterns.

Finally, the processed dataset was split into training and testing subsets to enable objective evaluation on unseen data, support generalization assessment, and reduce the risk of overfitting. Through this sequence of preprocessing steps, a higher-quality dataset was constructed to support robust machine learning models for predicting the remaining useful life (RUL) of batteries [4].

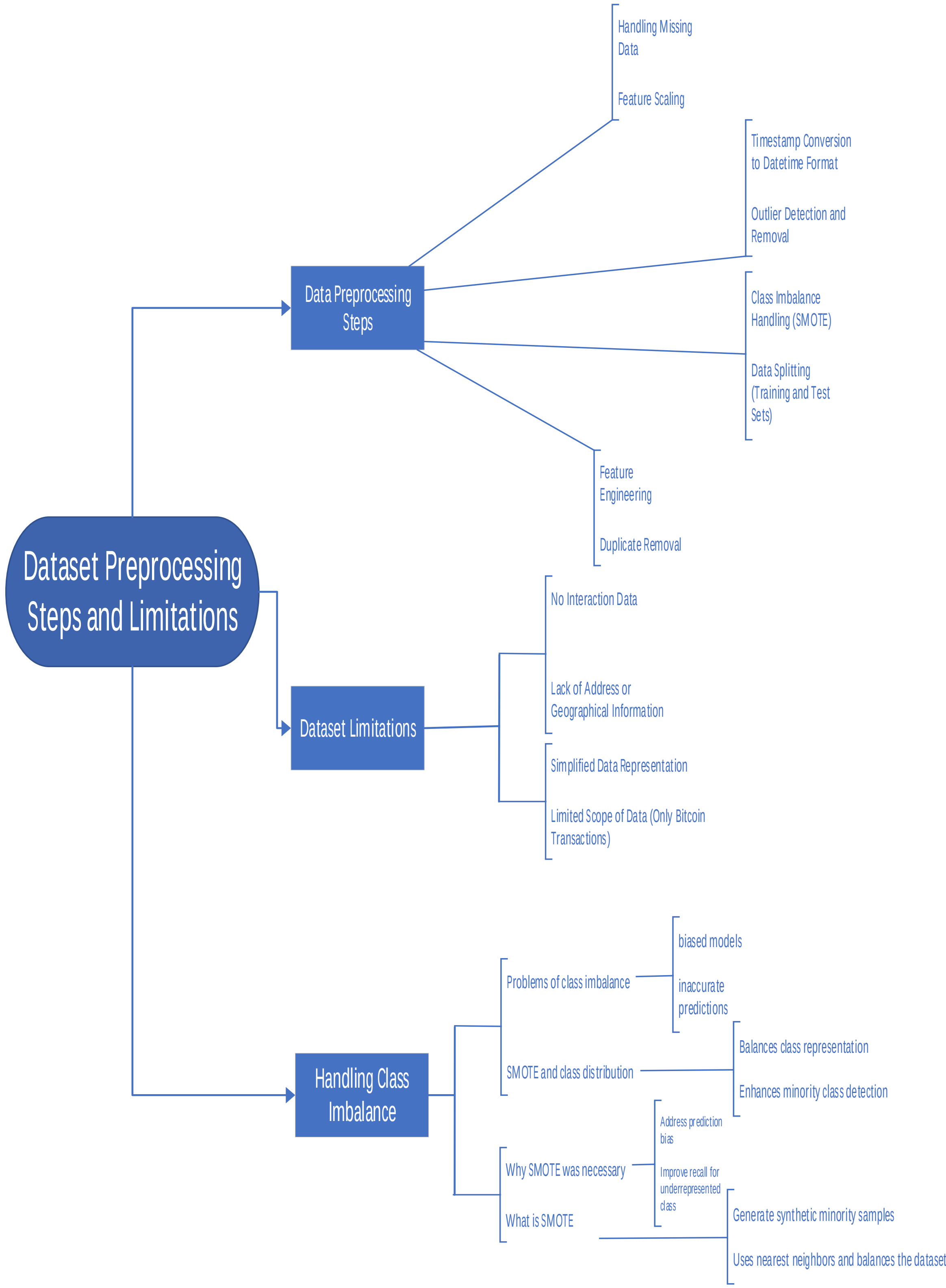


Figure 2: Correlation Matrix of Battery Management System Dataset

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

III. EVALUATION AND PERFORMANCE ANALYSIS

A. *Evaluation Methodology*

The proposed models were evaluated on real-time data collected from the distributed Battery Management System (BMS) described earlier. The dataset consists of timestamped measurements of cell voltage, current and temperature, together with derived indicators of battery State of Health (SoH) / Remaining Useful Life (RUL). Prior to modelling, the raw data underwent a structured preprocessing pipeline including removal of outliers, feature normalization and class balancing using SMOTE, ensuring that rare but safety-critical degradation states were adequately represented [5]. To obtain an unbiased estimate of generalization, the cleaned dataset was partitioned using an 80/20 train–test split (Table 1). The training portion was used exclusively for model fitting and hyperparameter tuning, while the held-out test subset—covering both low- and high-stress operating regimes—served only for final evaluation and cross-model comparison. To prevent data leakage, transformations were fitted on the training set and then applied to the test set. In particular, the StandardScaler used for SVR was calibrated only on training data. For the stacked ensemble, predictions from the base learners (SVR, XGBoost, Random Forest) were first generated independently on the test set and then supplied as inputs to a Random Forest meta-learner, preserving a strict separation between training and evaluation phases [6].

Table 1**:** Dataset Split and Description

Dataset Split	Percentage	Sample Count	Description
Training Set	80%	N_1 samples	Used for model fitting and hyperparameter tuning
Testing Set	20%	N_2 samples	Used exclusively for model evaluation and comparison

All experiments were conducted under a consistent hardware and software configuration (Table 2) on a workstation equipped with an Intel® Core™ i7-12700K CPU, 32 GB RAM and an NVIDIA RTX 3060 GPU, running Python 3.10 in a Jupyter/Anaconda environment. Each model was trained and evaluated multiple times with different random seeds; results were averaged to reduce stochastic variation arising from data partitioning and algorithm initialization.

Table 2: Hardware and Software Configuration

Component	Specification
Processor	Intel® Core™ i7-12700K (12 Cores, 3.6 GHz)
RAM	32 GB DDR5
GPU	NVIDIA RTX 3060 (12 GB VRAM)
Operating System	Windows 11 (64-bit)
Programming Language	Python 3.10
Libraries Used	scikit-learn, xgboost, numpy, pandas, matplotlib

B. *Evaluation Metrics and Mathematical Formulations*

Both Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), which are the most generally used indicators for regression problems involving time-series degradation prediction, were utilized in order to assess the effectiveness of each and every model [7].

1) *Mean Absolute Error (MAE)*

MAE provides a linear score that measures the average absolute difference between predicted and actual values, offering an intuitive interpretation of average model deviation.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

A smaller MAE indicates that predictions are closely aligned with actual values, reflecting overall model reliability across the operational range.

2) Root Mean Squared Error (RMSE)

RMSE measures the standard deviation of prediction errors and penalizes large deviations more strongly. This property makes RMSE more sensitive to sporadic or extreme mispredictions, which are particularly critical in degradation forecasting.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

The lower the RMSE, the more accurate and stable the prediction. Since real-world battery degradation often includes nonlinear and abrupt shifts, the use of RMSE provides insights into each model's robustness against noise and variability.

IV. COMPARATIVE RESULTS AND ANALYSIS

A. Overall Comparison of Models

(Table 3) summarizes the quantitative results for all four models—SVR, XGBoost, Random Forest and the Hybrid (Stacked) ensemble—on the test set, including error metrics, computational cost and qualitative interpretability:

- SVR attains an MAE of 0.1261 and RMSE of 0.1536, with a training time of about 20.1 s and fast inference (0.05 s per prediction).
- XGBoost improves accuracy with MAE 0.1035 and RMSE 0.1292, while training more quickly (10.5 s) but incurring slightly higher inference latency (0.10 s).
- Random Forest is the best individual learner, achieving MAE 0.0983 and RMSE 0.1206 with moderate training time (15.3 s) and low inference time (0.07 s).
- The Hybrid (Stacked) model, which uses a Random Forest meta-learner to fuse base-model predictions, achieves the lowest errors by a clear margin, with MAE 0.0388 and RMSE 0.0484. This corresponds to roughly a 53% reduction in MAE and 55% reduction in RMSE relative to the strongest base model (Random Forest), while incurring negligible additional runtime once base predictions are available.

Table 3: Detailed Model Comparison

Model Type	MAE	RMSE	Training Time (s)	Inference Time (s)	Key Hyperparameters	Interpretability	Feature Importance
SVR	0.1261	0.1536	20.1	0.05	C=100, ε=0.1	High	No
XGBoost	0.1035	0.1292	10.5	0.10	n_estimators=100, learning_rate=0.1	Low	Yes
Random Forest	0.0983	0.1206	15.3	0.07	n_estimators=100, max_depth=5	Moderate	Yes
Hybrid (Stacked)	0.0388	0.0484	–	–	Random Forest (meta-model)	Moderate	Yes

Residual distribution plots (Figure 3) further corroborate these findings. SVR and XGBoost exhibit broader, slightly skewed residual curves, indicating biased errors and a higher proportion of large deviations. Random Forest produces a narrower, more symmetric distribution around zero, reflecting improved stability. The hybrid model's residuals are the most sharply peaked and centered, demonstrating both low bias and low variance.

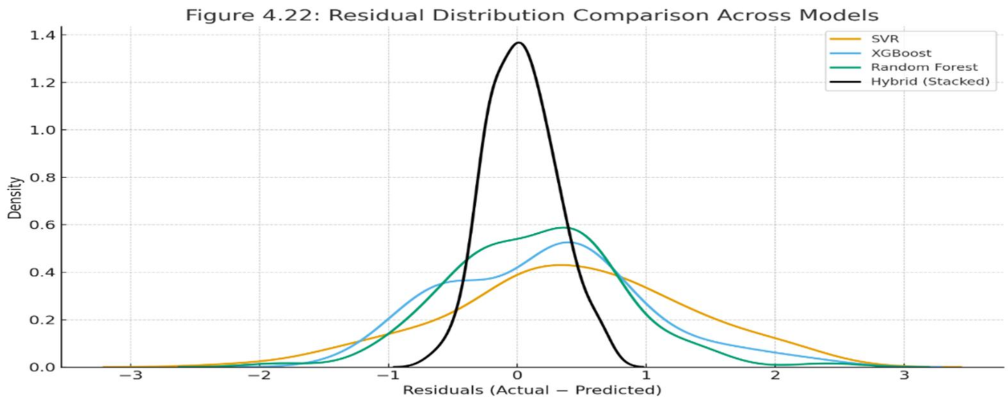


Figure 3: Residual Distribution Comparison Across Models

B. Behavior of Individual Models

SVR, while theoretically effective for non-linear regression with the RBF kernel, struggles with feature scaling sensitivity and kernel hyperparameter tuning, leading to underfitting in dynamic SoH transitions and large errors in regions with rapid voltage or temperature changes. Its simplicity and interpretability are outweighed by poor accuracy on high-dimensional, sensor-rich BMS data. XGBoost, an ensemble of gradient-boosted trees, performs better with complex feature interactions and non-linear degradation trends, but its lower interpretability and sensitivity to hyperparameters like tree depth and learning rate can lead to overfitting and instability under noisy conditions [8]. Random Forest, on the other hand, offers robust performance, efficiently handling non-linear relationships and noisy features, while providing meaningful feature importance as (Figure 4) show It closely tracks the SoH/RUL trajectory, even in high-variance regions, making it a reliable baseline for real-time BMS deployment due to its accuracy, robustness, and moderate interpretability [9].

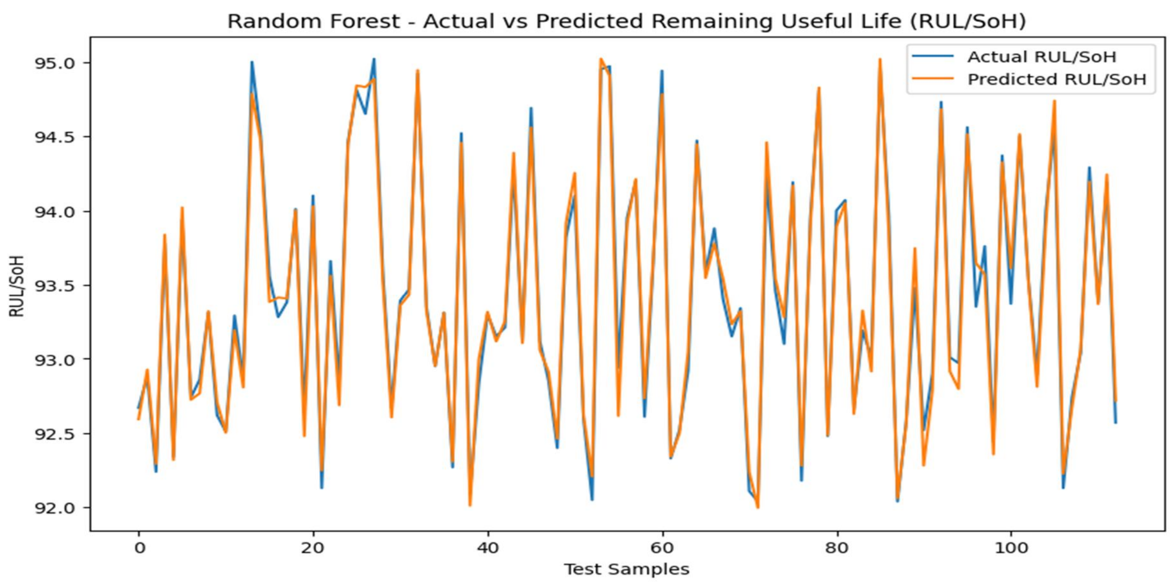


Figure 4: Random Forest - Actual vs Predicted Remaining Useful Life (RUL/SoH)

C. Hybrid Stacked Model

The hybrid architecture integrates the three base learners through a Random Forest meta-model that operates on their predicted outputs. Conceptually, if $f_i(X)$denotes the prediction of the i-th base model and $w_i$the weight learned by the meta-learner, the final hybrid prediction can be expressed as

$$\hat{y}_{\text{hybrid}} = \sum_{i=1}^{n} w_i f_i(X),$$

where the weights are not fixed a priori but are inferred from data in a non-linear fashion. The conceptual diagram in (Figure 5) illustrates this flow: SVR, XGBoost and Random Forest generate candidate predictions, which are then combined by the meta-learner to produce the final SoH/RUL estimate.

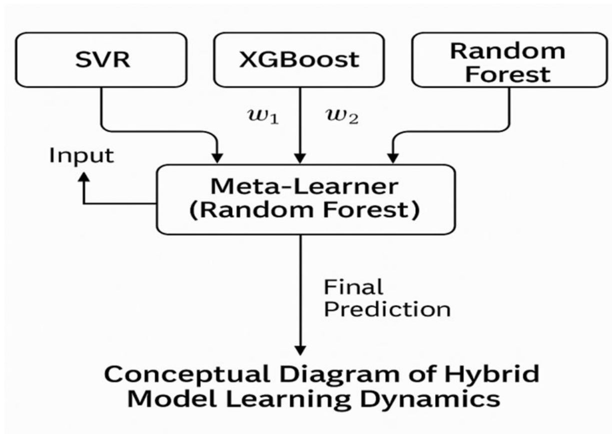


Figure 5: Conceptual Diagram of Hybrid Model Learning Dynamics

The hybrid model's superiority arises from three complementary mechanisms:

- Error diversification and bias cancellation – Each base learner exhibits distinct error patterns; by combining them, the ensemble attenuates individual biases and reduces overall variance.
- Hierarchical learning – The meta-learner models second-order relationships between base predictions and true outcomes, enabling systematic correction of under- or overestimation that single models cannot resolve.
- Dynamic weighting – The Random Forest meta-model automatically adjusts the relative influence of each base learner depending on the local operating regime. For example, XGBoost may dominate in highly non-linear mid-life regions, while Random Forest and SVR contribute more in stable phases.

(Figure 6) hybrid model predictions illustrates the impact of this design: the predicted curve almost completely overlaps the actual SoH/RUL trajectory, and residuals cluster tightly around zero. This behaviour confirms that stacking not only improves pointwise accuracy but also preserves temporal coherence, which is essential for online health monitoring.

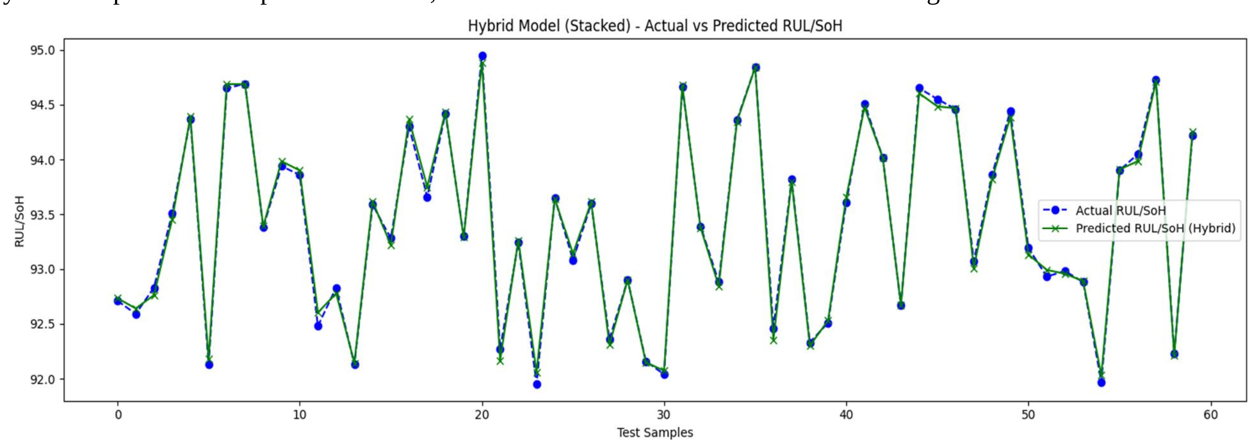


Figure 6: Hybrid Model (Stacked) - Actual vs Predicted Remaining Useful Life (RUL/SoH)

V. DISCUSSION AND IMPLICATIONS

Overall, the results show that ensemble methods, particularly the proposed stacked hybrid model, offer clear advantages over single-model approaches for RUL/SoH estimation from distributed BMS data [10]. Random Forest emerges as the most reliable base learner due to its noise robustness and strong generalization, while XGBoost contributes powerful gradient-driven refinement of complex patterns [11]. SVR, despite high theoretical appeal, is less suitable for large-scale, non-stationary sensor streams.

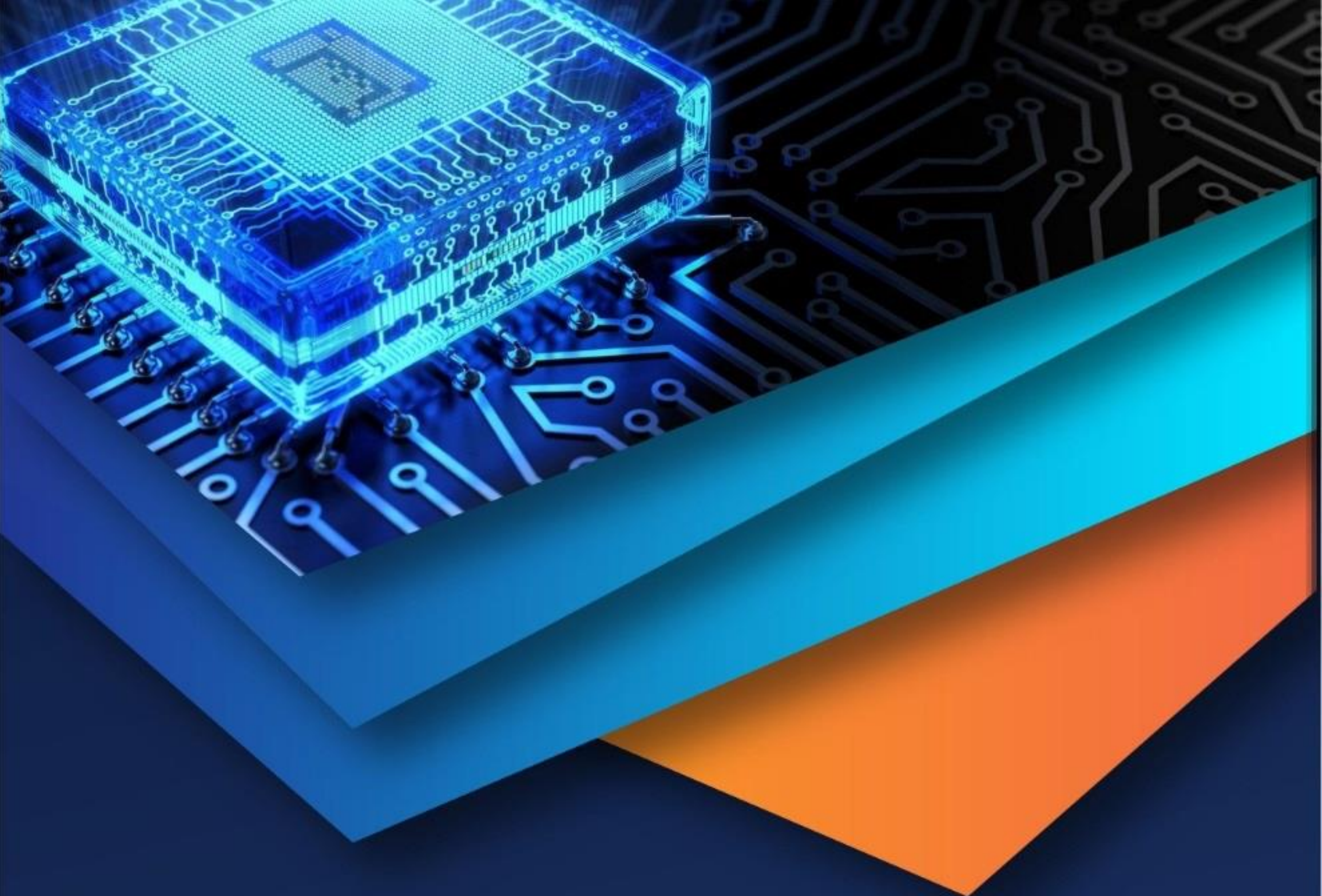
The hybrid stacked model effectively leverages the complementary strengths of all three. Its markedly lower MAE and RMSE, near-Gaussian and unbiased residual distribution, and close visual alignment with true SoH curves indicate that it can serve as a dependable decision-support component in real-time battery health management. Accurate and stable RUL/SoH predictions can support earlier fault detection, improved charging strategies and proactive maintenance scheduling in electric vehicles and stationary storage systems. At the same time, the study's limitations—such as dataset coverage and assumptions of largely stationary degradation patterns—suggest directions for future work, including larger multi-environment datasets and online or continual-learning extensions [12]. Nonetheless, within the evaluated setting, the proposed hybrid stacked ensemble provides a strong, practically deployable solution for data-driven battery life prediction.

VI. CONCLUSION

This study demonstrated that ensemble learning techniques, particularly the Hybrid Stacked model, offer significant advantages in predicting the Remaining Useful Life (RUL) and State of Health (SoH) of lithium-ion batteries compared to traditional single-model approaches [13]. The Hybrid Stacked model outperformed the base models—SVR, XGBoost, and Random Forest—in terms of both predictive accuracy and generalization ability, achieving a substantial reduction in error metrics (MAE and RMSE). The model's performance is further supported by its ability to leverage the complementary strengths of the base learners, reducing biases and improving robustness to noisy data. Additionally, the Hybrid model demonstrated practical applicability for real-time BMS deployment, with manageable training and inference times. However, limitations in the dataset, such as its focus on stationary degradation patterns, suggest future avenues for improvement, including the incorporation of more diverse operational conditions and online learning techniques [14]. The findings confirm that Hybrid Stacked models are a powerful and scalable solution for battery health monitoring, supporting predictive maintenance, optimized charging strategies, and enhanced decision-making in EVs and energy storage systems.

REFERENCES

[1] K. Bakirov, "Application of hybrid Model for Data Analysis in Hydroponic System," Technologies , 2025.
[2] A. Wilson, "Rceent Advances inThermal Imaging and its Applications Using Machine Learning: A Review," IEEE Sensors Journal, vol. 23, no. 4, 2023.1.10.
[3] A. R. Moumen, "Adaptive traffic lights based on traffic flow prediction using machine learning models," International Journal of Electrical and Computer Engineering , vol. 13, no. 5, p. 581, 2023.6.23.
[4] L. A. S. C. R. Riley, "Evaluation of clinical prediction models (part 2): how to undertake an external validation study," British medical journal, vol. 384, 2024.1.15.
[5] 4. H. N. K. S. A. D. Aida Brankovic1, "Explainable machine learning for real-time deterioration alert prediction to guide pre-emptive treatment," Scientific Reports , vol. 12, no. 1, 2022.7.11.
[6] S. M. A. S. L. Q. A. D. L. O. B. M. Matthias Heinrich1*, "Detection of cleaning interventions on photovoltaic modules with machine learning," Applied Energy, vol. 263, 2020.2.20.
[7] M. K. b. A. T. B. T. c. M. A. d. Ali Aldrees a, "Evaluation of water quality indexes with novel machine learning and SHapley Additive ExPlanation (SHAP) approaches," Journal of Water Process Engineering , 2024.1.16.
[8] F. H. ∗. Y. S. Mohammed Al Saleem, "Explainable machine learning methods for predicting water treatment plant features under varying weather conditions," Results in Engineering, vol. 21, 2024.3.1.
[9] O. A. Semra Taşabat, "Using Long-Short Term Memory Networks with Genetic Algorithm to PredictEngine Condition," GAZI UNIVERSITY JOURNAL OF SCIENCE, vol. 35, no. 3, p. 1210, 2021.10.6
[10] *. M. M. S. 2. S. B. 2. Nahar F. Alshammari 1, "Comprehensive Analysis of Multi-Objective OptimizationAlgorithms for Sustainable Hybrid Electric VehicleCharging Systems," mathematics , vol. 11, no. 7, p. 175, 2023.4.5.
[11] b. G. C. b. B. S. a. M. G. d. F. M. a. Leandro Masello a, "Using contextual data to predict risky driving events: A novel methodology from explainable artificial intelligence," Accident Analysis and Prevention , vol. 184, p. 106, 2023.2.26.
[12] H. M. A. M. M. S. Shimaa Barakat, "Eco-Efficient Mobility: Comparative Optimization of PV-Wind EV Charging Solutions," The Economist, 2023.12.19.
[13] E. S. C. T. G. Thodoris Garefalakis1 , "Predicting risky driving behavior with classification algorithms: results from a large-scale field-trial and simulator experiment," European TransportResearch Review, vol. 16, no. 1, 2024.11.21.
[14] F.-L. Luo, Machine Learning for Future WirelessCommunications, wiley, 2019.12.13.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY