INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Enhanced Prognostics for Lithium-Ion Batteries in eVTOL Aircraft Using CatBoost and LightGBM Algorithms

Srinivas Mallimoggala[1]*, **K Rama Devi**[2]

1 Research Scholar, ECE Department, University College of Engineering, JNTUK Kakinada, Andhra Pradesh, India
2 Assistant Professor, ECE Department, University College of Engineering, JNTUK Kakinada, Andhra Pradesh, India

OPEN ACCESS

Received: 30-05-2024
Accepted: 04-01-2025
Published: 31-01-2025

Citation: Mallimoggala S, Devi KR (2025) Enhanced Prognostics for Lithium-Ion Batteries in eVTOL Aircraft Using CatBoost and LightGBM Algorithms. Indian Journal of Science and Technology 18(2): 147-159. https://doi.org/10.17485/IJST/v18i2.1804

* **Corresponding author**.

mallimoggla@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: The goals of this study are as follows: The comprehension of advanced and efficient E-VTOL techniques is required to control batteries better as well as to increase their performance, safety, and lifespan. The data-driven machine learning models using CatBoost and LightGBM are developed to predict the battery charge, health, and RUL based on flight conditions. This paper aims to assess the impacts that charging/discharging rates as well as temperature have on batteries and determine characteristics that influence discharging. **Methods:** In the framework that has been proposed, the models used for the analysis and prediction of battery charge, health, and RUL include CatBoost and LightGBM. These algorithms define charging and discharging profiles and the effects of temperature on the battery. Different comparative studies on machine learning algorithms are performed to find an optimal solution for battery prediction on E-VTOL vehicles. It is analyzed in real-time using flight profiles to predict the takeoff, landing, and cruising phases of the flight. **Findings:** The suggested framework effectively estimates the battery charge, health, and RUL, outcompeting all previous algorithms. CatBoost is known for its ability to handle categorical data and its efficient training process, making it suitable for real-time applications. LightGBM is recognized for its speed and efficiency, particularly in handling large datasets, and its robustness in making accurate predictions. **Novelty:** Drawing from these methods, this has proposed a new battery control in E-VTOL vehicles using a machine learning framework that incorporates handling categorical features, robustness to overfitting, automatic feature scaling, built-in cross-validation, and robustness to noisy data.

Keywords: CatBoost; Electric Vertical Take Off and Landing; Gradient Boosting Algorithm; LightGBM; RUL; SOC; Battery Lifespan; Battery Management; SOH

1 Introduction

E_VTOL is something that reacts to the issue of future traffic patterns and pollution, primarily given the use of fossil fuels. Some of these advanced technologies are still emerging, but main industries and companies are already working on and experimenting with them. Batt_Mgmt is relevant in E_VTOLs that require immense Batt_Pwr; especially during Mis_Pfls that include take-off and landing. They induce further computation and decision-making requirements in the battery management systems which are instrumental in the safety, efficiency, and reliability of the E_VTOL. Many papers have been written concerning the applicability of machine learning algorithms (MLA) for anticipation of different characteristics of batteries.

In detail, Reference [1] envisaged the method of using MLAs for the prediction of Batt_EoL through the integration of attention mechanisms and transfer learning to ensure early-stage degradation detection. In [2], the eXtreme Gradient Boosting algorithm (Alg_XGB) was used to predict the accurate Pred_RUL of the NMC-LCO battery for the improvement of SOH and Batt_Mgmt. Likewise, [3] proposed an efficient solution for predicting the battery life (Pred_SOH) of lithium batteries (Li_Batt) at Batt_EoL using the GPNAR models. Reference [4] proposed a method for Pred_SOH of Li_Batt using relaxation voltage that has promising implications for online monitoring.

From the studies, it is noticed that MLAs are beneficial in predicting the several parameters of battery as well as enhancing battery management systems. Despite the advancements, several gaps remain in the literature: Despite the advancements, several gaps remain in the literature, Comparative Analysis: Conducting a comparison of various MLAs is missing in the literature, particularly in the context of battery prediction. The authors of papers tend to investigate separate reordering techniques and do not pay attention to the fact that such algorithms should be compared in similar circumstances. E_VTOL Specific Studies: Most prior investigations center on electric vehicles (E_Veh) and portable electronics, while few works investigate the issues related to E_VTOL batteries, which, nonetheless, possess higher power density and different application profiles. Real-time Implementation: There are limited numbers of research that focus on the actual time utilizations of these prediction models, yet this is extremely important for understanding real-life Batt_Mgmt in E_VTOLs. Integration with Flight Profiles: Research integrating battery prediction models with specific flight profiles (take-off, landing, cruising) is sparse, limiting the applicability of these models for E_VTOLs.

To address these gaps, the following steps are necessary, Comprehensive Comparative Analysis: Try to compare the same MLAs such as Alg_CB, Alg_XGB, Alg_LGBM, and many others, which would use the same input and output to know the best algorithm for battery prediction. Focus on E_VTOL Batteries: Design and create forecast models designed for the battery use in E_VTOL, since their need for the power is different and they can be charged and discharged differently. This also entails incorporating MLAs with flight profiles in a way that perfect predictions may be achieved particularly in the take-off, landing, and flee stages. Real-time Implementation and Validation: Launch applications of the most promising MLAs in real-time Batt_Mgmt systems and examine their effectiveness in real-life scenarios. This implies coming up with algorithms that could analyze the data and make the prediction in real-time. Holistic Approach: Hence, engage a multi-source approach at the feature level by fusing various important data features namely Batt_Curr, Batt_Vol, Batt_Chg, Batt_DsChg and Batt_Temp then using Complicated Multivariate Linear Algorithms to enhance the accuracy of the battery load forecasting models. For this model, define Batt_EoL as the state at which the battery is at 85% charge. Include the whole battery life cycle from the beginning of its use up to Batt_EoL.

Thus, this paper presents a gradient-boosting approach to obtaining achievable Pred_RUL, Pred_SOC, and Pred_SOH of E_VTOL batteries, mainly relying on CatBoost (Alg_CB) and LightGBM (Alg_LGBM). The rationale for the selection of such algorithms lies in their applicability to big data and the extremely high accuracy as demonstrated in past research.

Training Other Algorithms [5,6] can be onerous especially for deep architectures because of such complications as vanishing gradient problems, exploding gradients, and many others, and because many hyperparameters often must be adjusted. Other Algorithms can detrimentally overfit since when the model is large compared to the training data set then it will. Like other neural network structures, they are always intertwined with black-box phenomena that complicate the analysis of the obtained representations.

Here's how Proposed Algorithms overcome some of the limitations, Instead, LightGBM is a distributed and highly efficient model. For the splitting of the feature values, it makes use of a histogram which helps to save memory and enhance the training process. It grows trees leaf-wise and not level-wise, which results in possibly deeper trees and better performances with fewer nodes. This is because; it is optimized for distributed computation and has lower memory requirements than a traditional database management system. offers pre-embedded regularization tools that help in tackling the problem of overfitting, including L1 and L2. To solve such problems, Nestler has additional options for parallel and GPU training, which enhances its performance and capacity.

CatBoost integrates feature preprocessing specific to categorical variables – one does not have to one-hot encode the data beforehand. As for the feature transformation, it adopted target encoding and ordered boosting to address the problem of many categorical features. It also comes equipped with several geometrical ensembles such as random permutations of the rules, ordered boosting operations and learning rate annealing to avoid overfitting. eliminates the need to manually scale features as a preprocessing step which makes it an advantage. Cross-validation is one of the admirable features of Scikit-learn which helps in adjusting hyperparameters and testing models. It hardly affects the noisy data or the presence of outliers, due to the robustness of most of its loss functions and regularization [7–9].

All in all, LightGBM and CatBoost are superior in terms of performance, general practicality, and stability regarding the traditional methods. However, like any other algorithms, they also come with certain limitations and factors that should be adjusted to achieve the best results in certain tasks.

This framework intends to predict the RUL, SOC, and SOH by using Batt_Curr and Batt_Vol plots over time with respect to Batt_Chg, Batt_DsChg, and Batt_Temp. The reason for taking Batt_EoL at 85% is because; E_VTOLs demand sincere and precise power needs now and then to guarantee petroleum and electrical safety in flight operations. The outcomes reveal that the proposed Alg_CB is effective in predicting RUL, SOC, and SOH of the batteries and performs better than the Alg_LGBM and other existing algorithms.

2 Methodology

This study builds upon previous research [10] and datasets [11] to develop and evaluate MLA for battery performance, RUL, SOC, and SOH. The foundation of this research is reference [10], which serves as the base paper and primary motivation. Implemented and evaluated using several MLA, including Alg_SVM, Alg_RFR, Alg_XGB, Alg_GPR, and Alg_MLP. The study concluded that Alg_XGB provided the best results for Pred_RUL, while Alg_RFR was most effective for Pred_SOH [10]. These findings guide our choice of algorithms and set a benchmark for our performance evaluations.

Reference [11] provides the E_VTOL battery dataset, which includes comprehensive data on battery performance parameters such as Batt_Chg, Batt_Curr, Batt_DsChg, Batt_Temp, Batt_EChg, and Batt_EDC under various operating conditions. Additionally, [12] details the model of Airbus's E_VTOL, named Ava. Understanding this model is essential for contextualizing the dataset and ensuring the relevance of predictions to practical applications.

Research focuses on developing and implementing two MLA, Alg_CB and Alg_LGBM, for Pred_RUL, Pred_SOC, and Pred_SOH of batteries using the dataset from [11]. Alg_CB is developed by Yandex, known for its efficiency in handling categorical features without the need for extensive preprocessing, while mitigating the issue of overfitting through its use of ordered boosting, making it highly robust and accurate, especially with complex datasets. Alg_LGBM is developed by Microsoft, recognized for its speed and efficiency, particularly with large datasets [13,14]. Allowing for more complex and deeper trees compared to depth-wise growth used by many other algorithms [13,14].

Such that these two MLAs are chosen over Alg_SVM, Alg_RFR, AlgXGBoost, Alg_GPR, and Alg_MLP while considering their ability to provide faster training times and higher accuracy for real-time applications with large-scale datasets like the one used in this study. The process as shown in Figure 1, begins with data preprocessing, including data collection, and splitting the dataset into training, testing, and validation sets.

Next, the T_RUL, T_SOC, T_SOH are calculated, which later serve as the ground for evaluating the model predictions. The MLA is trained using Alg_CB and Alg_LGBM models. During testing set the MLA generate Pred_RUL, Pred_SOC, Pred_SOH and are compared with true values. Error metrics such as EM_MAE and EM_RMSE are compared to quantify model's performance, here mean absolute percentage error is not considered due to their high sensitivity for small changes resulting in high percentage error. The final step is to compare the performance metrics of our Alg_CB and Alg_LGBM models with the results from [10], to determine the effectiveness of the proposed algorithms in Pred_RUL, Pred_SOC, and Pred_SOH.

Finally, the Predicted Data stage represents the output of the models, which includes the Pred_RUL, Pred_SOC, and Pred_SOH for the testing set. This data is crucial for making informed decisions about Batt_Mgmt in applications like E_Veh's and E_VTOL's. The process concludes at the END, marking the completion of the predictive modeling workflow.

2.1 Dataset Considerations

Dataset [11] are considered using the reference [10] for optimal performance of MLA. The T_SOH of battery can be determined by its Cap_Test and Nom_Cap of the battery (initially measured as 3 Ah), but this condition differs with 01 and 23 Mis_Pfl, with 3.03 Ah and 2.71 Ah respectively [10]. Therefore, Nom_Cap is not fixed at 3Ah Batt_Cap.

The Batt_EoL threshold influences the selection of Mis_Pfl. Similarly, battery depletion reaching Batt_EoL, specifically, for 01, 02, 15, 16, 20, 23, 24 and 25 series of Batt_Cap measurements are terminated before reaching 85% of the initial Batt_Cap. To ensure that Mis_Pfl are performed without termination before reaching their Batt_EoL. Table 1 represents the total number of Mis_Cyc and Cap_Test with and without the Batt_EoL considerations [15,16].

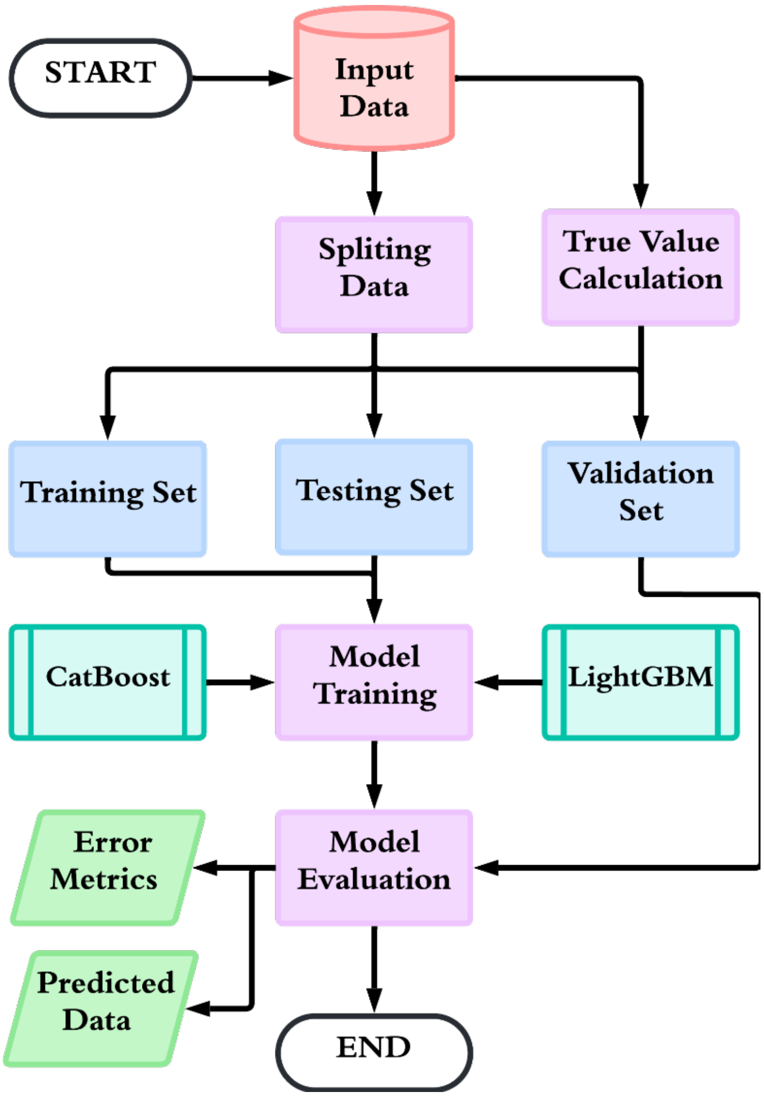

Fig 1. Flowchart for proposed MLA's

Table 1. Battery Data Metrics

Mis_Pfl	Mis_Cyc Profile	Cap_Test Profile	Mis_Cyc Batt_EoL	Cap_Test Batt_EoL	Batt_Vol Min (V)	Batt_Vol Max (V)	Batt_Curr Min (mA)	Batt_Curr Max (mA)	Batt_E Chg (Wh)	Batt_E DC (Wh)	Batt_Chg (mAh)	Batt_Ds Chg (mAh)	Batt_Temp Min (C)	Batt_Temp Max (C)
1	847.00	17.00	613.00	13.00	2.50	4.20	-21180.91	2995.80	11.77	-11.26	3027.44	3089.34	19.98	65.99
2	625.00	13.00	511.00	11.00	2.50	4.20	-21608.49	2994.19	11.77	-11.32	3031.22	3104.82	19.72	67.84
10	1431.00	28.00	614.00	13.00	2.50	4.21	-19883.94	3010.88	11.83	-11.28	3062.30	3085.54	29.15	69.83
12	2349.00	46.00	766.00	16.00	2.50	4.20	-18669.36	3016.85	11.88	-11.16	3066.23	3060.80	16.31	54.85
13	1042.00	20.00	562.00	12.00	2.50	4.20	-21594.12	3007.32	11.84	-11.12	3056.60	3045.31	20.43	61.13
15	554.00	11.00	460.00	10.00	2.50	4.20	-20565.66	3000.29	11.79	-11.39	3044.59	3124.82	21.43	61.28
16	559.00	11.00	460.00	10.00	2.50	4.20	-19818.61	4492.40	11.85	-11.44	3025.66	3138.38	22.00	58.12
17	1002.00	20.00	562.00	12.00	2.50	4.20	-21621.64	2995.01	11.76	-11.35	3032.23	3111.98	20.92	65.90
20	611.00	12.00	460.00	10.00	2.50	4.20	-21608.25	4492.20	11.85	-11.45	3026.10	3139.67	21.55	67.58
22	579.00	12.00	460.00	10.00	2.50	4.20	-21611.44	3001.48	11.79	-11.40	3042.73	3125.64	20.07	64.11

Continued on next page

Table 1 continued

23	697.00	13.00	562.00	12.00	2.50	4.20	-21594.39	3008.80	10.38	-11.01	2713.53	3021.30	15.31	61.29
25	554.00	11.00	513.00	11.00	2.50	4.20	-21594.26	3008.32	11.85	-11.03	3050.03	3027.02	18.81	61.61

Table 2 presents the average battery voltage and current metrics across variousMis_Pfl. The average Batt_Vol ranges from 2.4999751V to 4.4001856V. And average Batt_Curr ranges from -20924.014mA to 3201.9146mA, reflecting the peak current usage during charging or high-demand scenarios.

Table 2. Comparison between T_SOC & Pred_SOC , T_SOH & Pred_SOH and T_RUL & Pred_RUL

Mis_Pfl	Condition	Mis_Cyc 1	Mis_Cyc 3	Mis_Cyc 5	Mis_Cyc 7	Mis_Cyc 9
	T_SOC	61.1	70.14	70.87	71.32	71.99
	Pred_SOC for Alg_CB	60.76	70.35	71.09	71.15	72.78
	Pred_SOC for Alg_LGBM	61.27	75.02	75.02	75.02	75.02
	T_SOH	100	94.98	91.85	89.4186	87.4502
	Alg_RFR [12]	96.9	NA	91.31	NA	NA
1	Pred_SOH for Alg_CB	99.99	94.98	91.88	89.4067	87.4764
	Pred_SOH for Alg_LGBM	99.95	94.96	91.84	89.4319	87.4751
	T_RUL	612	510	408	306	204
	Alg_XGB [12]	484	458	401	279	67
	Pred_RUL for Alg_LGBM	609.95	510.33	407.34	306.17	206.23
	Pred_RUL for Alg_CB	612.73	507.82	408.65	307.55	201.62
	T_SOC	61.56	60.05	60.92	61.31	61.35
	Pred_SOC for Alg_CB	61.32	60.73	61.58	61.63	61.82
	Pred_SOC for Alg_LGBM	61.58	65.31	65.31	65.32	65.32
	T_SOH	100	94.472	91.1436	88.6236	86.6816
	Alg_RFR [12]	98.36	NA	91.88	NA	NA
2	Pred_SOH for Alg_CB	99.9835	94.4725	91.1265	88.639	86.6989
	Pred_SOH for Alg_LGBM	99.9529	94.4576	91.149	88.6439	86.7134
	T_RUL	510	408	306	204	102
	Alg_XGB [12]	568	451	375	273	78
	Pred_RUL for Alg_LGBM	507.85	408.32	306.54	206.1	100.22
	Pred_RUL for Alg_CB	509.34	406.6	304.67	200.88	104.35
	T_SOC	60.9	89.19	80.05	80.49	80.59
	Pred_SOC for Alg_CB	60.98	89.88	80.47	80.44	80.51
	Pred_SOC for Alg_LGBM	61.26	92.55	82.55	82.56	82.56
	T_SOH	100	95.7101	92.8858	91.034	89.2848
	Alg_RFR [12]	99.19	NA	92.01	NA	NA
12	Pred_SOH for Alg_CB	99.9865	95.6938	92.8752	91.015	89.2746
	Pred_SOH for Alg_LGBM	99.9252	95.6607	92.8531	91.0123	89.2735
	T_RUL	765	663	561	459	357
	Alg_XGB [12]	643	532	382	305	200
	Pred_RUL for Alg_LGBM	758.87	660.25	551.01	449.83	348.9
	Pred_RUL for Alg_CB	762.78	656.03	558.8	451.66	352.67
	T_SOC	61.25	69.98	60.62	61.71	62.36
	Pred_SOC for Alg_CB	61.37	70.6	60.59	61.63	61.96
	Pred_SOC for Alg_LGBM	60.06	74.62	64.62	64.62	64.62
	T_SOH	100	94.2712	90.8697	87.9455	85.7905
	Alg_RFR [12]	98.72	NA	90.69	NA	NA
15	Pred_SOH for Alg_CB	99.9859	94.27	90.9479	87.9481	85.7698
	Pred_SOH for Alg_LGBM	99.9473	94.2525	90.871	87.9641	85.822
	T_RUL	459	357	255	153	NA
	Alg_XGB [12]	460	365	238	141	NA
	Pred_RUL for Alg_LGBM	457.2	355.78	233.53	153.07	NA
	Pred_RUL for Alg_CB	458.99	354.82	235.06	154.61	NA
	T_SOC	61.29	76.09	77.03	78.2	79.16
	Pred_SOC for Alg_CB	60.85	76.35	77.68	78.29	78.78

Continued on next page

16

Table 2 continued

	Pred_SOC for Alg_LGBM	59.2	81.54	81.54	81.59	81.59
	T_SOH	100	94.7986	91.2354	88.655	86.3644
	Alg_RFR [12]	96.81	NA	91.97	NA	NA
	Pred_SOH for Alg_CB	100.0011	94.8078	91.2466	88.6716	86.3899
	Pred_SOH for Alg_LGBM	99.9497	94.7791	91.237	88.6718	86.3948
	T_RUL	459	357	255	153	NA
	Alg_XGB [12]	479	415	277	208	NA
	Pred_RUL for Alg_LGBM	456.48	357.02	256.07	151.43	NA
	Pred_RUL for Alg_CB	459.5	355.6	255.37	151.04	NA
	T_SOC	61.28	69.72	70.11	70.95	71.47
	Pred_SOC for Alg_CB	61.23	70.1	70.24	70.42	72.21
	Pred_SOC for Alg_LGBM	60.51	73.61	73.87	73.62	73.62
	T_SOH	100	94.7444	91.741	89.1892	87.1791
	Alg_RFR [12]	98.84	NA	91.75	NA	NA
17	Pred_SOH for Alg_CB	99.9932	94.7696	91.758	89.1613	87.2617
	Pred_SOH for Alg_LGBM	99.9323	94.7079	91.7222	89.1855	87.1873
	T_RUL	561	459	357	255	153
	Alg_XGB [12]	516	443	345	255	177
	Pred_RUL for Alg_LGBM	556.33	457.5	355.76	250.47	152.28
	Pred_RUL for Alg_CB	561.21	454.9	353.4	252.26	155.69
	T_SOC	61.28	76.31	77.42	78.4	79.5
	Pred_SOC for Alg_CB	61.59	77.12	78.25	78.88	78.96
	Pred_SOC for Alg_LGBM	60.37	80.99	80.99	81	81
	T_SOH	100	94.4024	90.8939	88.2235	85.2254
	Alg_RFR [12]	95.91	NA	89.48	NA	NA
20	Pred_SOH for Alg_CB	99.9985	94.4016	90.9521	88.2234	85.2415
	Pred_SOH for Alg_LGBM	99.942	94.3775	90.8897	88.2352	85.2549
	T_RUL	459	357	255	153	NA
	Alg_XGB [12]	450	349	250	199	NA
	Pred_RUL for Alg_LGBM	456.3	355.39	256.47	151.39	NA
	Pred_RUL for Alg_CB	458.83	355.23	254.97	151.84	NA
	T_SOC	61.15	59.84	60.71	61.62	62.34
	Pred_SOC for Alg_CB	60.65	60.41	60.96	61.41	61.47
	Pred_SOC for Alg_LGBM	59.52	64.45	64.45	64.45	64.45
	T_SOH	100	94.3489	90.8171	88.0192	85.6323
	Alg_RFR [12]	98.96	NA	90.38	NA	NA
22	Pred_SOH for Alg_CB	99.9945	94.3242	90.7876	88.0467	85.6399
	Pred_SOH for Alg_LGBM	99.9453	94.3276	90.8168	88.0354	85.6626
	T_RUL	459	357	255	153	NA
	Alg_XGB [12]	460	356	249	147	NA
	Pred_RUL for Alg_LGBM	456.32	355.81	254.16	153.89	NA
	Pred_RUL for Alg_CB	459.12	354.09	254.29	155.87	NA
	T_SOC	60.74	69.79	70.23	70.69	71.33
	Pred_SOC for Alg_CB	60.14	69.93	69.93	71.25	71.62
	Pred_SOC for Alg_LGBM	59.8	73.88	73.88	73.89	73.89
	T_SOH	100	93.7274	91.2538	88.8308	85.7316
	Alg_RFR [12]	96.18	NA	90.17	NA	NA
25	Pred_SOH for Alg_CB	100.0015	93.7253	91.236	88.8213	85.7797
	Pred_SOH for Alg_LGBM	99.9391	93.7036	91.2446	88.8361	85.7552
	T_RUL	512	410	307	205	103
	Alg_XGB [12]	454	413	299	161	36
	Pred_RUL for Alg_LGBM	510.17	407.53	307.94	204.8	NA
	Pred_RUL for Alg_CB	511.75	407.2	309.05	205.87	NA

Table 3 provides the average values for various Batt_EChg, Batt_EDC, Batt_Chg and Batt_DsChg metrics across different Mis_Pfl. The average Batt_EChg is 11.71Wh, while the average Batt_EDC is -11.278371Wh, indicating the typical amount of energy the battery can store and release, respectively. Additionally, the average Batt_Chg is 3022.05mAh, and the average

Batt_DsChg is 3092.42mAh. These averages highlight the overall performance of the battery on various Mis_Pfl's [Table 1].

Table 3. Error Metrics

Mis_Pfl	Alg_RFR [12]				Alg_LGBM						Alg_CB					
	SOH		RUL		SOC		SOH		RUL		SOC		SOH		RUL	
	EM_MAE	EM_RMSE	EM_MAE	EM_RMSE	EM_MAE	EM_RMSE	EM_MAE	EM_RMSE	EM_MAE	EM_RMSE	EM_MAE	EM_RMSE	EM_MAE	EM_RMSE	EM_MAE	EM_RMSE
1	1.2900	1.5100	62.6400	75.8500	0.1526	0.2829	0.0247	0.0302	1.3300	1.8700	0.1103	0.2260	0.0021	0.0043	0.8000	1.3700
2	0.8500	1.0100	52.7200	57.5400	0.1412	0.2727	0.0239	0.0292	1.0500	1.4600	0.1098	0.2342	0.0016	0.0030	0.5700	1.0200
12	0.6500	0.8400	97.4300	109.1600	0.2134	0.2920	0.0215	0.0270	1.7600	2.4400	0.1171	0.2427	0.0021	0.0042	0.7400	1.8400
15	0.3100	0.4600	18.1500	23.7000	0.1967	0.2719	0.0256	0.0300	1.0300	1.2400	0.1225	0.2109	0.0029	0.0048	0.4800	0.7900
16	1.2600	1.4500	30.7700	34.7700	0.1950	0.2761	0.0254	0.0295	1.0400	1.2600	0.1168	0.2105	0.0020	0.0035	0.4700	0.8000
17	0.4600	0.6300	26.1800	31.5700	0.2007	0.2798	0.0245	0.0293	1.3900	1.7000	0.1178	0.2152	0.0023	0.0045	0.5700	1.1400
20	1.8500	2.3400	19.6300	27.2100	0.1930	0.2733	0.0277	0.0323	1.1200	1.3300	0.1178	0.2138	0.0028	0.0047	0.4800	0.8400
22	1.7900	3.7400	8.7600	14.4100	0.1950	0.2718	0.0262	0.0307	1.0600	1.2700	0.1208	0.2139	0.0025	0.0044	0.5300	0.8500
25	2.0000	2.7100	65.2700	126.3100	0.2017	0.2739	0.0170	0.0205	0.7000	0.9000	0.1200	0.2149	0.0005	0.0010	0.2600	0.5900

Figure 2, represents a bar chart that illustrates the minimum, average, and maximum battery temperatures for various Mis_Pfl. Each Mis_Pfl is represented along the x-axis, while the y-axis shows the temperature (C). Table 1 presents the minimum, maximum, Batt_Temp across various Mis_Pfl's. Each Mis_Pfl indicates a different operational scenario[10], with Batt_Temp recorded at minimum and maximum levels. By calculating the average values for all Mis_Pfl, we find that the average minimum and maximum Batt_Temp are 20.98°C, and 63.66°C respectively.

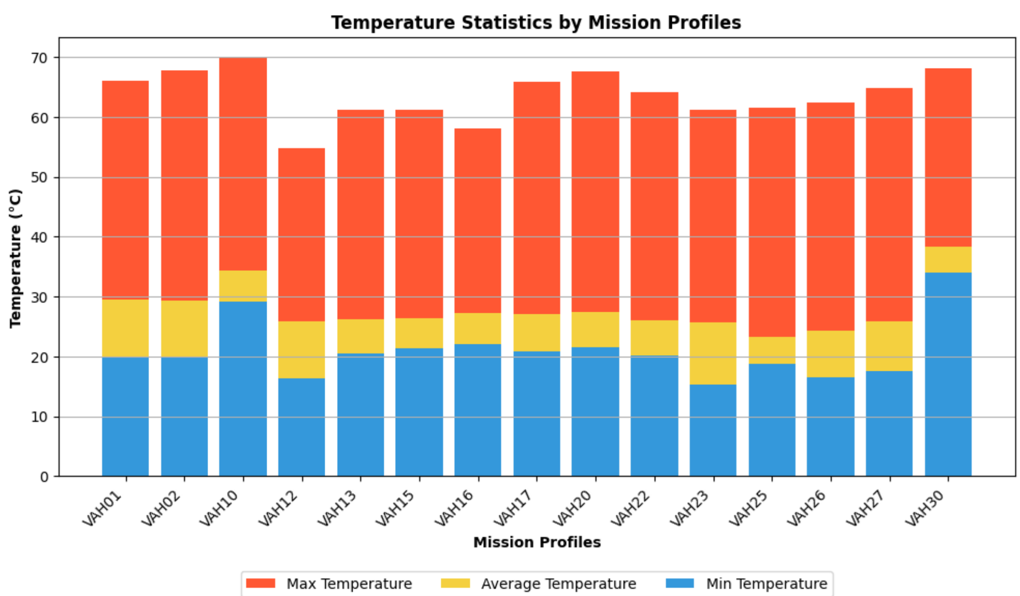

Fig 2. Batt_Temp Statistics by individual Mis_Pfl

2.2 Mathematical calculations

The T_RUL of the battery can be defined as Batt_Lfe before it reaching Batt_Eol[2,3,14,17,18] which can be calculated by Equation (1)

$$T\_RUL = T_{tc}\ T_{cc} \tag{1}$$

The T_SOC is the remaining charge in the battery represented in percentage of charge and T_SOH is the current health of the battery. The T_SOC and T_SOH can be calculated by the Equations (2) and (3) respectively. [19,20]

$$T\_SOC = \frac{B_{cc}}{B_{maxc}} \times 100\% \tag{2}$$

$$T\_SOH = \frac{B_{cc}}{B_{ic}} \times 100\% \tag{3}$$

The general formula for the Alg_CB and Alg_LGBM for Pred_Alg is given in Equation (4) and Equation (5) respectively. To calculate the performance metrics of Alg_CB and Alg_LGBM, the error metrics EM_MAE and EM_RMSE are calculated to determine the Pred_Alg performance, presented in the Equation (6) for EM_MAE and Equation (7) for EM_RMSE[21,22]

$$y = \phi\left(\sum_{i=1}^{N} f_i(x)\right) \tag{4}$$

$$y = \sum_{i=1}^{N} w_i f_i(x) \tag{5}$$

$$EM\_MAE = \frac{1}{n}\sum_{i=1}^{n} |w_i - \widehat{w_i}| \tag{6}$$

$$EM\_RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (w_i - \widehat{w}_i)^2} \tag{7}$$

3 Results and Discussion

The MLA using Alg_CB and Alg_LGBM are implemented, resulting in Pred_RUL, Pred_SOC, and Pred_SOH. A comparative analysis between predicted and true values of RUL, SOC, and SOH is performed for Alg_CB and Alg_LGBM techniques while comparing them to the outputs from previous research[10].

Figure 3 illustrates the T_SOC, Pred_SOC using Alg_CB and Alg_LGBM trends respectively for all Mis_Pfl. Table 2, presents the T_SOC values and the Pred_SOC values from, Alg_CB and Alg_LGBM across various Mis_Cyc's. For each condition, the T_SOC values are given for cycles 1 through 9, indicating the T_SOC measured at each Mis_Cyc and are compared with Pred_SOC values from Alg_CB and Alg_LGBM. For instance, T_SOC starts at 61.10 at Mis_Cyc 1 and increases to 79.16 by Mis_Cyc 11, reflecting the battery's SOC. Alg_CB predicts these values closely, with its predictions ranging from 60.76 to 79.42. Alg_LGBM, while generally close, tends to slightly overestimate in some cycles, with predictions like 61.27 at Mis_Cyc 1 and a consistent 75.02 from Mis_Cyc 3 to 9. Similar trends are observed across all Mis_Pfl's. Thus, providing consistent and close predictions of true values.

Figure 3 also illustrates the T_SOH, Pred_SOH using Alg_CB and Alg_LGBM trends for all Mis_Pfl. While a comparison of the T_SOH of a system across different Mis_Pfl with Pred_SOH values from three algorithms: Alg_RFR proposed in[10], Alg_CB, and Alg_LGBM are presented. The T_SOH value starts at 100 and gradually decreases across the cycles, indicating the degradation of the system's health over time.

The Pred_SOH values from Alg_CB and Alg_LGBM closely follow the T_SOH values, showing high prediction accuracy. For instance, T_SOH at Mis_Cyc 1 is 100, and the predictions by Alg_CB and Alg_LGBM are 99.9964 and 99.9507, respectively. These values are very close to the true value, indicating good predictive performance. The values of Alg_RFR[10] are also presented in the table.

Figure 3 also showcases a holistic view of the battery's degradation over time with T_RUL, Pred_RUL using Alg_CB and Alg_LGBM respectively across all Mis_Pfl's. Table X provides a comparison of the T_RUL of a system across different Mis_Cyc with the Pred_RUL values from three algorithms: Alg_XGB, Alg_LGBM, and Alg_CB.

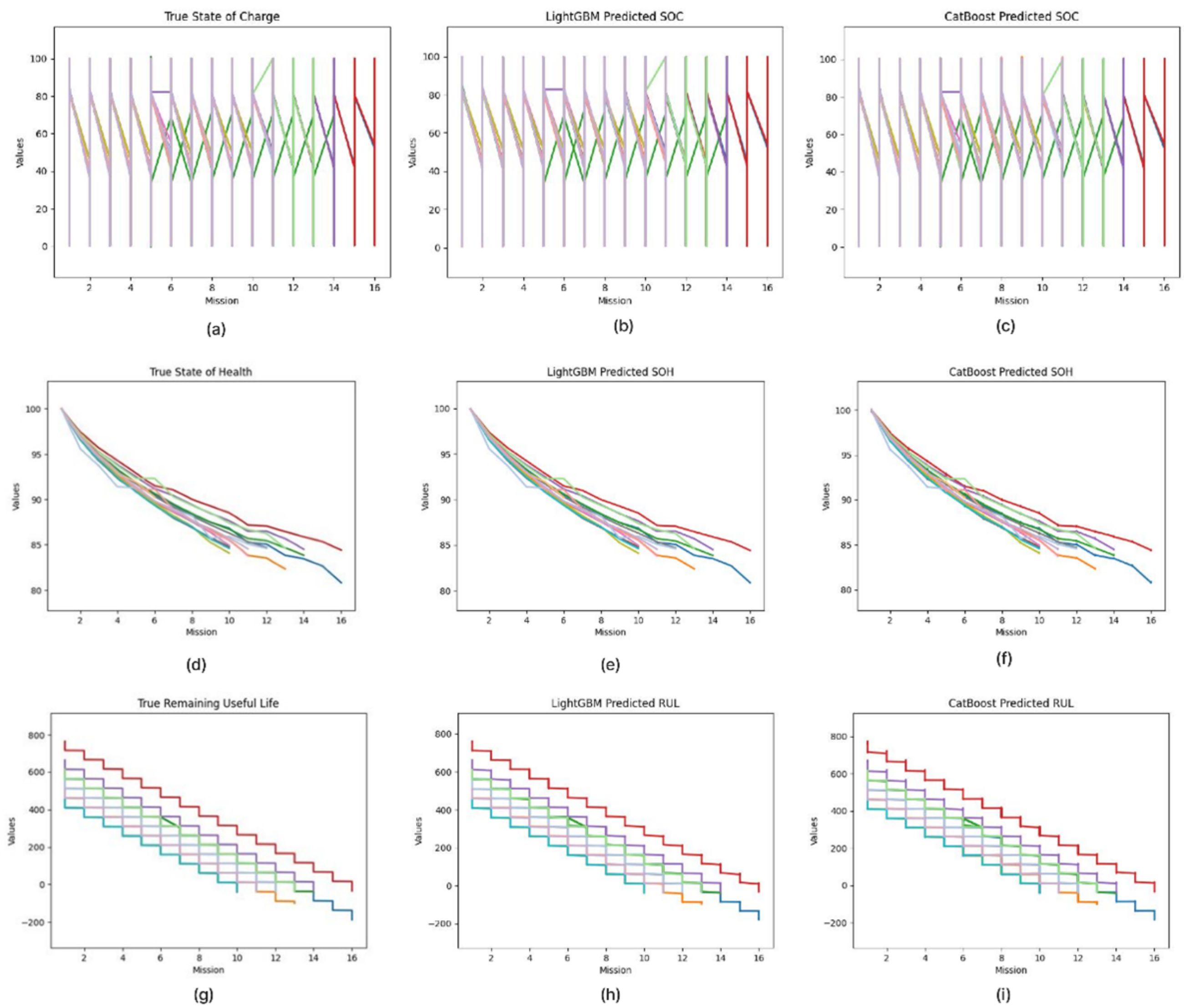

Fig 3. (a) T_SOC across all Mis_Pfl (b) Pred_SOC estimation with Alg_CB (c) Pred_SOC estimation with Alg_LGBM (d) T_SOH across all Mis_Pfl (e) Pred_SOH estimation with Alg_CB (f) Pred_SOH estimation with Alg_LGBM (g) T_RUL across all Mis_Pfl (h) Pred_RUL estimation with Alg_CB (i) Pred_RUL estimation with Alg_LGBM

The cycles range from Mis_Cyc 1 to 9, T_RUL values start at a certain point and decrease across the cycles, indicating the expected remaining life of the system over time. The Pred_RUL values from Alg_CB and Alg_LGBM closely follow the T_RUL values, demonstrating high prediction accuracy. At Mis_Cyc 1 T_RUL is 612, and the predictions by Alg_CB and Alg_LGBM are 612.73 and 609.95, respectively. These values are very close to the true value, indicating good predictive performance. The outputs of Alg_XGB predictions are included to provide a feasible comparison [Table 2].

Figure 4, represents the Pred_SOC error metrics: EM_MAE and EM_RMSE for the Alg_CB and Alg_LGBM. Table 3, provides a detailed comparison of the performance of, Alg_LGBM and Alg_CB, in Pred_SOC across various Mis_Pfl conditions, listing the EM_MAE and EM_RMSE for both algorithms. For Mis_Pfl 1, Alg_LGBM has an EM_MAE of 0.1525793 and an EM_RMSE of 0.2828813, while Alg_CB performs better with an EM_MAE of 0.1103394 and an EM_RMSE of 0.2260082. On average, Alg_LGBM has an EM_MAE of 0.191924525 and an EM_RMSE of 0.295422008, while Alg_CB shows better performance with a lower average EM_MAE of 0.117812275 and an EM_RMSE of 0.247490042. This demonstrates, Alg_CB consistently achieves higher accuracy and reliability, as evidenced by its lower EM_MAE and EM_RMSE values across all Mis_Pfl conditions [Table 3].

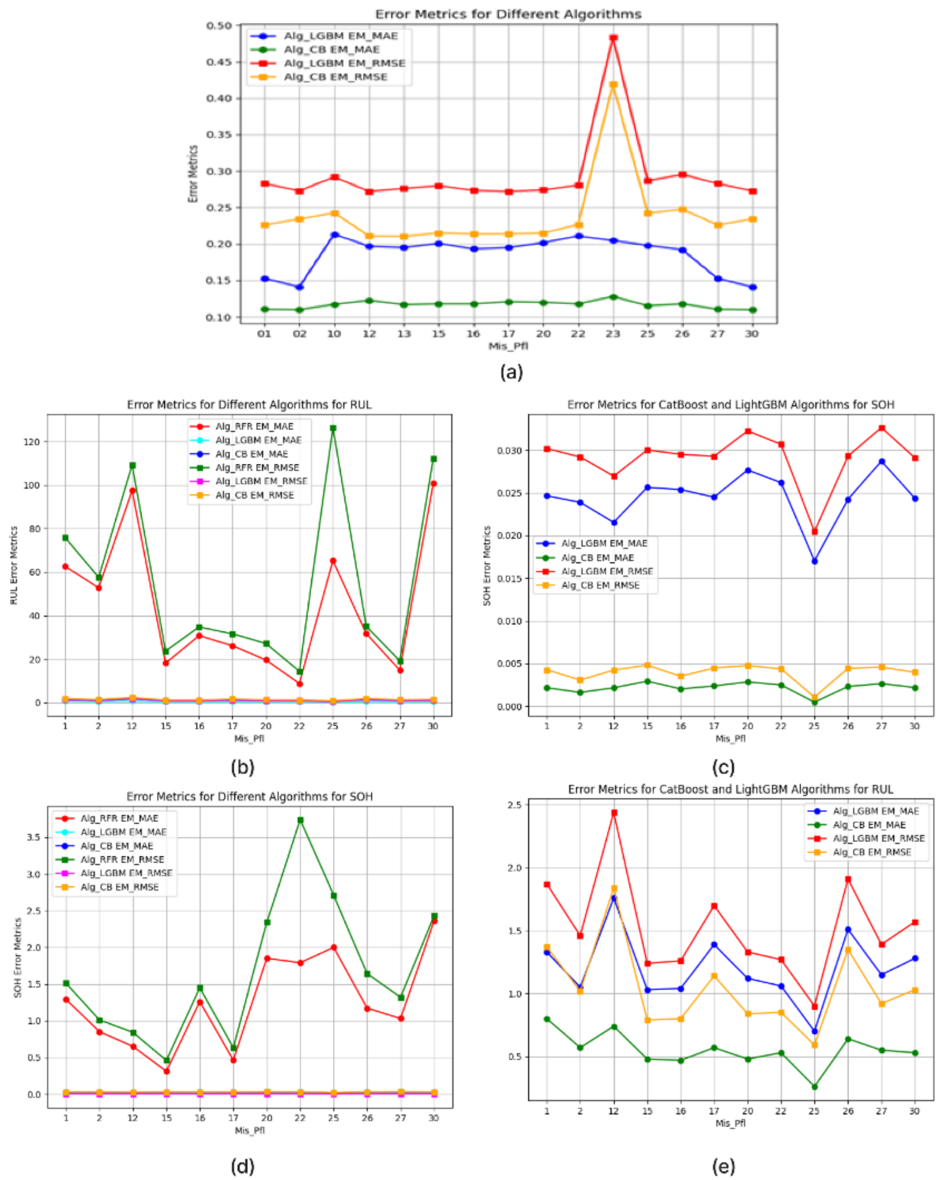

Fig 4. (a) ProposedPred_Alg EM_MAE and EM_RMSE for SOC (b) Proposed Pred_Alg EM_MAEand EM_RMSE for SOH (c) Proposed Algorithms vs Alg_RFR[5] error metrics for SOH (d) Proposed Pred_Alg EM_MAE and EM_RMSEfor RUL (e) Proposed Algorithms vs Alg_XGB [5] errormetrics for RUL

Similarly, Table 3 provides a comprehensive comparison of three different algorithms: Alg_RFR, Alg_LGBM, and Alg_CB in terms of their EM_MAE and EM_RMSE for Pred_SOH across various Mis_Pfl conditions. The average errors across all conditions further emphasize this, with Alg_RFR [10] showing the highest average EM_MAE and EM_RMSE with 1.2516 and 1.6733 respectively, while Alg_LGBM has a lower average EM_MAE of 0.0244 and EM_RMSE of 0.0291. Alg_CB again has the best performance with an average EM_MAE of 0.0021 and an EM_RMSE of 0.0039, highlighting its accurate prediction with significantly lower error metrics. Figure 4 presents the comparison of error metrics of Alg_Cb and Alg_LGBM, it also presents the comparison of proposed Pred_Alg and existing Alg_RFR from [10].

The Pred_RUL with EM_MAE and EM_RMSE for the proposed Pred_Alg's are illustrated in Figure 4 and the comparison with the outputs of existing research [10] is illustrated in Figure 4. Table 2, compares the performance of three different algorithms: Alg_XGB, Alg_LGBM, and Alg_CB in Pred_RUL of components across various Mis_Pfl conditions.

The evaluation metrics used are EM_MAE and EM_RMSE, the average values in the table highlight the overall performance of each algorithm. Alg_XGB has the highest average EM_MAE (44.08) and EM_RMSE (55.58), indicating less accurate predictions. Alg_LGBM performs better with an average EM_MAE of 1.20 and an EM_RMSE of 1.53. Alg_CB outperforms both, with the lowest average EM_MAE of 0.55 and EM_RMSE of 1.04, demonstrating superior predictive accuracy and reliability.

To summarize proposed algorithms when compared with existing reports [10] outperform in Predicting SOC, SOH and RUL and it is clear from the above error metrics.

4 Conclusion

This research work has shown an accurate and optimum MLA Pred_Alg for Pred_RUL, Pred_SOC, and Pred_SOH of the Li_Batt in E_VTOL by establishing the efficiency of Alg_CB and Alg_LGBM while comparing them to the existing Pred_Alg. The results show that Pred_Alg leads to an accurate estimation and, Alg_CB performs better than Alg_LGBM, although there is a minimal difference in their performance.

In Pred_RUL while comparing with the Alg_XGB [10] Alg_LGBM values are 97.27% and 97.25% reduction in EM_MAE and EM_RMSE respectively; similarly Alg_CB with a 98.75% and 98.13% in case of EM_MAE and EM_RMSE error reduction. This improvement to its performance highlights its Robustness to Overfitting, Automatic Feature Scaling, Built-in Cross-Validation, Robustness to Noisy Data.

Similarly, the findings for Pred_SOC reveal that Alg_CB outperforms Alg_LGBM by 27.76% in EM_MAE and 25.6% in EM_RMSE. In terms of Pred_SOH, when compared with Alg_RFR [10], the Alg_LGBM achieves a 98.04% reduction in EM_MAE and 98.26% reduction in EM_RMSE, similarly Alg_CB 99.83% EM_MAE reduction and 99.76% EM_RMSE reduction. The results indicate that while both Pred_Alg offer precise estimation, Alg_CB consistently surpasses Alg_LGBM in performance.

The study shows that the two algorithms, CatBoost and LightGBM, can successfully determine the state and state-of-health and performance of eVTOL batteries. Minimal error rates of EM_MAE and EM_RMSE, as well as very high-accuracy predictions of SOC, SOH, and RUL, are also presented by the algorithms. Fast computation and adaptability, hence, make these algorithms apt for real-time application in battery management. The framework, for example, can ensure credible data at each phase of the flight, taking into consideration temperatures and charge profiles. It is particularly known for handling complex datasets and minimizing errors, making it a very strong candidate for high-stakes applications.

Although the algorithms perform well in simulations, real-world validation is not that extensive, with few tests in actual operational environments. The requirement for very precise parameter tuning becomes difficult in dynamic applications and real-time systems. Another limitation of the study is its focus on eVTOL applications, which restricts the generalizability of the study to other battery-powered systems. The quality and representativeness of the dataset are highly crucial for the reliability of predictions, and robust data sources must be ensured for performance in varied scenarios.

To take eVTOL battery prognostics forward, real-world tests under different conditions should be conducted to validate algorithms. Adaptive models must be developed for real-time dynamic tuning, and more complex feature fusion methodologies will also further enhance the accuracy of prediction. Tailoring these models to specific flight profiles, while integrating them into whole energy management systems, would maximize power usage and, thus, improve operational effectiveness. These efforts will solve the current limitations and make it possible to extend its application to the field of aerospace and related areas.

Table 4. Acronyms

Acronym	Meaning
Alg_CB	Algorithm CatBoost
Alg_XGB	Algorithm eXtreme Gradient Boosting
Alg_GPR	Algorithm Gaussian Process Regression
Alg_LGBM	Algorithm LightGBM
Alg_MLP	Algorithm Multi-Layer Perceptron
Alg_RFR	Algorithm Random Forest Regression
Alg_SVR	Algorithm Support Vector Regression
Alg_CNN	Algorithm Convolutional Neural Network
Alg_LSTM	Algorithm Long Short-Term memory
Alg_SS	Algorithm Sparrow Search
Batt_Age	Battery Aging
Batt_Cap	Battery Capacity

Continued on next page

Table 4 continued

Batt_Chg	Battery Charge
Batt_Curr	Battery Current
Batt_DsChg	Battery Discharge
Batt_EoL	Battery End of Life
Batt_Lfe	Battery Life
Batt_Mgmt	Battery Management
Batt_Pwr	Battery Power
Batt_Pred	Battery Prediction
Batt_Temp	Battery Temperature
Batt_Vol	Battery Voltage
Batt_EChg	Battery Energy Charge
Batt_EDC	Battery Energy Discharge
Cap_Test	Capacity test
E_Veh	Electric Vehicle
E_VTOL	Electric Vertical Take Off and Landing
EM_MAE	Error Metrics of Mean Absolute Error
EM_RMSE	Error Metric of Root Mean Square Error
Li_Batt	Lithium Battery/Batteries
MLA	Machine Learning Algorithm
Mis_Cyc	Mission Cycle
Mis_Pfl	Mission Profile
Nom_Cap	Nominal Capacity
Pred_Alg	Prediction Algorithm
Pred_RUL	Predicted Remaining Useful Life
Pred_SOC	Predicted State of Charge
Pred_SOH	Predicted State of Health
RUL	Remaining Useful Life
SOC	State of Charge
SOH	State of Health
T_RUL	True Remaining Useful Life
T_SOC	True State of Charge
T_SOH	True State of Health

Table 5. Nomenclature

Symbol	Description
y	Output
φ	Transformation function
$f_i(x)$	i-th output for x
w_i	i-th true weight value
$\widehat{w_i}$	i-th predicted weight value
B_{ct}	Battery capacity with time
B_{maxc}	Maximum battery capacity (Full charge)
B_{ic}	Initial battery capacity
T_{tc}	Total battery charge and discharge before EOL
T_{cc}	Current battery charge and discharge cycle
N	Observation Number

References

1) Karthick K, Ravivarman S, Priyanka R. Optimizing Electric Vehicle Battery Life: A Machine Learning Approach for Sustainable Transportation. *World Electric Vehicle Journal*. 2024;15(2). Available from: https://doi.org/10.3390/wevj15020060.
2) Zhao W, Ding W, Zhang S, Zhang Z. A deep learning approach incorporating attention mechanism and transfer learning for lithium-ion battery lifespan prediction. *Journal of Energy Storage*. 2024;75. Available from: https://doi.org/10.1016/j.est.2023.109647.

3) Shah AA, Shah N, Luo L, Xing WW, Leung PK, Zhu X, et al. Nonlinear autoregressive models for high accuracy early prediction of Li-ion battery end-of-life. *Journal of Energy Storage*. 2023;73. Available from: https://doi.org/10.1016/j.est.2023.109014.
4) Zhang S, Zhu H, Wu J, Chen Z. Voltage relaxation-based state-of-health estimation of lithium-ion batteries using convolutional neural networks and transfer learning. *Journal of Energy Storage*. 2023;73. Available from: https://doi.org/10.1016/j.est.2023.108579.
5) Na HS, Numan-Al-Mobin AM. Machine learning approaches to estimate the health state of next-generation energy storage. *Green Sustainable Process for Chemical and Environmental Engineering and Science*. 2023;p. 343–363. Available from: https://doi.org/10.1016/B978-0-323-90635-7.00002-6.
6) Xue K, Yang J, Yang M, Wang D. An Improved Generic Hybrid Prognostic Method for RUL Prediction Based on PF-LSTM Learning. *IEEE Transactions on Instrumentation and Measurement*. 2023;72:1–21. Available from: https://ieeexplore.ieee.org/document/10057430.
7) Kim H, Park S, Park HJ, Son HG, Kim S. Solar Radiation Forecasting Based on the Hybrid CNN-CatBoost Model. *IEEE Access*. 2023;11:13492–13500. Available from: https://ieeexplore.ieee.org/document/10040681.
8) Yao XY, Chen G, Pecht M, Chen B. A novel graph-based framework for state of health prediction of lithium-ion battery. *Journal of Energy Storage*. 2023;58. Available from: https://doi.org/10.1016/j.est.2022.106437.
9) Chang W, Wang X, Yang J, Qin T. An Improved CatBoost-Based Classification Model for Ecological Suitability of Blueberries. *Sensors*. 2023;23(4). Available from: https://doi.org/10.3390/s23041811.
10) Mitici M, Hennink B, Pavel M, Dong J. Prognostics for Lithium-ion batteries for electric Vertical Take-off and Landing aircraft using data-driven machine learning. *Energy and AI*. 2023;12. Available from: https://doi.org/10.1016/j.egyai.2023.100233.
11) Bills A, Viswanathan V, Sripad S, Frank E, Charles D, Fredericks WL, et al. eVTOL Battery Dataset. 2022. Available from: https://kilthub.cmu.edu/articles/dataset/eVTOL_Battery_Dataset/14226830?file=26855063.
12) Airbus. ana our single-set evtol demonstrator. 2019.
13) Zhang M, Yin J, Feng T. Lithium Battery SOH Estimation Based on Manifold Learning and LightGBM. *Applied Sciences*. 2023;13(11). Available from: https://doi.org/10.3390/app13116540.
14) Jafari S, Byun YC. Optimizing Battery RUL Prediction of Lithium-Ion Batteries Based on Harris Hawk Optimization Approach Using Random Forest and LightGBM. *IEEE Access*. 2023;11:87034–87046. Available from: https://ieeexplore.ieee.org/document/10216301.
15) Bills A, Sripad S, Fredericks L, Guttenberg M, Charles D, Frank E, et al. A battery dataset for electric vertical takeoff and landing aircraft. *Scientific Data*. 2023;10. Available from: https://www.nature.com/articles/s41597-023-02180-5.
16) Dixit M, Bisht A, Essehli R, Amin R, Kweon CBM, Belharouak I. Lithium-Ion Battery Power Performance Assessment for the Climb Step of an Electric Vertical Takeoff and Landing (eVTOL) Application. *Ruhul Amin, Chol-Bum M Kweon, and Ilias Belharouak*. 2024;2024:934–940. Available from: https://pubs.acs.org/doi/10.1021/acsenergylett.3c02385.
17) Samanta A, Williamson S. Machine Learning-based Remaining Useful Life Prediction Techniques for Lithium-ion Battery Management Systems: A Comprehensive Review. *IEEJ Journal of Industry Applications*. 2023;12(4):563–574. Available from: https://www.researchgate.net/publication/368970467_Machine_Learning-based_Remaining_Useful_Life_Prediction_Techniques_for_Lithium-ion_Battery_Management_Systems_A_Comprehensive_Review.
18) Zhang M, Yin J, Chen W. SOH estimation and RUL prediction of lithium batteries based on multidomain feature fusion and CatBoost model. *Energy Science & Engineering*. 2023;11(9):3082–3101. Available from: https://doi.org/10.1002/ese3.1506.
19) Pang X, Yang W, Wang C, Fan H, Wang L, Li J, et al. A novel hybrid model for lithium-ion batteries lifespan prediction with high accuracy and interpretability. *Journal of Energy Storage*. 2023;61. Available from: https://doi.org/10.1016/j.est.2023.106728.
20) Chen M, Ma G, Liu W, Zeng N, Luo X. An overview of data-driven battery health estimation technology for battery management system. *Neurocomputing*. 2023;532:152–169. Available from: https://doi.org/10.1016/j.neucom.2023.02.031.
21) Geslin A, van Vlijmen B, Cui X, Bhargava A, Asinger PA, Braatz RD, et al. Selecting the appropriate features in battery lifetime predictions. *Joule*. 2023;7(9):1956–1965. Available from: https://doi.org/10.1016/j.joule.2023.07.021.
22) Rauf H, Khalid M, Arshad N. A novel smart feature selection strategy of lithium-ion battery degradation modelling for electric vehicles based on modern machine learning algorithms. *Journal of Energy Storage*. 2023;68. Available from: https://doi.org/10.1016/j.est.2023.107577.