INTERNATIONAL JOURNAL OF RESEARCH AND SCIENTIFIC INNOVATION (IJRSI)
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

Electric Vehicle Energy Consumption Prediction: A Physics-Informed Machine Learning Approach

[1] Emine Can*, [2]Elif Selay HAYAL, [3]Maksude Selina YAVUZ, [4]Çağdaş Alper YEGİT, [5]Efe SEVER, [6]Nafiseh Farajirad

[1]Faculty of Engineering and Natural Sciences, Physics Egineering Department, Istanbul Medeniyet University, Istanbul, Türkiye,

[2]Faculty of Engineering and Natural Sciences, Physics Engineering Department, Istanbul Medeniyet University, Istanbul, Türkiye

[3]Faculty of Engineering and Natural Sciences, Physics Engineering Department, Istanbul Medeniyet University, Istanbul, Türkiye

[4]Faculty of Engineering and Natural Sciences, Physics Engineering Department, Istanbul Medeniyet University, Istanbul, Türkiye

[5]Faculty of Engineering and Natural Sciences, Physics Engineering Department, Istanbul Medeniyet University, Istanbul, Türkiye

[6]Occupational Health and Safety Department, Faculty of Health Sciences, Uskudar University, Istanbul, Türkiye

DOI: https://doi.org/10.51244/IJRSI.2026.1307000026

Received: 05 July 2026; Accepted: 10 July 2026; Published: 23 July 2026

ABSTRACT

The increasing adoption of electric vehicles (EVs) as a sustainable alternative to internal combustion engine vehicles has intensified the need for accurate and interpretable energy consumption prediction models to support vehicle design, battery management, and charging infrastructure planning. This study presents a physics-informed machine learning framework for predicting EV energy consumption using a dataset of approximately 300 electric vehicle models sourced from publicly available vehicle specifications. A reduced-order physical model derived from the work–energy theorem and Newtonian mechanics was developed to bridge classical vehicle dynamics theory with data-driven modeling, incorporating vehicle mass, aerodynamic drag coefficient, and the mass-to-battery-capacity ratio as physically meaningful input features. Five regression algorithms Linear Regression, Random Forest, XGBoost, LightGBM, and Support Vector Regression were implemented and evaluated under a consistent 5-fold cross-validation framework using R^2, RMSE, and MAE as performance metrics. Random Forest achieved the highest predictive accuracy ($R^2 = 0.841$, RMSE = 1.593 kWh/100 km), followed by Linear Regression ($R^2 = 0.837$) and XGBoost ($R^2 = 0.820$), while LightGBM and SVR demonstrated substantially weaker performance. Feature importance analysis confirmed that battery capacity, vehicle mass, and driving range are the most influential predictors, consistent with the physics-informed framework and recent literature. The convergence between data-driven findings and physical interpretations validates the proposed approach as a robust, transparent, and scalable tool for EV energy modeling, with direct applications in sustainable transportation planning and evidence-based energy policy.

Keywords: Electric vehicles, Energy consumption, Machine learning, Physics-informed modeling.

INTRODUCTION

The transportation sector is one of the largest contributors to global greenhouse gas (GHG) emissions, accounting for approximately 23% of total energy-related CO_2 emissions worldwide (IPCC, 2022). Among the various mitigation strategies being pursued, the electrification of road transport has emerged as one of the most promising pathways toward decarbonization. Electric vehicles (EVs) eliminate tailpipe emissions, offer higher energy conversion efficiency compared to internal combustion engine (ICE) vehicles, and can be integrated with renewable energy sources to further reduce lifecycle carbon footprints (IEA, 2024). In response to these advantages, global EV adoption has accelerated substantially over the past decade. The International Energy Agency reported that the number of electric cars on the road surpassed 40 million in 2023, with sales continuing to grow at record pace across major markets including China, Europe, and North America (IEA, 2024). Policy frameworks such as the European Green Deal and national zero-emission vehicle mandates have further reinforced this trajectory, positioning electric mobility as a central element of the global energy transition (OECD, 2023). Despite this rapid growth, one of the most persistent technical challenges limiting the wider adoption of EVs is range anxiety the concern that a vehicle's battery will be depleted before the driver reaches their destination or a charging point. Accurate prediction of energy consumption is therefore critical, not only for real-time range estimation and driver assistance systems, but also for vehicle design optimization, battery management, charging infrastructure planning, and energy policy formulation (Guzzella & Sciarretta, 2013; Larminie & Lowry, 2003). However, EV energy consumption is a complex phenomenon influenced by a multitude of interacting factors, including vehicle mass, aerodynamic drag, battery capacity, powertrain efficiency, driving speed, road gradient, ambient temperature, and driver behavior. This multidimensional complexity makes it difficult to develop accurate predictive models using purely physics-based analytical approaches. From a physical standpoint, the energy demand of an electric vehicle is governed by the work–energy theorem and classical Newtonian mechanics. The total tractive force acting on the vehicle includes rolling resistance, aerodynamic drag, road slope effects, and inertial forces, all of which vary dynamically during real-world driving (Guzzella & Sciarretta, 2013). While these fundamental relationships are well established in vehicle dynamics literature, practical implementation is constrained by the unavailability of certain parameters such as instantaneous acceleration profiles, road inclination data, and real-time environmental conditions in standardized vehicle specification datasets. This gap between theoretical completeness and practical data availability motivates the use of data-driven approaches that can approximate complex physical relationships from observable vehicle characteristics. In recent years, machine learning (ML) has emerged as a powerful paradigm for modeling complex, high-dimensional systems where traditional analytical methods are limited. Unlike physics-based models, ML algorithms can identify nonlinear patterns and feature interactions directly from data, without requiring explicit specification of the underlying functional relationships. A growing body of literature has demonstrated the effectiveness of ML techniques for EV energy consumption prediction. Pokharel et al. (2021) applied multiple regression-based machine learning models to real-world EV data and used SHapley Additive exPlanations (SHAP) values to identify the most influential vehicle features, demonstrating that data-driven approaches can achieve high predictive accuracy while offering interpretability. Similarly, Ullah et al. (2022) conducted a comparative analysis of several ML algorithms including Random Forest, Support Vector Regression, and gradient boosting methods and reported that ensemble-based models consistently outperform conventional regression approaches for EV energy prediction tasks. Zhang et al. (2024) extended this line of inquiry by reviewing ML applications in urban EV energy modeling, emphasizing the importance of feature selection and model interpretability for practical deployment. More recently, Hussain et al. (2025) systematically benchmarked multiple ML algorithms for EV energy consumption prediction and confirmed the superiority of ensemble methods in handling nonlinear relationships within structured vehicle specification datasets. Beyond predictive performance, the literature has increasingly emphasized the role of specific vehicle characteristics as dominant drivers of energy consumption. Weiss et al. (2020) demonstrated that vehicle mass has a pronounced effect on energy demand across vehicle classes, with heavier vehicles exhibiting significantly higher consumption under standardized test cycles. Di Martino et al. (2022) conducted a comprehensive review of EV energy consumption modeling strategies and identified aerodynamic drag, vehicle weight, and powertrain efficiency as the three most consistently influential variables across studies. Complementing these findings, Hao et al. (2021) analyzed real-world driving data and showed that energy consumption variability is substantially explained by vehicle-level parameters rather than driving behavior alone under highway conditions. The interaction between aerodynamic design, vehicle mass, and battery capacity introduces nonlinear dependencies that simple regression models fail to capture adequately, a challenge consistently highlighted across the EV energy modeling literature (Di Martino et al., 2022; Ullah et al., 2022; Zhang et al., 2024). These findings collectively underscore the need for modeling frameworks that can accommodate both the physical interpretability of key vehicle parameters and the nonlinear complexity of their interactions. A particularly important development in this domain is the integration of domain knowledge with machine learning often referred to as physics-informed machine learning (PIML). Rather than treating ML models as black boxes, PIML frameworks incorporate physical laws or domain-specific constraints into the feature engineering or model structure, improving both interpretability and generalization (Latrach et al., 2023; Lin et al., 2025). In the context of EV energy modeling, this approach offers a compelling opportunity to bridge the gap between the theoretical rigor of vehicle dynamics and the flexibility of data-driven methods. However, despite growing interest in hybrid physics–data frameworks, relatively few studies have explicitly integrated physics-based reduced-order representations into a systematic comparison of multiple ML algorithms using publicly available EV specification datasets. Building on this foundation, the present study aims to develop and evaluate a physics-informed machine learning framework for predicting electric vehicle energy consumption using a dataset of approximately 300 EV models sourced from publicly available vehicle specifications (EVspecs.org, 2026). Vehicle-related features including battery capacity, vehicle mass, aerodynamic drag coefficient, driving range, top speed, and power output are used as input variables, while energy consumption (kWh/100 km) serves as the target variable. Five regression algorithms Linear Regression, Random Forest, XGBoost, LightGBM, and Support Vector Regression (SVR) are implemented and compared under a consistent cross-validation framework. In addition, a reduced-order physics-based formulation is introduced to connect the machine learning inputs with classical vehicle dynamics principles, providing a theoretically grounded interpretability layer. The study contributes to the growing body of research on data-driven EV analysis by demonstrating that integrating physical reasoning with machine learning leads to more interpretable and reliable energy consumption predictions, and by providing a systematic comparison of algorithm performance within this physics-informed context.

METHODS

Dataset, Preprocessing, and Study Design

This study is based on a dataset comprising approximately 300 electric vehicle models collected from a publicly available vehicle specification database (EVspecs.org, 2026). The dataset includes a range of technical and performance-related parameters commonly used to characterize EV behavior, namely battery capacity (kWh), vehicle mass (kg), aerodynamic drag coefficient (Cd), driving range (km), top speed (km/h), power output (kW), acceleration performance, and energy consumption expressed in kWh per 100 kilometers. Energy consumption was selected as the target variable, while the remaining specifications served as input features. The selection of these variables is consistent with previous studies that have identified vehicle specifications as the primary determinants of EV energy demand (Di Martino et al., 2022; Pokharel et al., 2021; Ullah et al., 2022). Prior to model development, the dataset was examined for missing values, inconsistencies, and potential outliers. Standard preprocessing procedures were applied to ensure data quality and comparability across models. The study was framed as a supervised machine learning regression problem, in which the goal is to learn a mapping from vehicle specification features to energy consumption values. All analyses were performed using Python, with core libraries including Pandas and NumPy for data management, Scikit-learn for model implementation and evaluation, XGBoost and LightGBM for gradient boosting, and Matplotlib for visualization.

Physics-Informed Feature Framework

To enhance the interpretability of the machine learning models and ensure consistency with fundamental vehicle dynamics principles, a physics-informed feature framework was developed based on the work–energy theorem and Newtonian mechanics. In classical vehicle dynamics, the total energy required to propel a vehicle is expressed as the integral of the total tractive force over displacement: $E = \int F \cdot dx$

where the total force F encompasses four physical components:

$$F = F_{rolling} + F_{aero} + F_{slope} + F_{inertia}.$$

Each component is derived from established physical laws:

rolling resistance, $F_{rolling} = Cr \cdot mg,$

aerodynamic drag, $F_{aero} = ½\rho C_d A v^2$,

road slope force, $F_{slope} = mg \cdot sin\theta$,

and inertial force, $F_{inertia} = ma$.

Substituting these into the energy equation yields the complete physics-based energy model:

$$E = \int (C_r \cdot mg + ½\rho C_d A v^2 + mg \cdot sin\theta + ma) \cdot v\, dt.$$

While this formulation provides a theoretically complete description of vehicle energy demand, several parameters including air density (ρ), frontal area (A), road inclination (θ), and instantaneous acceleration (a) are not available in standardized vehicle specification datasets. To address this limitation, a reduced-order representation is adopted by projecting the full physical model onto the observable feature space available in the dataset. Retaining the most physically influential and measurable variables vehicle mass (m), aerodynamic drag coefficient (C_d), and battery capacity (B) the simplified physics-inspired relationship takes the form:

$$E \approx f(m, C_d, m/B).$$

Where the interaction term m/B captures the relationship between vehicle load and available energy storage capacity. This reduced-order formulation was incorporated into the Linear Regression model as a physics-informed feature, allowing the model to retain physical interpretability while operating within a data-driven framework. The resulting regression equation takes the form:

$$E = \beta_0 + \beta_1 m + \beta_2 Cd + \beta_3 (\frac{m}{B}).$$

This approach bridges classical vehicle dynamics with statistical learning, enabling the machine learning models to learn patterns that remain physically consistent with the underlying energy consumption behavior.

Machine Learning Models and Evaluation

Five regression algorithms were implemented and compared to predict electric vehicle energy consumption: Linear Regression, Random Forest, XGBoost, LightGBM, and Support Vector Regression (SVR). These models were selected to represent a broad spectrum of algorithmic approaches from simple parametric methods to advanced ensemble and kernel-based techniques enabling a comprehensive assessment of predictive performance across different modeling paradigms. Linear Regression was included as an interpretable baseline model that estimates energy consumption through a linear combination of input features (James et al., 2021). Its coefficient structure directly reflects the contribution of each vehicle parameter, making it particularly well-suited for physics-informed interpretation. Random Forest is an ensemble method that constructs multiple decision trees on random subsets of the data and averages their outputs to reduce variance and improve generalization (Breiman, 2001). XGBoost employs a sequential gradient boosting strategy in which successive trees are trained to correct the residual errors of the previous ensemble, enabling the model to capture complex nonlinear dependencies with high efficiency (Chen & Guestrin, 2016). LightGBM follows a similar boosting framework but adopts a leaf-wise tree growth strategy, offering computational efficiency on structured datasets (Ke et al., 2017). Support Vector Regression maps the input features into a high-dimensional space via kernel functions and identifies an optimal regression hyperplane within a specified error tolerance, providing an alternative nonlinear modeling approach (Smola & Schölkopf, 2004). All five models were trained on the same input features under identical conditions to ensure a fair and consistent comparison. Hyperparameter tuning was applied where appropriate to optimize individual model performance. Model generalization was assessed using k-fold cross-validation (k = 5), which partitions the dataset into five equally sized subsets, trains the model on four folds, and evaluates it on the remaining fold iteratively. This procedure ensures that performance estimates are robust and not dependent on a single data split. Three complementary evaluation metrics were used to quantify predictive accuracy: the coefficient of determination (R^2), which measures the proportion of variance in energy consumption explained by the model; Root Mean Square Error (RMSE), which penalizes larger prediction errors more heavily; and Mean Absolute Error (MAE), which provides a straightforward measure of average prediction deviation. Together, these metrics offer a comprehensive and balanced assessment of model performance, consistent with established practice in EV energy prediction research (Hussain et al., 2025; Ullah et al., 2022).

RESULTS

Model Performance Comparison

The predictive performance of the five machine learning models evaluated in this study is summarized in Table 1. Results are reported based on 5-fold cross-validation to ensure robust and unbiased performance estimates across all algorithms.

Table 1. Performance metrics of the evaluated machine learning models

Model	R^2	RMSE (kWh/100km)	MAE (kWh/100km)	MSE
Linear Regression	0.837	1.6134	1.2152	2.6030
Random Forest	0.841	1.5926	1.2535	2.5365
XGBoost	0.820	1.6980	1.1835	2.8834
LightGBM	0.638	2.4065	1.7855	5.7914
SVR	0.615	2.4816	1.6390	6.1584

The results reveal clear and meaningful differences in predictive performance across the evaluated models. Among all algorithms, Random Forest achieved the highest overall performance with an R^2 of 0.841, indicating that approximately 84.1% of the variance in electric vehicle energy consumption is explained by the selected vehicle characteristics. This was closely followed by Linear Regression ($R^2 = 0.837$) and XGBoost ($R^2 = 0.820$), both of which demonstrated strong predictive capability. The proximity of these three models in terms of R^2 suggests that the dataset contains strong and consistent signal that can be captured by both linear and nonlinear approaches, provided the modeling framework is appropriately designed. In terms of error metrics, Random Forest also achieved the lowest RMSE (1.5926 kWh/100 km) and MSE (2.5365), reflecting its stability and reduced sensitivity to larger prediction deviations. XGBoost produced the lowest MAE (1.1835 kWh/100 km) among all models, indicating that its median prediction error is particularly small despite slightly higher variance in extreme cases.

In contrast, LightGBM and SVR exhibited substantially weaker performance, with R^2 values of 0.638 and 0.615 respectively, and considerably higher error values across all metrics. The relatively poor performance of LightGBM is noteworthy given its algorithmic similarity to XGBoost, and may be attributed to its leaf-wise tree growth strategy, which can be more sensitive to dataset size and feature distribution characteristics. SVR, while theoretically capable of capturing nonlinear relationships through kernel transformations, appears to be limited in this context by its sensitivity to hyperparameter settings and its reduced ability to generalize across the full range of vehicle specifications present in the dataset. Overall, the results indicate that ensemble tree-based methods — particularly Random Forest and XGBoost — provide the most reliable and accurate framework for predicting electric vehicle energy consumption from vehicle specification data, consistent with findings reported in recent comparative studies in this domain (Hussain et al., 2025; Ullah et al., 2022).

Physics-Informed Interpretation and Feature Importance

The empirical Linear Regression equation obtained in this study provides a valuable opportunity to validate the consistency of the data-driven results with fundamental vehicle dynamics principles. The fitted model takes the

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

following form: $Energy = 3.1087 + 0.0072 \cdot Weight - 0.0194 \cdot Range + 17.8930 \cdot C_d + 0.0128 \cdot TopSpeed$

where Weight represents vehicle mass (kg), Range denotes the WLTP driving range (km), Cd is the aerodynamic drag coefficient, and TopSpeed refers to the maximum vehicle speed (km/h). This empirical formulation can be directly mapped onto the physics-informed reduced-order representation introduced in the Methods section:

$$E = \beta_0 + \beta_1 m + \beta_2 Cd + \beta_3 (\frac{m}{B})$$

The direction and relative magnitude of the regression coefficients are physically consistent with classical vehicle dynamics theory. Vehicle mass carries a positive coefficient, reflecting its role in both rolling resistance and inertial force, both of which increase energy demand as vehicle weight grows. The aerodynamic drag coefficient exhibits the largest positive coefficient (17.893), underscoring the dominant influence of aerodynamic resistance on energy consumption, particularly at higher driving speeds — a finding that aligns with the quadratic velocity dependence of aerodynamic drag established in the physical model. Top speed also carries a positive coefficient, consistent with the well-established relationship between driving velocity and aerodynamic energy losses. In contrast, driving range is associated with a negative coefficient, which is physically interpretable as an efficiency proxy — vehicles designed to achieve greater range tend to be more energy-efficient by design, incorporating optimized powertrains, lightweight construction, and improved aerodynamics.Feature importance analysis conducted for the Random Forest and XGBoost models further corroborates these physical interpretations. Across both ensemble models, battery capacity, vehicle mass, and driving range consistently ranked as the most influential predictors of energy consumption, followed by the aerodynamic drag coefficient and top speed. This ranking is consistent with the reduced-order physical framework developed in this study, where mass, drag coefficient, and the mass-to-battery-capacity ratio are identified as the primary determinants of energy demand. The convergence between data-driven feature importance rankings and physics-based variable prioritization provides additional confidence in the validity and physical consistency of the modeling framework.Figure 1 presents the actual versus predicted energy consumption values generated by the SVR model, illustrating the general alignment between model predictions and observed values across the dataset.

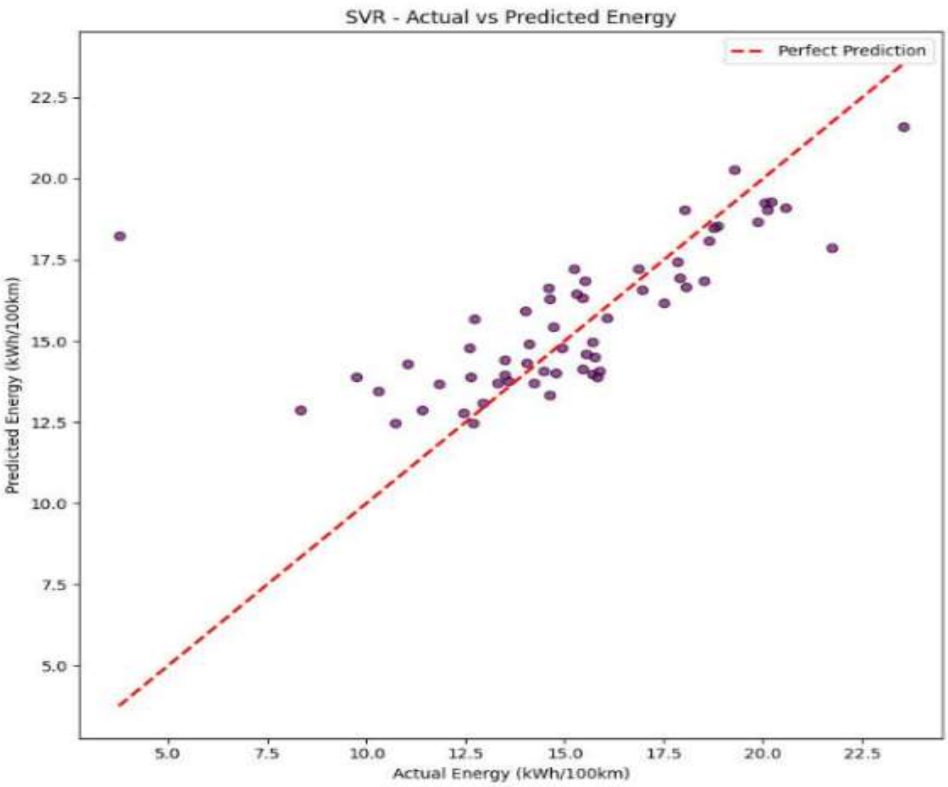

Figure 1. Actual vs. predicted energy consumption values (SVR model)

While the SVR model demonstrated lower overall accuracy compared to ensemble methods, the scatter plot reveals that predictions remain reasonably well-distributed around the diagonal reference line, with larger deviations concentrated at higher energy consumption values. This pattern suggests that the model captures the central tendency of energy consumption behavior but is less effective in representing vehicles with extreme specifications, where nonlinear physical interactions are most pronounced.

DISCUSSION

The results of this study confirm that machine learning algorithms can effectively predict electric vehicle energy consumption from standardized vehicle specification data, with ensemble-based methods achieving the strongest overall performance. The superiority of Random Forest ($R^2 = 0.841$) and XGBoost ($R^2 = 0.820$) over linear and kernel-based approaches reflects the inherently nonlinear and interaction-dependent nature of EV energy consumption, where vehicle mass, aerodynamic drag, battery capacity, and driving range contribute simultaneously and interdependently to total energy demand. Ensemble methods are particularly well-suited to this problem because they learn complex feature interactions without imposing any predefined functional form, while their aggregation mechanisms reduce overfitting and improve generalization to unseen vehicle configurations (Breiman, 2001; Chen & Guestrin, 2016). The comparatively weaker performance of LightGBM ($R^2 = 0.638$) and SVR ($R^2 = 0.615$) can be attributed to the sensitivity of leaf-wise boosting to modest dataset sizes and the strong dependence of SVR on kernel selection and regularization tuning, both of which limit their generalization capability in this context (Ke et al., 2017; Smola & Schölkopf, 2004). The close performance of Linear Regression ($R^2 = 0.837$) relative to the ensemble methods is particularly noteworthy and reflects the effectiveness of the physics-informed feature engineering applied in this study. By incorporating a reduced-order physical representation derived from the work–energy theorem specifically the mass-to-battery-capacity ratio (m/B) the model was equipped with a feature structure that captures physically meaningful interactions beyond simple additive effects, consistent with the growing literature on physics-informed machine learning (Latrach et al., 2023; Lin et al., 2025; Karniadakis et al., 2021). The empirical regression coefficients further confirm this physical consistency: vehicle mass and aerodynamic drag coefficient carry the expected positive relationships with energy consumption, the dominant coefficient of the drag term (17.893) underscores the critical role of aerodynamic resistance (Weiss et al., 2020; Di Martino et al., 2022), and the negative coefficient of driving range reflects the efficiency characteristics of long-range vehicles. Feature importance rankings from Random Forest and XGBoost corroborate these interpretations, with battery capacity, vehicle mass, and driving range consistently emerging as the most influential predictors — a finding directly supported by recent studies in this domain (Pokharel et al., 2021; Zhang et al., 2024). The convergence between data-driven importance rankings and physics-based variable prioritization strengthens confidence in the validity of the proposed framework, and aligns with the broader argument that interpretable and physically consistent models are essential for trustworthy deployment in engineering applications (Rudin, 2019; Arrieta et al., 2020; Lundberg & Lee, 2017). The findings are broadly consistent with recent literature. Hussain et al. (2025) and Ullah et al. (2022) both reported that ensemble tree-based methods consistently outperform linear and kernel-based approaches for EV energy prediction, while Pokharel et al. (2021) demonstrated that vehicle mass and battery-related parameters are the dominant predictors across multiple modeling frameworks. Zhang et al. (2024) further emphasized that feature selection and model interpretability are critical for real-world deployment, reinforcing the value of the physics-informed approach adopted here. From a sustainability perspective, the identification of vehicle mass and aerodynamic drag as primary energy drivers provides actionable guidance for vehicle designers, supporting the use of lightweight materials and aerodynamic optimization as engineering strategies for improving EV efficiency (Mayyas et al., 2017; Hofer et al., 2014). At the infrastructure level, accurate specification-based prediction models can support charging network planning, grid integration, and energy policy formulation without requiring real-time driving data (IEA, 2024; World Bank, 2023). Several limitations should be acknowledged. The dataset does not include real-world driving variables such as road gradient, ambient temperature, or driver behavior, which are known sources of energy consumption variability (Hao et al., 2021; Di Martino et al., 2022). The modest dataset size of approximately 300 vehicles may also constrain the generalizability of the results. Future work could address these limitations by incorporating real-world driving datasets, expanding vehicle diversity, and exploring physics-informed neural networks (Karniadakis et al., 2021; Cai et al., 2022) or SHAP-based explainability frameworks (Arrieta et al., 2020; Lundberg & Lee, 2017) to further enhance model transparency and predictive robustness.

CONCLUSION

This study presented a physics-informed machine learning framework for predicting electric vehicle energy consumption using standardized vehicle specification data. By integrating a reduced-order physical model derived from the work–energy theorem with five regression algorithms, the study demonstrated that combining domain knowledge with data-driven modeling leads to more interpretable and reliable energy consumption predictions than purely statistical approaches alone. Among the evaluated algorithms, Random Forest achieved the highest predictive performance ($R^2 = 0.841$, RMSE = 1.5926 kWh/100 km), followed closely by Linear Regression ($R^2 = 0.837$) and XGBoost ($R^2 = 0.820$). The strong performance of the physics-informed Linear Regression model is particularly significant, demonstrating that embedding physically meaningful features specifically the mass-to-battery-capacity ratio derived from classical vehicle dynamics can substantially enhance both predictive capability and interpretability. Feature importance analysis from ensemble models consistently identified battery capacity, vehicle mass, and driving range as the most influential predictors, in direct alignment with the reduced-order physical framework and with findings reported in recent literature (Pokharel et al., 2021; Hussain et al., 2025). The comparatively weaker performance of LightGBM and SVR further highlights the importance of model selection when working with moderately sized, physics-structured datasets. From a practical standpoint, the proposed framework is highly scalable and accessible, enabling accurate energy consumption estimation from publicly available vehicle specifications without requiring real-world driving data. This makes it directly applicable to vehicle design optimization, charging infrastructure planning, fleet energy management, and transport policy formulation (IEA, 2024; OECD, 2023). Future work should focus on incorporating real-world driving variables, expanding dataset diversity, and exploring physics-informed neural networks and SHAP-based explainability tools to further improve predictive robustness and model transparency (Karniadakis et al., 2021; Arrieta et al., 2020; Lundberg & Lee, 2017).

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

REFERENCES

1. Alamin, K. S. S., Chen, Y., Macii, E., Poncino, M., & Vinco, S. (2022). A machine learning-based digital twin for electric vehicle battery modeling. arXiv preprint. https://arxiv.org/abs/2206.08080
2. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
3. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD, 785–794. https://doi.org/10.1145/2939672.2939785
4. Chen, Y., Sun, R., & Wu, X. (2021). Estimating bounds of aerodynamic, mass, and auxiliary load impacts on autonomous vehicles: A powertrain simulation approach. Sustainability, 13(22), 12405. https://doi.org/10.3390/su132212405
5. Chen, Y., Wu, G., Sun, R., Dubey, A., Laszka, A., & Pugliese, P. (2021). A Review and Outlook on Energy Consumption Estimation Models for Electric Vehicles. SAE International Journal of Sustainable Transportation, Energy, Environment & Policy, 2(1), 79–96. https://doi.org/10.4271/13-02-01-0005
6. Dataset source: EVspecs.org (2026), Most aerodynamic electric cars database. https://www.evspecs.org/most-aerodynamic-electric-cars
7. Di Martino, A., Miraftabzadeh, S. M., & Longo, M. (2022). Strategies for the modelisation of electric vehicle energy consumption: A review. Energies, 15(21), 8115. https://doi.org/10.3390/en15218115
8. Fattahi, A., Sijm, J., & Faaij, A. (2020). A systemic approach to analyze integrated energy system modeling tools: A review of national models. Renewable and Sustainable Energy Reviews, 133, 110195. https://doi.org/10.1016/j.rser.2020.110195
9. Guzzella, L., & Sciarretta, A. (2013). Vehicle propulsion systems: Introduction to modeling and optimization (3rd ed.). Springer. https://doi.org/10.1007/978-3-642-35913-2

10. Hao, X., Wang, H., Lin, Z., & Ouyang, M. (2021). Seasonal effects on electric vehicle energy consumption and driving range: A case study on personal, taxi, and ridesharing vehicles. Journal of Cleaner Production, 278, 123415. https://doi.org/10.1016/j.jclepro.2019.119403
11. Hussain, I., Ching, K. B., Uttraphan, C., Tay, K. G., & Noor, A. (2025). Evaluating machine learning algorithms for energy consumption prediction in electric vehicles: A comparative study. Scientific Reports, 15, 16124. https://doi.org/10.1038/s41598-025-94946-7
12. Intergovernmental Panel on Climate Change. (2022). Climate change 2022: Mitigation of climate change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. https://doi.org/10.1017/9781009157926
13. International Energy Agency. (2024). Global EV outlook 2024: Moving towards increased affordability. IEA. https://www.iea.org/reports/global-ev-outlook-2024
14. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1
15. Jiang, Z., Palacios, A., Zou, B., Zhao, Y., Deng, W., Zhang, X., & Ding, Y. (2022). A review on the fabrication methods for structurally stabilised composite phase change materials and their impacts on the properties of materials. Renewable and Sustainable Energy Reviews, 159, 112134. https://doi.org/10.1016/j.rser.2022.112134
16. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154. https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
17. Larminie, J., & Lowry, J. (2003). Electric vehicle technology explained. John Wiley & Sons. https://doi.org/10.1002/0470090707
18. Latrach, A., Malki, M. L., Morales, M., Mehana, M., & Rabiei, M. (2023). A critical review of physics-informed machine learning applications in subsurface energy systems. https://arxiv.org/abs/2308.04457
19. Lin, Y., Tang, J., Guo, J., Wu, S., & Li, Z. (2025). Advancing AI-enabled techniques in energy system modeling: A review of data-driven, mechanism-driven, and hybrid modeling approaches. Energies, 18(4), 845. https://doi.org/10.3390/en18040845
20. Moawad, A., Gurumurthy, K. M., Verbas, O., et al. (2021). A deep learning approach for macroscopic energy consumption prediction with microscopic quality for electric vehicles. arXiv preprint. https://arxiv.org/abs/2111.12861
21. Organisation for Economic Co-operation and Development. (2023). ITF Transport Outlook 2023. OECD Publishing. https://doi.org/10.1787/b6cc9ad5-en
22. Pokharel, S., Sah, P., & Ganta, D. (2021). Improved Prediction of Total Energy Consumption and Feature Analysis in Electric Vehicles Using Machine Learning and Shapley Additive Explanations Method. World Electric Vehicle Journal, 12(3), 94. https://doi.org/10.3390/wevj12030094
23. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88
24. Ullah, I., Liu, K., Yamamoto, T., Al Mamlook, R. E., & Jamal, A. (2022). A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability. Energy & Environment, 33(8), 1583–1612. https://doi.org/10.1177/0958305X211044998
25. Weiss, M., Cloos, K. C., & Helmers, E. (2020). Energy efficiency trade-offs in small to large electric vehicles. Environmental Sciences Europe, 32(1), 46. https://doi.org/10.1186/s12302-020-00307-8
26. World Bank. (2023). The economics of e-mobility for passenger transportation. World Bank. https://www.worldbank.org/en/topic/transport/publication/the-economics-of-e-mobility-for-passenger-transportation
27. Zhang, X., Zhang, Z., Liu, Y., Xu, Z., & Qu, X. (2024). A review of machine learning approaches for electric vehicle energy consumption modelling in urban transportation. Renewable Energy, 234, 121243. https://doi.org/10.1016/j.renene.2024.121243