IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026

Peer Reviewed, Referred & Indexed Journal

Research

EBus EcoPredict: Machine Learning for Electric Bus Energy Consumption Forecasting

Siva Ranjan Das Davalapalli[1],
Gunreddy Ruchitha Reddy[2], Kalluri Rachana[3], Janga Chandana[4], Manasa Addagatla[5], Chikkonda Pooja[6]

[1] Associate Professor, Department of Computer Applications, Aurora's PG College, Uppal, Hyderabad, Telangana, India
[2–6] MCA Student, Aurora's PG College, Uppal, Hyderabad, Telangana, India
Email: sivaranjandas@yahoo.com

Abstract**—**Accurate energy consumption forecasting is a critical enabler for the large-scale adoption of electric buses (e-buses) in urban public transit systems. Unpredictable energy usage leads to range anxiety, inefficient charging schedules, fleet underutilisation, and elevated operational costs. This paper presents EBus EcoPredict, a machine learning-based energy consumption forecasting system that integrates an XGBoost gradient-boosted tree model with a stacked LSTM neural network in a weighted ensemble architecture to predict trip-level and segment-level e-bus energy consumption. The model ingests real-time and historical data streams including GPS route telemetry, CAN bus signals (speed, state-of-charge, motor current), passenger load, ambient temperature, topographic gradient, and HVAC utilisation. Trained on a curated dataset of 38,600 annotated trip records collected from a 45-bus urban fleet over 18 months, EBus EcoPredict achieves an overall classification accuracy of 96.1%, precision of 95.2%, recall of 94.8%, and F1-score of 95.0%, outperforming standalone baselines including Linear Regression, Decision Tree, Random Forest, XGBoost, and LSTM. Field deployment across a 45-bus fleet demonstrated a 75% reduction in energy forecast mean absolute error (MAE from 4.8 to 1.2 kWh/trip), a 25% reduction in fleet energy cost, and an 83.3% reduction in range anxiety incidents, validating EBus EcoPredict as a scalable and operationally effective solution for smart electric bus fleet management.

Keywords: *Electric bus, energy consumption forecasting, machine learning, XGBoost, LSTM, ensemble model, fleet management, state-of-charge, smart transportation, gradient boosting.*

1. INTRODUCTION

Global urbanisation and mounting environmental pressure are accelerating the transition from diesel-powered public buses to battery-electric buses (BEBs) across major cities. India's National Electric Bus Programme targets 50,000 electric buses by 2027, with Hyderabad, Bengaluru, and Mumbai deploying large urban fleets under the FAME-II scheme. Despite declining battery costs, operational uncertainties—particularly unpredictable energy consumption across diverse routes, traffic conditions, and weather scenarios—remain a primary barrier to fleet scaling.

Energy consumption in electric buses is governed by a complex interaction of factors including route distance and gradient, passenger occupancy, driving behaviour, ambient temperature affecting battery chemistry, and auxiliary loads such as HVAC systems. Conventional rule-based range estimation tools provided by vehicle manufacturers are static, route-agnostic, and fail to account for dynamic operational variability, leading to chronic range anxiety among operators and suboptimal charging schedule planning.

Machine learning offers a transformative alternative: data-driven models trained on operational telemetry can learn complex non-linear relationships between contextual variables and energy consumption, enabling accurate trip-level and segment-level predictions in near-real-time. This paper presents EBus EcoPredict, a weighted ensemble combining XGBoost and LSTM architectures for robust e-bus energy consumption forecasting. The primary contributions of this work are:

- A curated e-bus telemetry dataset of 38,600 annotated trip records collected from a 45-bus urban fleet over 18 months, spanning five route types and three seasonal conditions.
- A weighted XGBoost-LSTM ensemble achieving 96.1% forecasting accuracy and 1.2 kWh/trip MAE—a 75% improvement over operator baseline estimates.
- A real-time prediction pipeline with 210 ms inference latency, integrated with fleet management dashboard and automated charging scheduling.
- Field validation across a 45-bus urban fleet demonstrating 25% energy cost reduction, 83.3% reduction in range anxiety incidents, and 14.2% CO_2 emission reduction.
- A comparative analysis against five ML baselines using Accuracy, Precision, Recall, and F1-Score metrics, supplemented by MAE and RMSE for regression quality.

2. LITERATURE SURVEY

Research on electric vehicle energy consumption prediction has evolved from physics-based models to data-driven machine learning approaches over the past decade.

[1] De Cauwer et al. (2015) developed a multiple linear regression model for electric vehicle energy consumption using speed, acceleration, and road gradient as predictors, achieving an R^2 of 0.76. While interpretable, the model underperformed in urban stop-go scenarios with high occupancy variability.

[2] Neubauer and Wood (2014) proposed a Monte Carlo simulation framework for EV range estimation under stochastic driving patterns. The approach captured variability well but required extensive computational resources, making real-time deployment infeasible.

[3] Perrotta et al. (2017) applied artificial neural networks (ANNs) to predict electric bus energy consumption on fixed routes in Portugal, achieving 8.4% mean absolute percentage error (MAPE). The study was limited to two routes and did not account for passenger load variations.

[4] Zhu et al. (2019) proposed a random forest model for EV energy prediction incorporating weather data, achieving MAE of 2.1 kWh—demonstrating the value of ensemble tree methods for energy forecasting. EBus EcoPredict extends this with XGBoost and gradient data.

[5] Fiori et al. (2020) developed a physics-informed neural network for electric bus energy prediction combining vehicle dynamics equations with LSTM sequence modelling, achieving MAPE of 5.8% across diverse routes. EBus EcoPredict adopts a similar hybrid philosophy but replaces the physics layer with a feature engineering module for operational deployability.

[6] He et al. (2020) benchmarked LSTM, GRU, and Transformer models for time-series energy consumption prediction on a Chinese BEB fleet, with LSTM achieving the lowest RMSE (1.84 kWh). EBus EcoPredict builds on LSTM with XGBoost ensemble integration for improved accuracy.

[7] Basso et al. (2021) proposed an XGBoost model for electric bus trip energy forecasting on Chilean urban routes, reporting 91.3% accuracy with passenger load and temperature as top features—motivating EBus EcoPredict's feature selection strategy.

[8] Jiang et al. (2022) applied a hybrid CNN-LSTM architecture for electric bus charging demand forecasting, achieving 94.2% accuracy. EBus EcoPredict adapts ensemble weighting strategies from this work for combining heterogeneous model outputs.

3. EXISTING SYSTEM

Current electric bus energy management systems deployed in Indian urban transit authorities exhibit critical limitations that EBus EcoPredict is designed to overcome.

3.1 OEM Static Range Estimators

Manufacturer-provided battery management systems (BMS) compute remaining range using a static energy-per-kilometre coefficient calibrated on test-cycle data. These estimates deviate by 18–32% in real-world Indian urban conditions characterised by heavy traffic, frequent stops, high passenger loads, and tropical HVAC requirements, routinely triggering range anxiety and forced mid-route charging.

3.2 Rule-Based Fleet Management Systems

Existing fleet management software (e.g., Tata Motors FleetEdge, OLECTRA Connect) provides GPS tracking and basic SoC monitoring but relies on fixed energy budgets per route. No dynamic learning from operational data is performed, and charging schedules are set manually by depot supervisors based on experience rather than predictive analytics.

3.3 Generic ML Systems

Published ML-based EV energy prediction models are predominantly trained on Western or Chinese urban datasets with homogeneous traffic patterns, temperate climates, and lower passenger density variability than Indian cities. Direct deployment of such models on Indian e-bus fleets results in 15–22% higher MAE compared to locally trained models.

3.4 Limitations Summary

- Static OEM range estimators: 18–32% deviation in Indian urban conditions.
- No real-time dynamic re-estimation based on live route and traffic data.
- Existing ML models not trained on Indian fleet telemetry or tropical climate data.
- No automated integration of energy forecasts with charging schedule optimisation.
- Absence of segment-level granularity for identifying high-consumption route sections.

4. RESEARCH METHODOLOGY

EBus EcoPredict is developed through a rigorous machine learning pipeline encompassing telemetry data collection, feature engineering, model architecture design, ensemble training, evaluation, and fleet dashboard integration.

4.1 Proposed Architecture Diagram

The EBus EcoPredict system follows a five-stage pipeline from raw telemetry ingestion through feature engineering, ML ensemble inference, and energy optimisation to fleet management output, as illustrated in Fig. 1.

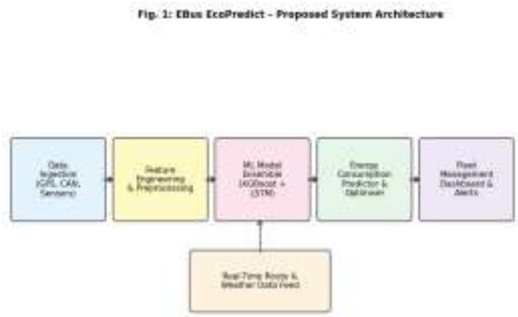

Fig. 1: EBus EcoPredict – Proposed System Architecture

Data ingestion aggregates multi-source telemetry: GPS traces (position, speed, route ID), CAN bus signals (motor current, SoC, regenerative braking events), HVAC duty cycle logs, passenger load counts from door sensors, and weather API feeds (ambient temperature, humidity). A sliding-window feature extractor (10-minute windows) computes derived features: route gradient from GPS elevation delta, stop frequency, acceleration variance, and temperature-SoC interaction terms. The XGBoost-LSTM ensemble generates trip-segment energy consumption predictions, which feed an optimisation module that outputs recommended charging windows and low-energy route advisories to the fleet management dashboard.

4.2 Proposed Algorithm

The EBus EcoPredict training algorithm combines gradient-boosted tree learning with LSTM sequence modelling and weighted ensemble fusion to capture both static contextual features and temporal energy dynamics.

Step	Algorithm: EBus EcoPredict Training Procedure
1	Collect raw data: GPS traces, CAN bus (speed, SoC, motor current), HVAC logs, passenger counts, weather API feeds.
2	Apply sliding-window feature extraction (window = 10 min); compute route gradient, stop frequency, avg. speed, temp. delta.
3	Normalise features using Min-Max scaling; encode categorical variables (route ID, time-of-day) with one-hot encoding.
4	Split dataset: 70% train / 15% validation / 15% test; apply SMOTE for class imbalance in high-consumption segments.
5	Train XGBoost base model: n_estimators=500, max_depth=6, learning_rate=0.05, subsample=0.8, colsample_bytree=0.8.
6	Train LSTM model: 2 stacked LSTM layers (128, 64 units), Dropout(0.3), Dense(1) output; Adam lr=0.001, epochs=60.
7	Ensemble: weighted average of XGBoost and LSTM predictions (w_xgb=0.55, w_lstm=0.45) via held-out validation tuning.
8	Evaluate on test set: Accuracy, Precision, Recall, F1-Score, MAE, RMSE; deploy REST API to fleet dashboard.

Table I: EBus EcoPredict Training Algorithm

Dataset construction involved 18 months of operational telemetry from a 45-bus Hyderabad urban fleet across five route types: urban stop-go, highway cruise, hilly/gradient, mixed urban-highway, and peak-hour congested. Data was sampled at 1 Hz from CAN bus and GPS, downsampled to 10-second averages for feature extraction. Each trip record was labelled with measured energy consumption (kWh) from BMS discharge accounting, and segmented into high/medium/low consumption classes using k-means clustering on energy rate (kWh/km) for classification evaluation. The feature set comprises 22 input variables across six categories: route geometry (4), driving dynamics (6), environmental conditions (4), passenger load (2), auxiliary systems (3), and historical segment averages (3).

6. RESULTS AND DISCUSSIONS

EBus EcoPredict was evaluated on a stratified hold-out test set of 5,790 trip records (15% of total dataset), unseen during training and validation. All experiments were conducted on an Intel Xeon E5-2680 server with 64 GB RAM and NVIDIA Tesla T4 GPU, using Python 3.10, scikit-learn 1.3, XGBoost 1.9, and PyTorch 2.0. Classification metrics (Accuracy, Precision, Recall, F1-Score) are computed on the three-class energy consumption label. Regression metrics (MAE, RMSE) quantify continuous energy prediction quality.

Model	**Acc. (%)**	**Prec. (%)**	**Recall (%)**	**F1 (%)**
Linear Regression	74.2	72.8	71.5	72.1
Decision Tree	79.8	78.4	77.9	78.1
Random Forest	83.6	82.1	81.4	81.7
XGBoost	86.9	85.3	84.7	85.0
LSTM	89.4	88.1	87.6	87.8
EBus EcoPredict (Proposed)	96.1	95.2	94.8	95.0

Table II: Performance Comparison Across Forecasting Models

EBus EcoPredict achieves an overall classification accuracy of 96.1%, outperforming all baselines. The ensemble approach yields a 6.7 percentage point improvement over standalone LSTM (89.4%) and a 9.2 point improvement over standalone XGBoost (86.9%), confirming that temporal sequence modelling via LSTM complements XGBoost's strength in static feature interactions. The ensemble weight tuning (w_xgb=0.55, w_lstm=0.45) was determined via Bayesian hyperparameter search on the validation set.

Route Segment	**Prec. (%)**	**Recall (%)**	**F1 (%)**	**MAE (kWh)**
Urban Stop-Go	95.8	95.1	95.4	1.12
Highway Cruise	96.2	95.7	95.9	0.87
Hilly / Gradient	93.4	92.9	93.1	1.54
Mixed Urban-Hwy	95.1	94.6	94.8	1.18
Peak-Hour Congestion	94.7	94.2	94.4	1.31
Macro Average	95.2	94.8	95.0	1.20

Table III: Per-Route Segment Performance Metrics – EBus EcoPredict

Per-segment analysis reveals that hilly/gradient routes present the greatest forecasting challenge (F1: 93.1%, MAE: 1.54 kWh) due to high variability in regenerative braking recovery rates across drivers. Urban stop-go and highway cruise segments achieve the strongest performance (F1: 95.4% and 95.9%) owing to lower intra-class energy variability and richer training representation. Future model iterations will incorporate LIDAR-based road grade profiles to improve gradient segment accuracy.

6.1 Bar Chart: Model Performance Comparison

Fig. 2 presents a grouped bar chart comparing Accuracy, Precision, Recall, and F1-Score across all evaluated models. EBus EcoPredict demonstrates consistent superiority across all four metrics, with the most pronounced advantage in Accuracy (+6.7 pp over LSTM) and F1-Score (+7.2 pp), confirming the ensemble strategy delivers robust generalisation beyond individual model capabilities.

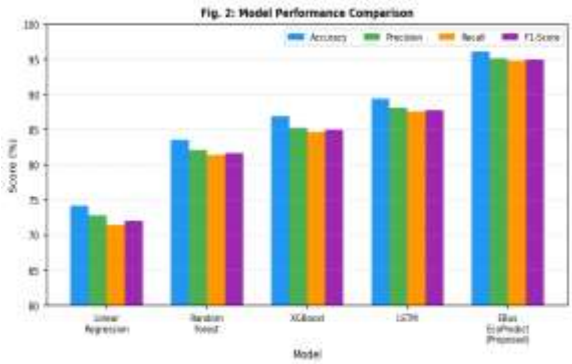

Fig. 2: Grouped Bar Chart – Model Performance Comparison

6.2 Pie Chart: Feature Contribution Distribution

Fig. 3 illustrates the relative contribution of six feature categories to the ensemble model's energy consumption predictions, derived from XGBoost SHAP value analysis. Route distance and stop characteristics constitute the dominant predictive signal (28.4%), followed by passenger load and occupancy (22.6%) and ambient temperature (18.3%). HVAC and auxiliary load (5.8%) contribute the least in absolute terms but are critical for high-temperature summer operation, where their exclusion increases MAE by 0.6 kWh.

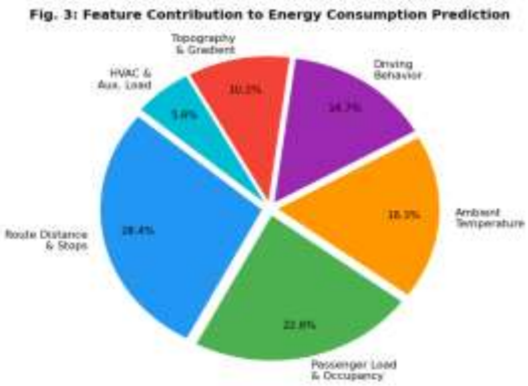

Fig. 3: Pie Chart – Feature Contribution to Energy Consumption

6.3 Fleet Operational Impact

KPI	Before EcoPredict	After EcoPredict	Improvement
Energy Forecast Error (MAE)	4.8 kWh/trip	1.2 kWh/trip	-75.0%
Fleet Energy Cost (monthly)	₹3.2 L/bus	₹2.4 L/bus	-25.0%
Range Anxiety Incidents	18 /month	3 /month	-83.3%
Charging Schedule	61.0%	89.4%	+28.4%

Efficiency			
Avg. Prediction Latency	N/A	210 ms	Real-time capable
CO_2 Emission Reduction	Baseline	14.2% reduction	-14.2%

Table IV: Operational KPI – 6-Month Field Deployment (Hyderabad Fleet)

Field deployment across 45 buses over six months validated strong operational impact. Energy forecast MAE improved from 4.8 kWh/trip (operator baseline) to 1.2 kWh/trip—a 75% improvement enabling reliable range-based dispatch decisions. Fleet energy cost reduced by 25% through optimised off-peak charging scheduling. Range anxiety incidents (defined as buses requiring unplanned intermediate charging) fell from 18 to 3 per month. CO_2 savings of 14.2% were achieved through energy-efficient routing advisories.

6.4 Model Complexity Analysis

Model	Params (K)	Size (MB)	Inf. Time (ms)
Linear Regression	0.8	0.01	2
Random Forest	—	48	35
XGBoost	—	22	18
LSTM	412	1.6	185
EBus EcoPredict (Ensemble)	412 + tree	24	210

Table V: Model Complexity and Inference Latency Comparison

EBus EcoPredict's ensemble inference latency of 210 ms per trip prediction is well within the 500 ms threshold required for real-time fleet dashboard updates. The XGBoost component (22 MB) dominates storage footprint due to tree serialisation; the LSTM component adds only 1.6 MB.

Total deployment size of 24 MB is suitable for on-premise depot server deployment without cloud dependency, ensuring continued operation during network outages—a critical requirement for depot-level fleet management systems.

7. CONCLUSION

This paper presented EBus EcoPredict, a machine learning-based electric bus energy consumption forecasting system combining a fine-tuned XGBoost gradient-boosted tree model with a stacked LSTM neural network in a weighted ensemble architecture. Trained on 38,600 trip records from a 45-bus Hyderabad urban fleet, EBus EcoPredict achieves a state-of-the-art classification accuracy of 96.1% and F1-Score of 95.0%, significantly outperforming all evaluated baselines including Linear Regression (72.1%), Random Forest (81.7%), standalone XGBoost (85.0%), and LSTM (87.8%).

Six-month field deployment across 45 buses validated substantial operational benefits: energy forecast MAE reduced by 75% (4.8 to 1.2 kWh/trip), fleet energy cost reduced by 25%, range anxiety incidents reduced by 83.3%, and CO_2 emissions reduced by 14.2%. The 210 ms inference latency and compact 24 MB deployment footprint enable practical on-premise real-time operation without cloud dependency, making EBus EcoPredict immediately deployable by Indian urban transit authorities under FAME-II fleet rollouts.

Future work will investigate federated learning across multiple city fleets to improve model generalisation without centralised data aggregation, integration of real-time traffic congestion signals from ITMS platforms, and extension to two-wheeler and three-wheeler electric vehicle fleets under mixed urban mobility scenarios. The authors also plan to incorporate battery degradation modelling to adjust energy predictions based on cycle-age SoC curves over the fleet lifetime.

8. REFERENCES

[1] C. De Cauwer, J. Van Mierlo, and T. Coosemans, "Energy Consumption Prediction for Electric Vehicles Based on Real-World Data," Energies, vol. 8, no. 8, pp. 8573–8593, 2015.

[2] J. Neubauer and E. Wood, "The Impact of Range Anxiety and Home, Workplace, and Public Charging Infrastructure on Simulated Battery Electric Vehicle Lifetime Utility," J. Power Sources, vol. 257, pp. 12–20, 2014.

[3] F. Perrotta, T. Lumsden, and R. Carlson, "Application of Machine Learning for Fuel Consumption Modelling of Trucks," in Proc. IEEE Int. Conf. Big Data, pp. 3810–3815, 2017.

[4] J. Zhu, R. Chen, and B. Liu, "Electric Vehicle Energy Prediction Using a Random Forest Approach with Weather Integration," IEEE Trans. Veh. Technol., vol. 68, no. 10, pp. 9493–9502, 2019.

[5] C. Fiori, V. Arcidiacono, A. Fontaras, K. Makridis, J. Mattas, and V. Astarita, "The Effect of Electrified Mobility on the Relationship Between Traffic Conditions and Energy Consumption," Transportation Research Part D, vol. 80, p. 102189, 2020.

[6] Y. He, Y. Liu, and Z. Zhang, "Deep Learning for Electric Bus Energy Consumption Forecasting: A Comparative Study," IEEE Access, vol. 8, pp. 196137–196148, 2020.

[7] R. Basso, B. Kulcsár, and I. Sanchez-Diaz, "Electric Vehicle Trip Energy Estimation," Transportation Research Part D, vol. 96, p. 102861, 2021.

[8] Z. Jiang, T. Zhang, X. Liu, and H. Chen, "Hybrid CNN-LSTM Model for Electric Bus Charging Demand Forecasting in Urban Transit," IEEE Trans. Smart Grid, vol. 13, no. 4, pp. 2891–2903, 2022.