

KATHMANDU UNIVERSITY
SCHOOL OF ENGINEERING
DEPARTMENT OF MECHANICAL ENGINEERING

PROJECT REPORT ON



**A DATA-DRIVEN APPROACH TO PREDICT DRIVING RANGE & ENERGY
CONSUMPTION IN ELECTRIC VEHICLES USING REAL-WORLD DATA**

Prashant Bhusal [32306]

Pradeep Hamal [32310]

Aaditya Shrestha [32315]

Prija Kafle [021141]

February, 2026

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our project supervisors, **Prof. Dr. Bivek Baral** and **Er. Sirapa Shrestha** for their continuous guidance throughout the completion of this project. Their support and encouragement played a vital role in the successful accomplishment of our work. We are also deeply thankful to the respected panel members for their insightful comments, constructive criticism, and valuable suggestions during the project defense, which helped us further improve our work. Finally, we would like to extend our appreciation to all our project team members for their dedication, coordination, and collaborative efforts throughout the project.

ABSTRACT

Accurate prediction of electric vehicle energy consumption is essential for reliable range estimation, particularly in hilly regions where road gradient significantly affects vehicle performance. Nepal's Kathmandu to Manthali route represents a challenging terrain profile characterized by steep inclines, elevation variability, and diverse driving behaviors. This study proposes a terrain-sensitive, data-driven framework for estimating EV energy consumption along this route by combining physics-based energy computation with combined machine learning techniques.

The raw driving data were first processed using outlier removal and cubic spline interpolation to obtain a per-second resolution. A Kalman Filter was applied to the per-second dataset to reduce sensor noise. After preprocessing, the dataset was divided into 75% training and 25% testing subsets, and incremental mechanical energy at the wheels was calculated using longitudinal vehicle dynamics principles. The incremental energy values were integrated over successive 3 km segments to obtain trip-level energy consumption. The dataset was further categorized according to driving states: idling, cruising, acceleration, and deceleration in order to capture operational variability and to feed the model. Finally, Random Forest models were implemented to learn the nonlinear relationship between driving parameters and energy consumption.

The Random Forest model achieved an R^2 of 0.8273 and an RMSE of 0.145 kWh per 3 km segment. The results confirm that ensemble learning approaches can effectively approximate physics-based energy estimation while offering computational efficiency suitable for EV energy management and infrastructure planning hilly road network.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	ii
ABSTRACT	iii
LIST OF FIGURES	vi
LIST OF TABLES	vii
LIST OF SYMBOLS.....	viii
LIST OF ABBREVIATIONS	ix
CHAPTER 1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Objective.....	2
1.3 Scope.....	2
1.4 Limitations	2
CHAPTER 2 LITERATURE REVIEW	3
2.1 Data Preprocessing for Time Series Data	3
2.2 Modeling Methods	3
2.3 Factors affecting EV range	5
2.4 Energy Consumption Equation.....	7
2.5 Driving Range.....	8
2.6 Critical Comparison of Existing EV Energy Consumption Models.....	8
CHAPTER 3: METHODOLOGY	10
3.1 Study	10
3.2 Data Description	11
3.3 Study Area and Route Description	11
3.4 Data Preprocessing	12

3.4.1	Outlier Removal	12
3.4.2	Interpolation	12
3.4.3	Filtering	13
3.5	Incremental Energy Calculation	14
3.6	Splitting Train and Test Data.....	15
3.7	Model Development	16
3.8	Model Evaluation.....	18
3.8.1	Error Definitions.....	18
3.9	Gantt Chart.....	20
Chapter 4: Findings, Discussion, and Recommendations		21
4.1	Results and Discussion	21
4.2	Recommendations.....	23
REFERENCES		24

LIST OF FIGURES

Figure 1 Random Forest Decision Tree[10]	4
Figure 2 Gradient Boosting[12].....	5
Figure 3 Road gradient effect on energy[17]	6
Figure 4 Rolling resistance effect on energy consumption[19].....	7
Figure 5 Workflow of the Project.....	10
Figure 6 Route-Plot for One Round Trip	12
Figure 7 Graph showing raw data and preprocessed data	13
Figure 8 Incremental Mechanical Energy along Varied Altitudes	15
Figure 9 Methodological Process	16
Figure 10 Energy Consumption Breakdown	17
Figure 11 Feature Importance of Data.....	19
Figure 12 Random Forest Model Accuracy	22

LIST OF TABLES

Table 1 Comparison of EV Energy Consumption Modeling Approaches	9
Table 2 Raw Data Extracted from GPS Logger	11
Table 3 Preprocessed Data	14
Table 4 Random Forest Hyperparameters	18
Table 5 Work Schedule of the Project.....	20
Table 6 Random Forest Model Performance across Iterations	21

LIST OF SYMBOLS

Symbol	Meaning	Unit
A	Frontal area of the vehicle	m^2
C_x	Aerodynamic drag coefficient	—
dE	Incremental mechanical energy at wheels	kWh
ds	Small distance segment	m
$\frac{dv}{dt}$	Vehicle acceleration	m/s^2
f	Rolling resistance coefficient	—
g	Acceleration due to gravity	m/s^2
m	Vehicle mass	kg
m_f	Fictive mass (rotational inertia equivalent mass)	kg
v_{EV}	Vehicle speed	km/h
v_x	Wind speed	km/h
ρ	Air density	kg/m^3
φ	Road gradient angle	rad or degree

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
EV	Electric Vehicle
FNN	Feedforward Neural Network
GIS	Geographic Information System
ICE	Internal Combustion Engine
IQR	Interquartile Range
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multilayer Perceptron
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SOC	State of Charge
SVR	Support Vector Regression
XGB	Extreme Gradient Boosting

CHAPTER 1 INTRODUCTION

1.1 Background

In recent years, transportation has been the major source of greenhouse gas emissions for many countries. Therefore, the automobile industry is actively transitioning toward more sustainable and environmentally friendly solutions[1]. EV's have much lower emissions than ICEs, and EV's can convert over 60% of the electrical energy from the grid to power at the wheels, while ICE vehicles typically convert only about 20% to 30% of the energy in gasoline into usable power[2]. Despite EV's environmental benefits, their widespread adoption is moderately progressing. Consumer adaptation of EV's is influenced by high initial purchase cost and limited driving range. Both high initial purchase cost and limited driving range are the results of the current development state of battery technology. Therefore, it is of great significance to increase the practicability and reliability of EV's by improving the driving range prediction accuracy to provide users with reliable information. Therefore, various models have been developed over the past few years for energy and driving range prediction. Energy consumption estimation models play a crucial role in forecasting and assessing the performance and environmental consequences of diverse vehicle types operating under varying traffic conditions. Analytical, statistical, and machine learning models are among the various approaches employed to model the energy consumption of EV's. The ML approach is a quickly expanding subject that uses big datasets and algorithms to let computers learn and make judgments or predictions[3]. There is multiple ML algorithm that helps in predicting the energy consumption more accurately; some of them are XGB, RF, MLP, and SVR[4]. The electric consumption of EV's is commonly calculated by measuring the corresponding battery current and voltage, typically in terms of watt-hour per kilometer (Wh/km). The determination of the electric energy consumed by EV's during a trip in the unit of watt-hour (Wh). The Energy efficiency and driving range are affected by many parameters such as gross vehicle weight, rolling resistance, Aerodynamic drag, Gradient resistance, etc[5]. All these parameters need to be collected from the sensors, and the data are then fed into a machine learning model, which helps in the prediction of data.

Despite extensive research on EV energy consumption modeling, most existing studies focus on standardized drive cycles and do not explicitly address the challenges posed by sustained gradients and elevation variability. In hilly terrains, gravitational forces dominate traction energy demand, and regenerative braking efficiency varies significantly across downhill segments. Consequently, models calibrated on flat terrain often exhibit poor transferability to mountainous regions. This study addresses this gap by developing a terrain-sensitive, physics-informed machine learning framework tailored to hilly road networks.

1.2 Objective

1. To develop a Machine learning-based model to predict the energy consumption and driving range of EV vehicles in Nepal.

1.3 Scope

1. Can predict EV energy use and driving range on hilly routes like Kathmandu to Manthali.
2. Focuses on estimating energy along the route based on road slopes and elevation, not on vehicle or battery details.

1.4 Limitations

1. The three-month dataset may not represent long-term and seasonal driving variations.
2. Absence of SOC and battery-related parameters reduces prediction accuracy.
3. The model does not consider external factors such as weather and traffic conditions.
4. Model is route-specific; results may not generalize to other roads without retraining.
5. The model accounts only for mechanical traction energy and neglects auxiliary power consumption, which may lead to an underestimation of total battery energy demand.

CHAPTER 2 LITERATURE REVIEW

2.1 Data Preprocessing for Time Series Data

Outlier detection is widely recognized as a fundamental preprocessing step for time-series and sensor data. Common statistical techniques include Z-Score methods, which flag values beyond a set number of standard deviations from the mean; Interquartile Range (IQR) methods, which identify values outside the 25th to 75th percentile range and are robust to skewed distributions; and percentile-based methods, which remove extreme values beyond chosen thresholds[6]. Studies consistently highlight that effective outlier handling improves model accuracy and prevents anomalous observations from biasing statistical analyses or machine learning predictions.

In many time-series applications, data are often missing or unevenly sampled, necessitating interpolation to generate regularly spaced inputs for modeling. Common techniques include linear interpolation, which is simple and computationally efficient but may introduce abrupt transitions, and cubic splines, which fit piecewise polynomials to produce smooth and continuous curves[7].

Noise reduction and smoothing techniques are employed to suppress random fluctuations while retaining meaningful trends and dynamics. Traditional methods include moving averages, which are simple but can lag and distort peaks. Savitzky–Golay filters fit local polynomials to preserve higher-order features. Model-based approaches, especially Kalman filters, use mathematical models to combine noisy measurements with predictions of how a system is expected to behave[8]. They are widely used in time-series analysis for real-time sensor data because they can adjust to changing noise levels while still following the true underlying system behavior.

2.2 Modeling Methods

Random Forest operates by constructing decision trees during training and outputting the average (for regression) or majority vote (for classification)[9]. Each tree is trained on a random subset of the data and features, which introduces diversity and reduces overfitting[9]. Because of this randomness, Random Forest is known

for its robustness and ability to handle noisy data well. It is particularly efficient in high-dimensional spaces and is less sensitive to outliers. While it performs strongly with minimal tuning, it may not capture very complex patterns as effectively as boosting or deep learning methods.

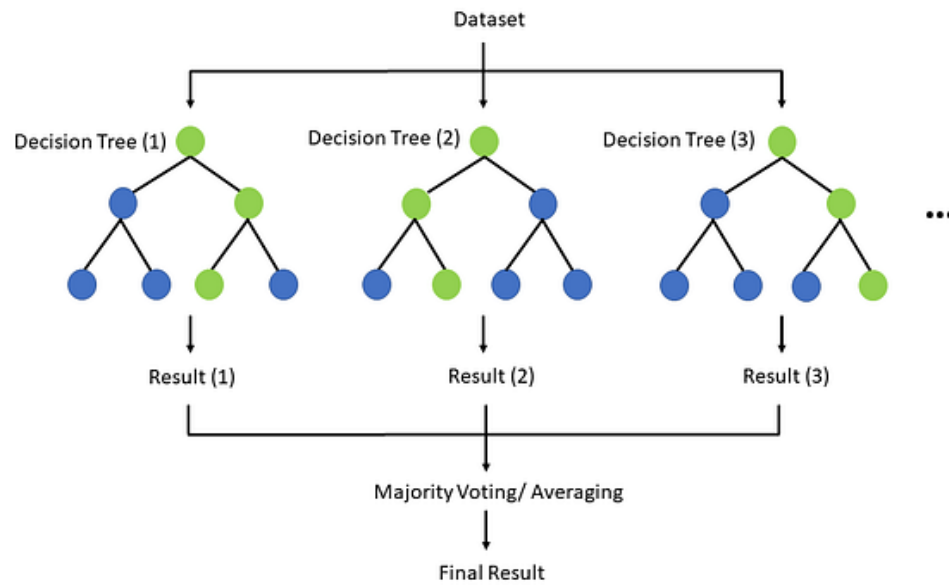


Figure 1 Random Forest Decision Tree[10]

Gradient Boosting, on the other hand, builds trees sequentially. Each new tree is trained to fix the errors of the previous ensemble. This makes Gradient Boosting very powerful for prediction tasks, as it learns from its mistakes in a focused manner[11]. Methods like XGBoost and LightGBM, which are advanced versions of gradient boosting, are widely used in machine learning competitions due to their high accuracy. However, this increased performance comes with the need for careful tuning of parameters such as learning rate, tree depth, and number of iterations. Gradient Boosting is more prone to overfitting than Random Forest, but when optimized, it often outperforms it, especially on structured data[11].

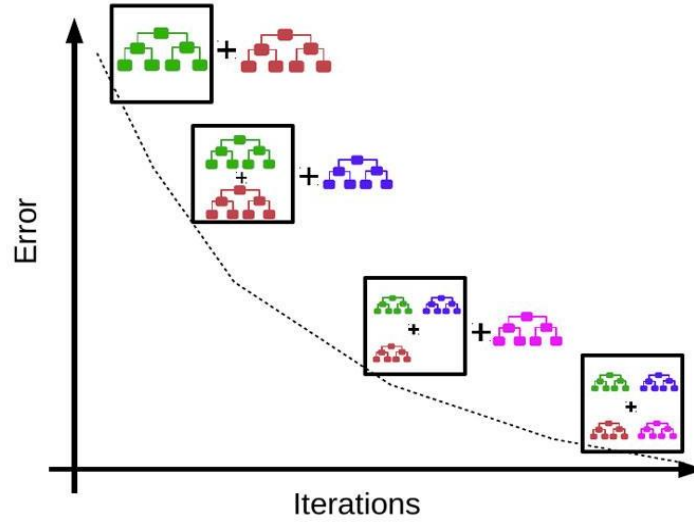


Figure 2 Gradient Boosting[12]

Deep Learning, unlike the tree-based methods above, uses layered neural networks to model complex patterns in data. It is suitable for tasks involving high-dimensional and unstructured data such as images, text, and time series. Deep learning models like feedforward neural networks (FNN), recurrent neural networks (RNN), and convolutional neural networks (CNN) can automatically extract features from raw data without manual intervention[13]. In EV applications, deep learning is used for tasks such as driving behavior prediction, battery degradation modeling, and sensor fusion[13]. While deep learning offers exceptional performance, it requires large amounts of data, longer training times, more computing power, and can be harder to interpret compared to ensemble methods like Random Forest and Gradient Boosting. Random Forest was selected due to its robustness to noisy inputs, ability to model nonlinear relationships, and lower overfitting risk on moderate datasets, compared to neural networks, which require larger datasets and extensive tuning.

2.3 Factors affecting EV range

The following are the factors that affect energy consumption in EV vehicles:

- i. **Battery Capacity:** It determines the total amount of electrical energy stored in the vehicle. A higher battery capacity allows the vehicle to store more energy and therefore travel a longer distance on a single charge[14].
- ii. **Vehicle Speed:** Driving speed has a major impact on range because aerodynamic drag increases rapidly with speed. As speed increases, the power required to overcome air resistance rises sharply, leading to higher

energy consumption. Therefore, electric vehicles typically achieve maximum range at moderate speeds rather than at high highway speeds[15].

- iii. **Aerodynamic Drag:** It depends on the vehicle's shape, frontal area, and drag coefficient. Streamlined vehicles experience less air resistance and require less energy to maintain motion. Poor aerodynamic design increases energy consumption, especially at higher speeds, thereby reducing overall driving range.
- iv. **Vehicle Weight:** Heavier vehicles require more energy for acceleration and climbing inclines. Increased mass also raises rolling resistance. Additional passengers or cargo further increase energy demand, which reduces the effective driving range of the electric vehicle[16].
- v. **Terrain and Road Gradient:** Driving uphill requires additional power to overcome gravitational force, which increases battery consumption. While regenerative braking can recover some energy during downhill driving, the energy recovered is usually less than the energy spent climbing. Hence, hilly terrain reduces the overall range[17].

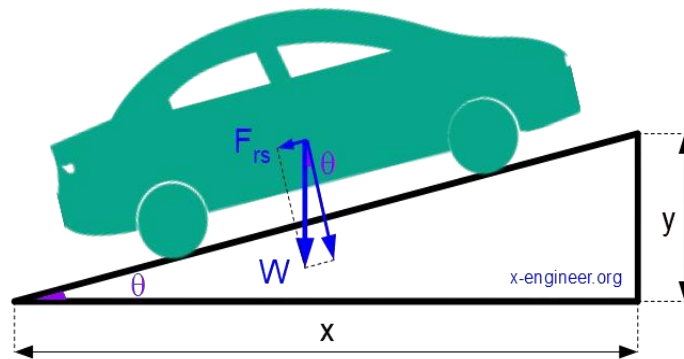


Figure 3 Road gradient effect on energy[17]

- vi. **Temperature:** Battery performance is sensitive to temperature. Extreme temperatures, especially cold weather, can reduce range by 10–30% due to reduced battery efficiency. In hot weather, air conditioning consumes extra power, and in cold weather, additional energy is required for cabin heating[18].
- vii. **Rolling Resistance:** It is influenced by tire type, tire pressure, and road surface conditions. Under-inflated tires or rough road surfaces increase

resistance and energy consumption. Maintaining proper tire pressure helps in improving driving range.

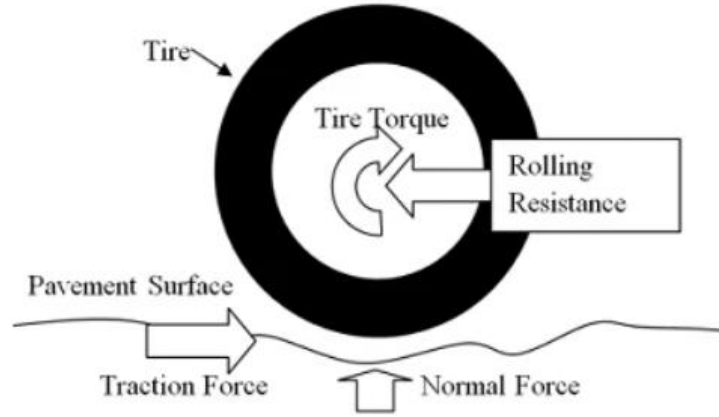


Figure 4 Rolling resistance effect on energy consumption[19]

- viii. **Auxiliary Loads:** Electrical accessories such as air conditioning, heating, lighting, infotainment systems, and other electronic components draw power from the battery. Although individually small, continuous usage can noticeably reduce the vehicle's driving range. Although there is availability of an auxiliary battery, the energy ultimately comes from the main traction battery, so using the AC/heater still reduces overall driving range.
- ix. **Driving Style:** Aggressive acceleration, sudden braking, and high-speed driving increase energy consumption. Smooth and steady driving allows better utilization of battery energy. Efficient driving habits can significantly improve the vehicle's range.

2.4 Energy Consumption Equation

The incremental mechanical energy formula is:

$$dE = \frac{1}{3600} \left[mg(f \cos \varphi + \sin \varphi) + \frac{1}{2} \rho C_x A \left(\frac{v_{EV} + v_w}{3.6} \right)^2 + (m + m_f) \frac{dv}{dt} \right] ds$$

Computes the mechanical energy required at the wheels over a small distance segment ds in kWh[5]. This formulation integrates three main contributions:

1. $mg(f \cos \varphi + \sin \varphi)$ accounts for rolling resistance and gravitational effects along slopes, capturing the influence of road gradient φ and vehicle weight.

2. $\frac{1}{2}\rho C_x A \left(\frac{v_{EV}+v_w}{3.6}\right)^2$ models the energy lost due to air resistance, considering vehicle speed, wind, frontal area, and drag coefficient.
3. $(m + m_f) \frac{dv}{dt}$ include vehicle mass and the fictive mass of rotational components, representing energy required for acceleration or energy generated during deceleration.

2.5 Driving Range

The range of EV vehicles is calculated by:

$$Range(in\ km) = \frac{Available\ Battery\ Capacity}{Energy\ Consumption\ per\ km}$$

2.6 Critical Comparison of Existing EV Energy Consumption Models

Existing approaches for electric vehicle energy consumption prediction can be broadly categorized into three groups: purely data-driven machine learning models, physics-based or drive-cycle models, and hybrid approaches that combine physical modeling with data-driven techniques. While each category has demonstrated strengths in specific applications, important limitations remain, particularly when applied to hilly road networks.

Pure machine learning models rely primarily on historical driving data to learn the relationship between input features and energy consumption. Although these models can achieve high prediction accuracy under conditions similar to the training data, they often lack physical interpretability and may fail to generalize when operating conditions differ significantly, such as in sustained uphill or downhill driving.

Physics-based and drive-cycle models explicitly incorporate vehicle dynamics and physical forces, providing strong interpretability and consistency. However, these models are typically developed using standardized drive cycles and are therefore not route-specific. As a result, their applicability to real-world routes with complex elevation profiles is limited.

To address these limitations, the present study adopts a hybrid, physics-informed machine learning approach. By computing incremental mechanical energy using longitudinal vehicle dynamics and combining this with data-driven modeling at the

segment level, the proposed framework captures terrain-induced energy variations while maintaining physical interpretability.

A comparative summary of the different modeling approaches and their limitations is presented in Table 1.

Table 1 Comparison of EV Energy Consumption Modeling Approaches

Study Type	Terrain Handling	Physics-Based	Segment-Level	Key Limitation
Pure machine learning models	Absent	Absent	Absent	Poor interpretability
Drive-cycle models	Implicit	Explicit	Absent	Not route-specific
Proposed approach	Explicit	Explicit	Explicit	Route-specific

CHAPTER 3: METHODOLOGY

3.1 Study

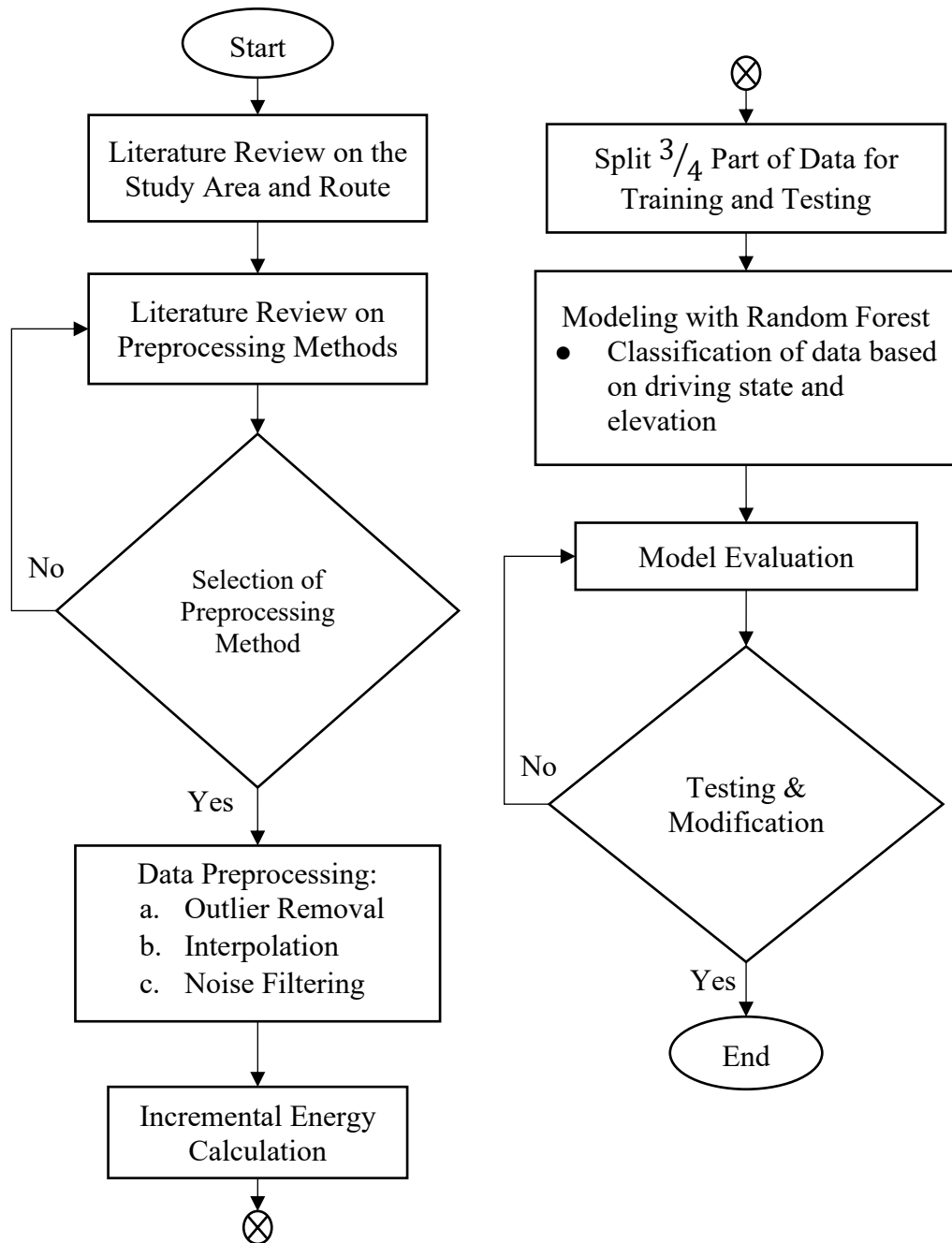


Figure 5 Workflow of the Project

This flowchart shows the step-by-step process of developing and evaluating an energy prediction model. It begins with a literature review of the study area, route, and preprocessing methods. After selecting a suitable preprocessing technique, the data is cleaned through outlier removal, interpolation, and noise filtering. Then,

incremental energy is calculated, and the dataset is split into $\frac{3}{4}$ part to train the Random Forest model. The model is developed, evaluated, and finally, testing and modification are performed until the desired accuracy is obtained.

3.2 Data Description

The dataset was collected from a passenger vehicle operating daily between Kathmandu and Manthali over a three-month period. GPS observations were recorded at irregular intervals, ranging from seconds to minutes. The raw data were extracted from the GPS, then preprocessed and feature-engineered for model development and analysis.

Table 2 Raw Data Extracted from GPS Logger

Latitude	Longitude	Speed	Angle	Altitude	Satellites	Time
27.380498 3	86.0594266	0	77	575	16	2024-01-07 05:53:19
27.380498 3	86.0594266	7	135	575	15	2024-01-07 05:53:58
27.38042	86.0595099	22	130	577	13	2024-01-07 05:54:01
27.380218 3	86.0598733	31	111	577	13	2024-01-07 05:54:06
27.380201 6	86.059965	31	98	578	13	2024-01-07 05:54:07
27.380153 3	86.0606133	25	110	575	15	2024-01-07 05:54:15
27.379866 6	86.0614866	27	88	571	14	2024-01-07 05:54:26

3.3 Study Area and Route Description

The study was conducted along the Kathmandu to Manthali highway in Nepal, a route characterized by significant elevation variation and hilly terrain conditions. The total route length is approximately 256 km and includes multiple ascending and descending segments.

The continuous alternation between uphill and downhill sections makes this route suitable for analyzing terrain-sensitive energy consumption. The terrain can be classified as predominantly hilly, with moderate to steep gradients that significantly influence traction energy demand.

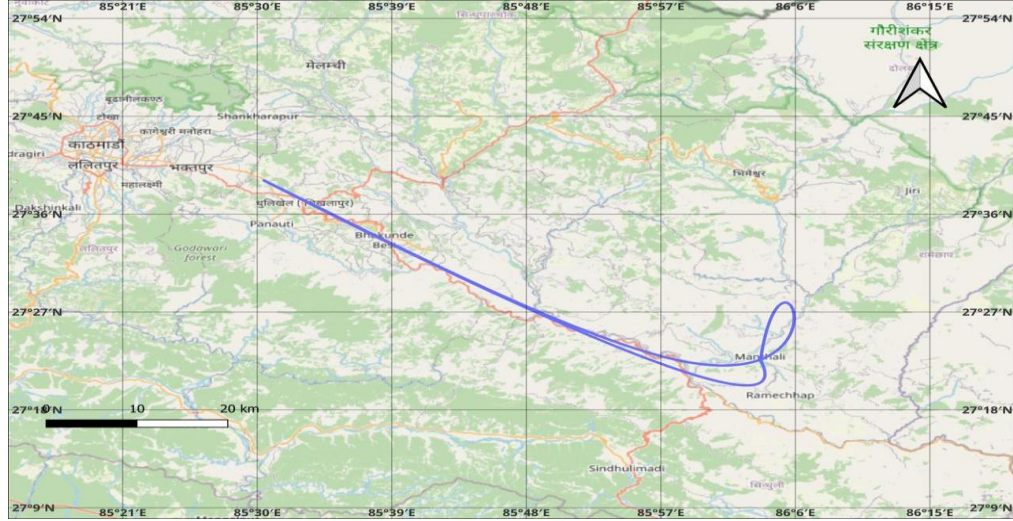


Figure 6 Route-Plot for One Round Trip

Figure 6 illustrates the driving route for the SRM vehicle. The smooth, curved path is plotted by using the interpolated latitude and longitude data using GIS software to visualize route behavior.

3.4 Data Preprocessing

3.4.1 Outlier Removal

Outliers in speed, gradient, and altitude data were removed to improve the quality of input features. For speed data, values outside the expected operational range, like (0–10) km/h for low-speed detection and extreme maximums, were filtered out. This simple approach effectively removed rare anomalies and sensor errors without assuming a specific data distribution, providing a clean dataset for subsequent processing.

3.4.2 Interpolation

To generate a uniformly spaced time series, cubic spline interpolation was applied. This method was chosen over linear interpolation, which tends to produce abrupt changes, and cubic polynomial fitting, which can generate unrealistic oscillations

for frequently varying EV data. Cubic splines provided smooth, continuous values for speed, gradient, and altitude, preserving natural variability in driving behavior while enabling accurate modeling of energy consumption and driving range.

3.4.3 Filtering

To reduce sensor noise and improve the signal quality, a Kalman Filter was applied to the interpolated data. Unlike simple moving averages or Savitzky-Golay filters, which either introduce lag or fail to fully capture rapid changes, the Kalman Filter dynamically adapts to real-time measurements and system uncertainties. This allowed for accurate and physically consistent smoothing of speed, gradient, and altitude, enhancing the reliability of machine learning predictions.

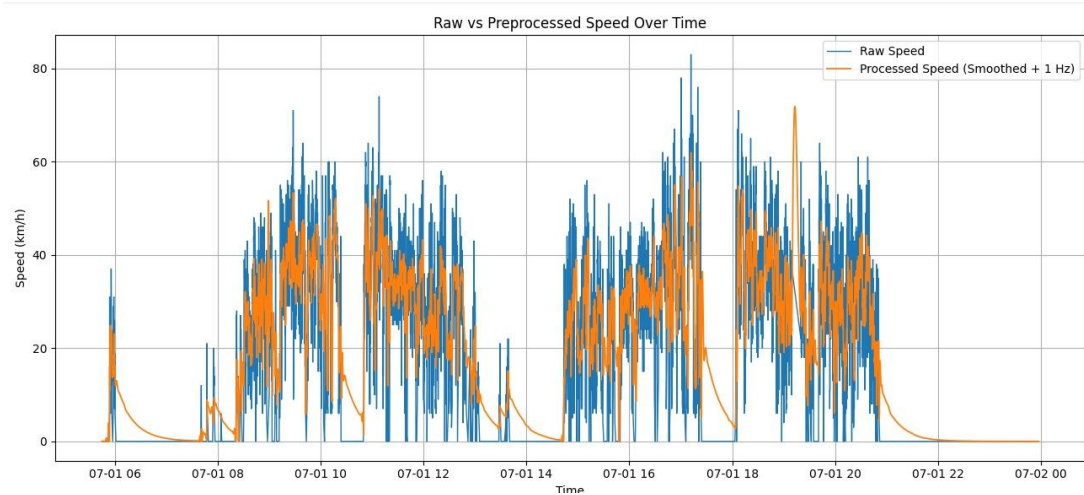


Figure 7 Graph showing raw data and preprocessed data

Figure 7 compares the raw speed data with the preprocessed speed signal over time. The raw measurements show noticeable noise, sudden spikes, and abrupt drops, indicating sensor irregularities and signal disturbances. After applying outlier removal, interpolation, and Kalman filtering, the processed signal becomes smoother and more continuous while preserving the overall driving pattern. The preprocessing effectively reduces high-frequency noise and unrealistic fluctuations, resulting in a stable speed profile suitable for accurate acceleration and energy consumption analysis.

Table 3 Preprocessed Data

t_seconds	Date_Time	Lat.	Long.	Alt.	Speed	Angle
0	2024-07-01 05:45:10	27.3804 983	86.0594 266	575	0	77
1	2024-07-01 05:45:11	27.3804 9829	86.0594 2661	575.000 1634	0.00144 546	77.0019 8449
2	2024-07-01 05:45:12	27.3804 9829	86.0594 2662	575.000 3267	0.00289 0318	77.0039 6815
3	2024-07-01 05:45:13	27.3804 9828	86.0594 2662	575.000 4899	0.00433 3971	77.0059 5016
4	2024-07-01 05:45:14	27.3804 9827	86.0594 2663	575.000 6529	0.00577 5817	77.0079 2969
5	2024-07-01 05:45:15	27.3804 9827	86.0594 2664	575.000 8156	0.00721 5254	77.0099 0591
6	2024-07-01 05:45:16	27.3804 9826	86.0594 2665	575.000 978	0.00865 1679	77.0118 7799
7	2024-07-01 05:45:17	27.3804 9825	86.0594 2666	575.001 14	0.01008 449	77.0138 4511

3.5 Incremental Energy Calculation

To estimate traction energy demand, incremental mechanical energy at the wheels was computed using longitudinal vehicle dynamics. The model accounts for rolling resistance, aerodynamic drag, gravitational force due to road gradient, and inertial force during acceleration. The incremental energy over a small distance segment for per-second resolution preprocessed data was calculated.

$$dE = \frac{1}{3600} \left[mg(f \cos \varphi + \sin \varphi) + \frac{1}{2} \rho C_x A \left(\frac{v_{EV} + v_w}{3.6} \right)^2 + (m + m_f) \frac{dv}{dt} \right] ds$$

Computes the mechanical energy required at the wheels over a small distance segment ds in kWh. This physics-based formulation ensures that terrain effects, vehicle mass, and speed variations are explicitly captured.

The incremental mechanical energy was computed using measured terrain and vehicle data. The dataset includes altitude, vehicle speed, terrain slope, and segment length for each section of the path.

Mechanical traction energy at the wheels was selected as the modeling target to isolate the effects of terrain, driving behavior, and vehicle dynamics from battery-specific characteristics. Battery energy consumption depends on additional factors such as state of charge, temperature, and auxiliary loads, which were not consistently available in the dataset. By modeling mechanical energy, the proposed framework maintains physical consistency and enables generalization across vehicles, while still providing a meaningful basis for range estimation and energy-aware routing.

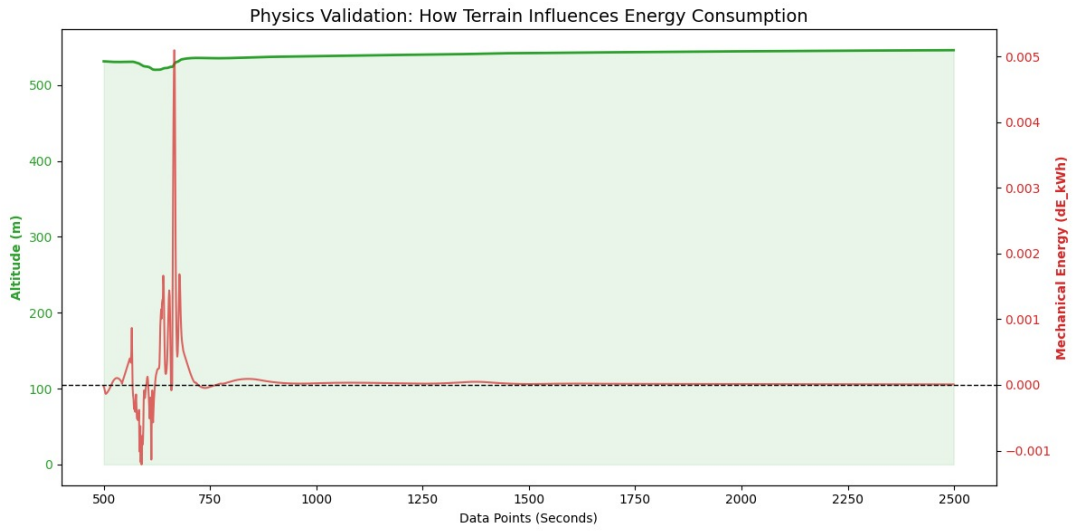


Figure 8 Incremental Mechanical Energy along Varied Altitudes

Figure 8 shows the incremental mechanical energy changes along varied terrain segments. The observed trends in energy align with theoretical expectations, confirming that the model appropriately accounts for terrain-induced energy variations.

3.6 Splitting Train and Test Data

The dataset was divided at the trip level, with approximately 75% of the trips randomly assigned to the training set and the remaining 25% reserved for testing. Random selection of trips avoids selection bias and ensures that both training and testing sets contain representative variations in driving behavior and terrain

conditions. This approach provides a fair and reliable evaluation of the model's generalization performance on unseen trips.

3.7 Model Development

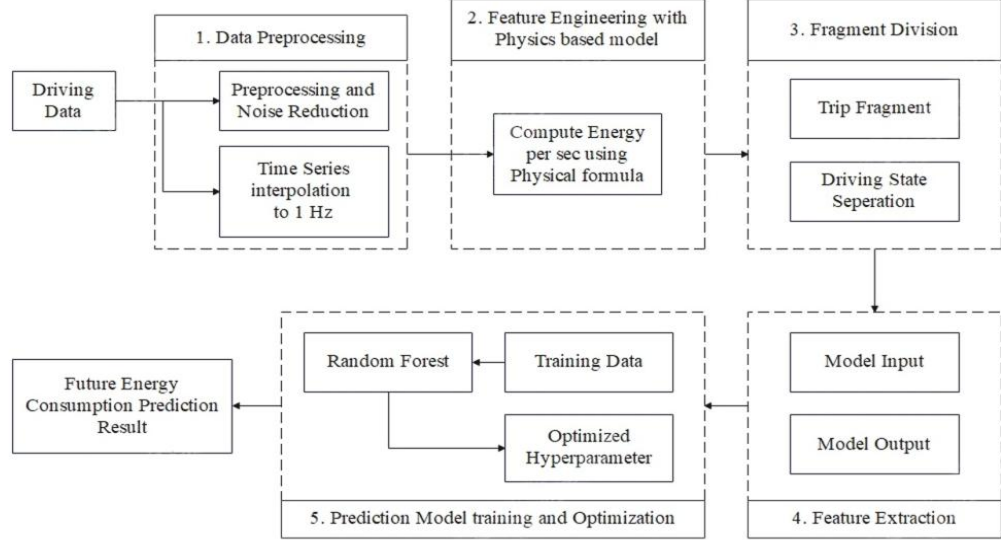


Figure 9 Methodological Process

Figure 9 presents the methodology adopted in this project for predicting vehicle energy consumption. The process begins with driving data collection and preprocessing, including noise removal and time-series interpolation to standardize the data to 1 Hz. Next, physics-based feature engineering is performed to compute energy per second using relevant physical formulas. The processed data is then divided into trip fragments and categorized based on driving states. Following fragment division, significant features are extracted and organized as model inputs and outputs. Finally, a Random Forest prediction model is trained and optimized using the training dataset, and the optimized model is applied to predict future energy consumption.

To capture operational variability in energy consumption, the preprocessed dataset was categorized into four distinct driving states: acceleration, deceleration, cruising, and idling, based on speed and acceleration thresholds. Kinematic drive states were separated based on:

- Idling: 0 to 3.6 km/h
- Acceleration: $> 0.3 \text{ m/s}^2$
- Deceleration: $< -0.3 \text{ m/s}^2$

- Cruising: remaining cruising

Random Forest regression was selected due to its robustness to noisy inputs, ability to model nonlinear relationships, and lower overfitting risk compared to neural networks for moderate-sized datasets. Input features included speed, gradient, acceleration, and derived energy-related variables.

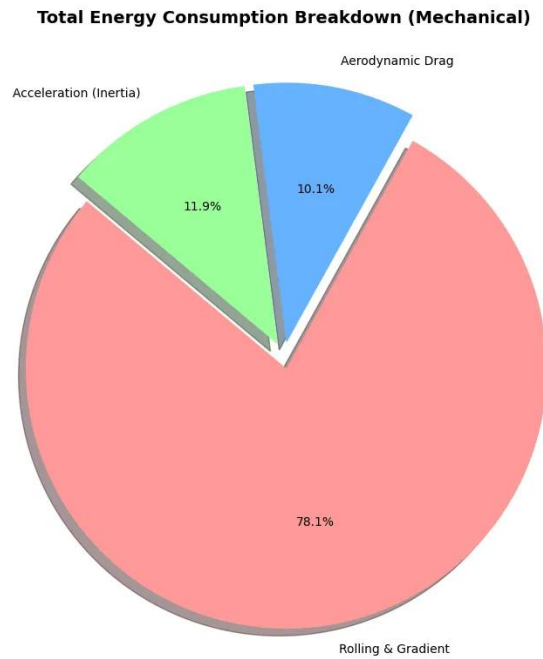


Figure 10 Energy Consumption Breakdown

Figure 10 presents the proportional contribution of each input feature to the predicted energy. Terrain slope and vehicle velocity account for the largest shares, consistent with their expected influence on mechanical energy, thereby validating the model's sensitivity to physically relevant variables.

Energy consumption was aggregated over 3 km segments before modeling. The 3 km segmentation was selected to:

- Reduce short-term noise effects
- Capture terrain influence over meaningful road sections
- Reflect practical range estimation intervals
- Balance data granularity and model stability

Shorter segments may introduce high variance, while longer segments may over-smooth terrain influence. The 3 km window provides an optimal trade-off between resolution and robustness.

Table 4 Random Forest Hyperparameters

Hyperparameters	Value	Description
n_estimators	100	The number of decision trees in the forest.
max_depth	10	Maximum depth of the trees to capture non-linear terrain relationships.
random_state	42	Ensures reproducibility of the model results across different runs.

3.8 Model Evaluation

The model's performance is assessed on the reserved test set using three complementary metrics: RMSE, MAE, and R^2 . RMSE emphasizes larger deviations, reflecting the model's ability to avoid physically unrealistic predictions. MAE provides a robust measure of average prediction error without overemphasizing outliers. R^2 quantifies the proportion of variance explained by the model, providing an intuitive measure of overall predictive quality. Together, these metrics offer a comprehensive evaluation of both accuracy and reliability

3.8.1 Error Definitions

Let the target values be v_t and the predicted values be v_p for n test samples. Define the prediction error for each sample as

$$\varepsilon = v_{t,i} - v_{p,i}$$

Where; $i=1,2, 3, \dots, n$

Coefficient of Determination (R^2): R^2 measures how much of the variance in the target variable is explained by the model. Values closer to 1 indicate better predictive performance.

$$R^2 = 1 - \frac{\sum_{i=1}^n (v_{t,i} - v_{p,i})^2}{\sum_{i=1}^n (v_{t,i} - \bar{v}_t)^2}$$

Where; $\bar{v}_t = \frac{1}{n} \sum_{i=1}^n v_{t,i}$ is the mean of the target values.

Root Mean Square Error (RMSE): RMSE penalizes larger errors more strongly than smaller ones, making it sensitive to outliers. Useful when large deviations from the

target are particularly undesirable (e.g., energy predictions that are physically unrealistic).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2}$$

Mean Absolute Percentage Error (MAPE): Expresses the error as a percentage of the target value. Useful when target values vary widely in scale, allowing comparison of errors across different magnitudes. Lower MAPE indicates better model performance.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{v_{t,i} - v_{p,i}}{v_{t,i}} \right|$$

Additionally, predicted versus actual energy consumption plots were analyzed to assess model performance.

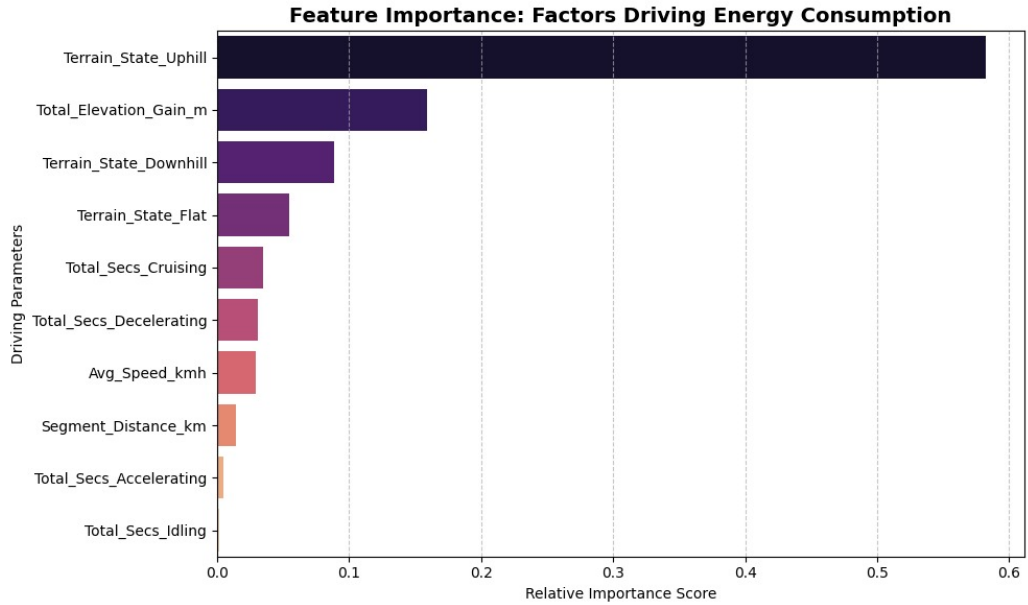


Figure 11 Feature Importance of Data

Figure 11 illustrates the relative importance of input features based on model evaluation. The highest importance is attributed to terrain slope and velocity, aligning with both physical expectations and the model's predictive accuracy, supporting the interpretability of the learned relationships.

3.9 Gantt Chart

Table 5 Work Schedule of the Project

Task Description/Month	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan	Feb
Literature Review on the Study area and route											
Literature Review on Preprocessing Methods											
Literature Review on Parameters Affecting Energy Consumption											
Literature Review on Machine Learning											
Selection of Preprocessing Methods											
Data Preprocessing											
Literature Review on Modelling Method											
Incremental Energy Calculation											
Selection of Appropriate Modeling Methods											
Model Development											
Model Evaluation											
Testing and Modification											

Chapter 4: Findings, Discussion, and Recommendations

4.1 Results and Discussion

Energy Prediction:

The Random Forest model was developed and evaluated through three sequential iterations to progressively improve prediction accuracy for 3 km terrain micro-segments. Each iteration involved training on the same dataset of terrain and vehicle features (altitude, speed, slope, segment length) while refining model parameters and assessing performance on the test set. The following sections summarize the results and observations from each iteration.

Table 6 Random Forest Model Performance across Iterations

Iteration	Description	R ²	RMSE (kWh / Trip)	MAPE (% / Trip)	Key Observations
1	Baseline model with default parameters	0.7505	3.1902	20.55	Lower RMSE and MAPE indicate consistent but slightly underfitted predictions.
2	Model refinement with optimized hyperparameters	0.8273	2.0725	25.49	Use of hyperparameter tuning improves the model.

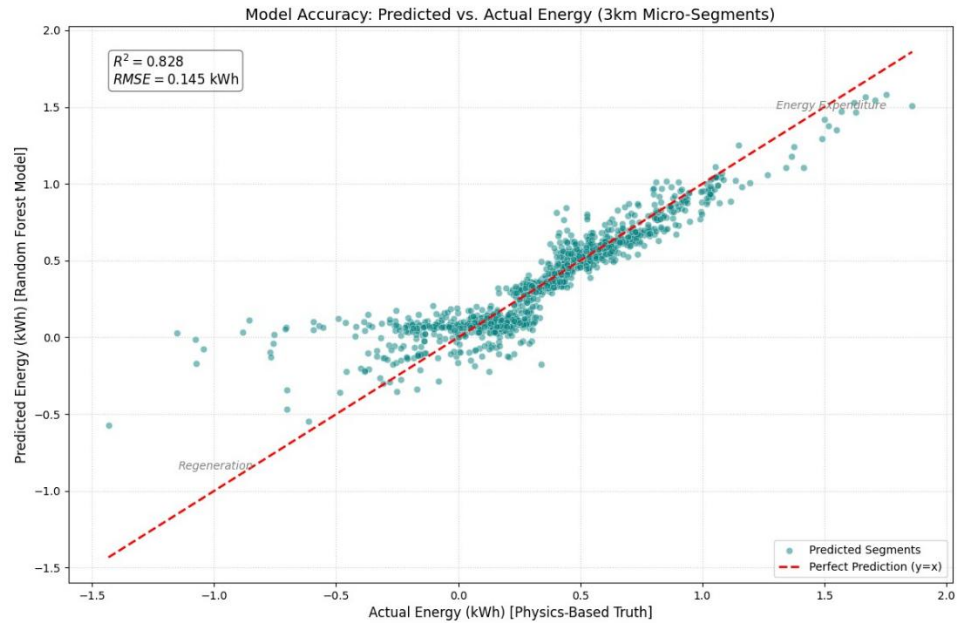


Figure 12 Random Forest Model Accuracy

The Predicted vs. Actual Energy plot for 3 km micro-segments demonstrates strong agreement between the physics-based ground truth and the Random Forest model predictions.

Key performance metrics:

- $R^2 = 0.8273$
- $RMSE = 0.145$ kWh per 3 km segment
- $RMSE = 2.0918$ kWh per full trip
- $MAPE = 25.49$ % per trip

It is important to note that error metrics are reported at two different aggregation levels. The RMSE value of 0.145 kWh corresponds to predictions at the 3 km segment level, while the RMSE value of 2.0735 kWh represents total energy prediction error at the full-trip level. Since trip energy is obtained by aggregating multiple 3 km segments, the trip-level RMSE naturally reflects accumulated segment-level deviations.

An R^2 of 0.8273 indicates that the model explains about 82.81% of the variation in energy consumption on the hilly route. Most data points lie close to the 45° perfect prediction line ($y = x$), confirming good overall accuracy.

However, the graph also shows:

- i. Slight dispersion in high-energy uphill segments
- ii. Larger spread in regenerative regions

iii. Minor underprediction at very high consumption levels

These deviations reflect areas where the model is less precise. This variation can be linked to the study's limitations. The absence of battery-related parameters such as SOC and efficiency means the model predicts only mechanical traction energy, not actual battery energy behavior. Additionally, external factors like temperature and weather conditions may contribute to inconsistencies in acceleration and deceleration phases observed in the graph.

Since the dataset is limited to three months and a single hilly route, the model's generalization capability is constrained. While it performs well for the Kathmandu to Manthali route, its accuracy may vary under different seasonal conditions or on other road types. Furthermore, excluding auxiliary loads such as air conditioning and heating may cause a slight underestimation of total energy demand. Therefore, while the model demonstrates strong predictive capability for terrain-based traction energy, incorporating additional real-world factors would improve robustness and practical applicability.

Driving Range:

The driving range of the vehicle was estimated by integrating the Random Forest energy consumption predictions with the vehicle's total usable battery capacity of about 41.86 kWh. Given the challenging hilly terrain of the Kathmandu to Manthali route, the model predicted an overall average driving range of 357.27 km, while the actual overall average driving range was 363.76 km.

Key performance metrics:

- $R^2 = 0.1479$ per trip
- RMSE = 20.22 km per trip
- MAPE = 3.92 % per trip

4.2 Recommendations

- i. Include SOC and battery parameters to improve uphill energy predictions.
- ii. Expand the dataset across multiple routes for better generalization.
- iii. Model auxiliary loads to combine traction and auxiliary energy for realistic range prediction.

REFERENCES

- [1] Y. Wang, W. Chenlong, and H. Hongwen, "Estimating the energy consumption and driving range of electric vehicles with machine learning," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Aug. 2021. doi: 10.1088/1742-6596/2005/1/012131.
- [2] H. Agoro, "Comparative Study of EV and Internal Combustion Engine (ICE) Vehicles," 2025. [Online]. Available: <https://www.researchgate.net/publication/390311945>
- [3] W. Achariyaviriya *et al.*, "Estimating Energy Consumption of Battery Electric Vehicles Using Vehicle Sensor Data and Machine Learning Approaches," *Energies (Basel)*, vol. 16, no. 17, Sep. 2023, doi: 10.3390/en16176351.
- [4] S. Sun, J. Zhang, J. Bi, Y. Wang, and M. H. Y. Moghaddam, "A Machine Learning Method for Predicting Driving Range of Battery Electric Vehicles," *J. Adv. Transp.*, vol. 2019, 2019, doi: 10.1155/2019/4109148.
- [5] C. De Cauwer, W. Verbeke, T. Coosemans, S. Faïd, and J. Van Mierlo, "A data-driven method for energy consumption prediction and energy-efficient routing of electric vehicles in real-world conditions," *Energies (Basel)*, vol. 10, no. 5, 2017, doi: 10.3390/en10050608.
- [6] "Outlier Detection Techniques for Time Series | by Alex Eslava | Medium." Accessed: Feb. 25, 2026. [Online]. Available: <https://medium.com/@alex.eslava96/outlier-detection-techniques-for-time-series-9868db2875c2>
- [7] M. Lepot, J. B. Aubin, and F. H. L. R. Clemens, "Interpolation in Time Series: An Introductory Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment," *Water 2017, Vol. 9, Page 796*, vol. 9, no. 10, p. 796, Oct. 2017, doi: 10.3390/w9100796.
- [8] T. Ono, H. W. Kim, M. Cho, and M. C. Lee, "A Study on Reducing the Noise Using the Kalman Filter in Digital Holographic Microscopy (DHM)," *Electronics 2025, Vol. 14, Page 338*, vol. 14, no. 2, p. 338, Jan. 2025, doi: 10.3390/electronics14020338.
- [9] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random Forests," in *Ensemble Machine Learning*, New York, NY: Springer New York, 2012, pp. 157–175. doi: 10.1007/978-1-4419-9326-7_5.
- [10] "DECISION TREE AND RANDOM FOREST CLASSIFIER IN ML.. | by Jivanjot | Medium." Accessed: Aug. 05, 2025. [Online]. Available: <https://medium.com/@jivanjotk.32/decision-tree-and-random-forest-classifier-in-ml-b4a54a5931a9>

- [11] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, vol. 7, no. DEC, 2013, doi: 10.3389/fnbot.2013.00021.
- [12] "A Quick Guide to Boosting Algorithms in Machine Learning." Accessed: Aug. 05, 2025. [Online]. Available: <https://brainalyst.in/boosting-algorithms-in-machine-learning/>
- [13] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 6, pp. 1–20, Nov. 2021, doi: 10.1007/S42979-021-00815-1/METRICS.
- [14] Y. Yang, "Optimizing Electric Vehicle Range Prediction Using Machine Learning: A Feature-Based Comparative Study," 2025, doi: 10.5220/0013822700004708.
- [15] "Study the impact of different parameters on driving range of electric vehicle using Matlab Simulink - ISJEM Journal." Accessed: Feb. 25, 2026. [Online]. Available: <https://isjem.com/download/study-the-impact-of-different-parameters-on-driving-range-of-electric-vehicle-using-matlab-simulink/>
- [16] M. Miotti, Z. A. Needell, S. Ramakrishnan, J. Heywood, and J. E. Trancik, "Quantifying the impact of driving style changes on light-duty vehicle fuel consumption," *Transp. Res. D Transp. Environ.*, vol. 98, Sep. 2021, doi: 10.1016/j.trd.2021.102918.
- [17] S. L. Prasad and A. Gudipalli, "An effective range estimation and state-of-charge to mitigate range anxiety in electric vehicles," *Heliyon*, vol. 11, no. 1, p. e41494, Jan. 2024, doi: 10.1016/j.heliyon.2024.e41494.
- [18] "6 Factors That Impact Your Electric Vehicle Range | MG Motors." Accessed: Feb. 25, 2026. [Online]. Available: <https://www.mgmotor.co.in/blogs/6-factors-that-impact-your-evs-range>
- [19] "Rolling Resistance Force Equations and Calculator." Accessed: Feb. 26, 2026. [Online]. Available: https://www.engineersedge.com/calculators/rolling_resistance_force_14728.htm