






# Data Driven Energy Economy Prediction of Electric Buses Using Machine Learning

Pooja GN \*<sup>1</sup>, Hanamant R. Jakaraddi <sup>†2</sup>, and Aditya U Diwan <sup>‡3</sup>

<sup>1</sup>Department Of MCA, Acharya Institute of Technology, Bangalore

<sup>2</sup>Assistant Professor, Department of MCA, Acharya Institute of Technology, Bangalore

<sup>3</sup>Assistant Professor, Acharya Institute of Graduate Studies, Bangalore

## Abstract

Electrification of transportation systems is increasing, in particular city buses raise enormous potential. Deep understanding of real-world driving data is essential for vehicle design and fleet operation. Various technological aspects must be considered to run alternative powertrains efficiently. Uncertainty about energy demand results in conservative design which implies inefficiency and high costs. Both, industry, and academia miss analytical solutions to solve this problem due to complexity and interrelation of parameters. Precise energy demand prediction enables significant cost reduction by optimized operations. This paper aims at increased transparency of battery electric buses' (BEB) energy economy. We introduce novel sets of explanatory variables to characterize speed profiles, which we utilize in powerful machine learning methods. We develop and comprehensively assess 5 different algorithms regarding prediction accuracy, robustness, and overall applicability. Achieving a prediction accuracy of more than 94%, our models performed excellent in combination with the sophisticated selection of features. The presented methodology bears enormous potential for manufacturers, fleet operators and communities to transform mobility and thus pave the

\*Email: [poojagn2002@gmail.com](mailto:poojagn2002@gmail.com) Corresponding Author

<sup>†</sup>Email: [hanamant2504@acharya.ac.in](mailto:hanamant2504@acharya.ac.in)

<sup>‡</sup>Email: [aditya2431@acharya.ac.in](mailto:aditya2431@acharya.ac.in)

way for sustainable, public transportation.

Keywords: BEB. GHG. Machine Learning. Support Vector Machines. Random Forest. Artificial Neural Networks.

## 1 Introduction

Traffic causes approximately 25% of greenhouse gas (GHG) emissions in Europe, and this percentage is increasing.(European Commission Directorate-General for Mobility and Transport, 2018). Therefore, widespread electrification of the mobility sector is one of the most positive actions that can be taken in relation to climate change and sustainability.(Gallo & Marinelli, 2020). It seems clear that electric buses, because of their low pollutant emissions, are set to play a key role in the public urban transportation of the future. Although the initial investment in electrification may be high - e.g. purchase costs of BEBs are up to twice as high as those of Diesel buses - it is quickly amortized because the inherent efficiency of electric vehicles far exceeds that of internal combustion engine vehicles (up to 77% ) and thus operational respectively life cycle costs are significantly lower. (Lajunen & Lipman, 2016).

In addition, electrification of the power train brings many other advantages, such as a reduced noise level or pollution. On the downside, the battery charging time of an electric bus is significantly longer than the refueling time of a diesel bus, while the opposite is true for the range. Ultimately, widespread electrification of the mobility sector is one of the most positive actions that can be taken in terms of climate change and sustainability, but more research is needed to ensure efficient operation, as it also poses significant challenges. The starting point for this study was a problem proposed by Seville's public bus operator. In short, they wanted to replace their diesel fleet with all-electric vehicles, but first they had to size the vehicles' batteries and determine the best charging locations around the city. In practice, this means using computers to predict consumption on each route. In order to improve the provision of services to citizens, emerging technologies such as artificial intelligence (AI) are strategically integrated.(Mittal & Gautam, 2023).

Unfortunately, this can currently only be done with complex physical models that require long simulation times, or with data-driven models that are less computationally intensive once trained, but require numerous driving, mechanical, and road measurements as inputs. This is where the present research comes in. In this paper we use the bus operator's database and a physics-based model of soon-to-be- deployed electric buses to develop data- driven models that predict the energy requirements of the vehicles. Amongst others, what distinguishes our contribution from previous data driven approaches is the small number of physical variables involved: we show that, to accurately predict the consumption on a route using machine learning, we only need to know the instantaneous

speed of the vehicle and the number of passengers on the bus. Specifically, our approach consists of three steps:

1. We calculate the energy consumed by the bus on each route using a physics-based model, validated by the vehicle manufacturer, that uses speed and mass as inputs, including the bus's weight and the weight of its payload. Both variables are taken from the operator's database.
2. We extract a comprehensive set of time and frequency features from the speed signal.
3. We train machine learning regression models to predict the energy consumption from bus payload mass and the above set of features and identify those with the best predictive value. Interestingly, the feature that turns out to be the most relevant, i.e., the spectral entropy of velocity, has so far gone unnoticed in this field of research.

## 2 Literature Survey

Zogaan's (2022) study focuses on predicting the energy consumption of electric buses using various machine learning algorithms. The authors compare the performance of multiple techniques, including Support Vector Machines (SVM), Random Forest (RF), and Artificial Neural Networks (ANN). Historical data on bus routes, passenger load, and environmental conditions are utilized to train the models. The results indicate that ANNs outperform other models in terms of accuracy and robustness. The study demonstrates the potential of machine learning in optimizing energy usage and improving the operational efficiency of electric buses, ultimately contributing to more sustainable urban transportation systems. In the paper Pamula and Pamula's (2022) propose a deep learning-based approach to forecast the energy consumption of electric urban buses. Leveraging Long Short-Term Memory (LSTM) networks, the model incorporates time-series data, including traffic conditions, weather patterns, and bus operation schedules. The LSTM model captures temporal dependencies and provides high-accuracy predictions compared to traditional methods. The study highlights the effectiveness of deep learning in addressing the complexities of energy consumption forecasting and suggests potential improvements in route planning and battery management for electric buses.

A data-driven predictive modeling framework for enhancing the energy efficiency of electric public transit buses. Using a combination of regression analysis and machine learning algorithms, the study assesses the impact of various factors such as speed, route topology, and passenger count on energy consumption. (Sennefelder, Martín-Clemente, & González-Carvajal, 2023). The Gradient Boosting Machine (GBM) model emerged as the most effective, providing detailed insights into the primary drivers of energy usage. The findings underscore the importance of data analytics in the operational planning of elec-

tric buses, aiming to reduce energy costs and environmental impact.(Ullah et al., 2022). Xiong, Cao, and Yu's (2018) paper presents machine learning approaches for real-time energy management in electric urban buses. By integrating real-time data on vehicle dynamics and external conditions, the proposed system uses Reinforcement Learning (RL) to optimize energy consumption dynamically. The RL model adapts to changing conditions, learning optimal driving strategies to minimize energy use. Simulation results demonstrate significant improvements in energy efficiency, validating the potential of machine learning for proactive and adaptive energy management in electric bus fleets.(Recalde et al., 2024).

The authors SRIDHAR and TARUN's (2024) explore predictive analytics for optimizing the energy consumption of electric bus fleets. They employ machine learning models, including Decision Trees and Ensemble Methods, to forecast energy needs based on historical and contextual data. The study by Ziliaskopoulos and Waller's (2000) integrates Geographic Information Systems (GIS) to enhance model accuracy by incorporating spatial data related to routes and traffic conditions. The predictive models enable better scheduling and charging strategies, reducing operational costs and improving the sustainability of urban bus systems. Ushakov et al.'s (2022) in his research highlights the role of advanced analytics in supporting smart transportation infrastructure and efficient energy use.Rapid innovation, digital capabilities, and IT skill integration has become a national norm that is critical to the general advancement of economies and communities. (Mittal, 2020).

### 3 Related Work

The prediction of energy demand for battery electric vehicles (BEVs) in general, and battery electric buses (BEBs) in particular, have been thoroughly investigated. BEBs are a viable replacement for conventional vehicles and are also less sensitive to variations in mission profiles than diesel buses. It is important to note also that the duty cycle and driving conditions of a BEB are very different from those of other BEVs, shifting the focus from kinematic relationships to route, schedule, and passenger load.(Sennefelder, Martín-Clemente, & González-Carvajal, 2023). The majority of previous studies utilize complex physics-based vehicle models, varying in focus and objective.(De Cauwer, Van Mierlo, & Coosemans, 2015; Marc, Tobias, & Thomas, 2018; Nijmeijer, Wang, & Besselink, 2017). For example, a study byAsamer et al.'s (2016) examines the impact of power train efficiency, rolling resistance, and auxiliary power on the energy consumption of battery electric vehicles (BEVs).A Sparse matrix with a significant number of zeros can better represent a transshipment model is proposed by Garg and Mittal's (2021).

While drive train efficiency and rolling resistance are relevant to the physical movement of the vehicles, auxiliary power demand is especially important at lower speeds where city buses typically operate, highlighting the need for accurate knowledge of auxiliary

power to predict overall energy consumption. Another study De Cauwer, Van Mierlo, and Coosemans's (2015) integrates a physical model of the vehicle with a data-driven methodology to detect and quantify correlations between kinematic parameters and the vehicle's energy consumption. Commonly used kinematic parameters are complemented by additional factors such as travel distance, time, and temperature.

Research by Nijmeijer, Wang, and Besselink's (2017) also studied the influence of rolling resistance, which depends on the road surface and various weather conditions, on power demand. One prediction model consists of a longitudinal dynamics model complemented by additional measurements from a dynamometer and coast-down tests to reduce the model's uncertainty. Another study introduces a novel and computationally efficient electro-mechanical model of a BEB to examine the influence of factors such as payload mass, temperature, and rolling resistance on consumption. These approaches provide valuable insights into the interrelation of influential factors; however, they involve intricate equations and require accurate modeling of the vehicles and their components to generate results. As with all physics-based models, they are of limited practical use due to long simulation times. Additionally, most previous research has focused primarily on light-duty vehicles, and scaling to the heavy-duty class is complex due to completely different driving profiles and dynamics.

Data-driven approaches, which use machine learning or deep learning algorithms and real-world driving data, or even mixed data-driven and physics-based approaches, have also been explored. (Aljohani, Ebrahim, & Mohammed, 2021; Chen et al., 2021; Li et al., 2021). For example, Pamuła and Pamuła's (2022) review covers state-of-the-art energy-consumption estimation models for electric vehicles and studies the case of electric buses using logistic regression and neural networks on real-world data. The study by Kontou and Miles's (2015) identifies a research gap for energy consumption models of heavy-duty vehicles such as city buses, supporting the motivation of our work. Another research used deep learning and classical neural networks to forecast the energy demand of electric buses using actual data from various bus lines. These models are based on input variables that fleet operators can easily measure, but also include operational information such as bus routes, stop locations, travel time between bus stops, schedules, and peak hour information. Other research investigates factors of influence such as the route and driver characteristics.

Some studies such as Ericsson's (2001) examined the effects of different driving patterns collected in real traffic on consumption and emissions of internal combustion vehicles. Starting with many features, a factorial analysis reduces this number significantly, demonstrating the influence of common kinematic driving pattern parameters, such as speed, acceleration, and deceleration, on energy consumption. They also evaluated the usefulness of feature analysis and selection. Simonis and Sennefelder's (2019) accurately

describes the behavior of drivers as a function of selected characteristics, which can be used to predict the energy demand of BEVs. Interestingly, some studies used a Simulink model to estimate the energy consumption of BEBs, with inputs carefully selected from a mix of operational, topological, vehicular, and external variables using machine learning algorithms and statistical models. They found that battery state of charge and road gradient were the most significant factors, while the vehicle's drag coefficients had a relatively minimal effect.

However, temperature and auxiliary power demand are not well covered, despite being crucial factors. One study investigates real-world data from a fleet of BEBs in Mihoko City, China, finding that ambient temperature significantly impacts energy consumption. Another recent study in Lancaster, California, examines BEBs' energy consumption and charging behavior, showing that temperature variability leads to increased energy use due to heating, cooling, venting, and air conditioning (HVAC). Results indicate relevant operational costs for the operator, which can increase significantly during summer. However, this cost analysis might differ in other situations, as cost assessment of BEBs is generally a vast field influenced by various factors. (Perugu et al., 2023).

Göhlich, Kunith, and Ly's (2014) studied a technology assessment for BEBs in Berlin, Germany, using an energy simulation model to forecast daily service consumption and analyze the system's economics in terms of total cost of ownership (TCO). Using a thermal model of the cabin, they find that heating by Positive Temperature Coefficient (PTC) elements is generally more critical than cooling, discovering a worst-case additional HVAC consumption that is substantial compared to the overall energy consumption.

1. Most approaches use data that standard vehicles are often not equipped to measure, such as the location of bus stops or road gradient. In addition, variables that are highly dependent on the particular conditions of the experiment are frequently taken into account, such as the length of the trip. The relationship of the latter with vehicle energy economy is obvious – e.g., the further you drive the more energy is consumed. However, it must be used with caution for prediction, as machine learning algorithms may focus on it and overlook other relevant factors. By contrast, our algorithms take as initial input only the mass (estimated from the curb weight plus number of passengers) and the vehicle speed, which can be easily obtained by the user. Furthermore, we characterize speed profiles by extracting 40 features at different levels of abstraction in the frequency and time domains. This way, we uncover hidden and valuable information that leads to higher prediction accuracy, improved generalization, and thus high application relevance. In addition, we implement an intelligent route segmentation algorithm that makes the prediction robust to data non-stationarity, making the final framework more transferable and even more applicable.

2. Despite the abundance of machine-learning techniques, only a few of them are commonly used. In this work, we consider the full range, from non-learning statistical approaches to supervised learning and probabilistic methods. Consequently, this work presents and comprehensively compares the full potential of novel machine learning methods for predicting the energy consumption of EVs. Ultimately, we investigate the performance of various powerful machine learning models, from the very technical detail to the long-term application.
3. Most studies use data from a single vehicle on a single route or use speed profiles from Standardized Driving Cycles (SDCs). Therefore, the variety and diversity within the data is comparatively low. However, a major challenge in this area is that the relevant factors are diverse and the interrelationships are complex. Thus, the larger the variety in the data, the better the machine learning predictions will be. In contrast, the underlying fleet data for this work is measured from an entire fleet of 30 vehicles, which operate various routes a day and drivers change frequently even during the day. This allows us to capture a wide variety of traffic situations and driving styles, containing much more valuable information.
4. Auxiliary power demand, including HVAC, is rarely considered in detail and often replaced by a constant term. However, we have considered complete energy profiles, including HVAC, recovery, etc., which allows this work to address accurate total energy consumption at the trip level, which is relevant to transit operators.

## 4 System Architecture

In this paper we use the bus operator's database and a physics-based model of soon to be deployed electric buses to develop data-driven models that predict the energy requirements of the vehicles. (see figure 1). Amongst others, what distinguishes our contribution from previous data driven approaches is the small number of physical variables involved: we show that, to accurately predict the consumption on a route using machine learning, we only need to know the instantaneous speed of the vehicle and the number of passengers on the bus. Specifically, our approach consists of We calculate the energy consumed by the bus on each route using a physics- based model, validated by the vehicle manufacturer, that uses speed and mass as inputs, including the bus's own weight and the weight of its payload. Both variables are taken from the operator's database. We extract a comprehensive set of time and frequency features from the speed signal. We train machine learning regression models to predict the energy consumption from bus payload mass and the above set of features, and identify those with the best predictive value. Interestingly, the feature that turns out to be the most relevant, i.e., the spectral entropy of velocity, has so far gone unnoticed in this field of reasearch. We propose a scalable and efficient hybridization

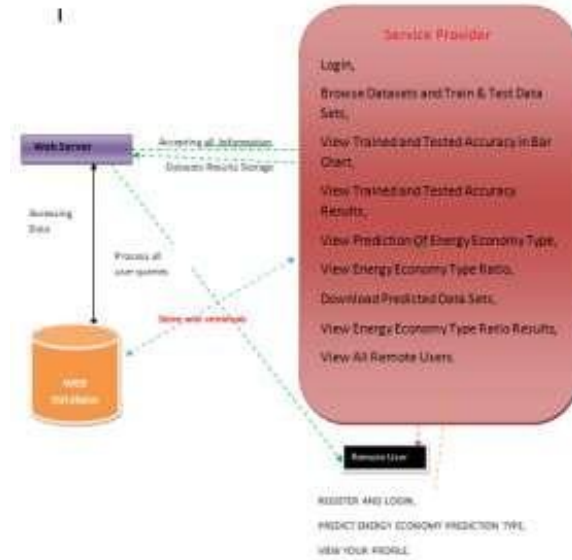


Figure 1. System Architecture

Machine Learning models for exact predictions. We conducted several hybridizations of genetic algorithm with filter and embedded feature selection methods, in the data pre-processing phase of Random Forest and Multivariate Linear Regression (MLR) predictive model, with the aim of improving its performance.

## 5 Module

### 5.1 Description Service Provider

The Service Provider must enter a valid user name and password to log in to this module. He can perform some tasks after logging in successfully, such browsing datasets and training and testing data sets. View the results of the trained and tested accuracy, view the bar chart showing the accuracy, view the prediction of the energy economy type, view the ratio of the energy economy type, and download the predicted data.

### 5.2 View and Authorize User

The administrator can see a list of all enrolled users in this module. This allows the administrator to examine user information such name, email address, and address, as well as the ability to authorise people. View All Remote Users, View Sets, and View Energy



Economy Type Ratio Results.

### 5.3 Remote Operator

There are n numbers of users present in this module. Prior to beginning any operations, the user must register. The user's information is saved in the database after they register. Upon successful registration, he must use his authorized user name and password to log in. After successfully logging in, the user will perform certain tasks, such as REGISTER AND LOGIN. ESTIMATE THE ENERGY ECONOMY.

## 6 Results

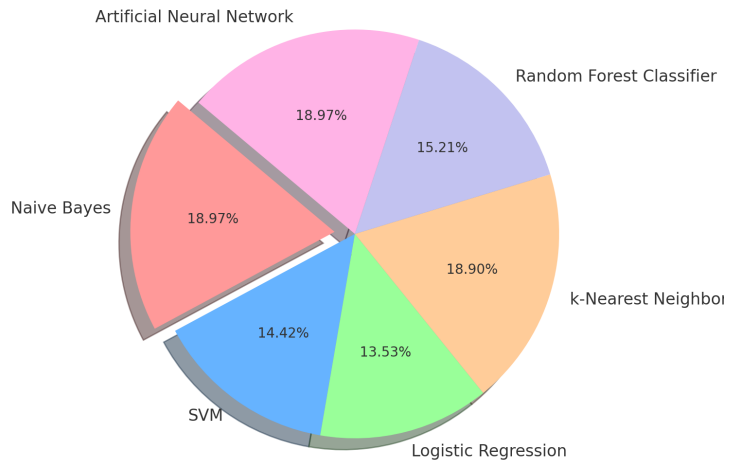


Figure 2. Chart

This paper offers a data-driven approach that uses both simulated and real-world data for planning problems and electrification of public transport. The results confirm that the energetic relevant features obtained by feature selection and regression analysis perfectly characterize the energy consumption of BEBs under different real driving conditions. (see figure 2). It is a practical approach for fleet operators who want to retrofit or replace their conventional buses with electric vehicles and build the corresponding infrastructure.

## 7 Conclusion

In order to improve robustness against non-stationarity, our article proposes a novel set of explanatory factors that integrate temporal and frequency features of the speed waveform. We found the strongest predictive variables by breaking down routes into micro trips; spectral entropy of velocity profiles was found to be significant. Subsequent investigations will expand this approach to diverse situations, primarily helping logistics and transportation firms, especially fleet owners of heavy-duty trucks. Other vehicles or transport systems could also be able to use the methodology. In order to provide reliable predictive analytics for a range of transportation situations, future research will concentrate on operational, road, and climatic aspects with the goal of predicting factors such as peak power and battery demands.

## References

- Aljohani, T. M., Ebrahim, A., & Mohammed, O. (2021). Real-Time metadata-driven routing optimization for electric vehicle energy consumption minimization using deep reinforcement learning and Markov chain model. *Electric Power Systems Research*, 192. <https://doi.org/10.1016/j.epsr.2020.106962>
- Asamer, J., Graser, A., Heilmann, B., & Ruthmair, M. (2016). Sensitivity analysis for energy demand estimation of electric vehicles. *Transportation Research Part D: Transport and Environment*, 46, 182–199. <https://doi.org/10.1016/j.trd.2016.03.017>
- 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). <https://doi.org/10.4271/13-02-01-0005>
- De Cauwer, C., Van Mierlo, J., & Coosemans, T. (2015). Energy consumption prediction for electric vehicles based on real-world data. *Energies*, 8(8), 8573–8593. <https://doi.org/10.3390/en8088573>
- Ericsson, E. (2001). Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transportation Research Part D: Transport and Environment*, 6(5), 325–345. [https://doi.org/10.1016/S1361-9209\(01\)00003-7](https://doi.org/10.1016/S1361-9209(01)00003-7)
- European Commission Directorate-General for Mobility and Transport. (2018). EU transport in figures. <https://doi.org/10.2832/729667>
- Gallo, M., & Marinelli, M. (2020). Sustainable mobility: A review of possible actions and policies. *Sustainability (Switzerland)*, 12(18). <https://doi.org/10.3390/su12187499>

- Garg, T. K., & Mittal, P. (2021). Logistics networks: a sparse matrix application for solving the transshipment problem. *Journal of Mathematical and Computational Science*. <https://doi.org/10.28919/jmcs/6654>
- Göhlich, D., Kunith, A., & Ly, T. (2014). Technology assessment of an electric urban bus system for berlin. *WIT Transactions on the Built Environment*, 138, 137–149. <https://doi.org/10.2495/UT140121>
- Kontou, A., & Miles, J. (2015). Electric Buses: Lessons to be Learnt from the Milton Keynes Demonstration Project. *Procedia Engineering*, 118, 1137–1144. <https://doi.org/10.1016/j.proeng.2015.08.455>
- Lajunen, A., & Lipman, T. (2016). Lifecycle cost assessment and carbon dioxide emissions of diesel, natural gas, hybrid electric, fuel cell hybrid and electric transit buses. *Proceedings of the ICE -Energy*, 106, 329–342. <https://doi.org/10.1016/j.energy.2016.03.075>
- Li, P., Zhang, Y., Zhang, Y., & Zhang, K. (2021). Prediction of electric bus energy consumption with stochastic speed profile generation modelling and data driven method based on real-world big data. *Applied Energy*, 298. <https://doi.org/10.1016/j.apenergy.2021.117204>
- Marc, G., Tobias, M., & Thomas, H. (2018). Estimation of the energy demand of electric buses based on real-world data for large-scale public transport networks. *Applied Energy*, 230, 344–356.
- Mittal, P. (2020). Impact of Digital Capabilities and Technology Skills on Effectiveness of Government in Public Services. *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020*. <https://doi.org/10.1109/ICDABI51230.2020.9325647>
- Mittal, P., & Gautam, S. (2023). Logistic Regression and Predictive Analysis in Public Services of AI Strategies. *TEM Journal*, 12(2), 751–756. <https://doi.org/10.18421/TEM122-19>
- Nijmeijer, H., Wang, J., & Besselink, I. (2017). Battery electric vehicle energy consumption modelling for range estimation. *International Journal of Electric and Hybrid Vehicles*, 9(2), 79. <https://doi.org/10.1504/ijehv.2017.10006163>
- Pamuła, T., & Pamuła, D. (2022). Prediction of Electric Buses Energy Consumption from Trip Parameters Using Deep Learning. *Energies*, 15(5). <https://doi.org/10.3390/en15051747>
- Perugu, H., Collier, S., Tan, Y., Yoon, S., & Herner, J. (2023). Characterization of battery electric transit bus energy consumption by temporal and speed variation. *Energy*, 263. <https://doi.org/10.1016/j.energy.2022.125914>
- Recalde, A., Cajo, R., Velasquez, W., & Alvarez-Alvarado, M. S. (2024). Machine Learning and Optimization in Energy Management Systems for Plug-In Hybrid Electric

- Vehicles: A Comprehensive Review. *Energies*, 17(13). <https://doi.org/http://dx.doi.org/10.3390/en17133059>
- Sennefelder, R. M., Martín-Clemente, R., & González-Carvajal, R. (2023). Energy Consumption Prediction of Electric City Buses Using Multiple Linear Regression. *Energies*, 16(11). <https://doi.org/10.3390/en16114365>
- Simonis, C., & Sennefelder, R. (2019). Route specific driver characterization for data-based range prediction of battery electric vehicles. 2019 14th International Conference on Ecological Vehicles and Renewable Energies, EVER 2019. <https://doi.org/10.1109/EVER.2019.8813579>
- SRIDHAR, S., & TARUN, G. (2024). Machine Learning-Based Data-Driven Energy Economy Forecasting for Electric City Buses. *International Journal For Advanced Research in Science and Technology*, 14(5).
- 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 and Environment*, 33(8), 1583–1612. <https://doi.org/10.1177/0958305X211044998>
- Ushakov, D., Dudukalov, E., Mironenko, E., & Shatila, K. (2022). Big data analytics in smart cities' transportation infrastructure modernization. *Transportation Research Procedia*, 63, 2385–2391. <https://doi.org/10.1016/j.trpro.2022.06.274>
- Xiong, R., Cao, J., & Yu, Q. (2018). Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle. *Applied Energy*, 211, 538–548. <https://doi.org/10.1016/j.apenergy.2017.11.072>
- Ziliaskopoulos, A. K., & Waller, S. T. (2000). An Internet-based geographic information system that integrates data, models and users for transportation applications. *Transportation Research Part C: Emerging Technologies*, 8(1-6), 427–444. [https://doi.org/10.1016/S0968-090X\(00\)00027-9](https://doi.org/10.1016/S0968-090X(00)00027-9)
- Zogaan, W. A. (2022). Power Consumption prediction using Random Forest model. *International Journal of Mechanical Engineering*, 7(5), 974–5823.