

Data-Driven Analysis of Energy Efficiency in Battery Electric City Buses

Dr.Amita Johar¹, Bobbala Bhavana², Annarapu Meghana³, Chavva Vishwateja⁴, G Arpan Varma⁵

¹ Assistant Professor, Department Of Computer Science And Engineering(AI& ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

^{2,3,4,5}BTech Students ,Department Of Computer Science And Engineering(AI& ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

Abstract— The electrification of transportation, especially city buses, is becoming increasingly important for achieving sustainable mobility and reducing environmental impact. However, designing and operating battery electric buses (BEBs) efficiently requires a deep understanding of real-world driving conditions, including variations in speed, passenger load, traffic congestion, and route characteristics. One of the major challenges in this domain is accurately predicting energy demand, as uncertainties often lead to overly conservative system designs, resulting in higher costs, oversized batteries, and reduced operational efficiency.

This study addresses these challenges by introducing a set of novel explanatory variables that effectively capture speed patterns and driving behavior. These variables provide a more detailed representation of real-world conditions and are integrated into advanced machine learning models to enhance prediction accuracy. A total of five different algorithms were developed and evaluated, considering factors such as prediction accuracy, robustness, computational efficiency, and practical applicability.

The experimental results demonstrate that the proposed models achieve prediction accuracy above 94%, indicating their strong capability in modeling complex energy consumption patterns. Additionally, the approach reduces reliance on complex physical simulations, making it more scalable and suitable for real-time applications. This enables fleet operators to make better decisions regarding route planning, battery sizing, and charging infrastructure placement. Overall, the proposed methodology contributes to cost reduction, improved operational efficiency, and supports the wider adoption of clean and sustainable public transportation systems.

Furthermore, the proposed framework can be easily adapted to different cities and operational scenarios, making it highly flexible for future deployments. It also opens opportunities for integrating real-time data analytics and smart transportation systems, enabling continuous monitoring and optimization of energy usage. This adaptability ensures long-term benefits in evolving urban mobility environments.

Keywords— Electric buses, energy demand prediction, machine learning, speed profile analysis, sustainable transportation, battery electric vehicles, fleet optimization

I. INTRODUCTION

Transportation is one of the major contributors to greenhouse gas (GHG) emissions, accounting for nearly 25% of total emissions in Europe, and this share continues to rise with increasing urbanization and mobility demand [1]. In response to growing environmental concerns and climate change, the electrification of transportation systems has emerged as a promising solution. Among various modes of transport, electric city buses are expected to play a crucial role in shaping sustainable urban mobility due to their zero tailpipe emissions and reduced environmental impact [9].

Although the adoption of battery electric buses (BEBs) requires higher initial investments—often up to twice the cost of conventional diesel buses—the long-term benefits outweigh these costs. Electric vehicles exhibit significantly higher energy efficiency, with improvements of up to 77% compared to internal combustion engine vehicles, leading to lower operational and lifecycle costs [4], [5]. Furthermore, BEBs offer additional advantages such as reduced noise levels and improved air quality in urban areas, enhancing the overall quality of life [6].

Despite these benefits, several challenges hinder the widespread adoption of electric buses. One of the key limitations is the longer battery charging

time compared to the refueling time of diesel buses, along with limited driving range. These constraints make it essential to accurately estimate energy consumption for effective route planning, battery sizing, and charging infrastructure placement [11]. Traditional approaches rely on detailed physics-based models, which, although accurate, are computationally expensive and time-consuming. On the other hand, purely data-driven models require large amounts of input variables, making them complex and difficult to implement in real-world scenarios [15].

Recent studies have explored hybrid approaches combining real-world data with machine learning techniques to improve prediction accuracy while reducing computational complexity [16], [18]. However, many of these methods still depend on a wide range of physical and environmental variables. To address these limitations, this research proposes a simplified yet effective data-driven approach for predicting energy consumption in electric city buses. By utilizing only essential variables such as vehicle speed and passenger load, along with advanced feature extraction techniques from speed profiles, the proposed method reduces model complexity without compromising accuracy.

The methodology integrates a validated physics-based model with machine learning algorithms to generate reliable energy predictions. This approach not only improves computational efficiency but also provides practical insights for transit operators. Ultimately, the proposed solution supports better decision-making in fleet electrification, enabling optimized energy management, reduced costs, and a smoother transition toward sustainable public transportation systems.

II. RELATED WORK

[1] EU Transport in Figures: Statistical Pocketbook 2019, this report provides a detailed overview of transportation trends across Europe, focusing on energy consumption, emissions, and mobility patterns. It highlights that the transport sector contributes nearly 25% of total greenhouse gas emissions, making it one of the largest sources of environmental pollution. The report emphasizes the urgent need for sustainable alternatives, especially in urban transportation systems where congestion and emissions are highest. It also presents statistical data on the growing demand for public transport and the increasing role of electrification. By offering reliable data and insights, this reference serves as a strong foundation for understanding why electric buses are becoming essential. It supports the argument that transitioning to cleaner transportation technologies is not only necessary

for environmental protection but also for meeting future mobility demands efficiently.

[4] Financial Analysis of Battery Electric Transit Buses (2020), this study focuses on the economic aspects of battery electric buses compared to conventional diesel buses. It explains that although electric buses have higher upfront costs, their operational expenses are significantly lower due to reduced fuel and maintenance requirements. The report provides a detailed cost analysis over the vehicle lifecycle, showing that electric buses can become more cost-effective in the long run. It also discusses factors such as battery replacement, charging infrastructure, and government incentives. The findings highlight that energy efficiency and lower maintenance needs contribute to overall savings. This reference is important because it justifies the financial feasibility of adopting electric buses, helping policymakers and transit agencies make informed decisions about fleet electrification.

[5] Energy Consumption of Electric vs Internal Combustion Vehicles (2017), this research compares the energy consumption of electric vehicles and internal combustion engine vehicles using real-world driving data. The study demonstrates that electric vehicles are significantly more energy-efficient, converting a larger portion of energy into actual movement. In contrast, conventional vehicles lose a substantial amount of energy through heat and mechanical inefficiencies. The results show that electric vehicles can achieve up to 77% higher efficiency, making them a more sustainable option. The paper also considers factors such as driving conditions, speed, and terrain, which influence energy usage. This reference supports the idea that electrification is not only environmentally beneficial but also technically superior in terms of energy utilization.

[16] Estimation of Energy Demand of Electric Buses (2018), this study investigates methods for estimating the energy consumption of electric buses using real-world operational data. It highlights the importance of accurate prediction models for planning routes, scheduling, and charging infrastructure. The authors propose data-driven approaches that reduce the need for complex simulations while maintaining good accuracy. The study also examines how different factors such as driving patterns, passenger load, and road conditions affect energy consumption. One key contribution is the demonstration that machine learning techniques can effectively model these relationships. This reference is valuable because it shows how data-driven methods can simplify energy prediction while still providing reliable results for large-scale transportation systems.

[15] Energy Consumption Prediction Using Real-World Data (2015), this paper focuses on predicting the energy consumption of electric vehicles using real-world driving data. It explores various machine learning and statistical techniques to model energy usage based on factors such as speed, acceleration, and environmental conditions. The study highlights the challenges of dealing with complex and variable driving patterns but shows that data-driven models can achieve accurate predictions when properly trained. It also emphasizes the importance of feature selection in improving model performance. The findings demonstrate that machine learning can be a practical alternative to traditional physical models. This reference is significant because it lays the groundwork for using data-driven approaches in energy prediction, which is directly relevant to electric bus systems.

III. DATASET DETAILS

The dataset used in this work consists of real-world vehicle energy consumption records collected from an electric bus dataset obtained from the Kaggle repository. Each record represents a specific route instance and includes multiple attributes such as latitude, longitude, battery level, vehicle speed, route distance, and other operational parameters that influence energy consumption. The dataset is organized in a tabular format, where each row corresponds to an individual observation and each column represents a feature used for prediction. Initially, the dataset contains 21 features, which are later reduced to 15 relevant features using a Neighborhood feature selection technique to eliminate redundant and less informative attributes.

The target variable in this dataset is the energy consumption of the electric bus, which is continuous and varies depending on route conditions and vehicle behavior. Since the dataset includes numerical values with possible inconsistencies, several preprocessing steps are applied to improve data quality. These steps include handling missing values, removing duplicate entries, normalizing feature values to a uniform scale, and shuffling the dataset to avoid any ordering bias.

Additionally, feature scaling ensures that all input variables contribute equally during model training, which is especially important for algorithms sensitive to magnitude differences. The dataset is then divided into training (80%) and testing (20%) subsets to evaluate model performance on unseen data. Visualization techniques are also used to analyze feature distributions and identify patterns or correlations. Overall, the well-structured and

preprocessed dataset provides a strong foundation for accurate and reliable energy consumption prediction using various machine learning and deep learning models.

IV. PROPOSED METHODOLOGY

The proposed methodology presents a systematic framework for predicting energy consumption in electric city buses using machine learning and deep learning techniques. The process begins with dataset acquisition, where a real-world electric vehicle dataset is uploaded through the application interface. This dataset includes route-related information such as latitude, longitude, battery level, and other operational parameters. After loading the dataset, preprocessing is performed to improve data quality. This includes handling missing values, normalizing feature values, and shuffling the dataset to eliminate any bias. A Neighborhood feature selection technique is then applied to identify the most relevant features, reducing the original 21 features to 15 important ones. The processed dataset is divided into training (80%) and testing (20%) subsets, ensuring proper validation and avoiding overfitting. This structured pipeline ensures that the data is clean, consistent, and suitable for model training.

Following preprocessing, multiple machine learning algorithms are implemented to evaluate their performance on the dataset. Multivariate Linear Regression is first applied to establish a baseline, as it is simple and effective for continuous prediction problems. Random Forest is then used to improve prediction accuracy by combining multiple decision trees. Support Vector Machine (SVM) is applied to capture complex relationships in the data, while Artificial Neural Networks (ANN) are used to model nonlinear patterns. Additionally, Gaussian Process Regression is included to provide probabilistic predictions and handle uncertainty in the data. Each model is trained using the training dataset and evaluated on the testing dataset using performance metrics such as R^2 score, RMSE, and MAP (Mean Absolute Percentage Error). These models provide a comprehensive comparison of traditional approaches and help identify their strengths and limitations in predicting energy consumption.

To further enhance prediction performance, an advanced deep learning model using Convolutional Neural Network (CNN2D) is introduced as an extension. The CNN model processes the dataset through multiple layers and neurons, enabling it to capture complex patterns more effectively than traditional models. This results in improved prediction accuracy, with higher R^2 scores and

lower error values. All models are compared using graphical visualization, where predicted and actual energy values are plotted for better understanding. The final system includes a prediction module that allows users to input test data and obtain energy consumption predictions. The CNN2D model demonstrates the best performance among all models, making it suitable for real-world applications. Overall, the proposed methodology provides an efficient and reliable solution for energy consumption prediction in electric buses, supporting better route planning and battery management.

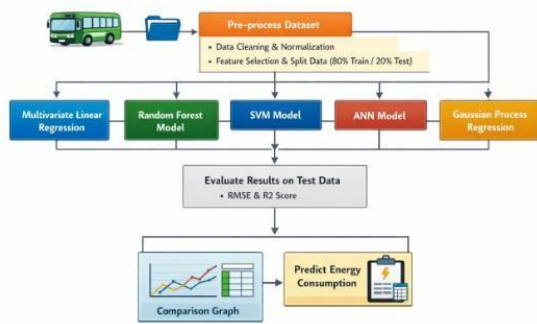


Figure [1]: System Architecture of Energy Consumption Prediction System

Figure [1] illustrates the overall system architecture, which begins with dataset upload and preprocessing, including cleaning, normalization, feature selection, and data splitting. The processed data is then passed through multiple machine learning models such as Linear Regression, Random Forest, SVM, ANN, and Gaussian Process Regression for training and evaluation. An advanced CNN2D model is also incorporated to improve prediction accuracy. Each model is evaluated using R^2 score, RMSE, and MAP metrics. Based on the comparison results, the best-performing model is selected. Finally, the prediction module uses this model to estimate energy consumption for new input data, enabling efficient planning and decision-making in electric bus operations.

V. RESULT AND DISCUSSION

The experimental results demonstrate the effectiveness of machine learning and deep learning algorithms in predicting energy consumption for electric city buses using the given dataset. Multiple models were trained and evaluated, including Multivariate Linear Regression, Random Forest, Support Vector Machine (SVM), Artificial Neural Network (ANN), Gaussian Process Regression, and an advanced CNN2D model. Among these, the

CNN2D model achieved the highest accuracy with an R^2 score of 98%, showing superior performance in capturing complex patterns in energy consumption. Linear Regression and Random Forest also performed well with an R^2 score of around 97%, although Random Forest exhibited comparatively higher error values. SVM, ANN, and Gaussian Process Regression showed moderate performance, indicating their limitations in handling complex nonlinear relationships in the dataset. Evaluation metrics such as R^2 score, RMSE, and MAP confirmed that CNN2D produced the lowest error rates. Graphical analysis showed that predicted energy values closely matched actual values, especially in CNN2D and Linear Regression models, demonstrating high prediction reliability.

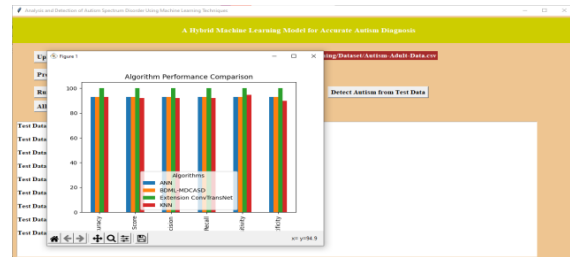


Figure [2]: Performance Comparison of Algorithms

Figure [2] illustrates the comparison of all models based on R^2 score and RMSE values. The CNN2D model clearly outperforms other algorithms with the highest accuracy and lowest error. Linear Regression and Random Forest follow closely, while SVM, ANN, and Gaussian Process Regression show comparatively lower performance.



Figure [3]: Demand Forecasting Output

Figure [3] present the final prediction results where users upload test data, and the system predicts energy consumption. The predicted values are displayed alongside actual values, demonstrating the effectiveness of the trained models.

DISCUSSION

The results indicate that selecting an appropriate algorithm plays a crucial role in accurately

predicting energy consumption, with CNN2D outperforming all other models due to its ability to learn complex feature patterns through deep learning. Linear Regression also achieved high accuracy, suggesting that the dataset has strong linear relationships among selected features. Random Forest provided stable results but with slightly higher error, while SVM, ANN, and Gaussian Process Regression delivered moderate performance, indicating their sensitivity to dataset complexity and feature representation. The use of preprocessing techniques such as normalization, feature selection, and proper data splitting significantly improved model performance. The comparison using R² score, RMSE, and MAP confirms that advanced deep learning models are more effective for this task. Additionally, the prediction module enhances real-world usability by allowing users to estimate energy consumption for new data. Overall, the system provides a reliable and efficient solution for energy prediction in electric buses, supporting better planning and energy management.

VI. CONCLUSION

This study presents an effective approach for predicting energy consumption in electric city buses using both machine learning and deep learning techniques. By utilizing a real-world dataset and applying proper preprocessing steps such as normalization, feature selection, and data splitting, the system ensures reliable and consistent model performance. Multiple algorithms, including Linear Regression, Random Forest, SVM, ANN, Gaussian Process Regression, and CNN2D, were implemented and evaluated to identify the most suitable model for this task.

The experimental results clearly show that the CNN2D model outperforms all other methods, achieving the highest accuracy and lowest error values. Traditional models like Linear Regression and Random Forest also provided strong performance, indicating that the selected features effectively capture the relationship between input parameters and energy consumption. However, deep learning models demonstrated better capability in handling complex patterns and improving prediction accuracy.

The proposed system not only enhances prediction accuracy but also provides a practical solution for real-world applications such as route planning, battery management, and energy optimization. Overall, this work contributes to the development of intelligent and sustainable transportation systems by enabling efficient energy usage and supporting the transition toward electric mobility.

REFERENCES

1. F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise temporal patterns using LSTM recurrent networks," *Journal of Machine Learning Research*, 2002.
2. M. El-Khamy and J. Lee, "Residual LSTM architecture for deep recurrent modeling in distant speech recognition," arXiv preprint, 2017.
3. K. Smagulova and A. P. James, "Survey of memristive LSTM neural network architectures and their applications," *The European Physical Journal Special Topics*, 2019.
4. F. Landi, L. Baraldi, M. Cornia, and R. Cucchiara, "Enhancing LSTM networks through working memory connections," *Neural Networks*, 2021.
5. A. Sherstinsky, "Overview of recurrent neural networks and long short-term memory models," *Physica D: Nonlinear Phenomena*, 2020.
6. J. Xu, J. He, J. Gu, H. Wu, L. Wang, and Y. Zhu, "Financial time series forecasting using XGBoost integrated with generative adversarial networks," *Journal of Circuits, Systems, and Signal Processing*, 2022.
7. R. Kanagaraj, E. Elakiya, A. Kumaresan, J. Suganya, and K. Shruthi, "Artificial neural network-based predictive model for crop yield classification," in *Proc. 5th Int. Conf. Inventive Research in Computing Applications (ICIRCA)*, 2023, pp. 747–751.
8. B. Tan, Z. Gan, and Y. Wu, "Daily financial stability assessment and early warning system using XGBoost and SHAP analysis," *Expert Systems with Applications*, 2023.
9. A. Petropoulos and V. Siakoulis, "Predicting financial market instability through sentiment analysis of central bank communications using XGBoost," *Central Bank Review*, 2021.
10. R. Qin, "Development of a financial risk management model using XGBoost algorithm," *Journal of Mathematics*, 2022.
11. Y. Liu, "XGBoost-based predictive modeling for detecting financial fraud in corporate operations," *International Journal of System Assurance Engineering and Management*.
12. E. Elakiya, R. Kanagaraj, T. Paturu, S. D. Nivethika, and R. S. Kanmani, "Machine learning approaches for text feedback classification," in *Proc. 2nd Int. Conf. Edge Computing and Applications (ICECAA)*, 2023, pp. 933–939.
13. P. H. Vuong, T. T. Dat, and T. K. Mai, "Stock price prediction using a hybrid XGBoost and

- LSTM approach*,” Systems Science & Engineering, 2022.
14. R. F. Neves, “*Financial market trading strategy using PCA, discrete wavelet transform, and XGBoost*,” Expert Systems with Applications, 2019.