Electrochemical Performance Prediction in Lithium-Ion Batteries Using Physics-Informed Features and Ensemble Learning

Samarth P Patel
Material and Process QC Engineer
Eurofins Professional Scientific Services Mid Atlantic, LLC
Samarth.Patel@PSS.EurofinsUS.com

***Abstract*—The difficulty in predicting the electrochemical performance is the complexity of the battery system and the lack of physically relevant information in current models. This paper addresses these shortcomings by introducing an ensemble model that integrates the electrochemical and thermal response to prediction, with the use of Physics-Informed feature engineering. This work leverages a NASA dataset on batteries and advanced feature engineering to model the non-linear battery behavior. R^2, MAE, RMSE, and MAPE are the most used measures to compare ML models including XGBoost, Random Forest, stacking and LightGBM. The stacking model has better results with R^2 = 0.9871 and minimal error values. Time Series Split (k = 5) cross-validation and paired statistical significance analysis were performed to validate the robustness and reliability of the proposed framework. Moreover, SHAP and permutation importance are used to improve model interpretability. The proposed framework contributes to the precision, strength, and dependability of prediction in battery management systems.**

Keywords—Lithium-Ion Batteries, Physics-Informed, Electrochemical Performance Prediction, Ensemble Learning, Battery Management System.

I. INTRODUCTION

Over the last few years, battery technologies have become key components in decarbonizing the world across both stationary and mobile applications [1]. The lithium-ion battery is significant due to its high specific energy, operating temperature range and the possibility of using it in electric cars to reduce the number of CO2 emissions [2]. Battery Management Systems (BMS) are suitable monitoring and control strategies that guarantee safety, reliability, and efficiency in the real-world operating circumstances; this is required since the functions of batteries rely on the complex aspects such as chemistry, temperature, and other operating circumstances [3].

The voltage, current and temperature measurements recorded by the BMS [4] can be used to predict important battery metrics such as state of charge (SOC) and state of health (SOH). In these systems, the safety of operations is encouraged and the lifetime of the batteries is prolonged with the help of such techniques as cell balancing and controlled charging [5]. However, traditional estimating methods often rely on oversimplified models and lack real-world data, which results in lower accuracy under dynamic operating situations. The suggested advances in the sphere of ML, such as SVM, NN, and DL models, have augmented the prediction capability, and the issues of interpretability and generalization remain [6].

To overcome these limitations, Physics-Informed modeling can assist in improving the issues associated with battery models by allowing for an integrated method of analyzing the electrochemical, thermal and transport properties of batteries [7].

It improves the usage of a single-domain model or models that are solely data-driven and allows for a deeper understanding of how batteries behave under various operating situations. The research will create a Physics-Informed framework to predict the electrochemical performance of batteries, ultimately providing a reliable, accurate and practical framework for the management of advanced battery systems. The most significant contributions to this work:

- A Physics-Informed framework is proposed with feature engineering based on electrochemical, thermal and transport phenomena in batteries.
- A stacking ensemble model using the combination of RF and XGBoost with a Linear Regression meta-learner is built to improve the prediction accuracy, robustness and generalization ability.
- SHAP-based explainability and permutation feature importance are integrated, enhancing the model's transparency and reliability by offering interpretable insights into the contribution of the Physics-Informed features.
- A comprehensive experimental study that incorporates a Time Series Split (k = 5) cross validation, multiple performance metrics, statistical significance analysis and computational efficiency evaluation, shows the effectiveness of a proposed BMS framework.

The study is structured as follows: Section II provides an overview of the corpus of current publications. Section III provides a description of the approach. A detailed explanation of the experimental setup and results is given in Section IV; A comparative study, model performance, debate, and limitations are examined in Section V. Section VI concludes the research with recommendations for future scope.

II. LITERATURE REVIEW

Recent studies in battery performance prediction focus on improving accuracy, robustness, and generalization through machine learning and electrochemical model-driven approaches. However, the incorporation of Physics-Informed visions is still not extensive, and there is a necessity to better comprehensive and understandable prediction models.

A year earlier, developments in battery prediction model indicate better accuracy and generalization. Ouyang et al. (2026) suggested an transfer learning model is able to forecast battery knee points with high accuracy with MAPE of 1.12, minimum deviation of 2 cycles and accuracy is improved by 39% compared to current deep learning techniques [8]. The deep forest model of EV battery SOH created by Gao et al. (2026) based on the real driving data demonstrated that the

model obtained the error rate of 0.9% with capacity labeling, user behavior effects, and scalability and generalization [9].

Recent studies in decision-making models of battery investments are done to maximize performance indicators. Kou et al. (2025) proposed essential solid-state battery investment indicators based on an integrated multi-criteria model of entropy game theory, Q-learning, LSO, and MORAN, which could improve the accuracy of decisions in the conditions of uncertainty [10].

Research on thermal analysis and sensors development focuses on monitoring and performance. Yang et al. (2024) demonstrated that Lithium-Ion batteries have a higher resistance and heat production rate than NMC/LFP and thus necessitate a sophisticated thermal control, owing to prevailing polarization heat effects [11].

Jafari, Byun and Ko (2023) proposed combining kNN, RF and XGBoost model to obtain the RMSE of 0.0168 and MAE of 0.0089 when predicting RUL [12]. Sun et al. (2023) electrochemical model that predicts battery SOP by GWO optimization which has a low error (MAPE 4%), and is robust, since it is able to capture the lithium-ion dynamics and thermal dependencies [13].

Research Gaps: Despite recent advancements, the majority of research is based on data-driven or electrochemical, and there is limited integration of Physics-Informed feature engineering to have comprehensive battery prediction. Current models are not very robust when trying to operate under different conditions and also have problems with generalization, dealing with uncertainty as well as real time usage. Therefore, more scalable, reliable and hybrid frameworks are required to make the accurate prediction of battery performance.

III. Methodology

As illustrated in Fig. 1, the proposed pipeline combines features of physics-based engineering and advanced models of ensemble learning to predict electrical performance safely and accurately. Physics-based Feature Engineering is created from three types of data related to electrochemical, thermal, and diffusion properties. By utilizing three forms of data, this method develops a more robust model that provides improved interpretability and generalization to conditions that exist in the real world. This organized approach improves the control of battery management systems (BMS), allows early detection of degradation, and helps optimize the performance of lithium-ion batteries while minimizing risk.

A. Dataset Collection and Data Quality Analysis

The research is carried out on the available dataset from NASA battery dataset [14] comprising of 35317 records and seven different time-series discharge cycle records. A subset of 100 discharge files is selected and merged for analysis. The NASA Battery dataset was selected because it is a representative benchmark dataset for forecasting electrochemical performance and degradation in lithium-ion batteries. It offers a standardized set of voltage, current, temperature, and discharge cycles under controlled operating conditions, which is very suitable for testing Physics-Informed machine learning models. The subset was carefully selected to balance the computation time with the capability to train the model while retaining the typical battery degradation and variation of battery operations. The set of batteries selected may represent typical degradation properties, but using a single benchmark set might restrict the ability of the proposed framework to be generalized to batteries with different chemistries and operating conditions. Therefore, validation on additional public battery datasets and real-world Battery Management System (BMS) data is considered an important direction for future work. The main features of the system are measured voltage, current and temperature, load variables, time and file names, which quantify the electrochemical and thermal properties. The rows represent different time instances in the battery cycle, enabling the investigation of battery degradation and performance over time.

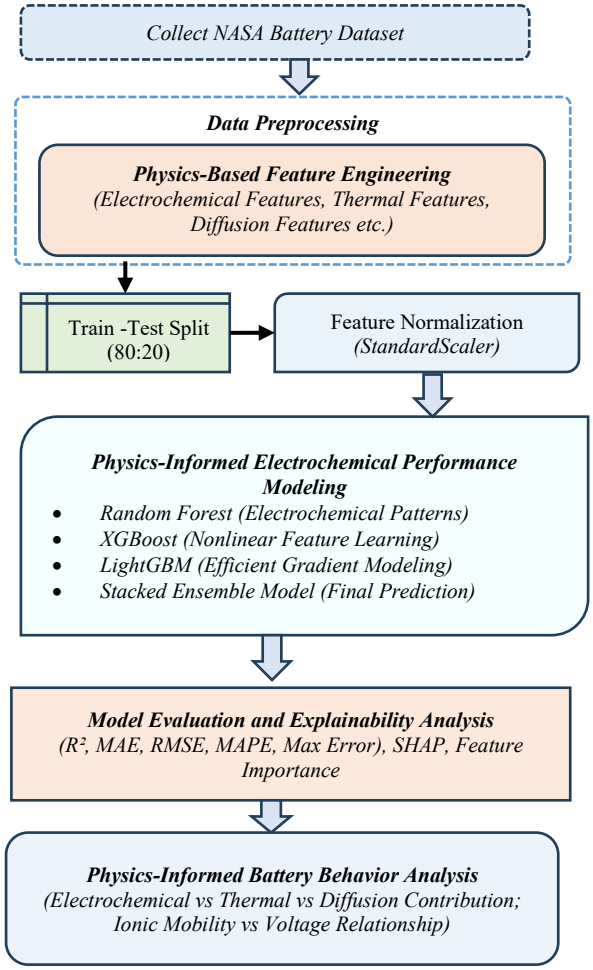

Fig. 1. Proposed Physics-Informed Electrochemical Performance Prediction Framework.

B. Physics-Informed Feature Engineering

A Physics-Informed Feature Engineering method is used to create features that describe the thermal and temporal behaviors of lithium-ion batteries using measured quantities like current, temperature and time. This approach differs from conventional physics-based methods, which use physics-inspired features such as temperature gradient, heat generation, current change, time difference, and Ion Mobility, which capture the physical processes of thermal dynamics and system dynamics. The electrical power ($P = V \times I$) is used to calculate the heat generation, which uses only the measured voltage. The Ion Mobility is calculated from the absolute current divided by temperature to provide non-negative values. The missing data are filled with the previous value. Specifically, all the developed features, mathematical expressions and definitions are listed in Table I.

TABLE I. PHYSICS-BASED ENGINEERED FEATURES AND THEIR PHYSICAL INTERPRETATIONS

Feature	Formula	Physical Meaning
Temperature Gradient	Δ Temperature	Thermal variation over time
Heat Generation	Voltage × Current	Heat produced during operation
Current Change	Δ Current	Change in current flow
Time Difference	Δ Time	Time interval between measurements
Ion Mobility	$\lvert Current \rvert / Temperature$	Ion movement efficiency

C. Feature Analysis and Visualization

To understand the underlying electrochemical and thermal behavior of the battery system, statistical and graphical representations are also carried out on the extracted features. These plots are helpful in the analysis of the pattern of heat generation and in the analysis of the distribution and variation of the internal resistance and voltage with each cycle of discharge.

Fig. 2 shows the results of heat production and voltage during the dataset. The histogram of heat generation indicates that the majority of values were concentrated within the range of 0.009 to 0.010, with the smaller cluster near 0.001, indicating consistent thermal behavior. The specified values of voltage indicate that 59.8% of the entire observations were in the range of 3.343 to 3.79 V, implying that the battery functions at a high stable voltage and the values of lower voltage ranges contribute insignificantly to the overall distribution of the values.

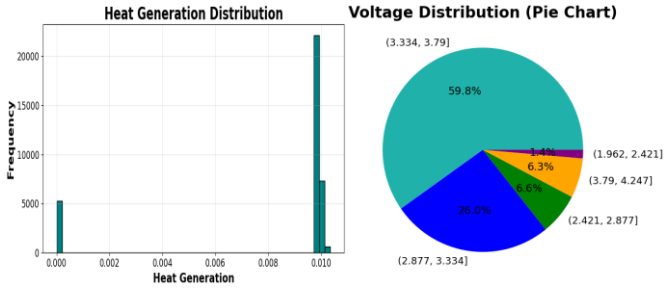

Fig. 2. Voltage and Heat generation analysis.

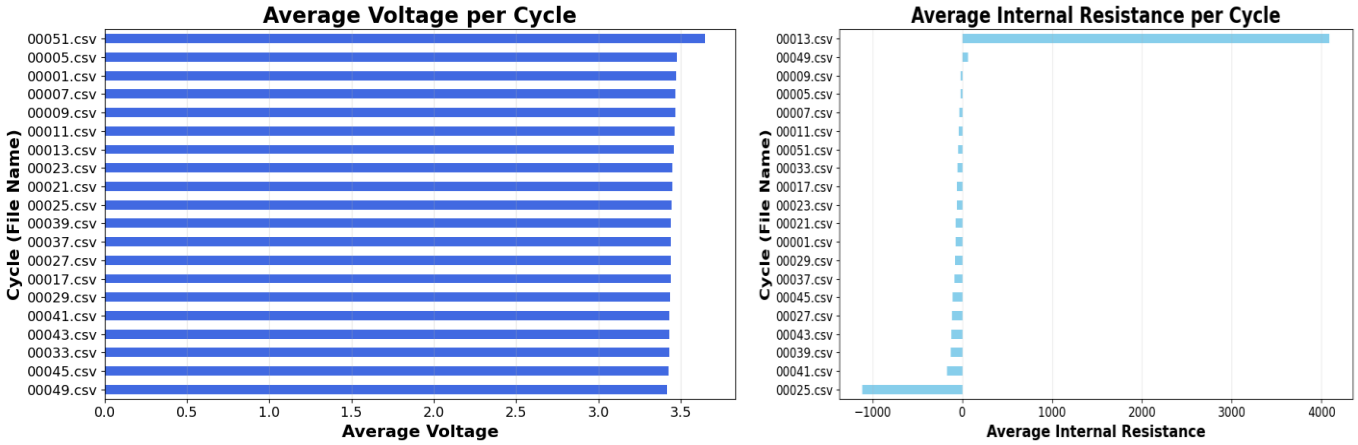

Fig. 3. Average Voltage and Internal Resistance per Battery Cycle.

Fig. 3 indicates that the battery voltage does not vary during the testing cycles but is recorded within the range of 3.2 and 3.6 V, which implies that the battery's operational performance is constant. The internal resistance exhibits significant variability since the majority of readings come close to zero and a few are above 4000 indicating possible battery malfunction or abnormal operation. The internal pattern of resistance is important in determining the thermal performance and the overall performance of a system and the pattern of internal resistance is a critical determinant in evaluation of the battery performance and durability.

D. Data Splitting

Divide the data into 80:20% training and testing of the learning process and the model. They are both used to test and train the model. Table II shows the sample and feature dimensions.

TABLE II. TRAIN-TEST DATASET DISTRIBUTION AND DIMENSIONAL ANALYSIS

Dataset	Dimensions	No. of Samples	No. of Features	Purpose
X_train	(28253, 9)	28,253	9	Model training (input features)
X_test	(7064, 9)	7,064	9	Model validation (unseen data)
y_train	(28253,)	28,253	—	Training target (Voltage)
y_test	(7064,)	7,064	—	Testing target (Voltage)

E. Feature Normalization Using StandardScaler

The method of normalizing input features is referred to as the StandardScaler procedure and it converts the input features into unit variance and zero mean [15]. To avoid data leaking and to offer the same transformation, the scaler is first applied to the TrainingData and then to the training and testing data. The StandardScaler formula is as shown (Equation 1).

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

In this case, μ represents the mean of the feature, σ the StandardDeviation of a feature's values, and z the standardized (scaled) output for the original input value x.

F. Proposed Models

This section covers ensemble machine learning models: The chosen models, RF, XGBoost, LightGBM, and stacking, that use Physics-Informed features to predict the battery's electrochemical performance. Each model is assessed by Time Series Split **(k=5)** to maintain the time order of battery discharge data. This ensures that prior knowledge is only used for training, and that no information leaks from the future to the past, enabling a more proper assessment of the prediction performance.

1) Random Forest (RF) Model

This model is a process called ensemble learning where it creates a number of decision trees and uses these to vote on the results for improved accuracy, as well as reducing overfitting [16].

The final predicted value is obtained by averaging the per-tree output, according to (Equation 2).

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \quad (2)$$

Where $\boldsymbol{N}$ is the total numberof trees and $\boldsymbol{T_i(x)}$ represents the prediction of the $\boldsymbol{i^{th}}$ tree.

In this study, the implementation with trees 50, a max depth of 10, and min split size of 5 to ensure efficient and robust prediction of battery performance.

2) Extreme Gradient Boosting (XGBoost) Model

XGBoost is a gradient-based boosting technique that has the capability of dealing with complex interactions and nonlinear relations [17]. Equation (3) defines the objective function of XGBoost, which enhances model generalization by combining the prediction loss with a regularization term.

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (3)$$

Where $l(y_i, \hat{y}_i)$ denotes the lossfunction between the actual value y_i and forecasted value $\hat{y}_i$, while $\Omega(f_k)$ represents the regularization term applied to the kth decision tree.

The study uses XGBoost with 50 trees and learning rate 0.1 together with maximum depth 5 and regularization parameter λ=1 to achieve efficient learning of complex nonlinear battery behavior.

3) Light Gradient Boosting Machine (LightGBM) Model

It is an efficient Gradient Boosting framework that creates decision trees through its leaf-wise growth method to deliver better training speed and increased accuracy, making it suitable for processing extensive datasets with many dimensions. The split gain used for selecting the optimal leaf node is calculated using Equation (4).

$$Gain = \frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda} \quad (4)$$

Where G and H denote the first-order and second-order gradients, respectively, and λ is the regularization parameter.

In this study, LightGBM is implemented with the number of trees 100, the Learning Rate 0.1, and the maximum tree depth 5 and regularization parameter λ=1 to ensure fast computation.

4) Stacking Ensemble Model

Stacking combines various ML algorithms to create a prediction system that performs better than any single algorithm. In stacking, a different model known as a meta learner receives the outputs of the separate base models and learns how to integrate each base model's predictions into a final prediction. The final prediction of the stacking ensemble is computed using Equation (5).

$$\hat{y} = \beta_0 + \beta_1 \hat{y}RF + \beta_2 \hat{y}XGB \quad (5)$$

Where $\hat{y}RF$ and $\hat{y}XGB$ are the predictions generated by the Random Forest and XGBoost models, respectively, while $\beta_0, \beta_1, \beta_2$ are the regression coefficients learned by the Linear Regression meta-learner.

G. Hyperparameter Tuning Strategy

Table III summarizes the optimized hyperparameter settings of the proposed ensemble models, selected through empirical tuning using Time Series Split (k = 5) cross-validation. The selected configurations provide an efficient trade-off among Computational Efficiency, model generalization, and prediction accuracy for electrochemical performance prediction.

TABLE III. HYPERPARAMETER CONFIGURATION OF THE PROPOSED MODELS

Model	Hyperparameter	Selected Value
Random Forest	Number of Trees (n_estimators)	50
	Maximum Depth	10
	Minimum Samples Split	5
XGBoost	Number of Trees (n_estimators)	50
	Learning Rate	0.1
	Maximum Depth	5
	Regularization Parameter (λ)	1
LightGBM	Number of Trees (n_estimators)	100
	Learning Rate	0.1
	Maximum Depth	5
	Regularization Parameter (λ)	1
Stacking	Base Learners	Random Forest + XGBoost
	Meta Learner	Linear Regression

In this study, XGBoost and RF as the base models and Linear Regression as the meta-learner; this configuration takes advantage of the strengths associated with each individual model, which build upon one another's ability to generalize by creating a more reliable and accurate prediction of electrochemical performance through the stacking of multiple algorithms. The proposed stacking framework will be explained in detail step-by-step in Algorithm 1.

Algorithm 1: Proposed Stacking Prediction Model

Input: Dataset D (Physics-Informed features), Target y
Output: Predicted electrochemical performance ŷ

Step 1: Data Preparation
Create a training set (X_train, y_train) and a testing set (X_test, y_test) from dataset D.
Step 2: Feature Scaling
Fit StandardScaler on X_train
Transform X_train and X_test
Step 3: Train Base Models
Train Random Forest model (RF) on X_train
Train XGBoost model (XGB) on X_train
Train LightGBM model (LGBM) on X_train

Step 4: Generate Base Predictions
Predict outputs using RF → y_RF
Predict outputs using XGB → y_XGB

Step 5: Stacking (Meta-Learning)
Combine base predictions [y_RF, y_XGB]
Train Linear Regression model on combined predictions

Step 6: Final Prediction
Use trained stacking model to predict ŷ on X_test

Step 7: Model Evaluation
Compute performance metrics:
R^2, MAE, RMSE, MAPE, Explained Variance, Max Error

Return ŷ and evaluation results

H. Performance Evaluation Metrics

Model analysis is performed by comparing the estimated values with real values using various statistical indicators such as R^2, MAE, MSE, RMSE, MAPE, Explained Variance, and Maximum Error. These measures provide a comprehensive assessment of the quality of prediction, the size of the error, and the reliability of the model for Equations (6)-(12). In the equations below, $\hat{y}_i$ is a forecasted value,y_i is a real value, n is the number of observations and $\bar{y}$ is the mean of the real values [18].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (7)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i \quad (8)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (10)$$

$$Explained\ Variance = 1\ \frac{Var\ (y - \hat{y})}{Var\ (y)} \quad (11)$$

$$Max\ Error = \max(|y_i - \hat{y}_i|) \quad (12)$$

The purpose of these assessment metrics is to evaluate the models' performance and identify the one that best predicts electrochemical performance with the least amount of error and the greatest explanation for variation.

IV. Result Analysis and Discussion

The performance prediction system for electrochemical reactions was built using the Scikit-learn, XGBoost, LightGBM, Pandas, NumPy, Matplotlib, and Seaborn libraries in the GoogleColab environment. The comparison of RF, XGBoost, LightGBM and stacking ensemble models with three evaluation approaches are based on prediction accuracy and error measures and computational efficiency test.

A. Results of Proposed Models

The improved performance of the ensemble models is as shown in Table IV where the model evaluation is based on several statistical metrics such as R^2, MAE, MSE, RMSE, MAPE, explained variance, maximum error and time to train to ensure a comprehensive evaluation. The stacking model has the highest R^2 =0.9871 and the lowest error values (MAE = 0.02270, RMSE =0.03716), which means that it has improved predictive performance. The improved performance is due to the optimal representation of features and ensemble learning, which is confirmed by cross-validation. LightGBM is also a strong model with low computational cost thus efficient for large-scale application. On the other hand, in contrast, the error variations of XGBoost and Random Forest are higher, thus lower prediction accuracy. The results show how well the suggested stacking model predicts intricate electrochemical processes.

TABLE IV. Performance Evaluation of Ensemble Models for Electrochemical Prediction

Model	CV	R^2	MAE	MSE	RMSE	MAPE	Explained Variance	Max Error	Training Time (s)
Stacking	0.9866	0.9871	0.02270	0.00138	0.03716	0.00730	0.9871	0.3213	35.26
LightGBM	0.9850	0.9855	0.02457	0.00155	0.03941	0.00788	0.9855	0.3348	0.22
Random Forest	0.9826	0.9829	0.02453	0.00183	0.04281	0.00795	0.9829	0.3542	8.58
XGBoost	0.9776	0.9785	0.03071	0.00231	0.04809	0.00981	0.9785	0.3841	0.14

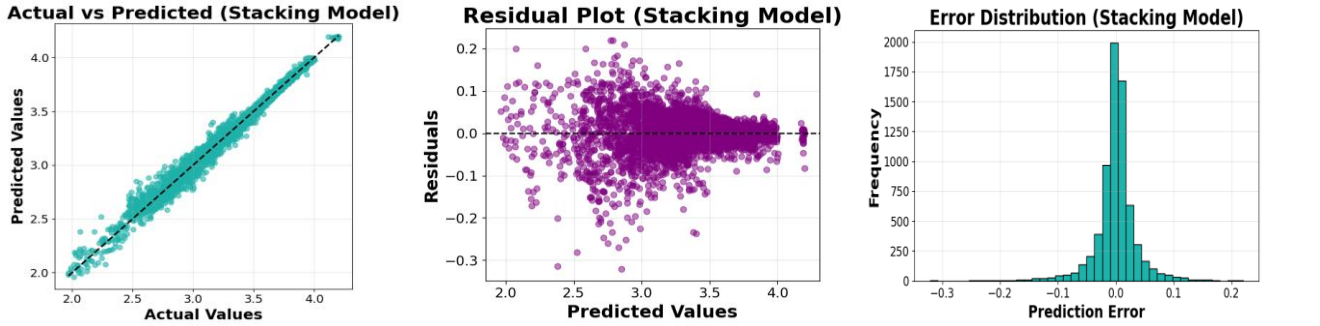

Fig. 4. Stacking Model Validation Using Residual and Error Distribution Analysis.

Fig. 4 demonstrates the model validation of the stacking model using actual vs predicted plot, residual plot and error distribution. The actual vs predicted plot is linear and shows accurate predictions across the entire voltage range (2.0 to 4.0). The residual plot reveals that these errors are very close to zero and have low variance, signifying low bias and good generalization. Finally, the distribution of errors is highly concentrated at zero, showing that prediction errors are very small and the model makes stable and accurate predictions of electrochemical performance.

B. Computational Complexity and Deployment Feasibility

Table IV shows that the suggested ensemble models find a good middle ground between being computationally efficient and having accurate predictions. While the stacking model does incur the most computational cost during offline training, it takes the longest time to train (35.26 s) due to combining numerous BaseLearners with a meta-learner. The trained model is well-suited for real-time BMS applications since it conducts inference during deployment with minimum computational overhead. The computational efficiency of the suggested framework for practical battery monitoring and predictive maintenance is further shown using XGBoost and LightGBM, two algorithms with very rapid training times (0.22 s and 0.14 s, respectively).

C. Statistical Significance Analysis

The cross-validation performance scores of the suggested stacking model and the individual ensemble models were used in a paired t-test to determine if the performance gains

of the suggested stacking model are statistically significant. The test evaluates whether the observed improvements are statistically significant rather than resulting from random variation. As shown in **Table V**, all p-values are below the significance level of 0.05, indicating that the proposed stacking model significantly outperforms the individual ensemble models.

TABLE V. STATISTICAL SIGNIFICANCE ANALYSIS OF THE PROPOSED STACKING MODEL

Comparison	Statistical Test	p-value	Interpretation
Stacking vs Random Forest	Paired t-test	< 0.001	Statistically Significant
Stacking vs XGBoost	Paired t-test	< 0.001	Statistically Significant
Stacking vs LightGBM	Paired t-test	0.013	Statistically Significant

D. SHAP Explainability

A method for model interpretation called SHAP uses cooperative game theory to explain the contribution of the model to its prediction. It assigns each feature a SHAP score, which represents the feature's contribution to the expected outcome, relative to the average prediction. SHAP is globally and locally interpretable, which is valuable when dealing with complex models. As illustrated in Fig. 5, the SHAP values show that the top three features are Temperature_measured (~0.16), Time (~0.11) and Current_measured (~0.07). Features like time_diff, Current_load, and ion_mobility have moderate impact, and temp_gradient, heat_generation and current_change have relatively low contributions (< 0.03). It is noteworthy that raw voltage channel inputs (such as Voltage_measured and Voltage_load) are not part of the feature set, because the prediction target is the voltage. Derived features, like generation (P = V × I), that express the physical power behavior and not the direct voltage readout, are kept as they contain information on the electrochemical processes and not the target signal.

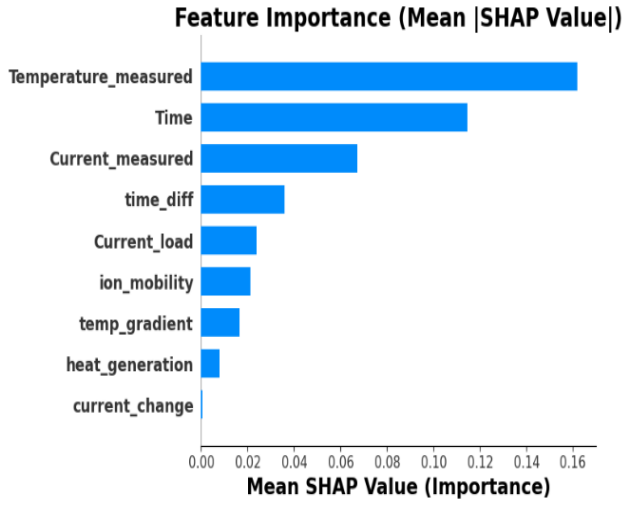

Fig. 5. SHAP-Based Feature Importance Analysis.

E. Physics-Informed and Behavior Analysis

In this section, the dependence of ion mobility on voltage is studied to understand battery performance. Fig. 6 presents a high positive correlation and the high-density region is located around the range of Ion Mobility **0.10-0.15** and voltage **3.2-3.7 V**, indicating good performance of battery and high Ion Mobility.

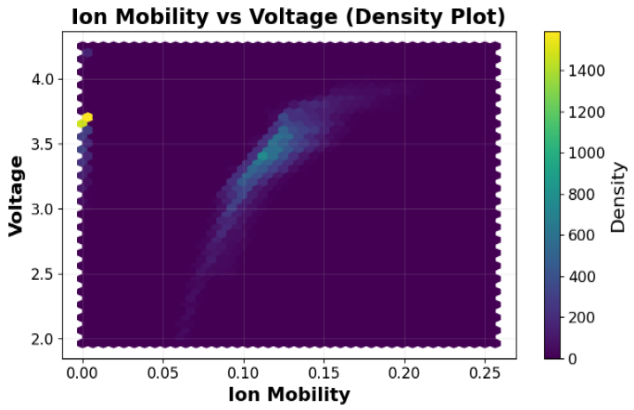

Fig. 6. Ion Mobility and Voltage Density Analysis.

V. COMPARATIVE ANALYSIS AND DISCUSSION

The performance measures, R^2, MAE, MSE, and RMSE, have been used to perform a Comparative Analysis of the existing models and the proposed ensemble models (LightGBM, XGBoost, Random Forest, and Stacking) as shown in Table VI. The results indicate that the proposed models are superior to the conventional ones and the stacking model is having the highest accuracy (R^2 = 0.9871), implying that it has good prediction power and better generalization ability than the individual models. This underscores the success of the Physics-Informed feature engineering-based ensemble method for electrochemical performance prediction.

Baseline models MLP [19] and SVR [20] are compared to make sure that the comparison is scientifically valid and that both are compared on the same data using the same preprocessing and validation. The MLP uses hidden layers (100, 50) with 500 iterations, while SVR employs an RBF kernel (C = 10, ε = 0.1). This ensures a fair and robust comparison of their performances.

TABLE VI. COMPARATIVE ANALYSIS OF PROPOSED AND EXISTING MODELS FOR ELECTROCHEMICAL PERFORMANCE PREDICTION

Ref	Models	R^2	MAE	MSE	RMSE
Baseline	MLP	0.9556	0.0512	0.0047	0.0689
Baseline	SVR	0.9530	0.0519	0.0050	0.0709
Proposed	Stacking	0.9871	0.02270	0.00138	0.03716
Proposed	LightGBM	0.9855	0.02457	0.00155	0.03941
Proposed	Random Forest	0.9829	0.02453	0.00183	0.04281
Proposed	XGBoost	0.9785	0.03071	0.00231	0.04809

A. Discussion

This ensemble framework proposed shows high predictive accuracy in the behavior of electrochemical battery with the stacking model showing the high accuracy (R2 = 0.9871, RMSE = 0.03716). Combining electrochemical, thermal and diffusion properties greatly improve the accuracy of the models and their interpretability. The stacking model outperforms individual models by fusing the good attributes of the models and getting them to work together to achieve superior generalization and robustness results. Moreover, the results of paired statistical significance analysis showed that the performance improvements obtained by the proposed stacking model from the individual ensemble models were Statistically Significant ($p < 0.05$), which further proved the effectiveness of the proposed framework and eliminated the possibility of the improvement being caused by Random Variation. Additionally, the methods of explainability like

permutation importance and SHAP give valuable information on the feature contribution, making the proposed framework more transparent and trustworthy.

The proposed study has, however, a number of limitations. The dataset used by NASA is an accepted standard for the study of lithium-ion batteries, but the experimental study is restricted to one public dataset that contains 100 representative discharge files. Therefore, the future potential of the proposed framework in generalizing to different battery chemistries, manufacturers, battery sizes and operating conditions has not yet been fully validated. The proposed approach would be more robust and applicable in practice if validated externally with other benchmark datasets, such as the CALCE Battery Dataset, the Oxford Battery Degradation Dataset or real-world data from Battery Management Systems (BMSs). The models rely on historical data and may not fully represent operational changes under a range of actual operational conditions. The future research agenda will focus on improving the accuracy of the predictions, scaling the framework, and real-time deployment, including the development of computationally efficient models that are hybrid in nature, testing the suggested framework on different battery datasets, integration of real-time BMS data, and exploring the transfer learning concept for different battery chemistries.

VI. Conclusion and Future Work

A physics-aware feature engineering-based ensemble learning model for lithium-ion battery voltage prediction is presented in this paper. The approach proves to be reliable and accurate in predicting by utilizing operational parameters including current, temperature, and time, and derived thermal and dynamic characteristics. Ensemble models, especially stacking, increase the generalization and strength of models. In order to provide scientific validity, it was necessary to exclude features based on the target variable and also use k-fold cross-validation to conduct strong evaluation. Furthermore, statistical significance analysis confirmed the robustness and reliability of the proposed framework. The findings affirm that the proposed framework has a viable and understandable answer to battery performance prediction. Future work will focus on validating the proposed framework using multiple publicly available battery datasets and real-world Battery Management System (BMS) data to further improve its generalization, robustness, and practical applicability. The incorporation of real-time data capture and the use of IoT-based monitoring systems can be used in the practical implementation of BMS. Additional enhancements can involve the use of further developed DL models to boost the accuracy of prediction and scalability.

References

[1] S. Jafari and Y.-C. Byun, "Prediction of the Battery State Using the Digital Twin Framework Based on the Battery Management System," *IEEE Access*, vol. 10, no. November, pp. 124685–124696, 2022, doi: 10.1109/ACCESS.2022.3225093.

[2] B. Senapati *et al.*, "Quantum Computing and Its Potential Disruption to Data Centers and Edge Computing in Battery Cell Manufacturing Sites," in *2025 IEEE International Conference on Electro Information Technology (eIT)*, 2025, pp. 126–131.

[3] A. Mittal, I. Kumar, and S. Singh, "Physics-Grounded Multi-Task Machine Learning for Photovoltaic Power Forecasting and Solar-Panel Health Monitoring," in *2026 International Conference on Electronics and Renewable Systems (ICEARS)*, Feb. 2026, pp. 1074–1079. doi: 10.1109/ICEARS67481.2026.11416796.

[4] J. F. Plumeyer, F. Moesle, S. Wolf, H. Born, H. H. Heimes, and A. Kampker, "Optimisation of Solid-State Batteries: A Modelling Approach to Battery Design," *Batteries*, vol. 11, no. 4, p. 153, Apr. 2025, doi: 10.3390/batteries11040153.

[5] S. A. Pushkala, "Explainable AI for Trustworthy Predictions: The Comparative Analysis of SHAP, LIME, and Counterfactuals," in *2025 2nd International Conference on Recent Trends in Electrical, Electronics and Computing Technologies (ICRTEECT)*, IEEE, Oct. 2025, pp. 1–5. doi: 10.1109/ICRTEECT67512.2025.11448658.

[6] M. R. B. Domalanta and J. A. D. R. Paraggua, "A Multiphysics Model Simulating the Electrochemical, Thermal, and Thermal Runaway Behaviors of Lithium Polymer Battery," *Energies*, vol. 16, no. 6, 2023, doi: 10.3390/en16062642.

[7] M. R. B. Domalanta, M. T. Castro, J. D. Ocon, and J. A. D. del Rosario, "An Electrochemical-Thermal Multiphysics Model for Lithium Polymer Battery," *Chem. Eng. Trans.*, vol. 94, no. April, pp. 145–150, 2022, doi: 10.3303/CET2294024.

[8] T. Ouyang, Y. Su, W. Peng, and L. Su, "A Hybrid Algorithm for Direct Knee Point Prediction of Lithium-Ion Battery in Electric Vehicle," *IEEE Trans. Power Electron.*, vol. 41, no. 3, pp. 3440–3451, Mar. 2026, doi: 10.1109/TPEL.2025.3616138.

[9] Y. Gao, D. Chu, Q. Chen, and L. Lu, "Predicting Battery State-of-Health in Electric Vehicles: A Deep Forest Approach Using Real-Driving Data," *IEEE Trans. Transp. Electrif.*, vol. 12, no. 1, pp. 998–1012, Feb. 2026, doi: 10.1109/TTE.2025.3624578.

[10] G. Kou, H. Dınçer, S. Yüksel, E. Ergün, and S. Eti, "A Novel Smart Molecular Fuzzy Decision Support System for Solid-State Battery Investments in Grid-Level Renewable Energy Storage," *IEEE Access*, vol. 13, pp. 121815–121837, 2025, doi: 10.1109/ACCESS.2025.3588345.

[11] R. Yang *et al.*, "Comparative Study on the Thermal Characteristics of Solid-State Lithium-Ion Batteries," *IEEE Trans. Transp. Electrif.*, vol. 10, no. 1, pp. 1541–1557, Mar. 2024, doi: 10.1109/TTE.2023.3289997.

[12] S. Jafari, Y.-C. Byun, and S. Ko, "A Novel Approach for Predicting Remaining Useful Life and Capacity Fade in Lithium-Ion Batteries Using Hybrid Machine Learning," *IEEE Access*, vol. 11, pp. 131950–131963, 2023, doi: 10.1109/ACCESS.2023.3329508.

[13] X. Sun *et al.*, "State of Power Capability Prediction of Lithium-Ion Battery From the Perspective of Electrochemical Mechanisms Considering Temperature Effect," *IEEE Trans. Transp. Electrif.*, vol. 9, no. 2, pp. 2453–2463, Jun. 2023, doi: 10.1109/TTE.2022.3206452.

[14] A. Pat, "NASA Battery Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/patrickfleith/nasa-battery-dataset/data

[15] J. M. H. Pinheiro *et al.*, "The Impact of Feature Scaling in Machine Learning: Effects on Regression and Classification Tasks," *IEEE Access*, vol. 13, pp. 199903–199931, 2025, doi: 10.1109/ACCESS.2025.3635541.

[16] X. Yin, J. Ma, S. Zhang, and G. Cui, "Machine learning drives a new paradigm in inorganic solid-state electrolytes research," *Comput. Mater. Today*, vol. 10, p. 100051, Jun. 2026, doi: 10.1016/j.commt.2026.100051.

[17] I. Alhasan and M. H. S. Alrashdan, "XGBoost-Powered Predictive Analytics for Early Identification of Thermal Runaway in Lithium-Ion Batteries," *World Electr. Veh. J.*, vol. 17, no. 2, p. 68, Jan. 2026, doi: 10.3390/wevj17020068.

[18] M. Iftikhar *et al.*, "A deep learning approach to optimize remaining useful life prediction for Li-ion batteries," *Sci. Rep.*, vol. 14, no. 1, p. 25838, Oct. 2024, doi: 10.1038/s41598-024-77427-1.

[19] V. Safavi, A. Mohammadi Vaniar, N. Bazmohammadi, J. C. Vasquez, and J. M. Guerrero, "Battery Remaining Useful Life Prediction Using Machine Learning Models: A Comparative Study," *Information*, vol. 15, no. 3, 2024, doi: 10.3390/info15030124.

[20] M. F. H. Shiblee and H. Laaksonen, "Battery-Insight-PSO: A machine learning model for accurate prediction of state of health and remaining useful life in lithium-ion batteries," *Futur. Batter.*, vol. 8, p. 100114, Dec. 2025, doi: 10.1016/j.fub.2025.100114.