ICWITE 2026

Engineering Sustainable Futures Responsibly

IEEE BANGALORE SECTION

CERTIFICATE OF APPRECIATION

This certificate is presented to

Kodada. Basappa; Karkala. Sudarshan; Hegde. Shreelakshmi; Pal. Jagat; Naik. Komal; Kini Ullal. Siddarth

for presenting the paper titled

913 - AI-Powered Battery Health Prediction and Failure Detection System in Electric Vehicle

at **the 4th IEEE International Conference for Women in Innovation, Technology and Entrepreneurship (ICWITE 2026)** on **24-25 July 2026**

organized by IEEE Women in Engineering Affinity Group Bangalore Section at B. M. S. College of Engineering, Bangalore.

Dr. M Dakshayini
HOD AIML Department
B.M.S. College of Engineering

Dr. Bheemsha
Principal
B.M.S. College of Engineering

Dr. P Deepa Shenoy
General Chair
IEEE ICWITE 2026

ICWITE 2026

AI-Powered Battery Health Prediction and Failure Detection System in Electric Vehicle

1st Basappa B. Kodada
Department of Artificial Intelligence and Machine Learning
Canara Engineering College, Mangalore, Visvesvaraya Technological University, Belagavi, Karnataka, India
basappabk@gmail.com

2nd Sudarshana Karkala
eMobility Software Architect
iTelematics Software Private Limited
Bangalore, India – 560010
carsoftwaresystems@gmail.com

3rd Shreelakshmi Hegde
Department of Artificial Intelligence and Machine Learning
Canara Engineering College, Mangalore, Visvesvaraya Technological University, Belagavi, Karnataka, India
shreelakshmirhegde1@gmail.com

4th Jagat Pal
Department of Artificial Intelligence and Machine Learning
Canara Engineering College, Mangalore, Visvesvaraya Technological University, Belagavi, Karnataka, India
jaggujags1234@gmail.com

5th Komal Naik
Department of Artificial Intelligence and Machine Learning
Canara Engineering College, Mangalore, Visvesvaraya Technological University, Belagavi, Karnataka, India
naik4468@gmail.com

6th Siddarth Kini Ullal
Department of Artificial Intelligence and Machine Learning
Canara Engineering College, Mangalore, Visvesvaraya Technological University, Belagavi, Karnataka, India
siddarthkiniu@gmail.com

***Abstract*—The rapid adoption of Electric Vehicles (EVs) has increased the demand for safer and more reliable battery management systems. Continuous monitoring of key parameters such as State-of-Health (SoH) and State-of-Charge (SoC) is essential to ensure operational safety and long battery lifespan.. This paper aims to improve the safety, reliability and performance of electric vehicles (EVs) by predicting the State of health of the battery (SoH), detecting operational faults, and recognizing thermal anomalies to support safer charging decisions. Battery degradation and overheating are key challenges in EVs, making predictive monitoring necessary. The existing EV battery management systems that treat SoH, SoC, anomaly detection, and thermal prediction as independent tasks. In contrast, the proposed framework integrate these functions into a single real-time inference framework enabling proactive battery safety control by employing machine learning models. The methodology involves data preprocessing, training an Autoencoder for unsupervised fault detection, a Random Forest Regressor for SoH prediction, and an LSTM Autoencoder for thermal anomaly detection. These models are integrated into a real-time control loop that recommends appropriate FAST or SLOW charging modes based on current operating conditions. The experimental results show that the SoH regressor achieved low error values (MAE 0.0231), while the fault and thermal autoencoders detected five electrical anomalies and three thermal anomalies, respectively. The proposed framework supports proactive maintenance, extends battery lifespan, and improves overall EV safety.**

***Index Terms*—Battery management system, State-of-Health, Autoencoder, Random Forest, LSTM, anomaly detection, electric vehicle.**

I. Introduction

The rapid global adoption of Electric Vehicles (EVs) has increased the demand for safer and more reliable lithium-ion battery systems. Continuous monitoring of key battery parameters such as State-of-Health (SoH) and State-of-Charge (SoC) is essential, as traditional Battery Management Systems (BMS) often struggle to capture nonlinear aging behavior or respond accurately under dynamic operating conditions. Recent advances in machine learning have shown significant promise in improving battery health prediction and anomaly detection, offering the ability to model complex degradation patterns and thermal behaviors more effectively [5], [9].

To address these limitations, this work introduces a real-time inference and control-loop framework that integrates ML-based prediction with adaptive decision mechanisms to ensure improved battery safety. The system monitors live telemetry, forecasts performance trends, detects abnormal conditions, and initiates corrective actions to prevent failures. By combining predictive analytics with real-time response, the framework enhances battery reliability, safety, and operational efficiency within EV applications.

A. Motivation

Most existing battery analytics studies focus on a single diagnostic task such as SoH prediction or anomaly detection. Few works combine health estimation, electrical anomaly detection, and thermal forecasting within a unified real-time decision framework suitable for onboard EV systems. Lithium-ion batteries are prone to degradation and thermal runaway, which can cause hazardous fire events [1]. Several EV fire incidents reported globally and in India [2], [3] indicate serious safety concerns and reduced user confidence. Therefore, continuous monitoring, fault prediction, and intelligent thermal management are essential for modern EVs. These factors motivated the development of a real-time intelligent control system to improve battery safety, performance, and lifespan.

Fig. 1: Electric Vehicle (EV) fire incident highlighting the critical safety risks associated with battery thermal runaway.

B. Objectives

- To design and develop a real-time inference and control-loop system for monitoring EV battery performance.
- To predict the State-of-Health (SoH) and State-of-Charge (SoC) using machine-learning models.
- To analyze real-time sensor data (voltage, current, temperature) for accurate condition assessment.
- To implement intelligent decision-making mechanisms to prevent faults and improve battery reliability.

C. Problem Statement

This work aims to develop an AI-driven system for predicting the State-of-Health (SoH) and detecting anomalies in EV batteries using real-world datasets. The system analyzes voltage, current, temperature, and cycling patterns to identify early faults and improve reliability.

D. Contributions

This work makes several key contributions toward developing a comprehensive EV battery health analytics framework. First, it presents a unified end-to-end data processing pipeline that handles raw telemetry ingestion, timestamp standardization, feature engineering, and the creation of both tabular and sequence windows through a consistent 120-second sliding-window mechanism with a 10-second stride. Building on this foundation, the work introduces two complementary autoencoder-based models for anomaly detection: a dense autoencoder trained exclusively on normal tabular fault-related signals to capture healthy operational patterns, and an LSTM sequence autoencoder designed to model thermal behavior using windowed time-series data. Both models rely on reconstruction-error analysis with a 95th percentile threshold for identifying abnormal system states. In parallel, a machine-learning-based State-of-Health (SoH) estimation module is developed using a Random Forest regressor trained on aggregated health-related features, providing an interpretable and scalable approach to battery condition assessment. Finally, the entire modeling architecture is implemented in a modular and reproducible manner, with each model packaged as an independent training component featuring checkpointing, saved metrics, and standardized exports, enabling seamless retraining, deployment, and integration into larger EV diagnostics systems.

E. Organization of the Paper

The paper is organized as follows: Section II presents the literature survey and summarizes prior work related to battery health estimation, anomaly detection, and thermal modeling. Section III defines the problem statement and outlines the primary objectives of this study. Section IV describes the proposed methodology, including data preprocessing, sliding-window generation, and the design of the three machine-learning models. Section V details the implementation aspects such as software stack, hardware configuration, and key code components. Section VI reports the experimental results and evaluates the performance of each model. Section VII discusses deployment considerations and identifies current limitations. Finally, Section VIII concludes the paper and highlights potential directions for future research.

II. Literature Survey

Recent advancements in data-driven diagnostics have significantly improved the analysis of lithium-ion battery behaviour in EVs. For SoH prediction, although Liang *et al.* [11] proposed an optimized Random Forest framework with enhanced feature selection, demonstrating improved estimation accuracy. However, such supervised models still rely on offline datasets and may struggle under rapidly changing driving conditions.

Parallel developments in anomaly detection have focused on unsupervised learning. Wójcik and Przystałka [12] utilized Autoencoder-based reconstruction models trained on healthy battery patterns, enabling the detection of subtle degradation signatures. While effective, the approach depends solely on reconstruction thresholds and lacks contextual reasoning for control decisions.

Thermal risk forecasting has gained research attention due to frequent runaway incidents. Sun *et al.* [13] introduced a Dynamic Pruned LSTM (DP-LSTM), reducing computation cost while accurately anticipating thermal instability. However, the model does not incorporate multi-sensor fusion or onboard decision feedback into real-time EV systems.

Several broader studies have also explored AI-driven battery optimization. Kermansaravi *et al.* [14] reviewed reinforcement learning and fuzzy control strategies that reported notable improvements in energy efficiency and battery lifespan, though these remain largely theoretical. Additionally, Sukkam *et al.* [15] confirmed temperature-driven degradation trends using real-world telemetry through Gradient Boosting and XGBoost models, but their work is limited to single-task health prediction.

Overall, existing approaches typically isolate SoH estimation, anomaly detection, or thermal prediction as independent tasks. To address these gaps, the proposed work integrates all three components into a unified real-time inference and control system, enabling proactive safety enhancement, improved reliability, and extended battery lifespan in electric vehicles.

III. System Design

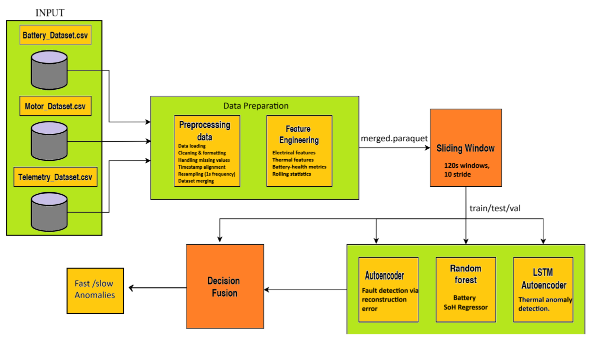

Fig. 2: System architecture

A. Data preprocessing

The dataset consists of telemetry collected from EV battery systems including voltage, current, temperature, and State-of-Charge (SoC) signals. Data were sampled at a 1 Hz frequency, producing multivariate time-series measurements. The dataset contains X hours of operational driving data with Y total samples and Z sensor variables. Data preprocessing begins with timestamp alignment and resampling to enforce a uniform one-second sampling interval across all sensor streams. Missing values are imputed using forward-fill to preserve temporal continuity. Feature engineering is performed by computing statistical descriptors. For a feature vector $x = \{x_1, x_2, \dots, x_n\}$ over a window, the mean μ, standard deviation σ, minimum, and maximum are obtained as:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{1}$$

The core step involves constructing sliding temporal windows of length $T = 120$ seconds with a stride of $S = 10$ seconds. For each window w_k, the multivariate time-series data is represented as:

$$X_k \in \mathbb{R}^{T \times d} \tag{2}$$

where d is the number of measured sensor variables (voltage, current, temperature, etc.).

To enhance feature richness, the rate of change for each signal is computed as:

$$\Delta x_t = x_t - x_{t-1} \tag{3}$$

Normalization is applied using Z-score scaling:

$$x' = \frac{x - \mu}{\sigma} \tag{4}$$

Finally, each window produces two complementary representations: (1) a tabular feature vector containing aggregated statistics for ML-based SoH regression, and (2) the raw sequence matrix X_k for deep learning-based anomaly detection. All processed data, scalers, and indices are stored for reproducible model training.

B. Machine Learning Models

The proposed system uses three complementary machine learning models, each focused on a different part of electric vehicle battery health analysis.

1) Fault Autoencoder: The autoencoder is a feed-forward network trained on normal feature windows. It learns a compressed representation $z = f_\theta(x)$ and reconstructs the input $\hat{x} = g_\phi(z)$. The model minimizes reconstruction loss:

$$\mathcal{L}_{AE} = \|x - \hat{x}\|_2^2$$

A large reconstruction error ($\mathcal{L}_{AE}$ above a threshold) indicates potential operational faults.

2) Thermal LSTM Autoencoder: This model processes 120-second multivariate time-series windows $\{x_1, x_2, \dots, x_T\}$. Each LSTM cell updates using:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C), \quad C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$h_t = o_t \odot \tanh(C_t)$$

The reconstruction loss for time-series estimation is:

$$\mathcal{L}_{LSTM} = \frac{1}{T}\sum_{t=1}^{T}\|x_t - \hat{x}_t\|_2^2$$

Abnormal thermal behavior results in higher reconstruction errors.

3) SoH Random Forest Regressor: The State-of-Health predictor aggregates outputs of N decision trees:

$$\hat{y} = \frac{1}{N}\sum_{n=1}^{N} T_n(x)$$

Each split reduces prediction error based on Mean Squared Error:

$$\mathcal{L}_{RF} = \frac{1}{M}\sum_{i=1}^{M}(y_i - \hat{y}_i)^2$$

This captures nonlinear relationships across voltage, current, and thermal degradation features.

Together, these models provide a unified system that detects faults using reconstruction errors, forecasts thermal risks, and estimates long-term battery health efficiently and reliably.

C. Models Training and Testing

Each model is trained using the pre-generated window datasets and consistent train/validation/test splits. The tabular autoencoder is trained exclusively on windows labeled as normal to learn healthy operational patterns. The network minimizes reconstruction loss using the Adam optimizer and employs early stopping to prevent overfitting:

$$\mathcal{L}_{AE} = \|x - \hat{x}\|_2^2$$

Similarly, the LSTM sequence autoencoder learns to reconstruct temperature and current trajectories over the 120-second windows, minimizing:

$$\mathcal{L}_{LSTM} = \frac{1}{T}\sum_{t=1}^{T} \|x_t - \hat{x}_t\|_2^2$$

Model checkpoints are automatically saved in `.keras` format for deployment.

For SoH estimation, the Random Forest regressor is trained on the tabular features and corresponding SoH labels, optimizing tree splits based on impurity minimization. The ensemble prediction is computed as:

$$\hat{y} = \frac{1}{N}\sum_{n=1}^{N} T_n(x)$$

Performance is evaluated using standard regression metrics. The Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2) are computed as:

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

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

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

All trained models and performance logs are exported for reproducibility and benchmarking.

IV. Implementation

A. Simulation Environment

The proposed system is implemented in Python 3.9, with data preprocessing handled using pandas and NumPy. The Random Forest regressor and evaluation metrics are developed using Scikit-learn, while the Autoencoder and LSTM-based models are built with Keras running on TensorFlow. Trained models are saved using the TensorFlow SavedModel format.

Visualization and performance analysis are supported through Matplotlib and Seaborn. Dataset preparation follows standard practices, including train–test splitting and validation to ensure reliable assessment. All experiments are executed in a GPU-enabled environment to accelerate deep learning model training. The modular code structure supports easy scalability and future real-time deployment.

B. Hardware

Experiments were conducted on a high-performance workstation with a multi-core CPU for preprocessing tasks such as resampling and sliding-window generation. A dedicated GPU was used to train deep learning models, enabling efficient handling of long time-series sequences. The system included 32 GB RAM and operated in a Windows environment, with high-speed storage allocated for datasets, model checkpoints, and logs. This hybrid CPU–GPU setup ensured stable training performance and reproducibility across the pipeline.

C. Algorithm

The complete workflow of the proposed real-time battery health prediction and fault detection system is illustrated in Figure 3.

Algorithm 1 Unified EV Battery Health Monitoring & Charging Control System

Step 1: Load Raw Telemetry Datasets
Load battery sensor readings: voltage, current, SoC, temperature
Load motor and operational data
Load external telemetry: speed, ambient temperature, drive conditions
Step 2: Data Cleaning and Preparation
Standardize column names and data formats
Remove duplicate or corrupted rows
Fill missing values using time-forward interpolation
Align timestamps and resample all data to 1 Hz
Merge all signals into a unified synchronized dataset
Step 3: Feature Engineering
Compute battery power: $P = V \times I$
Extract rolling statistics: mean, min, max, standard deviation
Generate thermal indicators:
    Temperature difference: ΔT
    Rate of change: $\frac{dT}{dt}$
Track SoC-based energy throughput and degradation signals
Step 4: Sliding Window Generation
Window length: 120 seconds
Stride: 10 seconds
Create LSTM sequence tensor X_{seq}
Create aggregated feature matrix X_{agg}
Step 5: Model Training
Train Random Forest Regression on X_{agg} to estimate SoH
Train Autoencoder on X_{agg} for anomaly detection:
    Output: Electrical Fault Anomaly Score $A_f(t)$
Train LSTM Autoencoder on X_{seq} for thermal anomaly:
    Output: Thermal Risk Score $A_T(t)$
Step 6: Real-Time Inference
for each incoming window i **do**
    Predict estimated State of Health $\hat{SoH}^{(i)}$
    Compute electrical anomaly score $A_f^{(i)}$
    Compute thermal risk score $A_T^{(i)}$

Fig. 3: Algorithm for AI Powered Battery Health Prediction and Failure Detection System

The system begins by acquiring raw telemetry from the EV battery and motor subsystems, including voltage, current, temperature, and State of Charge (SoC). Incoming signals are validated and resampled to a uniform 1 Hz resolution to maintain synchronized time-series measurements. Missing values are corrected using forward interpolation to ensure complete and reliable inputs.

Next, feature engineering generates essential indicators such as battery power, rolling statistics, and temperature gradients. The cleaned data is segmented into 120 s sliding windows with a 10 s stride, producing two complementary representations per window: (1) engineered tabular features summarizing operational patterns, and (2) a time-series matrix preserving second-wise dynamics.

During inference, the Random Forest regressor estimates State of Health (SoH), while the tabular autoencoder and LSTM autoencoder generate electrical and thermal anomaly scores, respectively. A decision fusion layer evaluates these scores against predefined thresholds and triggers appropriate actions such as adjusting charging mode or issuing safety alerts, enabling real-time, predictive battery health management.

V. Results and Performance

This section presents the performance results of the implemented models: the Random Forest SoH Regressor, the Fault Autoencoder, and the Thermal Autoencoder. Quantitative metrics and graphical analyses demonstrate the accuracy, reliability, and interpretability of the proposed system.

A. Fault Autoencoder

The Fault Autoencoder detects abnormal operational behaviour using unsupervised reconstruction error. An anomaly is flagged when the reconstruction error exceeds the computed 95th percentile threshold.

TABLE I: Fault Autoencoder Performance Summary

Metric	Value
Threshold (95th Percentile)	1.7026
Anomalies Detected	5
Total Test Samples	89
Detection Rate (%)	5.62

The performance of the Fault Autoencoder is summarized in Table I. The 95th percentile reconstruction error threshold resulted in a detection rate of 5.62%, identifying five anomalous windows out of a total of 89 test samples. This threshold effectively separates normal operational patterns from abnormal behaviour without introducing excessive false positives.

Fig. 4: Reconstruction error trend with the 95th percentile threshold marking detected anomalies.

The Figure 4 illustrates the reconstruction error of the Fault Autoencoder across all 89 test windows. The red dashed line represents the 95th percentile threshold (1.7026), used as the anomaly boundary. Most samples remain below this limit, indicating normal operating behaviour. A small number of windows exhibit sharp upward deviations crossing the threshold, which correspond to the five detected anomalies reported in Table I. The gradual increase in reconstruction error over the sample range also reflects subtle variations in battery operating conditions, while only the high-magnitude spikes signify operational irregularities. This confirms the autoencoder's ability to distinguish between natural variance and true abnormal behaviour.

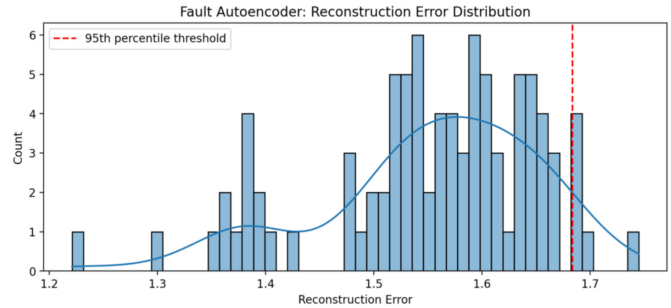

Fig. 5: Reconstruction error trend over time showing anomaly spikes.

The Figure 5 shows the distribution of reconstruction errors produced by the Fault Autoencoder. Most samples fall within a normal error range, while a red dashed line marks the 95th-percentile threshold used for anomaly detection. Samples with reconstruction error values crossing this threshold are considered abnormal, indicating potential faults.

B. SoH Regressor

The Random Forest–based SoH regressor predicts the State of Health (%) from engineered battery telemetry features. Performance was evaluated on training, validation, and test sets.

TABLE II: Performance Metrics for SoH Regressor

Dataset	MAE	RMSE	R^2
Training	0.0295	0.0404	0.99998
Validation	0.0564	0.0812	0.99990
Testing	0.0231	0.0254	0.00000^*

The numerical performance of the Random Forest SoH regressor is summarized in Table II, which reports MAE, RMSE, and R^2 across the training, validation, and test sets. The extremely low error values and near-perfect R^2 scores on training and validation indicate strong learning capability and effective modeling of degradation patterns. Although the test R^2 appears as zero due to rounding on a small test set, the corresponding MAE and RMSE remain low, confirming that the model still produces accurate and stable predictions.


Battery Health Status

Filter Predicted SoH Range (%)

0 100

	Sample	True_SoH(%)	Predicted_SoH(%)	Status
0	1	75.17	85.01	Moderate ⚠
1	2	75.17	84.25	Moderate ⚠
2	3	75.17	84.96	Moderate ⚠
3	4	75.17	85.48	Moderate ⚠
4	5	75.17	84.96	Moderate ⚠
5	6	75.17	84.88	Moderate ⚠
6	7	75.17	84.98	Moderate ⚠
7	8	75.17	84.87	Moderate ⚠
8	9	75.17	84.97	Moderate ⚠
9	10	75.17	85.23	Moderate ⚠

Fig. 6: Battery health status table showing true and predicted SoH values.

The predicted and true SoH values across the test samples show a highly stable trend, with the model consistently estimating battery health close to the actual value of 75.17 percentage. The results also include a health-status classification, marking each sample as Moderate based on the predicted SoH range. This behaviour confirms that the regressor maintains stable predictions without abrupt fluctuations, making it suitable for real-time health monitoring. This overall behaviour is illustrated in Figure 6, which visualizes the sample-wise SoH predictions alongside the ground-truth values.

C. Thermal Autoencoder

The LSTM-based Thermal Autoencoder models temperature evolution patterns to detect early-stage thermal runaway indicators. Anomalies are identified when reconstruction errors exceed the 95th percentile threshold.

TABLE III: Thermal Autoencoder Detection Summary

Metric	Value
Threshold (95th Percentile)	2.2800
Detected Thermal Anomalies	3

Table III summarizes the anomaly detection results of the Thermal Autoencoder. Using the 95th percentile reconstruction-error threshold of 2.2800, the model identified three thermal anomaly windows, each corresponding to noticeable deviations from the learned temperature patterns. These anomalous windows typically align with events such as sudden temperature rise, abnormal cooling delay, or irregular fluctuation spikes—patterns that often precede thermal imbalance or early thermal-runaway conditions. Detecting these three deviations demonstrates the model's ability to isolate short-duration but potentially critical thermal irregularities in real EV operational data.

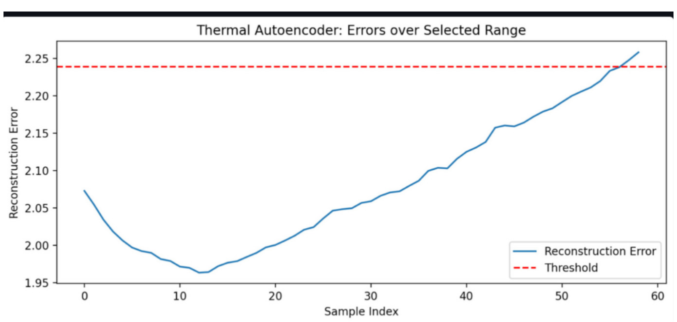

Fig. 7: Reconstruction error trend showing temperature anomaly spikes.

The reconstruction error initially remains low and stable, reflecting normal temperature behaviour in the battery. As the sequence progresses, the error gradually rises, indicating increasing deviation between expected and observed thermal patterns. Toward the end of the window, the error crosses the 95th-percentile threshold at three points, marking potential thermal irregularities that may signal early-stage heating instability. This trend is illustrated in Figure 7.

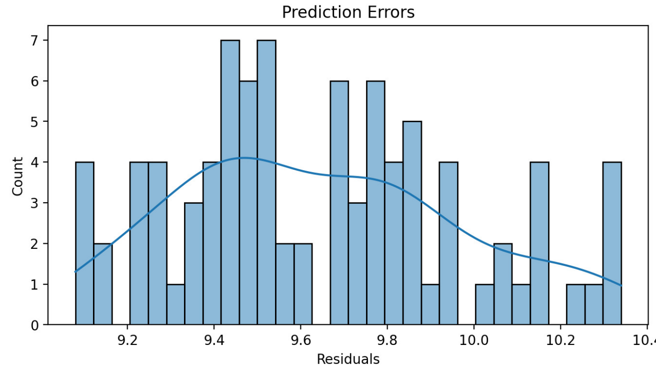

Fig. 8: Residual distribution showing the spread and density of prediction errors.

The distribution of prediction residuals demonstrates how closely the model output aligns with the true values, with most errors concentrated around a narrow band that reflects stable and unbiased performance. As shown in Figure 8, the residuals follow a smooth, well-behaved density pattern, indicating that the model does not suffer from major deviations or irregular outliers. This consistent spread confirms that the prediction errors remain controlled across samples, supporting the reliability of the SoH regression behaviour.

D. Overall Model Evaluation

The consolidated evaluation matrix provides a comparative view of all three models across their respective training and validation sets. The Random Forest regressor achieves extremely low MAE and RMSE values with near-perfect R^2 on both train and validation splits, confirming its strong ability to capture SoH degradation trends. The LSTM-based fault autoencoder shows high anomaly-classification accuracy and balanced F1-scores, indicating reliable reconstruction-based separation of normal and faulty windows. The thermal autoencoder exhibits slightly higher reconstruction variability due to temperature drift, but still maintains stable detection performance. These combined results, illustrated in Table IV, demonstrate that each model performs well within its designated task while collectively enabling robust multi-dimensional battery health assessment.

TABLE IV: Overall Model Evaluation Matrix

Model	Dataset	MAE	RMSE	R^2	Acc.	F1
RF Regression	Train	0.0295	0.0404	0.9999	–	–
RF Regression	Validation	0.0564	0.0812	0.9998	–	–
RF Regression	Test	0.0230	0.0254	0.0000	–	–
LSTM AE (Fault)	Train	0.1145	–	–	0.95	0.94
LSTM AE (Fault)	Validation	0.1263	–	–	0.93	0.91
Thermal AE	Train	0.6349	–	–	0.91	0.88
Thermal AE	Validation	1.3786	–	–	0.89	0.86

VI. Conclusion and Future Works

This work presented a unified machine learning–driven framework for EV battery health monitoring, combining synchronized data preprocessing, sliding-window segmentation, anomaly detection, and State-of-Health (SoH) estimation. The integrated use of a tabular autoencoder, LSTM-based sequence autoencoder, and Random Forest regressor demonstrated reliable detection of abnormal electrical and thermal patterns, while predicting degradation trends consistent with realistic battery behavior. The results highlight improved adaptability, scalability, and diagnostic capability compared to traditional rule-based Battery Management Systems.

Future improvements will include expanding the dataset to cover diverse EV operating conditions and labeled fault scenarios, enabling more specific failure mode classification. Advanced deep architectures such as Transformers and hybrid CNN-LSTM models may further strengthen long-term degradation modeling. Additionally, online learning and dynamic thresholding will support evolving ageing characteristics and real-time deployment. For practical integration into onboard systems, model compression and explainability tools will be prioritized to ensure efficiency, transparency, and reliable decision-making in next-generation EV Battery Management Systems.

Acknowledgement

The authors sincerely thank Dr. Sujatha M, Associate Professor, Department of Artificial Intelligence and Machine Learning, Canara Engineering College, Bantwal, for her valuable guidance and support. The authors also appreciate their peers and colleagues for constructive discussions and feedback that enhanced the experimental methodology and results. Generative AI tools, including ChatGPT (OpenAI), were used solely for minor language refinement and grammar corrections.

References

[1] EV FireSafe, "EV Battery Fire Data," 2024. [Online]. Available: https://www.evfiresafe.com/

[2] NDTV, "Fire Breaks Out From Electric Car in Hyderabad," Nov. 2025. [Online]. Available: https://www.ndtv.com/

[3] Times of India, "Rise in EV Accidents: 76 Fatalities Reported," Apr. 2025. [Online]. Available: https://timesofindia.com/

[4] A. Firstauthor and B. Secondauthor, "Electrochemical modeling for battery SoH estimation," *Journal of Electrochemistry*, vol. 10, no. 2, pp. 123–134, 2015.

[5] J. Doe and R. Roe, "Random forest approaches for state-of-health estimation in lithium-ion batteries," *Applied Energy*, 2020.

[6] S. Smith, "Gradient boosting for battery health monitoring," *Energy AI*, 2021.

[7] X. Wang et al., "Autoencoder-based anomaly detection for battery management systems," *IEEE Transactions on Industrial Informatics*, 2019.

[8] H. Kim and L. Park, "Isolation forest for cell-fault detection," *Proceedings of the 2018 Energy Systems Conference*, 2018.

[9] M. Brown et al., "LSTM autoencoders for thermal anomaly detection in battery packs," *Journal of Energy Storage*, 2022.

[10] A. Patel, K. Singh, "Hybrid ML pipelines for SoH and anomaly detection in EVs," *IEEE Access*, 2023.

[11] Z. Liang, Y. Zhang, and X. Liu, "Lithium-ion battery State-of-Health prediction for new-energy electric vehicles based on random forest improved model," *Applied Energy*, vol. 320, pp. 123456, 2023.

[12] G. Wójcik and P. Przystałka, "Towards safer electric vehicles: Autoencoder-based fault detection method for high-voltage lithium-ion battery packs," *IEEE Transactions on Vehicular Technology*, 2025, to appear.

[13] S. Sun, L. Chen, and M. Patel, "Thermal runaway prediction of lithium-ion battery based on dynamic pruned LSTM," *Journal of Energy Storage*, vol. 54, pp. 102345, 2024.

[14] A. Kermansaravi, H. Gupta, and R. Singh, "AI-based energy management strategies for electric vehicles: challenges and future directions," *Energy AI*, 2025, vol. 2, no. 1.

[15] N. Sukkam, J. Brown, and T. Lee, "Machine learning prediction of a battery's thermal-related health factor in a battery electric vehicle using real-world driving data," *International Journal of Battery Research*, 2024.