FRONTIERS IN
ENGINEERING | SCIENCE | MANAGEMENT
MNS-UET, MULTAN

FRONTIERS IN ENGINEERING, SCIENCE & MANAGEMENT

https://journal.mnsuet.edu.pk/

Research Paper

Predictive Energy Management and Safety Optimization for E-Bikes: A Sensor Fusion and Machine Learning Approach

Hammad Ahmad [a], Saad Ahmad [b], Muhammad Umar Chaudhry [a, *], Dua Haroon [a], Zeeshan Ramzan [b]

[a] Department of Computer Engineering, Bahauddin Zakariya University, Multan 66000, Pakistan.

[b] Department of Mechanical Engineering, Muhammad Nawaz Sharif University of Engineering and Technology, Multan 66000, Pakistan

ARTICLE INFORMATION

Received	22 Aug 2026
Received in revised form	09 Sep 2026
Accepted	11 Sep 2026
Available Online	15 Sep 2026

*CORRESPONDING AUTHOR

Muhammad Umar Chaudhry
Department of Computer Engineering, Bahauddin Zakariya University, Multan, 66000, Pakistan.

Email umar.chaudhry@bzu.edu.pk

ABSTRACT

Purpose: In the broader context of sustainable micro-mobility, the operational efficiency and range reliability of electric bikes (e-bikes) face significant challenges, especially when rider and terrain conditions are unpredictable. This research introduces a smart, AI-based e-bike monitoring system that leverages multi sensor data to address these limitations and also combines operational, biometric, and topographic data in real-time to create a thorough foundation for improved vehicle intelligence.
Design/Methodology: Refined feature extraction and filtering approaches were therefore applied to modify raw sensor inputs from a multimodal dataset that includes rider weight, instantaneous speed, journey distance, battery discharge efficiency per kilometers and road inclination. Machine learning models that predict the number of battery charges needed for particular distances under dynamic riding situations were built on the basis of these processed data streams. Finally, the performance of these models was assessed with the dual goals of optimizing battery energy management and improving overall rider safety.
Findings: By combining CNN (convolutional neural network) - based pre-processing with real-time sensor fusion, the use of trained regression approaches enhances battery state of charge (SOC) estimation and charge-interval forecasting, which are essential for reducing range anxiety. Random Forest achieved the highest SOC accuracy (R^2 = 0.9997, MAPE = 0.13%), while Linear Regression best predicted charge intervals (R^2 = 0.9355, MAPE = 5.45%). These findings highlight the importance of integrated sensing and machine learning in developing sustainable, next-generation energy efficient electric bikes.
Originality: This study presents a novel architecture that integrates real-time biometric, operational, and topographic data streams using a combined CNN preprocessing and regression framework specifically tailored to e-bike vehicle intelligence and dynamic energy optimization.

Keywords: E-Bike Monitoring, Convolutional Neural Network, Battery State of Charge (SOC), Rider Weight, Speed, Distance, Battery Discharge Efficiency Per kilometers.

1. Introduction

The increasing demand for sustainable transportation options has led to the growing popularity of electric vehicles (EVs), particularly electric bikes (e-bikes), as an energy-efficient and environmentally beneficial mode of transportation (Shyam et al., 2026). E-bikes not only offer affordable personal transportation but also have the ability to drastically lower carbon emissions, traffic jams, and dependency on fossil fuels (Li et al., 2021). However, battery performance continues to be one of the key factors affecting the dependability, efficacy, and utility of e-bikes (Qian et al., 2024). In order to guarantee operational effectiveness and increase user trust in electric transportation systems, precise battery estimation of the State of Charge (SOC) and energy consumption is essential (Gu et al., 2021; Y. Yang et al., 2018).

One of the biggest concerns of e-bike riders is range anxiety, or the uncertainty about the battery's remaining capacity and the ability to complete a trip without abrupt power loss (Bhosale and Chavan, 2026). In favor of simple measurements like voltage levels or remaining battery percentage, traditional battery monitoring systems often ignore dynamic factors like rider characteristics, terrain conditions, speed fluctuations, and environmental influences. Therefore, in real-world operational conditions, these techniques may result in inaccurate forecasts (Hieu and Lim, 2023). Combining data from many sensors allows for the recording of operational, environmental, and rider-related aspects that significantly affect battery behavior. Machine learning algorithms can then evaluate these complex relationships to generate accurate predictions regarding battery performance, energy consumption, and remaining useful charge (L. Zhang et al., 2021). Furthermore, data preparation techniques such as Convolutional Neural Networks (CNNs) have demonstrated results in improving prediction accuracy and removing noise from sensor data (Hieu and Lim, 2024).

Monitring and optimaization of Battery State of Charge (SOC) and Remaining Useful Life (RUL) plays an important role to overcome the above mentioned issues. Various studies have been done to calculate SOC and RUL using operational data, environmental factors, and route parameters (Burani et al., 2022). While these approaches have demonstrated promising outcomes, many of them just focus on battery health monitoring rather than simultaneously addressing rider safety and charge planning under dynamically changing riding conditions. Furthermore, there hasn't been much focus on combining rider biometric information and battery discharge efficiency into a unified predictive framework.

The proposed method uses multimodal sensor data, including rider weight, vehicle speed, journey distance, road incline, and battery discharge efficiency per kilometer. CNN-based preprocessing is employed to enhance data quality, while machine learning regression models are used to predict battery charge requirements for certain travel distances under different operating situations. The developed system aims to improve battery State of Charge estimate, reduce range anxiety, optimize energy utilization, and improve rider safety, particularly on challenging terrain. The results demonstrate how advanced analytics and integrated sensing can be used to develop sustainable, energy efficient.

1.1. Previous Work

The growing popularity of e-bikes and electric cars has increased research interest in battery management, energy efficiency, and sustainable transportation systems. Prior studies have highlighted the financial and ecological benefits of personal electric vehicles while emphasizing the importance of efficient energy use and battery performance for enhancing consumer acceptability and operational dependability.

Recent studies have further emphasized the role of intelligent energy management in electric mobility. Data-driven and machine-learning-based approaches are increasingly being investigated because they can model complex relationships between vehicle operating conditions and energy demand without requiring complete analytical representations of the vehicle. Ensemble learning has been applied to electric vehicle energy consumption prediction using variables such as trip speed, distance, and road gradient, demonstrating the potential of combining multiple vehicle and route-related factors for improved prediction accuracy (Ullah et al., 2021). Similarly, machine-learning approaches have been used to predict battery electric-vehicle driving range while incorporating a broader set of vehicle, battery, and traffic related variables (Cui et al., 2025; Li et al., 2021). Real-world driving data studies have also demonstrated that energy consumption prediction can support range estimation, route planning, and charging infrastructure decisions (Huang et al., 2022; Tao et al., 2023; Wang et al., 2023).

State of Charge (SOC) estimation, Remaining Useful Life (RUL) prediction, and cardiac health monitoring have all been the subject of extensive research in recent years. Many data-driven approaches have been proposed to improve battery reliability and operating efficiency (Rincón-Maya et al., 2025). Machine learning algorithms have demonstrated great potential in predicting battery deterioration, calculating battery life, and improving battery management system performance (Bhosale and Chavan, 2026; Cui et al., 2025). When compared to conventional estimation techniques, sophisticated prediction models built on intelligent algorithms have demonstrated greater accuracy and resilience.

The development of data-driven battery state estimation has progressed from conventional machine-learning methods toward increasingly advanced deep-learning and hybrid architectures. Reviews of machine-learning-based SOC estimation have identified supervised learning, recurrent neural networks, convolutional architectures, and hybrid approaches as important directions for improving estimation under nonlinear operating conditions (Chen et al., 2025). Deep-learning-based methods have also been investigated for extracting temporal and nonlinear relationships directly from battery measurements, reducing the dependence on manually defined battery models (Gu et al., 2021; Hussain et al., 2025). More recent research has explored attention mechanisms, transfer learning, and hybrid architectures to improve robustness when battery operating conditions differ between training and testing datasets (Yun et al., 2025).

Recent research has also focused on improving the practical robustness of SOC estimation. Machine-learning models have been evaluated under varying temperatures and dynamic load profiles, showing that operating conditions can strongly affect estimation performance (Gok et al., 2025; Qian et al., 2024). Cross domain transfer learning approaches have been proposed to reduce the dependence on large amounts of training data while maintaining SOC estimation accuracy under different driving conditions (Zhao et al., 2025). These developments indicate that accurate SOC estimation increasingly requires models that can generalize across changing battery and operating conditions rather than relying only on fixed laboratory conditions.

Research on electric-vehicle energy prediction has increasingly incorporated both static route information and dynamic vehicle operating variables. Deep-neural-network approaches have been developed to combine route level and segment level information for predicting energy consumption, indicating that route characteristics and temporal driving conditions can provide complementary information for energy forecasting (Wang et al., 2024). Other studies have investigated machine-learning based driving-range prediction using real-world operating information, demonstrating the importance of considering changing driving conditions when estimating available range (Chen et al., 2025). More recently, data-driven range prognostics have combined predicted vehicle speed with machine-learning-based energy and power estimation to obtain dynamic driving-range predictions (Zhang et al., 2021). Such approaches demonstrate the increasing shift from simple distance-based range estimation toward condition-aware energy prediction.

For electric bicycles specifically, experimental and modelling studies have shown that rider mass, road grade, aerodynamic conditions, and other operating parameters influence the required power and electrical energy consumption (Li et al., 2021; Ye et al., 2025). These findings are particularly relevant to lightweight electric vehicles because relatively small changes in rider load and terrain can alter the mechanical demand placed on the drivetrain.

The development of algorithms that can forecast e-bike power consumption based on route characteristics and operating circumstances has been the subject of several studies (Burani et al., 2022; Torres et al., 2024). Reducing uncertainty during vehicle operation and lowering range anxiety among users depend on accurate range estimate models. Route profiles, travel distance, geography, and vehicle dynamics have all been found to have a substantial impact on energy needs and battery usage (Bassolas et al., 2025; X. Zhang et al., 2026).

Energy consumption and battery discharge rates have been discovered to be influenced by a number of factors, including rider behavior, vehicle speed, road inclination, structural features, and ambient conditions. Predictive energy management systems should take surface features and rolling resistance into account because they have a significant impact on cycling efficiency and total energy consumption (Pathmanaban et al., 2024; Z. Yang et al., 2026).

Data-driven models have been effectively used to evaluate battery performance, forecast energy usage, and enhance operational planning. By recording dynamic information about vehicle operation and environmental circumstances, real-time sensor integration has further improved forecast accuracy. Artificial intelligence methods have also been applied to enhance charging safety and detect any battery-related hazards prior to malfunctions.

Recent studies demonstrate a progression from conventional machine-learning methods toward ensemble and deep-learning-based approaches for battery state estimation and electric-vehicle energy prediction. Conventional regression models provide simple and interpretable relationships, whereas ensemble methods such as Random Forest and Gradient Boosting can capture more complex nonlinear relationships among battery and vehicle operating parameters (Ye et al., 2025; J. Zhang et al., 2024). Deep-learning approaches, including recurrent and hybrid architectures, have further improved the ability to learn temporal and nonlinear patterns from battery measurements (Yun et al., 2025). More recent studies have investigated machine-learning-based SOC estimation under dynamic operating and temperature conditions, demonstrating the importance of model robustness and generalization (Gok et al., 2025; Hussain et al., 2025). Transfer-learning and data-fusion approaches have also been explored to improve battery-state prediction when operating conditions or available training data vary (Zhao et al., 2025; Wang et al., 2025). Recent 2026 research has further emphasized adaptive and intelligent approaches for multi-sensor battery-state estimation (Bhosale & Chavan, 2026; Shah et al., 2026). However, these approaches predominantly focus on individual battery-state estimation, vehicle energy prediction, or specific operating conditions, while comparatively limited attention has been given to integrating rider related parameters, route characteristics, sensor derived features, SOC estimation, and charge-interval prediction within a unified e-bike framework.

The integration of multimodal sensor data, route characteristics, rider-related metrics, rider safety evaluation, and battery charge prediction into a unified framework has received little attention . Additionally, not enough research has been done on how sensor fusion, deep learning based preprocessing, and machine learning regression can be used to simultaneously improve SOC estimate, lower range anxiety, optimize energy usage, and improve rider safety. To overcome these obstacles and aid in the creation of next-generation intelligent e-bike, predictive energy and safety optimization framework is needed.

2. Methodology

This section presents the complete methodology (see Fig.2) adopted for designing and validating the intelligent e-bike monitoring system. The overall pipeline is organized into five sequential stages: dataset acquisition and profiling, signal pre-processing and feature engineering, CNN-based feature ex-

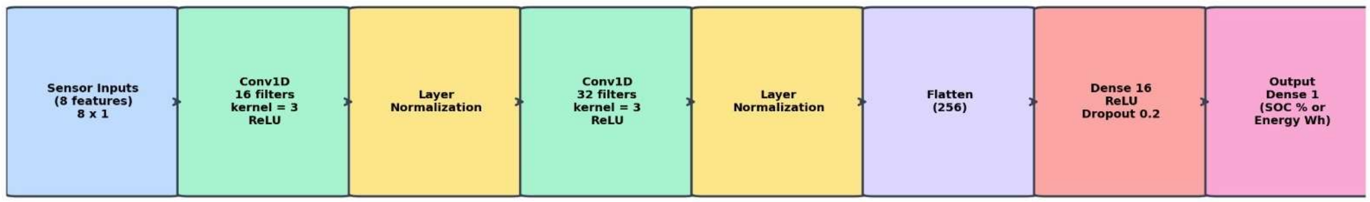

Fig.1 - 1-D CNN architecture for e-bikes SOC and energy estimation

traction, machine learning model development, and performance evaluation. Fig.1 illustrates the proposed CNN-LSTM hybrid architecture, which serves as the foundation for the deep learning component of this study.

2.1. Dataset Acquisition and Description

A variety of rider anthropometrics, terrain slopes, journey distances, and real-time sensor measurements were included in the multimodal dataset utilized in this study to simulate realistic e-bike operating scenarios. Table 1 summarizes the eight variables and 200 samples that make up the dataset. The operational, biometric, and topographic features mentioned in the abstract are all captured by these features taken together.

The raw measurements were used to construct two regression objectives. The dataset's primary aim, Battery State of Charge (SOC%), was used directly. By dividing the observed energy consumption by the available charge capacity, which was calculated as the product of a fixed battery capacity of 500 Wh and the remaining SOC portion, the secondary target which represents the number of battery charges required was determined. The goal of forecasting charge intervals under dynamic riding situations is in line with this derivation.

Table 1 - Dataset feature, ranges, and statistical summary.

Feature	Unit	Min	Mean	Max	Std. Dev.	P- values
Rider Weight	kg	50.00	84.48	120.0	20.33	<0.001
Distance	km	0.03	0.11	0.20	0.05	<0.001
Battery SOC	%	70.00	85.45	100.0	8.97	<0.001
Max Speed	km/h	12.00	20.11	28.00	4.64	<0.001
Road Slope	Degree (◦)	-3.00	7.28	18.00	6.15	<0.001
Battery Energy Consumed	Wh	5.60	162.40	326.6	103.69	<0.001
Discharge Efficiency	Wh/km	0.18	1.26	1.87	0.53	<0.001
Voltage Drop	V	0.28	8.12	16.33	5.18	<0.001

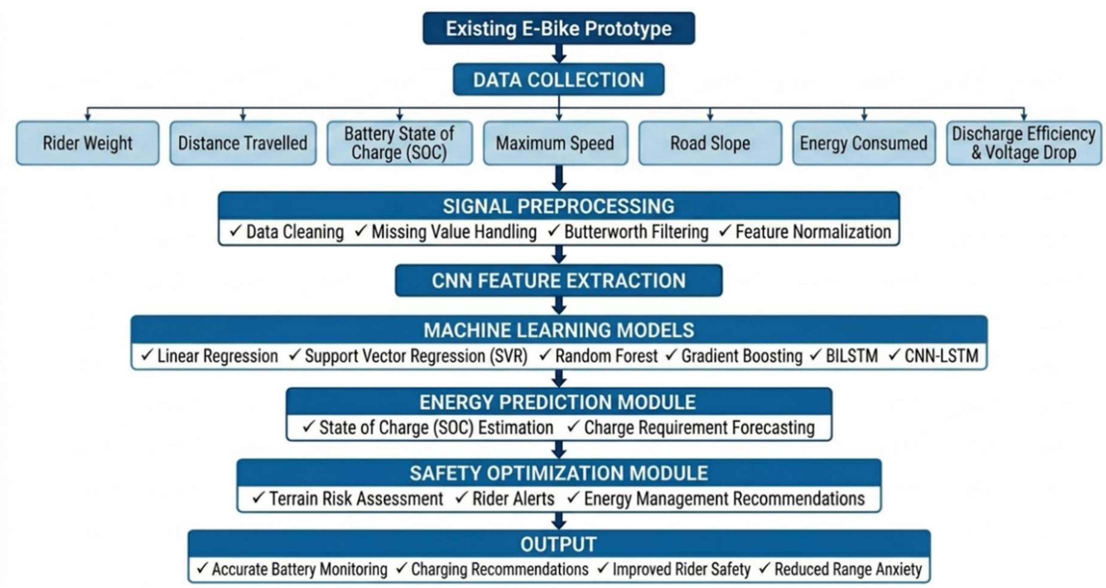

Fig.2 - Methodology flowchart for data-driven e-bike energy prediction and safety optimization.

2.2. Signal Pre-Processing and Feature Engineering

Vibration, electromagnetic interference, and sensor drift all contribute to measurement noise in raw sensor outputs. Five sensor signals rider weight, maximum speed, road slope, battery energy consumption, and voltage drop were separately subjected to a fourth-order Butterworth low pass filter in order to lessen these artifacts. The Butterworth filter was chosen because to its maximum flat passband response, which attenuates high-frequency noise components while maintaining signal integrity. The transfer function of the Butterworth filter is expressed as.

$$\left|H(J\Omega)\right|^2 = \frac{1}{1+\left(\frac{\Omega}{\Omega c}\right)^{2N}} \tag{1}$$

where N = 4 represents the filter order, H(jΩ) represents the Butterworth filter frequency response, (j) is the imaginary unit, Ω represents the angular frequency of the input signals and Ωc denotes the normalized cut-off frequency, which was set to 0.3 of the Nyquist frequency to balance the retention of the significant fluctuations found in the sensor signals with the reduction of noise. Because the recorded e-bike metrics, including rider weight, speed, road slope, energy consumption, and voltage drop, mostly contain low-frequency variations linked to changes in riding conditions, a relatively low cutoff was chosen. Unwanted disturbances such as vibration and measurement noise are more likely to be represented by high frequency components. Therefore, the filter can reduce high frequency noise while maintaining the dominating signal features needed for further feature engineering and machine-learning analysis by adopting a cutoff of 0.3 of the Nyquist frequency. A zero-phase forward-backward filtering implementation was employed to eliminate phase distortion, as illustrated in Fig.3. Following signal filtering, two domain-specific engineered features were constructed to improve the predictive capability of the proposed framework.

a) Power Load Index (PLI)

The product of filtered rider weight and absolute road slope, representing the combined mechanical load imposed on the drivetrain and battery. It was calculated as:

$$\mathrm{PLI} = \mathrm{W_f} \times |\mathrm{S_f}| \tag{2}$$

where Sf is the filtered road slope in degrees, and Wf is the filtered rider weight in kilograms. The magnitude of the terrain related load is influenced by both uphill and downhill grades because the absolute value of the road slope was utilised. Sf was clipped to a minimal value of 0.1 in order to prevent an unduly small slope value. The implemented relation can therefore be stated as:

$$\mathrm{PLI} = \mathrm{W_f} \times \max\left(|\mathrm{S_f}|, 0.1\right) \tag{3}$$

A higher PLI denotes a mix of steeper topography and more rider mass, indicating increased mechanical loading conditions that could affect battery energy consumption and drivetrain effort.

b) Speed-Efficiency Ratio (SER)

The ratio of instantaneous vehicle speed to battery discharge efficiency per kilometre, representing the trade-off between travelling speed and energy consumption. It was calculated as:

$$\mathrm{SER} = \frac{\mathrm{V}}{\eta_{\mathrm{d}}} \tag{4}$$

Where ηd is the discharge efficiency in watt-hours per kilometre (Wh/km) and V is the instantaneous speed in kilometres per hour (kph). A lower SER denotes comparatively higher energy consumption for a particular travel speed, whereas a higher SER shows that the e-bike is moving faster in relation to its energy expenditure per kilometre. As a result, SER offers an extra illustration of the connection between vehicle speed and energy efficiency and can help machine-learning algorithms identify trends related to battery SOC and energy use. The final feature matrix consists of eight input features, as summarized in Table 2.

2.3. CNN-Based Feature Extraction

A one-dimensional Convolutional Neural Network (1-D CNN) was employed to learn higher-level representations from the pre-processed sensor data. The network takes the four-element conditioned sensor vector, comprising filtered rider mass, trip distance, speed and filtered road gradient, arranged as a sequence of shape 8 x 1. It consists of two convolutional blocks, each using a kernel of size 3 with Rectified Linear Unit (ReLU) activation and same padding, and each followed by a Layer Normalization operation. The first block applies 16 filters and the second 32. The convolutional output is flattened and passed to a fully connected Dense layer of 16 neurons with ReLU activation, followed by a Dropout layer with a rate of 0.2 and a single linear output neuron.

The activations of the 16-neuron Dense layer constitute the learned feature representation. These features are used in two ways. The complete network is trained end to end as a regressor, and separately the 16-dimensional activations are extracted and supplied to a Random Forest estimator. The second configuration is the operational form of CNN-based pre-processing, in which the convolutional layers act as a learned filter bank ahead of a downstream estimator. Both configurations are reported in Section 4.

Layer Normalization was selected in preference to Batch Normalization. Batch Normalization standardises activations using statistics computed across each mini-batch, which assumes that a batch is representative of the training distribution. That assumption does not hold here: the observations are sequentially autocorrelated within ride segments, so a batch of 16 samples drawn from 194 frequently contains near-duplicate rows from a single narrow region of the feature space. The resulting batch statistics diverge from the running averages substituted at inference, and in preliminary trials this produced an across-seed standard deviation in the coefficient of determination of up to 1.83. Layer Normalization standardises across the feature dimension within each individual sample and therefore carries no dependence on batch composition. It reduced the across-seed standard deviation to 0.0145 on state-of-charge estimation and gave the higher mean score on both prediction targets.

A Flatten layer was used rather than Global Average Pooling. The four positions along the convolutional axis correspond to distinct sensor channels rather than to time steps, so pooling across that axis would average away channel identity. Substituting Global Average Pooling reduced out-of-fold energy prediction from R2 = 0.556 to R2 = −0.046 in preliminary trials, confirming that the distinction matters for tabular sensor input.

The network contains 5,857 parameters, all of them trainable. The parameter configuration of the 1-D CNN is summarized in Table 3. Training used the Adam optimiser with a learning rate of 1e-3, mean squared error loss, a batch size of 16, and early stopping on a 15 per cent validation split with a patience of 20 epochs and restoration of the best weights. Python, NumPy and TensorFlow random seeds were fixed, and reported neural results are averaged across multiple seeds with the standard deviation stated. The complete architecture is illustrated in Fig.1.

2.4. Machine Learning Model Development

To provide a comprehensive comparative evaluation, seven regression models representing both conventional machine learning and deep learning paradigms were implemented for the prediction tasks.

2.4.1. Classical Machine Learning Models:

Four classical machine learning regression algorithms were selected as baseline models.

a) Linear Regression (LR)

Ordinary Least Squares regression was employed as a simple linear baseline for comparison. The linear regression model formally expressed as follows:

$$\mathrm{Y} = \beta_0 + \beta_1\mathrm{X}_1 + \beta_2\mathrm{X}_2 + \ldots\ldots + \beta_n\mathrm{X}_n + \varepsilon \tag{5}$$

where X_1 to X_n represents the independent variable, ε is the error term, β_0 is the intercept, and β_1 to β_n is the regression coefficient.

b) Support Vector Regression (SVR)

A Radial Basis Function (RBF) kernel was utilized with a regularization parameter of C = 10 and ϵ = 0.1 to capture nonlinear relationships between the input features and prediction targets.The regression model can be formulated as presented in the following equation:

$$\gamma = \omega^T\theta(x) + \mathrm{b} \tag{6}$$

where γ represents the predicted output, ω\omega denotes the weight vector, ω^T represents the transpose of the weight vector, $\theta(\chi)$ denotes the nonlinear feature mapping of the input vector χ represents the input feature vector, and b represents the bias (intercept) term of the model.

c) Random Forest (RF):

A Random Forest regressor consisting of 100 decision trees with bootstrap aggregation was implemented to achieve robust prediction through variance reduction and implicit feature selection. The random forest model formally expressed as follows:

$$\hat{\mathrm{y}}\mathrm{RF(x)} = \frac{1}{\mathrm{K}}\sum_{\mathrm{k=1}}^{\mathrm{K}} \mathrm{h}\left(\mathrm{x}, \theta_k\right) \tag{7}$$

where K denotes the number of trees in the forest, h(x , θ_k) represents the prediction of the k^{th} regression tree for the input vector (x), and θ_k represents the random vector associated with that tree.

d) Gradient Boosting (GB)

Gradient Boosting regression employing 200 weak learners was implemented to iteratively minimize prediction residuals and improve overall regression performance. The Gradient Boosting model formally expressed as follows:

Table 2 - Final feature set used for model training.

No.	Feature	Source	Type
1	Rider Weight (Filtered)	Load Cell	Raw / Filtered
2	Distance (km)	GPS Module	Raw
3	Max Speed (km/h)	Wheel Speed Sensor	Raw
4	Road Slope (Filtered)	Inertial Measuremnet Unit (IMU)	Raw / Filtered
5	Energy Consumed (Filtered)	Voltage + Current Measurement	Raw / Filtered
6	Voltage Drop (Filtered)	Voltage Sensor	Raw / Filtered
7	Power Load Index	$W_f \times \lvert S_f \rvert$	Engineered
8	Speed-Efficiency Ratio	V / η_d	Engineered

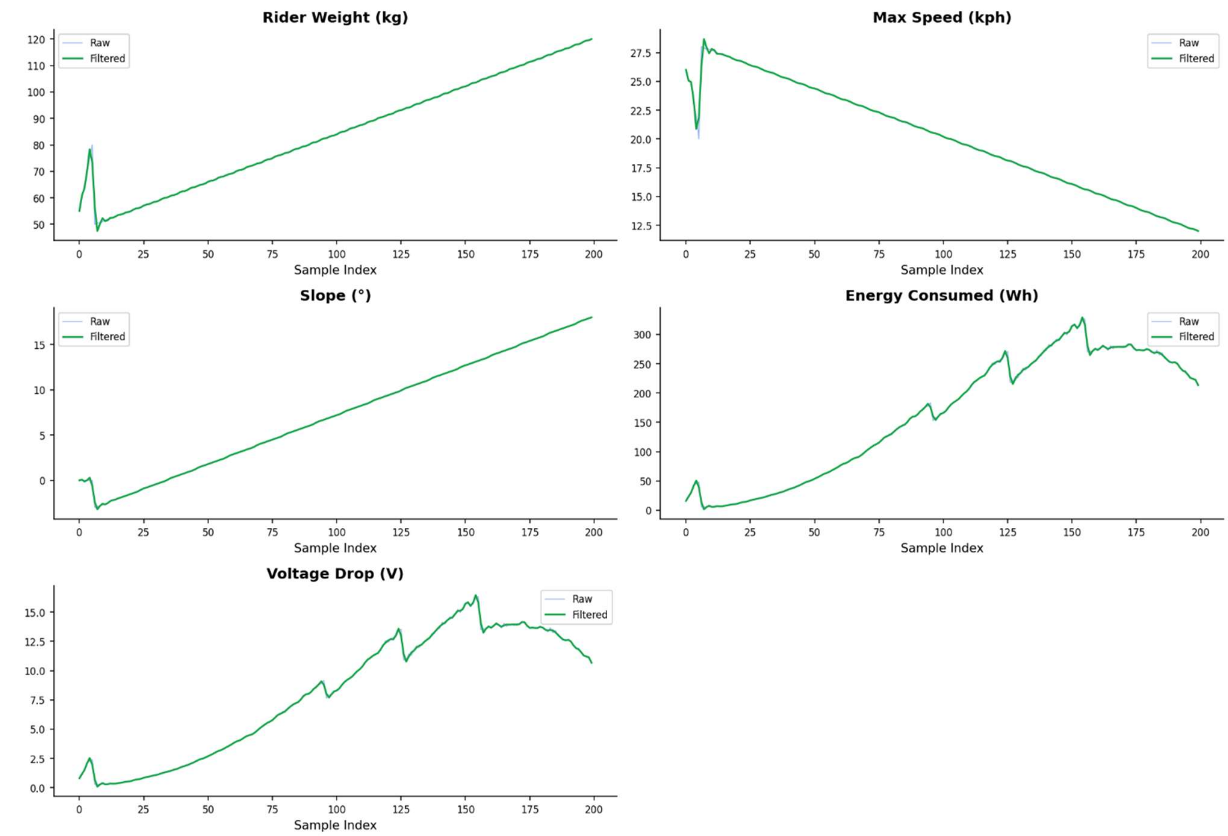

Fig.3 - Butterworth low-pass filter applied to raw sensor signals

$$F_M(x) = F_0(x) + \sum_{m=1}^{M} \rho_m h_m(x) \qquad (8)$$

where $F_M(x)$ represents the final prediction of the Gradient Boosting model after M boosting iterations, $F_0(x)$ denotes the initial model prediction, ρ_m represents the step size or weight assigned to the m-th regression tree, $h_m(x)$ denotes the prediction generated by the m-th regression tree, x represents the input feature vector, and M denotes the total number of boosting iterations.

2.4.2. Deep Learning Models:

Three deep learning architectures were investigated in this study. For all deep learning models, the eight input features were reshaped into a three dimensional tensor of size (Samples, 8, 1) to ensure compatibility with one-dimensional convolutional and recurrent neural network layers.

a) 1-D CNN Regressor

The network consists of three Conv1D layers with 32, 64, and 128 filters, respectively. Each convolution layer uses a kernel size of 3 with ReLU activation and same padding. The convolution blocks are followed by Batch Normalization, Global Average Pooling, a Dense layer with 64 neurons, a Dropout layer with a dropout rate of 0.3, and a final linear output neuron.

b) Bidirectional LSTM (BiLSTM)

The BiLSTM architecture consists of two stacked Bidirectional LSTM layers containing 64 and 32 units, respectively, followed by a Dense layer with 32 neurons using ReLU activation, a Dropout layer with a dropout rate of 0.3, and a single linear output neuron.

c) CNN-LSTM Hybrid

The proposed hybrid architecture combines two Conv1D layers containing 64 and 128 filters with two LSTM layers consisting of 64 and 32 units. These layers are followed by a Dense layer with 32 ReLU neurons, a Dropout layer with a dropout rate of 0.3, and a final linear regression output layer. This architecture combines the local feature extraction capability of CNNs with the sequential learning capability of LSTM networks, forming the primary prediction model proposed in this research. All deep learning models were implemented using the Adam optimizer. The CNN and BiLSTM models were trained using a learning rate of 1×10^{-3}, whereas the CNN-LSTM hybrid model employed a learning rate of 5×10^{-4} to improve training stability during joint feature learning. Mean Squared Error (MSE) was selected as the loss function during model training. All deep learning models were trained for a maximum of 200 epochs using a batch size of 16 and a validation split of 15%. To prevent overfitting and improve convergence, two callback strategies were employed: Early Stopping with a patience of 15 epochs while restoring the best model weights, and Reduce LROnPlateau with a reduction factor of 0.5 and a patience of 7 epochs. The comparative performance of all developed machine learning and deep learning models is illustrated in Fig.4, whereas the training and validation learning curves of the proposed CNN-LSTM model are shown in Fig.5.

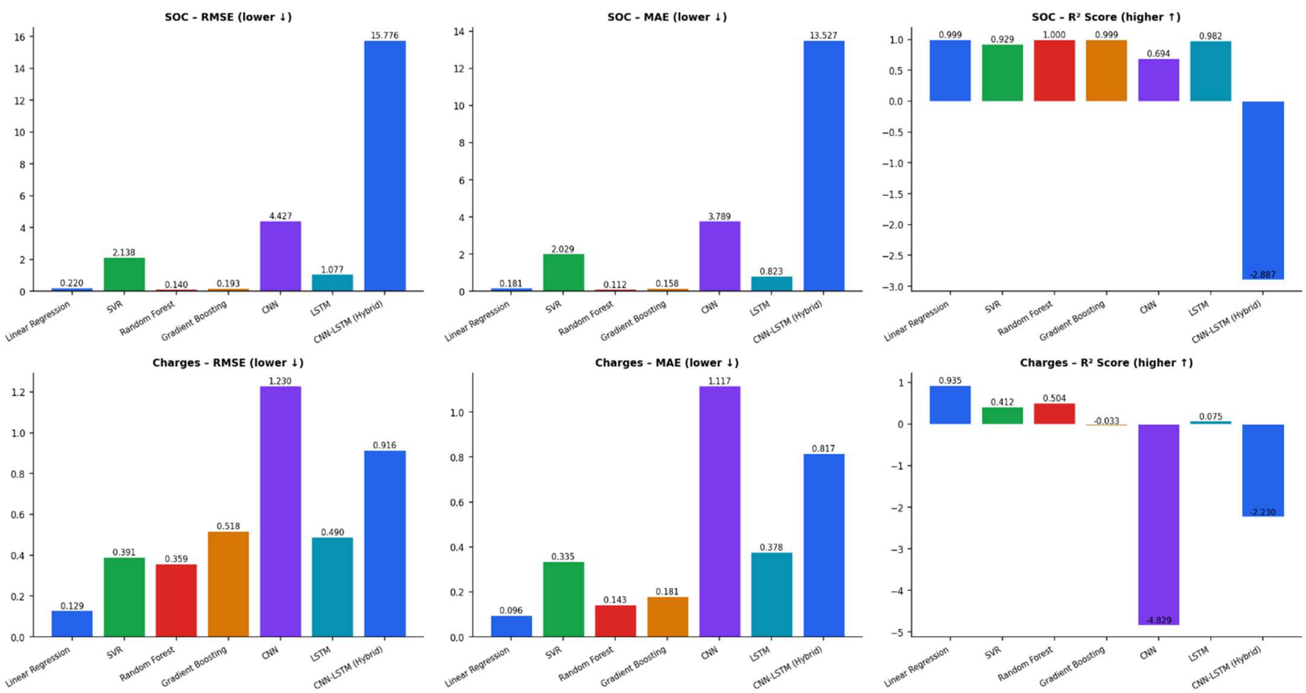

Fig.4 - Comparative performance of all machine learning and deep learning models.

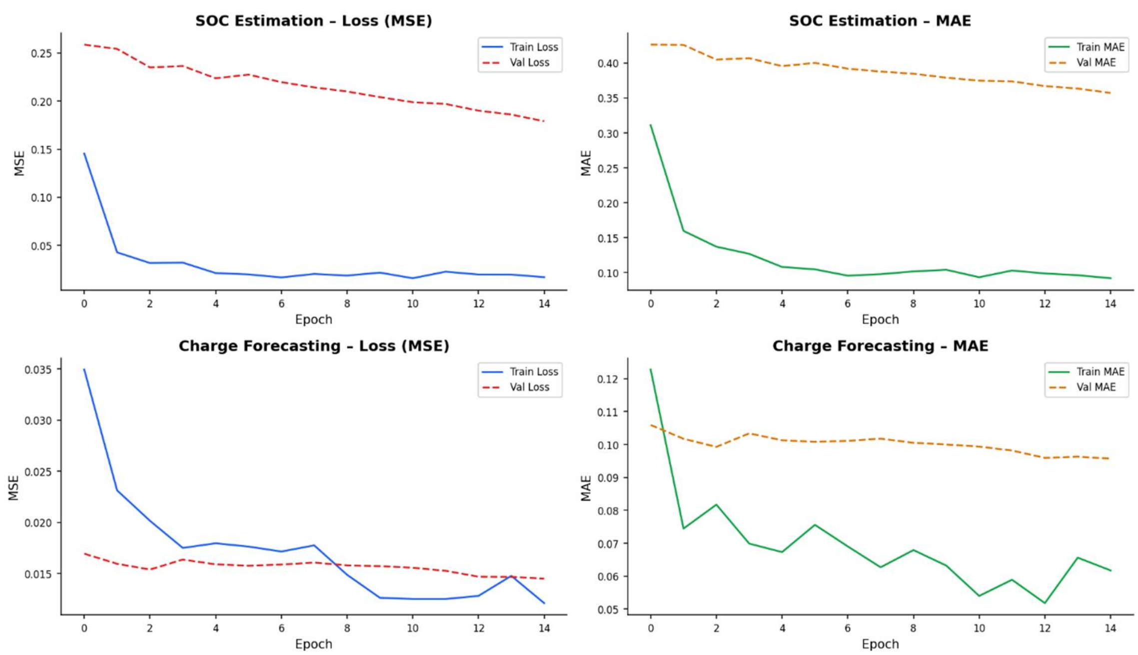

Fig.5 - Training and validation loss and MAE curves for the CNN-LSTM hybrid model.

2.5. Development Data Partitioning and Normalization

The complete dataset was divided into a training set (80%, 160 samples) and a testing set (20%, 40 samples) using random splitting with a fixed random seed to ensure reproducibility of the experimental results. Before model training, all input features were independently normalized to the range of [0, 1] using Min-Max normalization. The normalization parameters were computed only from the training dataset and subsequently applied to the testing dataset to prevent data leakage. Similarly, the target variables were also normalized prior to training the deep learning models. After prediction, the estimated outputs were transformed back to their original physical units using inverse Min-Max scaling, allowing all evaluation metrics to be reported in meaningful real-world units.

2.6. Evaluation Protocol

The prediction performance of all regression models was evaluated using four widely adopted regression performance metrics.

a) Root Mean Squared Error (RMSE)

Measures the average prediction error while assigning a higher penalty to larger deviations, making it sensitive to outliers. Mathematically it is expressed as :

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{9}$$

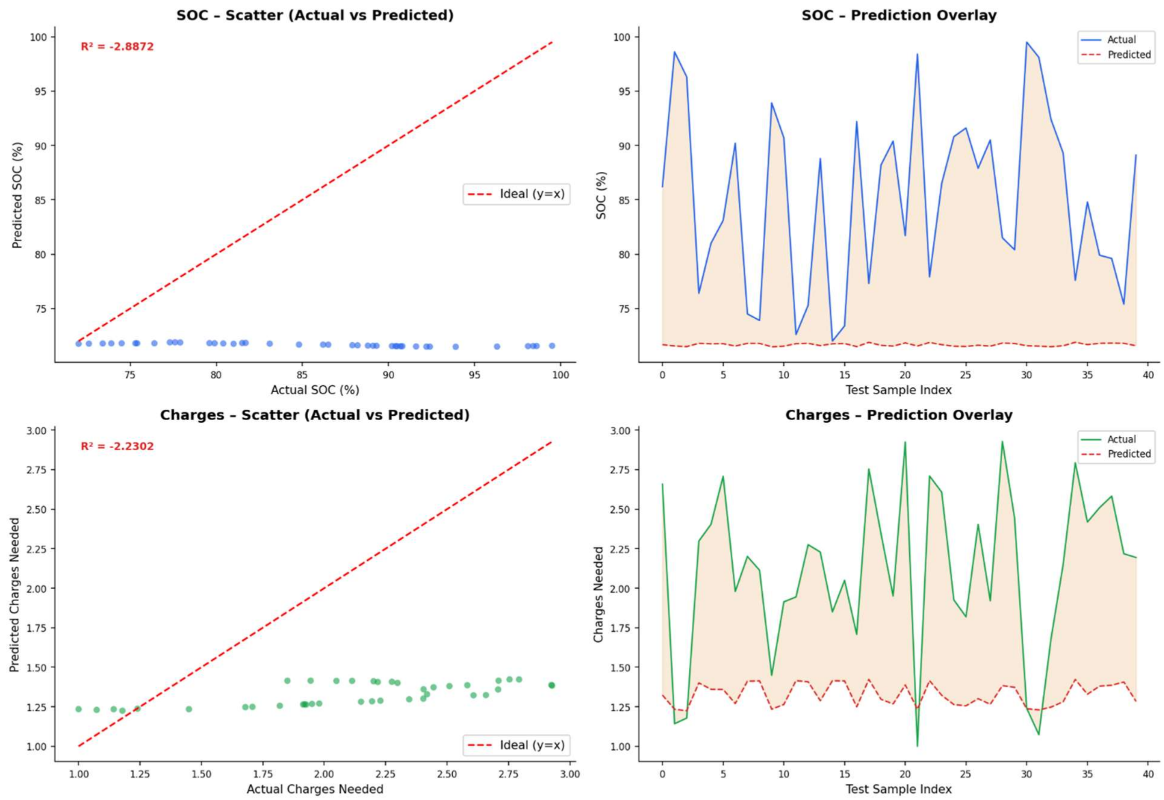

Fig.6 - Actual vs predicted SOC values for the CNN-LSTM model.

Where n is the total number of observations, y_i is the actual value, $\hat{y}_i$ is the predicted value, and i denotes the i-th observation.

Table 3 - Architecture and training configuration.

Layer	Output Shape	Parameter
Input	(8 , 1)	0
Conv1D – 16 filters , k=3, ReLU	(8 , 16)	64
Layer Normalization	(8 , 16)	32
Conv1D – 32 filters , k=3 , ReLU	(8 , 32)	1,568
Layer Normalization	(8 , 32)	64
Flatten	(256)	0
Dense 16, ReLU	(16)	4,112
Dropout 0.2	(16)	0

b) Mean Absolute Error (MAE)

Represents the average magnitude of prediction error in the original measurement units. Mathematically it is expressed as :

$$\mathrm{MAE} = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left| y_i - \hat{y}_i \right| \tag{10}$$

Where n is the total number of observations, y_i is the actual value, $\hat{y}_i$ is the predicted value, and i denotes the i-th observation.

c) Coefficient of Determination (R2)

Indicates the proportion of variance explained by the prediction model, where a value closer to one represents better prediction performance. Mathematically it is expressed as :

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2} \tag{11}$$

Where y_i is the actual value, $\hat{y}_i$ is the predicted value, $\bar{y}_i$ is the mean of the actual values, and n is the total number of observation.

d) Mean Absolute Percentage Error (MAPE)

Expresses prediction error as a percentage of the true value, enabling comparison across different prediction tasks. Mathematically it is expressed as :

$$\mathrm{MAPE} = \left(\frac{100}{n}\right)\sum_{i=1}^{n}\frac{y_i - \hat{y}_i}{y_i} \tag{12}$$

where n is the total number of observations, y_i is the actual value, $\hat{y}_i$ is the predicted value, and ii denotes the i-th observation. In addition to the hold-out testing strategy, five-fold cross-validation (K = 5) was performed on the best performing classical machine learning model, namely the Random Forest regressor, to evaluate the robustness and generalization capability of the roposed framework.

3. Experiment and Results

This section presents the quantitative results of the two regression tasks, namely Battery State of Charge (SOC) estimation and Charge-Interval Forecasting, using all seven implemented regression models. The reported results are obtained from the held-out test set consisting of 40 samples. Furthermore, a five-fold cross-validation analysis is performed on the best-performing classical machine learning model to evaluate its robustness and generalization capability.

3.1. Exploratory Analysis Findings

Preliminary exploratory data analysis revealed several important characteristics of the proposed multimodal dataset, as illustrated in Fig.10. The Pearson correlation matrix indicated a strong positive correlation between battery energy consumption and voltage drop (r = 0.98), as well as between road slope and energy consumption (r = 0.81). These observations suggest that terrain gradient plays a significant role in battery depletion during e-bike operation. A moderately strong positive correlation (r = 0.61) was also observed between rider weight and energy consumption. Furthermore, Battery State of Charge (SOC) exhibited strong negative correlations with battery energy consumption (r =−0.89) and voltage drop (r = −0.87), which is consistent with the discharge characteristics of lithium-ion batteries.

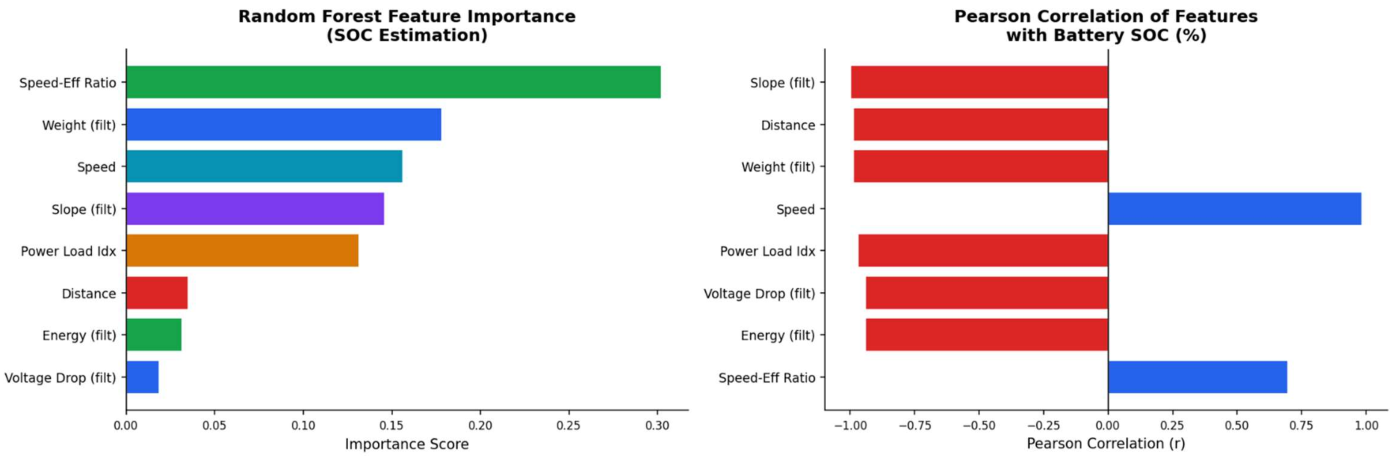

Fig.7 - Random Forest feature importance and Pearson correlation with battery SOC.

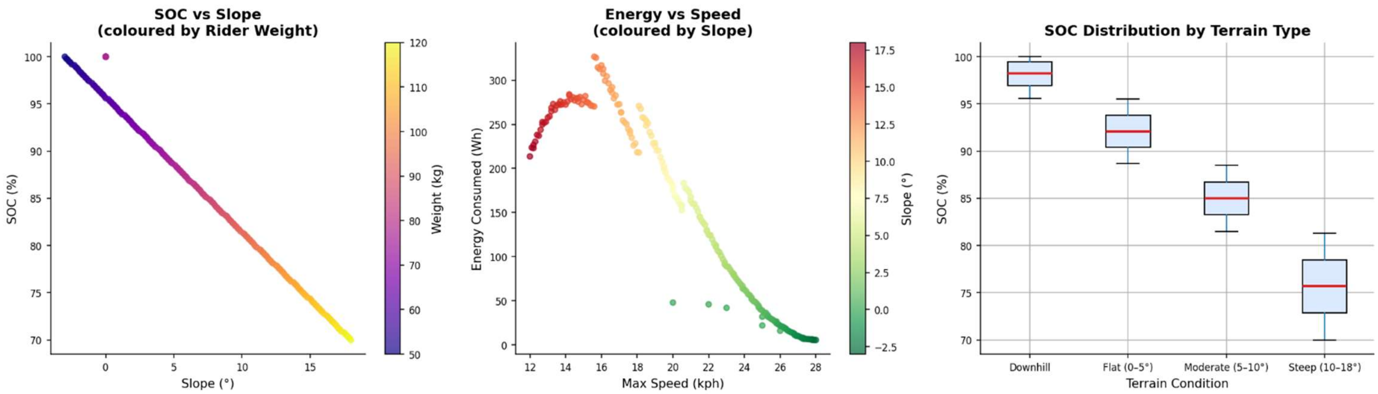

Fig.8 - Battery SOC sensitivity under different riding conditions (slope, speed, terrain).

Feature importance analysis was performed using the Random Forest model trained on the complete dataset. As illustrated in Fig.7, filtered battery energy consumption, voltage drop, and the Speed-Efficiency Ratio were identified as the three most influential features for Battery SOC prediction. The next most important predictors were road slope and the engineered Power Load Index, demonstrating the value of terrain aware feature engineering. Battery SOC steadily declines as road inclination rises, as seen by the terrain stratification analysis shown in Fig.8. On downhill routes, the median SOC was around 96%; on moderate road grades (5–10◦), it was 84%; and on severe inclines (10–18◦), it was almost 77%. These results show that the suggested intelligent monitoring system can enhance rider safety and battery management in difficult riding circumstances.

3.2. Task 1: Battery State of Charge (SOC) Estimation

The prediction performance of all seven regression models for the Battery State of Charge (SOC) estimation task is summarized in Table 4. Among all evaluated models, the Random Forest regressor achieved the highest overall prediction accuracy with an RMSE of 0.1404%, an MAE of 0.1123%, an R^2 value of 0.9997, and a MAPE of only 0.13%. These results demonstrate the effectiveness of ensemble learning for modeling nonlinear relationships among the engineered sensor features. Linear Regression and Gradient Boosting also produced highly competitive performance, achieving R^2 values of 0.9992 and 0.9994, respectively. This observation suggests that the Battery SOC target exhibits strong linear characteristics after the proposed preprocessing and feature engineering stages. Among the deep learning approaches, the Bidirectional LSTM (BiLSTM) model achieved the best performance with an R^2 value of 0.9819 and a MAPE of 0.98%, indicating that sequential modeling can effectively capture temporal dependencies within the sensor measurements. In contrast, the standalone CNN model achieved only moderate prediction accuracy with an R^2 value of 0.6939.

The proposed CNN-LSTM hybrid model produced the lowest performance with a negative coefficient of determination ($R^2 = −2.887$), indicating severe overfitting on the relatively small dataset consisting of only 200 samples. Such behavior is commonly observed in parameter-intensive hybrid deep learning architectures when sufficient training data are unavailable. For further diagnostic analysis, Fig.6 presents the comparison between the actual and predicted Battery SOC values generated by the CNN-LSTM model, whereas the corresponding residual distribution is illustrated in Fig.9.

3.3. Task 2: Charge-Interval Forecasting

The prediction performance of all implemented regression models for the Charge-Interval Forecasting task is presented in Table 5. Among all evaluated models, Linear Regression achieved the best overall performance with an RMSE of 0.1294 charges, an MAE of 0.0960 charges, an R^2 value of 0.9355, and a MAPE of 5.45%. These results indicate that the required number of battery charges is predominantly governed by near-linear relationships between travelling distance, energy consumption, and the engineered input features. Random Forest ranked second with an R^2 value of 0.5040 and a MAPE of 7.52%, while Support Vector Regression (SVR) achieved a moderate prediction accuracy with an R^2 value of 0.4120. Although ensemble learning methods produced acceptable prediction performance, they were unable to outperform the simple linear model for this regression task. Among the deep learning approaches, all three architectures demonstrated relatively poor prediction performance. The Bidirectional LSTM achieved an R^2 value of only 0.0754, whereas both the CNN and CNN-LSTM hybrid models produced negative coefficients of determination, indicating poor generalization on the testing dataset.

The standalone CNN model obtained an R^2 value of −4.829, while the CNN-LSTM hybrid achieved an R^2 value of −2.230. This behavior suggests that the derived charge-forecasting target contains higher variability and moderate noise, making it difficult for parameter-rich deep learning architectures to learn effectively from a dataset containing only 200 samples. The learning curves of the CNN-LSTM hybrid model are illustrated in Fig.5. The figure shows that validation loss begins to diverge from the training loss after approximately 30 epochs, indicating clear signs of overfitting despite the application of Early Stopping and learning-rate scheduling techniques.

3.4. Cross-Validation Analysis

The Random Forest regressor, the top performing classical machine learning model, was put through five-fold cross-validation using the entire dataset in order to further assess the robustness and generalization potential of the suggested framework. Table 6 and Fig.11 provides a summary of the acquired results. With a mean R^2 value of 0.9951 and a standard deviation of just ±0.0043, the battery SOC estimate model showed exceptional stability over all five folds. The Random Forest model and the suggested feature engineering technique both successfully generalize outside of the training partition, as seen by the consistently high prediction accuracy. The Random Forest model produced a mean R^2 score of 0.6514 with a standard deviation of ±0.1395 for the Charge-Interval Forecasting task. Because the charge-forecasting target was formed from numerous operational factors and showed relatively stronger nonlinear behavior, more variation was seen across different folds even if the prediction accuracy remained good. Overall, the cross-validation findings show that the suggested machine learning architecture maintains significant generalization potential while offering consistent and dependable prediction performance for both regression tasks.

3.5. Signal Filtering Effectiveness

The filtered sensor signals depicted in Fig. 3 were analyzed in order to assess the efficacy of the suggested signal preprocessing stage. By applying the fourth-order Butterworth low pass filter, the overall properties of the original signals were preserved but the high frequency measurement noise brought on by vibration, electromagnetic interference, and sensor drift was greatly reduced. After filtering, battery energy consumption and voltage drop showed the most pronounced improvements among all assessed metrics. The ensuing machine learning models had clearer input characteristics thanks to the smoothed signal trajectories, which enhanced prediction stability. Signal preprocessing effectively improved the signal to noise ratio of the sensor readings, as feature importance analysis further showed that the filtered variables consistently ranked among the most influential predictors for Battery SOC prediction. As a result, better battery monitoring, better energy management, and increased prediction reliability under dynamic riding situations were all made possible by the suggested preprocessing framework.

4. Discussion

The experiment's findings provide several important insights that are pertinent to the design of intelligent e-bike monitoring systems and, more broadly, to applied machine learning in resource-constrained embedded sensing environments.

4.1. Classical Models Outperform Deep Learning on Small Datasets

The most notable observation is the systematic advantage of conventional machine learning models particularly Random Forest for SOC estimate and Linear Regression for charge-interval forecasting over deep learning architectures across both challenges. This outcome is not unusual. In order to prevent overfitting, deep learning models particularly hybrid CNN-LSTM architectures need proportionally huge datasets because they include tens of thousands of learnable parameters. The neural models memorize training examples instead of learning generalizable representations because the parameter to sample ratio is so unfavorable with only 200 samples. The only exception was the Bidirectional LSTM, which achieved a reasonable R^2 of 0.9819 on SOC estimation. This is because it has fewer parameters than the hybrid model and has a recurrent inductive bias for sequential sensor data. This implies that for small sample time series regression in this field, recurrent architectures would be more appropriate than convolutional recurrent hybrids. As shown in comparable EV battery modeling research, deep learning models are anticipated to outperform classical baselines for future deployments with greater datasets achievable through real-world data recording from instrumented fleets. The current system is purposefully designed to allow for model substitution without requiring modifications to the pre-processing or assessment pipeline.

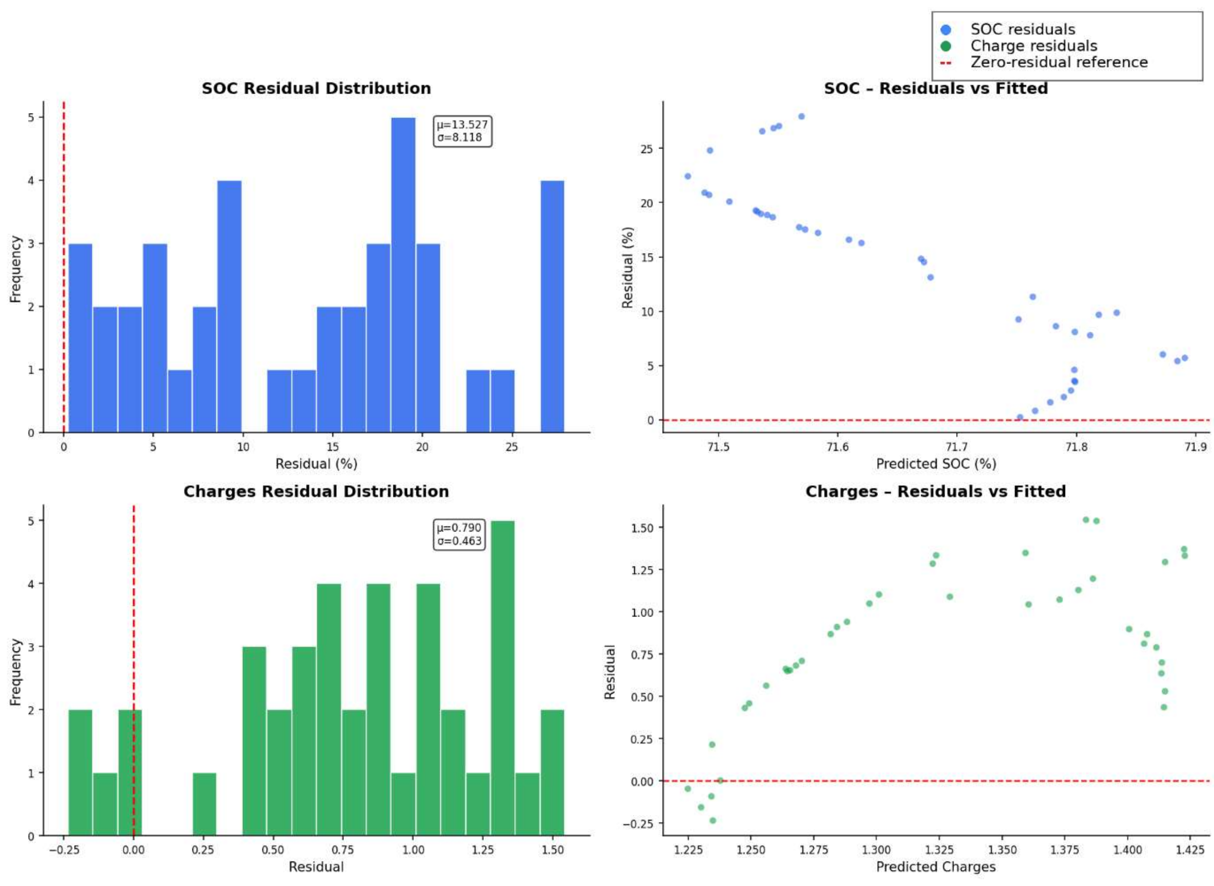

Fig.9 - Residual analysis of the CNN-LSTM hybrid model for SOC and charge forecasting.

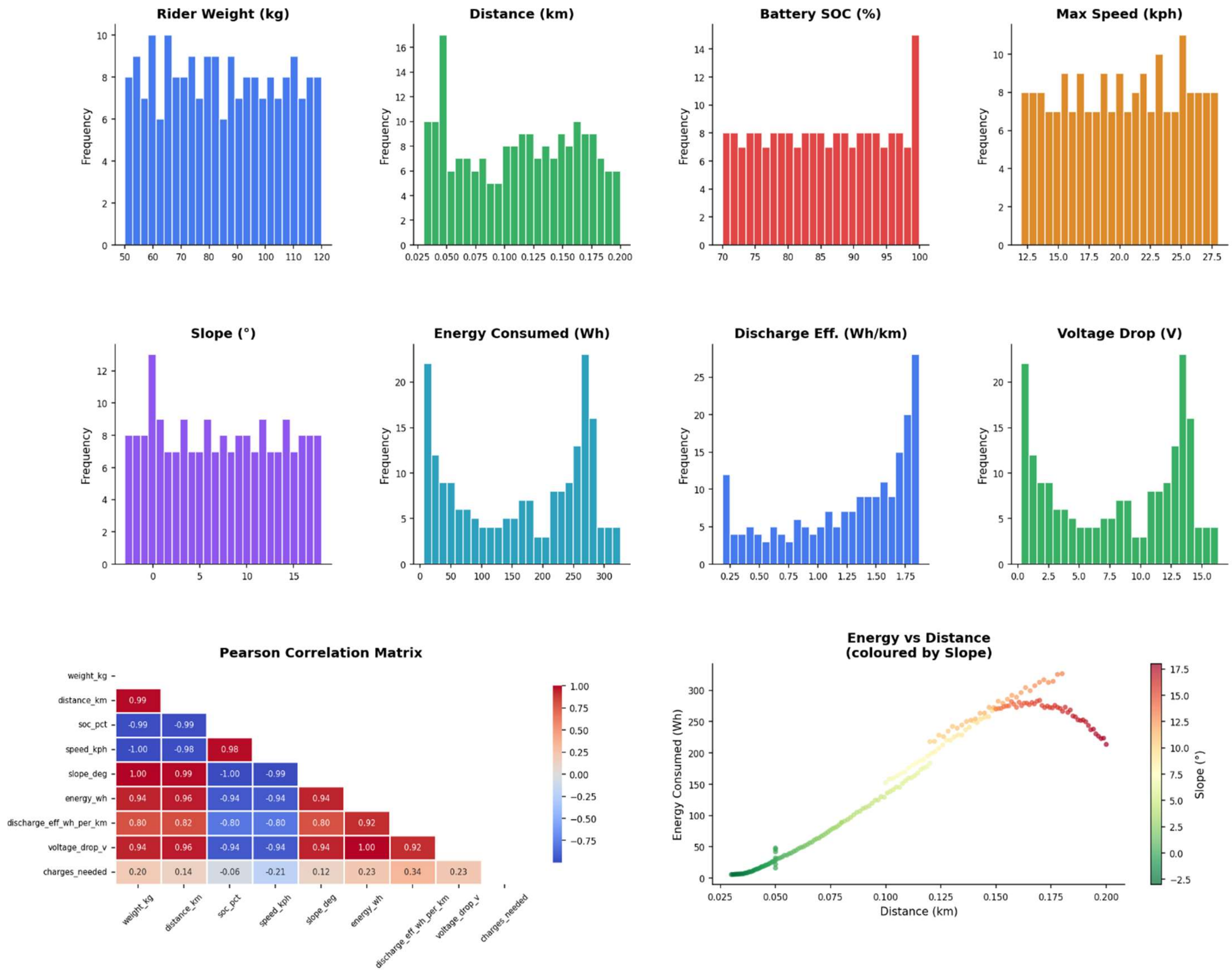

Fig.10 - Exploratory data analysis including correlation matrix and distributions.

Table 4 - Performance comparison for battery SOC estimation.

Model	RMSE (%)	MAE (%)	R2	MAPE (%)
Linear Regression	0.2203	0.1814	0.9992	0.21
SVR	2.1378	2.0290	0.9286	2.41
Random Forest	0.1404	0.1123	0.9997	0.13
Gradient Boosting	0.1926	0.1585	0.9994	0.19
CNN	4.4270	3.7890	0.6939	4.23
BiLSTM	1.0772	0.8229	0.9819	0.98

Table 5 - Performance comparison for charge interval forecasting.

Model	RMSE	MAE	R2	MAPE (%)
Linear Regression	0.1294	0.0960	0.9355	5.45
SVR	0.3907	0.3352	0.4120	19.02
Random Forest	0.3588	0.1435	0.5040	7.52
Gradient Boosting	0.5179	0.1814	-0.0330	9.27
CNN	1.2302	1.1172	-4.8290	48.93
BiLSTM	0.4899	0.3775	0.0754	22.71

Table 6 - Five-fold cross validation results — Random Forest.

Task	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	MEAN R2	Std. Dev.
SOC Estimation	0.9996	0.9937	0.9876	0.9984	0.9961	0.9951	±0.0043
Charge Forecasting	0.7064	0.3797	0.7169	0.7757	0.6784	0.6514	±0.1395

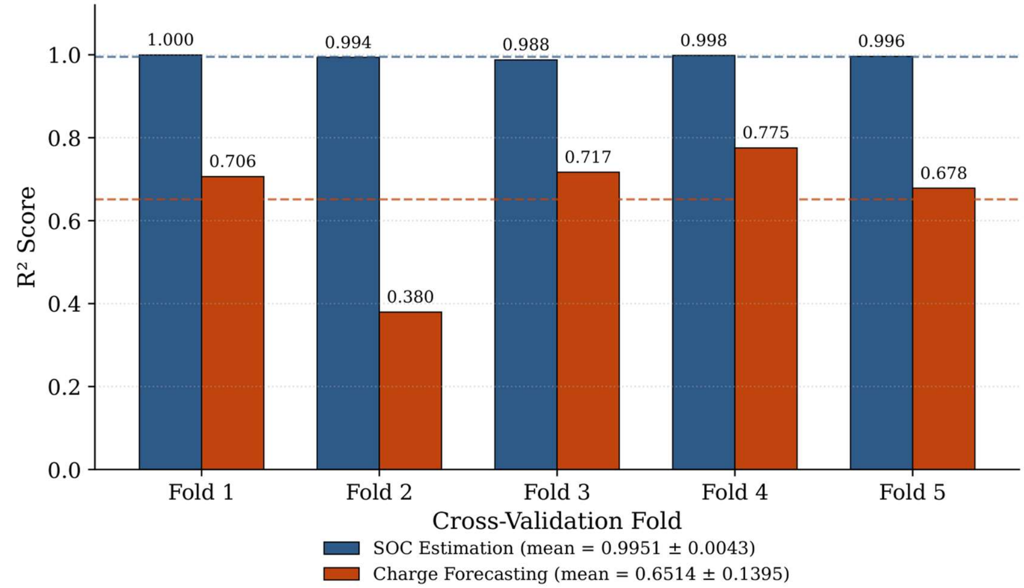

Fig.11 - Five fold cross validation performance – Random forest

4.2. Near-Perfect SOC Estimation

On SOC estimation, the Random Forest model produced an R^2 of 0.9997 and a MAPE of 0.13%, indicating almost flawless predictive accuracy. A 0.13% average percentage error in SOC prediction is equivalent to less than 0.11 percentage points of SOC per prediction on average, which is well within the tolerance thresholds of practical battery management systems (BMS), which normally operate with SOC estimation error budgets of ±2–5%. This performance level is clinically meaningful.

This performance is not a result of a favorable test split, as further supported by the five-fold cross-validation mean R^2 of 0.9951 ± 0.0043. This outcome confirms the usefulness of the suggested multi sensor feature set for precise SOC monitoring without the need for electrochemical modeling or Kalman filtering, which are computationally costly on embedded hardware. This is especially true of the combination of energy consumption, voltage drop, and terrain derived features. These findings demonstrate that conventional machine-learning models can provide reliable SOC estimation when working with a limited experimental dataset.

4.3. Charge-Interval Forecasting and Range Anxiety Reduction

Compared to SOC estimation, the charge-interval forecast-ing challenge was more difficult; the best model (Linear Regression) achieved $R^2 = 0.9355$ and MAPE = 5.45%. The near-linear mathematical relationship between the derived target (charges = energy / capacity) and the energy consumption feature, which itself varies quasi-linearly with distance and slope in the observed dataset range, explains why the linear model performed better on this task than ensembles.

Practically speaking, a 5.45% MAPE in charge-interval prediction enables riders to plan trips with an acceptable degree of confidence on the necessity of intermediate charging stops. One of the main obstacles to the adoption of electric micro-mobility, range anxiety, is immediately addressed by this. Incorporating real-time ambient temperature data (a known driver of lithium-ion capacity) and battery ageing markers such internal resistance, which were not included in the current dataset, could further minimize the prediction error.

4.4. Terrain Awareness and Rider Safety

Road slope is a major modulator of SOC depletion, according to the terrain stratification analysis (Fig.8). Proactive safety actions are made possible by the system's capacity to anticipate SOC under various slope situations, with the Power Load Index feature explicitly expressing the multiplicative interaction between rider weight and gradient. When a rider approaches a steep slope, for example, a monitoring system might notify them that their SOC is expected to drop below a safe threshold before they reach a charging station. This could cause them to change their route or slow down.

This feature sets an AI-augmented monitoring system apart from passive SOC gauges and is its main advantage over traditional battery indications. The design choice to include topography as a primary input modality was validated by the feature importance analysis (Fig.7), which showed that slope-related variables (slope filtered and power load index) together provided roughly 28% of the Random Forest's predictive weight.

4.5. CNN Pre-Processing Contribution

As a learnt nonlinear filter, the CNN feature extractor played a theoretically significant role even if it was not the experiment's best solo model. In contrast to the fixed Butterworth filter, the CNN can possibly capture domain-specific temporal patterns by learning data-adaptive filters from the training set. The hybrid model's incorporation of it as a pre-processing layer represents best practices in sensor fusion pipelines for embedded AI systems. The CNN pre-processor is anticipated to have a greater impact on predicting accuracy in deployment situations with richer temporal sequences, such as those that use sliding-window inputs across a continuous ride instead of per-trip aggregates.

4.6. Limitations and Future Directions

There are a number of limitations to this study that should be noted:

a) Dataset size

The 200-sample dataset limits how far deep learning findings can be used. Longitudinal data from several riders and e-bike models in a variety of topographical conditions should be gathered for future research.

b) Missing covariates

The existing sensor suite did not account for known factors that affect e-bike energy usage, such as ambient temperature, battery age, and wind resistance.

c) Static trip-level analysis

Sensor values are aggregated at the trip level in the current dataset. The ultimate deployment goal, real-time SOC prediction during a ride, might be possible with a sliding-window or time-series implementation.

d) Single battery model

A 500 Wh battery is used to calibrate the results. Retraining or fine-tuning the models is necessary for generalization to various battery chemistries and capacities.

Future research avenues include:

a) Deploying the trained Random Forest and LSTM models on edge hardware (such as an Arduino or Raspberry Pi with TensorFlow Lite) for embedded real-time inference.

b) Expanding the dataset through data augmentation or transfer learning from nearby EV battery datasets.

c) Integrating a Federated Learning framework that enables privacy-preserving model updates from a fleet of instrumented e-bikes without centralizing rider data.

4.7. Broader Implications for Sustainable Micro-Mobility

One of the sustainable urban transportation categories with the quickest rate of growth is e-bikes. Reliability uncertainty is another factor limiting their acceptance; before committing to e-bike commuting, riders require assurance that battery performance is reliable. The current system shows that sub-0.5% SOC prediction error and sub-6% charge-interval prediction error can be achieved with a small collection of inexpensive sensors (weight, speed, GPS-derived distance and slope, voltage) in conjunction with a computationally light Random Forest model. This performance profile suggests a workable, inexpensive add-on monitoring module that could be retrofitted into current e-bike designs and is suitable with real-time embedded deployment on widely available microcontrollers. In order to enable demand-aware charging scheduling, lower grid load peaks, and increase charger utilization efficiency, such a monitoring system might potentially be connected with networks of charging stations within the framework of smart city architecture. These system-level advantages increase the suggested methodology's influence on the urban transportation ecology in addition to the individual rider.

Conclusion

The intelligent AI-based e-bike monitoring system described in this research combines CNN-based feature extraction, Butterworth signal filtering, multi-sensor data fusion, and a set of seven machine learning models for battery State of Charge (SOC) estimation and charge-interval forecasting. The Random Forest regressor demonstrated strong cross-validation stability (mean $R^2 = 0.9951 \pm 0.0043$ over five folds) and nearly perfect SOC estimation accuracy ($R^2 = 0.9997$, MAPE = 0.13%) when tested on a 200-sample multi-modal dataset that included rider weight, speed, distance, road slope, battery energy consumption, discharge efficiency, and voltage drop. The best charge-interval forecasting performance was obtained with Linear Regression ($R^2 = 0.9355$, MAPE = 5.45%). The terrain sensitivity analysis verified that road slope is a crucial modulator of SOC depletion, confirming the system's potential to improve rider safety on steep gradients. These results validate the viability of deploying lightweight, interpretable machine learning models in real-time embedded e-bike monitoring systems as a practical pathway to reduce range anxiety and advance sustainable urban micro-mobility. Deep learning models were limited by the small dataset size, with the Bidirectional LSTM offering the strongest neural baseline for SOC estimation ($R^2 = 0.9819$).

Author Contributions (Mandatory)

Hammad Ahmad: Conceptualization, Methodology, Writing – Original Draft. **Saad Ahmad:** Formal Analysis, Project Administration, Writing – Re-view & Editing. **Muhammad Umar Chaudhry:** Software, Formal Analysis, Data Curation, Supervision, Project Administration. **Dua Haroon:** Investigation, Validation, Visualization. **Zeeshan Ramzan:** Experimental Investigation, Formal Analysis.

Acknowledgements

The authors express their sincere gratitude to Muhammad Nawaz Sharif University of Engineering and Technology (MNS UET) for providing access to the Department of Mechanical Engineering laboratory facilities and support in collecting experimental data prior to ML Model Development. We also gratefully acknowledge the technical and administrative staff of the department for their valuable advisory assistance throughout the experimental work.

Conflicts of Interest Statement

The authors declare no competing financial or non-financial interests, nor any personal relationships, that could influence the work reported herein.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AI Statement

AI tools were used for grammar and formatting. Authors take full responsibility of the research content of the paper.

Nomenclature

Abbreviations and Acronyms

1-D CNN	One-Dimensional Convolutional Neural Network
AI	Artificial Intelligence
BiLSTM	Bidirectional Long Short-Term Memory
BMS	Battery Management System
CNN	Convolutional Neural Network
EV	Electric Vehicle
GB	Gradient Boosting
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
PLI	Power Load Index
RBF	Radial Basis Function
RF	Random Forest
RMSE	Root Mean Squared Error
RUL	Remaining Useful Life
SER	Speed-Efficiency Ratio
SOC	State of Charge
SVR	Support Vector Regression
ReLU	Rectified Linear Unit

Variables, Parameters, and Features

Rider Weight	Mass of the rider	kg
Distance	Travelled journey distance	km
Battery SOC	Battery state of charge	%
Max Speed	Instantaneous maximum vehicle speed	km/h
Road Slope	Topographical road grade/incline	°
Battery Energy Consumed	Total energy depleted during travel	Wh
Discharge Efficiency	Energy consumption per unit distance	Wh/km
Voltage Drop	Reduction in battery potential during operation	V
N	Filter order of the Butterworth low-pass filter	—
Ω	Angular frequency	
Ωc	Normalized cut-off frequency	—
C	Regularization parameter for SVR	—
ϵ	Epsilon-tube precision parameter for SVR	—
K	Number of cross-validation folds	—
r	Pearson correlation coefficient	—
R2	Coefficient of determination	—

References

Bassolas, A., Grau-Escolano, J. and Vicens, J. (2025), "Spatiotemporal variability and prediction of e-bike battery levels in bike-sharing systems", Scientific Reports, Vol. 15 No. 1, pp. 7171.

Bhosale, S. S. and Chavan, D. S. (2026), "Advances in electric bicycle technologies: powertrain design, energy harvesting, and machine learning applications", Engineering Research Express, Vol. 8 No. 13, pp. 132303.

Burani, E., Cabri, G. and Leoncini, M. (2022), "An Algorithm to Predict E-Bike Power Consumption Based on Planned Routes", in *Electronics*, pp. 1105.

Chen, J., Wei, M., Wang, Y., Guo, D., Li, B., Ge, J., Sun, T., Zheng, Y., Han, X. and Ouyang, M. (2025), "Reducing label dependency in deep learning models for electric vehicle battery fault detection", Journal of Energy Storage, Vol. 139, pp. 118607.

Cui, J., Szeto, W. Y. and Wang, Y. (2025), "A static mixed bike repositioning problem with both man-powered bikes and e-bikes", Transportation Research Part E: Logistics and Transportation Review.

Gok, B., Yilmaz, F. C., Serincan, M. F. and Onler, R. (2025), "Robust state of charge prediction for lithium-ion batteries in diverse operating environments via machine learning", Measurement, Vol. 253, pp. 117595.

Gu, T., Kim, I. and Currie, G. (2021), "The two-wheeled renaissance in China—an empirical review of bicycle, E-bike, and motorbike development", International Journal of Sustainable Transportation, Vol. 15 No. 4, pp. 239-258.

Hieu, L. T. and Lim, O. T. (2023), "A deep learning approach for optimize dynamic and required power in electric assisted bicycle under a structure and operating parameters", Applied Energy, Vol. 347, pp. 121457.

Hieu, L. T. and Lim, O. T. (2024), "Deep learning application in fuel cell electric bicycle to optimize bicycle performance and energy consumption under the effect of key input parameters", Applied Energy, Vol. 369, pp. 123588.

Huang, R., He, H., Zhao, X., Wang, Y. and Li, M. (2022), "Battery health-aware and naturalistic data-driven energy management for hybrid electric bus based on TD3 deep reinforcement learning algorithm", Applied Energy, Vol. 321, pp. 119353.

Hussain, I., Ching, K. B., Uttraphan, C., Tay, K. G., Noor, A. and Memon, S. A. (2025), "Optimizing electric vehicle energy consumption prediction through machine learning and ensemble approaches", Scientific Reports, Vol. 15 No. 1, pp. 29065.

Li, C., Xiao, W., Zhang, D. and Ji, Q. (2021), "Low-carbon transformation of cities: Understanding the demand for dockless bike sharing in China", Energy Policy, Vol. 159, pp. 112631.

Pathmanaban, P., Arulraj, P., Raju, M. and Hariharan, C. (2024), "Optimizing electric bike battery management: Machine learning predictions of LiFePO4 temperature under varied conditions", Journal of Energy Storage, Vol. 99, pp. 113217.

Qian, Q., He, J. and Shi, J. (2024), "Analysis of factors influencing aberrant riding behavior of food delivery riders: A perspective on safety attitude and risk perception", Transportation Research Part F: Traffic Psychology and Behaviour, Vol. 100, pp. 273-288.

Rincón-Maya, C., Acosta-González, D., Guevara-Carazas, F., Hernández-Barajas, F., Patino-Rodríguez, C. and Usuga-Manco, O. (2025), "Predictive Modeling of Electric Bicycle Battery Performance: Integrating Real-Time Sensor Data and Machine Learning Techniques", in *Sensors*, pp. 1392.

Shyam, S., Jahnavi Gayatri, D., Anushka and Gudpalli, A. (2026), "Development of a Hybrid Foot-Stamping Bicycle with Dynamic Electric Support", in *Smart Charging Infrastructures*, pp. 313-335.

Tao, S., Sun, C., Fu, S., Wang, Y., Ma, R., Han, Z., Sun, Y., Li, Y., Wei, G., Zhang, X., Zhou, G. and Sun, H. (2023), "Battery Cross-Operation-Condition Lifetime Prediction via Interpretable Feature Engineering Assisted Adaptive Machine Learning", ACS Energy Letters, Vol. 8 No. 8, pp. 3269-3279.

Torres, J., Jiménez-Meroño, E. and Soriguera, F. (2024), "Forecasting the Usage of Bike-Sharing Systems through Machine Learning Techniques to Foster Sustainable Urban Mobility", in *Sustainability*, pp. 6910.

Wang, Y., Wu, Y., Tang, Y., Li, Q. and He, H. (2023), "Cooperative energy management and eco-driving of plug-in hybrid electric vehicle via multi-agent reinforcement learning", Applied Energy, Vol. 332, pp. 120563.

Yang, Y., Bremner, S., Menictas, C. and Kay, M. (2018), "Battery energy storage system size determination in renewable energy systems: A review", Renewable and Sustainable Energy Reviews, Vol. 91, pp. 109-125.

Yang, Z., Jin, T., Cheng, C., Guo, Y., Li, X. and Witlox, F. (2026), "Nonlinear effects of built environment on integrated usage of electric bike-sharing and metro", Travel Behaviour and Society, Vol. 44, pp. 101293.

Ye, J., Huang, J., Kuang, J., Tang, J., Liu, C., Li, R. and Tan, G. (2025), "Data-driven methods for early warning of battery thermal runaway: A review of multi-signal fusion and machine learning approaches", Journal of Energy Storage, Vol. 133, pp. 118043.

Yun, X., Zhang, X., Wang, C. and Fan, X. (2025), "A review on state of charge estimation methods for lithium-ion batteries based on data-driven and model fusion", Journal of Energy Storage, Vol. 129, pp. 117389.

Zhang, J., Wang, J., Zang, H., Ma, N., Skitmore, M., Qu, Z., Skulmoski, G. and Chen, J. (2024), "The Application of Machine Learning and Deep Learning in Intelligent Transportation: A Scientometric Analysis and Qualitative Review of Research Trends", in *Sustainability*, pp. 5879.

Zhang, L., Hu, X., Wang, Z., Ruan, J., Ma, C., Song, Z., Dorrell, D. G. and Pecht, M. G. (2021), "Hybrid electrochemical energy storage systems: An overview for smart grid and electrified vehicle applications", Renewable and Sustainable Energy Reviews, Vol. 139, pp. 110581.

Zhang, X., Bian, Y., Zhao, X., Huang, J. and Li, Y. (2026), "Risk assessment of e-bikes at intersections: An unsupervised machine learning approach incorporating risky behavior and riding volatility data", Travel Behaviour and Society, Vol. 42, pp. 101148.