AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 1(2026)

REAL-WORLD ELECTRIC VEHICLE ENERGY CONSUMPTION PREDICTION USING MACHINE LEARNING ALGORITHMS

MUDDASANI SHRAVANI[1], M SAI TEJA[2], KOMMULA DEEPAK REDDY[3], KARAGIRI MANIKANTA[4], KATIPALLY SPOORTHI[5], KAJRANA DEEPAK KUMAR[6]

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3,4,5&6]

DEPARTMENT OF CSE, NARSIMHA REDDY ENGINEERING COLLEGE (UGC- AUTONOMOUS) MAISAMMAGUDA (V), KOMPALLY, SECUNDERABAD, TELANGANA-500100

ABSTRACT

The rapid adoption of electric vehicles (EVs) has increased the need for accurate prediction of energy consumption to improve battery management, driving efficiency, and route planning. Electric vehicle energy usage is influenced by multiple factors such as driving behavior, vehicle characteristics, road conditions, traffic patterns, and environmental variables. Traditional estimation methods often fail to capture these complex relationships, leading to inaccurate predictions. This study proposes a **machine learning–based approach for predicting real-world electric vehicle energy consumption using field data** collected from EV sensors and telematics systems. The system analyzes various input features including vehicle speed, acceleration, battery state of charge, temperature, road gradient, and driving patterns. Machine learning algorithms such as Random Forest, Support Vector Machine, and Gradient Boosting are applied to model the nonlinear relationships between these factors and energy consumption. The proposed framework preprocesses the collected data, extracts relevant features, and trains predictive models to estimate energy usage under different driving conditions. The trained models are evaluated using performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and prediction accuracy. Experimental results demonstrate that machine learning algorithms can significantly improve the accuracy of EV energy consumption prediction compared to traditional methods. The developed system can support **intelligent energy management, optimal route planning, and improved battery utilization** in electric vehicles. This research contributes to the development of smarter and more efficient EV transportation systems.

INTRODUCTION

The rapid growth of **electric vehicles (EVs)** has become an important step toward reducing greenhouse gas emissions, minimizing dependence on fossil fuels, and promoting sustainable transportation. Governments and automotive industries worldwide are increasingly encouraging the adoption of EVs due to their environmental benefits and improved energy efficiency compared to conventional internal combustion engine vehicles. However, one of the major challenges faced by EV users and manufacturers is accurately predicting **energy consumption and driving range** under real-world operating conditions.

Electric vehicle energy consumption is influenced by several factors such as **driving behavior, vehicle speed, road conditions, traffic patterns, weather conditions, battery characteristics, and vehicle load**. Traditional mathematical and physics-based models often struggle to capture these complex relationships, especially when dealing with large and dynamic real-world datasets. As a result, predicting EV energy usage with high accuracy remains a challenging task.

In recent years, **machine learning algorithms** have emerged as powerful tools for analyzing large-scale transportation and vehicle datasets. These algorithms can learn hidden patterns from historical driving data and provide accurate predictions of energy consumption under varying conditions. By utilizing real-world field data collected from sensors, onboard diagnostics (OBD) systems, and telematics devices, machine learning models can better understand the complex interactions between vehicle operation and environmental factors.

This study focuses on developing a **machine learning-based framework** for predicting electric vehicle energy consumption using real-world operational data. The proposed approach applies advanced machine learning techniques to analyze driving patterns and environmental parameters to estimate energy usage more accurately. Accurate energy consumption prediction can help improve **route planning, battery management, charging infrastructure planning, and range estimation**, ultimately enhancing the reliability and efficiency of electric vehicles.

Therefore, integrating machine learning techniques with real-world EV data provides a promising solution for improving **energy efficiency, smart transportation systems, and sustainable mobility** in modern electric vehicle ecosystems.

LITERATURE REVIEW

1. Traditional Energy Consumption Modeling

Early research on electric vehicle (EV) energy consumption relied mainly on **physics-based and vehicle dynamics models**. These models estimate energy usage using parameters such as vehicle mass, road gradient, aerodynamic drag, rolling resistance, and battery characteristics. Although these approaches provide interpretable results, they often struggle to capture complex driving behaviors and real-world environmental variations such as traffic congestion or weather conditions. As a result, prediction accuracy may decrease in practical applications.

2. Statistical and Regression-Based Prediction Methods

To improve prediction accuracy, researchers introduced **statistical and regression models**, including multiple linear regression and probabilistic regression techniques. These approaches analyze relationships between variables such as vehicle speed, acceleration, road conditions, and temperature. Regression-based models can estimate trip-level energy consumption but often fail to represent nonlinear relationships present in real-world driving data.

3. Machine Learning Approaches for EV Energy Prediction

With the increasing availability of **large-scale driving datasets and field data**, machine learning (ML) algorithms have been widely adopted for predicting EV energy consumption. Algorithms such as **Support Vector Regression (SVR), K-Nearest Neighbors (KNN), Random Forest, Gradient Boosting, and Extreme Gradient Boosting (XGBoost)** have been applied to capture nonlinear relationships between driving factors and energy consumption. Comparative studies show that advanced ML models like XGBoost and Random Forest generally outperform traditional regression models in prediction accuracy when trained on real-world datasets.

4. Deep Learning-Based Energy Consumption Prediction

Recent studies focus on **deep learning models** because of their ability to automatically extract complex features from large datasets. Neural network architectures such as **Deep Neural Networks (DNN), Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Transformer models** have been applied to EV energy prediction tasks. These models can capture temporal patterns in driving data and improve prediction accuracy for dynamic conditions such as varying traffic patterns and driving behaviors. For example, LSTM and hybrid transformer models have achieved lower prediction errors compared with conventional ML algorithms.

5. Ensemble and Hybrid Machine Learning Models

To further improve predictive performance, researchers have developed **ensemble learning frameworks** that combine multiple machine learning models. Techniques such as **stacking, bagging, and boosting** integrate the strengths of different algorithms. For instance, a stacking model combining KNN, Random Forest, and Extra Trees with a linear regression meta-model demonstrated improved robustness and prediction accuracy compared to single-model approaches. Ensemble models help reduce prediction variance and handle complex feature interactions more effectively.

6. Challenges and Research Gaps

Despite the progress made in EV energy prediction research, several challenges remain. Real-world energy consumption depends on many factors including **driver behavior, traffic conditions, road topology, vehicle specifications, and environmental factors**. Accurately capturing these variables in predictive models remains difficult. Additionally, some advanced deep learning models require large training datasets and high computational resources, which may limit their use in real-time applications. Future research aims to develop **more interpretable, scalable, and real-time machine learning models** that can efficiently predict EV energy consumption using diverse real-world datasets.

SYSTEM ANALYSIS

EXISTING SYSTEM

Traditional electric vehicle (EV) energy consumption prediction methods mainly rely on **physics-based models and simple statistical techniques**. These models estimate energy usage based on parameters such as vehicle speed, battery capacity, road gradient, driving distance, and environmental conditions. Many early systems use predefined mathematical formulas or rule-based approaches to estimate energy consumption.

Some existing studies also apply **basic machine learning algorithms** such as Linear Regression, Support Vector Machines (SVM), and Decision Trees to predict EV energy usage. These models analyze historical driving data and vehicle parameters to estimate energy consumption for a specific trip or driving condition.

However, these systems have several limitations. Physics-based models often fail to capture **complex real-world driving behavior**, including sudden acceleration, traffic congestion, and driver habits. Similarly, traditional machine learning models may struggle with **large-scale and high-dimensional datasets**, which can reduce prediction accuracy. Many existing systems also rely on **limited datasets or simulated driving conditions**, which do not fully represent real-world scenarios. As a result, prediction performance may be inconsistent and less reliable.

PROPOSED SYSTEM

The proposed system introduces a **Real-World Electric Vehicle Energy Consumption Prediction Framework using Advanced Machine Learning Algorithms**. This system utilizes **large-scale real-world field data** collected from EV sensors, telematics systems, and onboard diagnostic devices.

The collected data includes parameters such as **vehicle speed, acceleration, battery state of charge (SOC), driving distance, road slope, temperature, traffic conditions, and driving patterns**. After collecting the data, a **data preprocessing stage** is performed to clean, normalize, and prepare the dataset for model training.

The system then applies **advanced machine learning models**, such as Random Forest, Gradient Boosting, Artificial Neural Networks (ANN), or Deep Learning models, to learn complex relationships between driving conditions and energy consumption. These models analyze historical patterns and generate accurate predictions for future trips.

The trained model can be integrated into **intelligent transportation systems or EV navigation applications** to estimate energy usage, optimize route planning, and improve battery management. By using real-world field data and advanced algorithms, the proposed system significantly improves prediction accuracy and reliability.

Advantages of the Proposed System

- Higher prediction accuracy using real-world driving data
- Ability to handle large and complex datasets
- Improved battery management and trip planning
- Better estimation of EV driving range
- Support for intelligent transportation and smart mobility systems

IMPLEMENTATION

1. Data Collection

The first step in the implementation is collecting **real-world electric vehicle (EV) operational data**. The data can be obtained from onboard vehicle sensors, telematics systems, charging stations, or publicly available EV datasets. Important parameters include **vehicle speed, battery state of charge (SOC), acceleration, distance traveled, road conditions, temperature, and driving behavior**. This data provides the foundation for training machine learning models to predict energy consumption.

2. Data Preprocessing

In this phase, the collected raw data is cleaned and prepared for model training. Data preprocessing includes **removing missing**

values, handling outliers, normalizing numerical features, and encoding categorical variables. Feature selection techniques are also applied to identify the most relevant parameters influencing EV energy consumption. Proper preprocessing improves the accuracy and efficiency of machine learning models.

3. Feature Engineering

Feature engineering involves creating meaningful input variables from the available dataset. Derived features such as **average speed, energy consumption per kilometer, acceleration patterns, and road gradient effects** can be generated. These features help the model better understand the relationship between driving conditions and energy consumption.

4. Model Selection and Training

Different **machine learning algorithms** are used to build predictive models for EV energy consumption. Common algorithms include:

- Linear Regression
- Random Forest
- Support Vector Machine (SVM)
- Gradient Boosting
- Artificial Neural Networks (ANN)

The dataset is divided into **training and testing sets**, and the models are trained using the training data to learn patterns in energy consumption behavior.

5. Model Evaluation

After training, the models are evaluated using performance metrics such as:

- **Mean Absolute Error (MAE)**
- **Mean Squared Error (MSE)**
- **Root Mean Square Error (RMSE)**
- **R-squared (R^2) score**

These metrics help determine how accurately the model predicts energy consumption under different driving conditions.

6. Model Optimization

Hyperparameter tuning techniques such as **Grid Search or Cross-Validation** are used to improve the performance of the selected machine learning model. This step helps in selecting the best model configuration for accurate prediction.

7. Deployment and Real-Time Prediction

The final optimized model is deployed in a **vehicle monitoring system or cloud platform**. Real-time vehicle data from sensors is fed into the trained model to predict **future energy consumption and driving range**. The system can assist drivers in planning efficient routes and managing battery usage.

8. Visualization and Monitoring

The predicted energy consumption results are displayed through **dashboards or mobile applications**. Visualization tools help drivers, fleet managers, and researchers monitor energy usage trends and optimize EV performance.

9.Algorithms

DECISION TREE CLASSIFIERS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

GRADIENT BOOSTING Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

K-NEAREST NEIGHBORS (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)

- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

LOGISTIC REGRESSION CLASSIFIERS

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar. Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does. This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

NAÏVE BAYES

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique. Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance. The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg. An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance. Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space. SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

CONCLUSION

This paper has introduced a field data-based ML pipeline for the prediction of EV energy consumption. The first novelty arises from the proposed data processing method tailored for a large amount of real-world EV data that was inherently plagued by various issues and outliers. Then, a new feature set was constructed from physical insights and picked meticulously through systematic correlation analyses. Based on these data and features, four quantile-based machine learning algorithms were pertinently formulated and innovatively applied for the EV energy prediction, enabling accurate and reliable prediction of both the energy consumption and associated uncertainties. Finally, the best-performing global ML models were adapted online for individualized predictions, leadingto consistently improved accuracy and tightened confidence internals.The developed ML models as well as their underpinning data processing and feature engineering were validated extensively for EV energy prediction. Comprehensive comparisons were conducted for different steps of data processing, between global models and online adaptive models, and with models in the literature. The online adaptive QRNN models outperformed all other models with an average prediction error of 5.04%, corresponding to an over 35% improvement over the state-of-the-art models. Substantial advantages have also been observed from different steps of data processing and online model adaptation.

REFERENCES

[1] G. J. Offer, "Automated vehicles and electrification of transport," Energy Environ. Sci., vol. 8, no. 1, pp. 26–30, 2015.

[2] E. M. Bibra, E. Connelly, S. Dhir, M. Drtil, P. Henriot, I. Hwang, J.- B. L. Marois, S. McBain, P. Leonardo, and T. Jacob, "Global EV outlook 2022: Securing supplies for an electric future," Int. Energy Agency, 2022.

[3] G. Crabtree, "The coming electric vehicle transformation," Science, vol. 366, no. 6464, pp. 422–424, 2019.

[4] D. Tan, "Transportation electrification: Challenges and opportunities," IEEE Power Electron. Mag., vol. 3, no. 2, pp. 50–52, 2016.

[5] X. Hu, S. J. Moura, N. Murgovski, B. Egardt, and D. Cao, "Integrated optimization of battery sizing, charging, and power management in plugin hybrid electric vehicles," IEEE Trans. Contr. Syst. Technol., vol. 24, no. 3, pp. 1036–1043, 2015.

[6] Y. Xu, S. C¸ olak, E. C. Kara, S. J. Moura, and M. C. Gonz´alez, "Planning for electric vehicle needs by coupling charging profiles with urban mobility," Nat. Energy, vol. 3, no. 6, pp. 484–493, 2018.

[7] D. Baek, Y. Chen, A. Bocca, L. Bottaccioli, S. Di Cataldo, V. Gatteschi, D. J. Pagliari, E. Patti, G. Urgese, N. Chang et al., "Battery-aware
operation range estimation for terrestrial and aerial electric vehicles," IEEE Trans. Veh. Technol., vol. 68, no. 6, pp. 5471–5482, 2019.

[8] R. Basso, B. Kulcs´ar, and I. Sanchez-Diaz, "Electric vehicle routing problem with machine learning for energy prediction," Transp. Res. B, vol. 145, pp. 24–55, 2021.

[9] Y. Zhang, X. Qu, and L. Tong, "Optimal eco-driving control of autonomous and electric trucks in adaptation to highway topography:
Energy minimization and battery life extension," IEEE Trans. Transport. Electrif., vol. 8, no. 2, pp. 2149–2163, 2022.

[10] J. Zhang, T.-Q. Tang, Y. Yan, and X. Qu, "Eco-driving control for connected and automated electric vehicles at signalized intersections
with wireless charging," Appl. Energy, vol. 282, p. 116215, 2021.

[11] M. Sabri, K. A. Danapalasingam, and M. F. Rahmat, "A review on hybrid electric vehicles architecture and energy management strategies," Renew. Sustain. Energy Rev., vol. 53, pp. 1433–1442, 2016.

[12] G. Correa, P. Mu˜noz, and C. Rodriguez, "A comparative energy and environmental analysis of a diesel, hybrid, hydrogen and electric urban bus," Energy, vol. 187, p. 115906, 2019.

[13] H. Abdelaty, A. Al-Obaidi, M. Mohamed, and H. E. Farag, "Machine learning prediction models for battery-electric bus energy consumption in transit," Transp. Res. D, vol. 96, p. 102868, 2021.

[14] M. Xylia, S. Leduc, P. Patrizio, F. Kraxner, and S. Silveira, "Locating charging infrastructure for electric buses in stockholm," Transp. Res. C, vol. 78, pp. 183–200, 2017.

[15] L. Li, H. K. Lo, and F. Xiao, "Mixed bus fleet scheduling under range and refueling constraints," Transp. Res. C, vol. 104, pp. 443–462, 2019.