Energy Engineering

Tech Science Press

https://doi.org/10.32604/ee.2026.084024 Published Online: 16 June 2026

ARTICLE

Explainable Machine Learning for Electric Bus Energy Prediction under Tropical Urban Conditions: A Physics-Informed Parametric Framework

Mohammed Almajed[1], **Mahbub Hassan**[2], **Turjoy Das Turjo**[3], **Md Ashequl Islam**[4,*], **Md Ehtesamul Haque**[1] and **M. M. Hafizur Rahman**[5]

[1]Department of Computer Science, College of Computer Sciences and Information Technology (CCSIT), King Faisal University, Al-Ahsa, Saudi Arabia

[2]Department of Civil Engineering, Faculty of Engineering, Chulalongkorn University, Pathumwan, Bangkok, Thailand

[3]Department of Civil Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh

[4]School of Electrical and Mechanical Engineering, College of Engineering and Information Technology, Adelaide University, Adelaide, Australia

[5]Department of Computer Networks & Communications, College of Computer Sciences and Information Technology (CCSIT), King Faisal University, Al-Ahsa, Saudi Arabia

*Corresponding Author: Md Ashequl Islam. Email: mdashequl.islam@adelaide.edu.au

Received: 15 April 2026; Accepted: 20 May 2026

ABSTRACT: Accurate per-kilometer energy consumption prediction is essential for effective electric bus (EB) fleet management in urban transit networks. Existing studies report mean absolute percentage errors of 5%–8% and rarely cross-validate explainability attributions or quantify prediction uncertainty. This study presents a physics-informed, interpretable machine learning framework for EB energy consumption prediction under Bangkok-like tropical urban conditions. Due to the institutional inaccessibility of operational telemetry, a parametric dataset of 4000 trip-level scenarios was constructed across 29 input features anchored to published primary sources. Four algorithms were benchmarked under Bayesian hyperparameter optimization: Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Random Forest, and Support Vector Regression. XGBoost achieved the best performance (R^2 = 0.9439, MAPE = 3.51%), improving on the 5%–8% range of comparable prior studies within the defined parametric scenario space. SHapley Additive exPlanations (SHAP) identified average speed, HVAC load, and road gradient as the dominant energy drivers. Cross-method validation against Local Interpretable Model-Agnostic Explanations (LIME) provided near-threshold support for the dominant attribution hierarchy ($\bar{\rho}$ = 0.686), with both methods agreeing on the top-three predictors. An ablation study across four feature groups supported the non-redundant predictive contribution of speed, congestion, thermal, and infrastructure variables within the parametric scenario space. Quantile regression-based prediction intervals achieved 79.4% empirical coverage, indicating useful but under-calibrated exploratory uncertainty estimates. A companion XGBoost classifier assigned trips to ordinal efficiency categories with 88.25% accuracy and macro-F1 of 0.8423, with no misclassifications across extreme class boundaries. These findings provide a transparent, quantitatively grounded framework that may support future fleet energy management, route optimization, and driver training after validation on live fleet data, while advancing interpretable EB energy modeling for tropical urban contexts.

KEYWORDS: Electric bus; energy consumption prediction; energy efficiency modeling; explainable machine learning; sustainable transit

1 Introduction

The decarbonization of urban public transport has become a major policy priority. Road-based bus services contribute substantially to urban greenhouse gas (GHG) emissions, and many cities are accelerating electric bus (EB) deployment as a mitigation strategy [1–4]. Thailand follows this transition. Bangkok has introduced electric buses into its Bus Rapid Transit and Electric Vehicle (BRT/EV) network, while national targets call for a substantially electrified public transit fleet by 2030 [5]. Although research on EB charging, scheduling, and infrastructure planning has advanced, accurate per-kilometer energy prediction under complex urban conditions remains challenging [6–8].

EB energy consumption depends on many interacting factors, including speed, congestion, stop frequency, passenger load, road gradient, ambient temperature, HVAC demand, regenerative braking, and driver behavior [9–15]. Inaccurate energy estimation can lead to scheduling mismatches, premature battery depletion, unplanned charging, and reduced service reliability [16–18]. Reliable energy prediction is therefore important for fleet planning, charging management, and operational efficiency.

Physics-based drivetrain models can represent traction demand using first-principle equations. However, they often require detailed vehicle parameters that are not available in routine transit operations. They may also generalize poorly across different routes, fleets, and traffic conditions [19,20]. Machine learning provides a complementary approach by learning nonlinear relationships from operational or physics-informed data. Gradient boosting methods such as LightGBM [21–23] have shown strong performance in structured transport energy prediction because they can capture nonlinear effects and feature interactions [24,25].

Prediction accuracy alone is not sufficient for transit applications. Fleet operators and planners also need interpretable explanations that show why energy consumption increases or decreases under specific conditions. Explainable artificial intelligence (XAI) methods address this need. SHapley Additive exPlanations (SHAP) provide consistent global and local feature attributions [26], while Local Interpretable Model-Agnostic Explanations (LIME) can be used as an independent local explanation method [27]. However, SHAP-based EB energy studies rarely validate attributions using another XAI method. Most studies also do not report prediction intervals or test the independent value of feature groups through ablation analysis [28,29].

Bangkok is a distinctive case for EB energy modeling. The city has a tropical savanna climate, with high ambient temperatures and high relative humidity for much of the year [30]. These conditions create near-continuous HVAC demand, which separates Bangkok from many temperate-climate settings studied in previous EB research [31]. Warm-climate studies show that HVAC systems can account for 15% to 30% of EB energy use [31], and Thai evidence from Chiang Mai confirms that high temperature increases HVAC-related energy penalties [32]. Bangkok also has severe congestion, dense stop and intersection patterns, and multilevel road infrastructure such as bridges, flyovers, and elevated sections [33]. These conditions create strong interactions among speed, thermal load, gradient, and stop-and-go operation.

High-resolution trip-level telemetry from Bangkok transit authorities is not available for this study. Therefore, this paper develops a physics-informed parametric dataset of 4000 trip-level scenarios that represent Bangkok-like EB operating conditions. All variable ranges are anchored to published sources, including Thai Meteorological Department climate records, BMTA route information, Bangkok traffic data, and international EB performance studies. This approach is consistent with prior EB energy studies that use simulation or parametric data when live telemetry is limited [15,29,34]. Accordingly, the results should be interpreted as evidence from a calibrated parametric modeling framework, not as direct validation on live Bangkok fleet telemetry.

This study addresses four main gaps. First, tropical Southeast Asian EB operations remain underrepresented in explainable machine learning research. Second, SHAP attributions in EB energy modeling are rarely checked against an independent explanation method. Third, physics-informed parametric studies often do not test whether feature groups provide non-redundant predictive information beyond the target-generation equation. Fourth, prediction uncertainty is rarely reported, even though it is important for battery range planning and fleet scheduling.

The objectives of this study are to develop and compare machine learning models for per-kilometer EB energy prediction under Bangkok-like tropical urban conditions, to interpret the best model using SHAP, to assess attribution reliability using LIME, to quantify uncertainty using quantile regression intervals, to evaluate feature groups through ablation analysis, and to develop an ordinal efficiency classifier for operational screening. The study asks which model performs best, which variables contribute most to energy consumption, whether SHAP and LIME provide consistent attribution patterns, how speed and HVAC load interact, and how much predictive value is lost when key feature groups are removed.

The main contributions are as follows. This study provides a Bangkok-calibrated physics-informed EB energy modeling framework with 29 empirically anchored input variables. It compares XGBoost, LightGBM, Random Forest, and SVR under Bayesian hyperparameter optimization. The study applies SHAP and LIME together to examine attribution reliability, uses ablation analysis to address circularity in the physics-informed dataset, and reports quantile regression prediction intervals to characterize uncertainty.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the dataset construction, preprocessing, modeling, and explainability workflow. Section 4 presents the results. Section 5 discusses implications and limitations. Section 6 summarizes the conclusions and future research directions.

2 Literature Review

2.1 Energy Consumption Modelling for Electric Buses

The literature on electric bus (EB) energy modelling has developed along two complementary methodological traditions. Physics-based simulation models decompose energy demand into its constituent mechanical components: rolling resistance, aerodynamic drag, gravitational gradient force, and inertial kinetic energy at each acceleration event [13,19,35]. These models provide mechanistic transparency and are well suited to controlled parametric analysis under defined driving cycles. Vepsäläinen et al. [14] coupled longitudinal dynamics simulations with empirically measured HVAC loads, demonstrating that auxiliary electrical systems can contribute up to 40% of total per-kilometre energy consumption under cold Nordic operating conditions. The same group subsequently proposed a computationally efficient simulator that reproduces these effects at substantially reduced computational cost [34]. Fiori et al. [36] developed and validated a microscopic energy consumption model at second-by-second resolution, confirming that stop-and-go dynamics are the primary driver of per-kilometre consumption variability across urban routes. Hjelkrem et al. [20] validated a strategic EB energy model against fleet data from China and Norway, demonstrating transferability across climate and route contexts. Despite their physical rigour, physics-based models require comprehensive vehicle parametric inputs that are rarely available in operational data records, and their generalisation across heterogeneous fleets and diverse route profiles remains limited.

A complementary approach bridges physics-based modelling and data-driven prediction through the construction of physics-informed parametric datasets. These datasets apply validated energy equations and empirically anchored parameter ranges to generate large-scale scenario libraries for machine learning training, without requiring direct access to live telematics. Abdelaty et al. [15] constructed a fractional-factorial simulation dataset and demonstrated that machine learning models trained on such data can explain over 96% of energy consumption variance. Dabčević et al. [37] developed a trip-based regression model derived directly from a validated backward-looking physical simulation of a 12 m electric city bus, incorporating separate powertrain and HVAC submodels. Their final eight-input quadratic model achieved $R^2 = 0.981$ on unseen data and operated approximately six orders of magnitude faster than the underlying physical simulator, demonstrating that physics-derived models can deliver both accuracy and computational tractability for large-scale fleet planning. Wang et al. [29] extended this approach by generating a high-fidelity simulation dataset of 169,344 scenarios from a validated MATLAB/Simulink model, covering extreme operational conditions including road gradients of ±8%, passenger loads up to 75 passengers, and HVAC demands from 1.25 to 22.3 kW. These studies collectively validate the methodology of constructing physics-informed parametric datasets as a reliable and well-established pathway for EB energy model development, particularly in operational contexts where high-resolution telemetry is institutionally inaccessible [15,37].

Data-driven approaches using live operational records have also produced significant advances. Lajunen [13] applied regression tree models to cross-fleet EB consumption data, identifying battery capacity and route speed profile as primary discriminating predictors. Chen et al. [38] applied Long Short-Term Memory networks to EB energy estimation, achieving high accuracy using trip-level variables including distance, duration, speed, and temperature. Pamuła and Pamuła [39] attained precise trip-level predictions using deep learning incorporating weather conditions. Li et al. [40] analysed over 160,000 trips to examine the effects of dynamic traffic conditions and environmental variables on EB electricity consumption. Blades et al. [41] systematically evaluated route feature selection methods for zero-emission bus energy prediction, confirming that stop density and average speed are the most predictive route-level features. Liu and Liang [25] proposed a data-driven framework validated across multiple transit networks.

A critical lesson emerging from large-scale real-world studies is the necessity of multi-category feature engineering. Ma et al. [42] analysed 95,204 trips from 36 electric buses across three routes in China, extracting 17 features from four categories: traffic condition, vehicle status, driving behaviour, and environmental condition. After variance inflation factor screening and stepwise regression, all four categories retained significant predictors, confirming that EB energy consumption cannot be explained by any single factor family. LightGBM achieved the best cross-validated performance in this study. Tian et al. [43] extended this evidence to cold-climate conditions using 66,102 trip records from 29 New Flyer XE40 battery-electric buses in Montreal. XGBoost achieved $R^2 = 0.96$ and RMSE of 1.33 kWh, and SHAP analysis revealed that winter operating conditions increase energy consumption by up to 26% due to heating loads, while regenerative braking efficiency declines from 53.4% in summer to 32.2% in winter.

The role of driving behaviour and operational micro-dynamics has emerged as a distinct area of investigation. Lyu et al. [44] focused specifically on energy consumption at intersections, demonstrating that stopping cases consume approximately 71.5% more energy than non-stopping cases, at 1.386 vs. 0.808 kWh/km, because the recovered braking energy is smaller than the additional energy required for re-acceleration. Dong et al. [45] developed a real-time prediction framework that revealed scale-dependent accuracy: mean absolute error of average energy consumption rate decreases from 43.9% at 100 m prediction horizons to 7.5% at 16 km, with kinematic features dominating at short distances and environmental features becoming more important at longer horizons. These findings suggest that a single global model may conceal important sub-mechanisms, and that route-sensitive or horizon-sensitive modelling approaches can improve both predictive accuracy and operational relevance.

2.2 Gradient Boosting and Ensemble Methods in Transport Energy Prediction

Gradient boosting ensemble methods have established themselves as the benchmark approach for tabular transport prediction tasks [46]. The foundational gradient boosting framework was introduced by Friedman [24] as a general function approximation procedure that builds additive models by sequentially fitting base learners to residuals. Ke et al. [23] subsequently proposed LightGBM, which employs histogram-based split finding and gradient-based one-side sampling to achieve substantially faster training with competitive accuracy. Both XGBoost and LightGBM have been validated on vehicle energy prediction tasks [47]. Random Forest [48] serves as a widely used performance baseline, while Support Vector Regression [49] remains a competitive alternative for moderate-dimensional feature spaces under careful hyperparameter tuning.

Recent EB-specific benchmarking studies confirm this hierarchy. Zhao et al. [12] integrated traction modelling, transmission system efficiency evaluation, and dynamic passenger load estimation with LightGBM, achieving R^2 = 0.995 and MAPE of 3.92% for a BRT fleet in Xiamen, China. The study demonstrates that embedding physical realism into the feature construction process, rather than treating prediction as a purely statistical problem, yields models that are both more accurate and physically interpretable. Hussain et al. [50] benchmarked 11 ML algorithms for EV energy prediction and confirmed that tree-based ensembles, including Extra Trees, XGBoost, CatBoost, and LightGBM, consistently outperform kernel-based and neural methods on structured energy datasets, with Extra Trees achieving R^2 = 0.959. Ullah et al. [51] compared MLR, ANN, XGBoost, and LightGBM on 38,362 EV trips from Japan, finding that LightGBM performed best (R^2 = 0.98, RMSE = 16.34), followed by XGBoost (R^2 = 0.96, RMSE = 27.16), and that direct HVAC usage variables captured much of the temperature-related energy effect. Comparative studies consistently show that gradient boosting methods outperform bagging ensembles and kernel methods on large, feature-rich transportation datasets [25,52,53]. The primary reason is structural: gradient boosting's sequential residual correction allows it to model high-order feature interactions that bagging methods, which train trees independently, cannot capture with the same fidelity. This property is particularly consequential for EB energy prediction, where the interaction between speed, gradient, HVAC load, and congestion produces consumption patterns that are inherently non-additive.

2.3 Explainable Artificial Intelligence for Electric Bus Energy Modelling

The operational deployment of ML models in public transit management has been constrained by interpretive opacity. Fleet operators, route planners, and policy analysts require not only accurate numerical predictions but transparent, mechanistically coherent explanations that enable informed interventions rather than passive reliance on opaque algorithms [54–57]. The SHAP framework [26], derived from cooperative game theory, provides theoretically principled, consistent attributions by computing each feature's marginal contribution averaged across all possible feature coalitions. TreeSHAP [58] adapts this framework to tree ensembles in polynomial time, enabling exact attribution computation for XGBoost and LightGBM without approximation. The LIME framework [27] provides local explanations by fitting sparse linear models in a perturbed neighbourhood around individual predictions. While LIME lacks SHAP's global consistency guarantees, these two methods are theoretically complementary: SHAP provides both global and local attributions grounded in cooperative game theory, whereas LIME provides local, instance-level diagnostic interpretations. The structural differences between SHAP and LIME mean they may assign different importance rankings to features for the same prediction, and concordance testing between the two methods serves as a practical validation that reported attributions reflect genuine model behaviour rather than artefacts of a specific explanation algorithm [59,60].

Applications of SHAP to EB energy modelling have grown but remain limited in both geographic and methodological scope. Nan et al. [28] employed Shapley values to identify electricity-saving driving behaviours for electric buses in Nanjing, using micro-trip segmentation and Random Forest classification to distinguish energy consumption states across road links, stop entries, and stop exits. Their SHAP and PDP results established clear eco-driving thresholds: an optimal speed range of 30–40 km/h for road-link operation, regenerative braking benefits above 5% brake pedal depth at deceleration proportions exceeding 40%, and recommended acceleration profiles below 1.5 m/s^2 when departing from stops. Wang et al. [29] presented the most methodologically developed SHAP application in the EB domain to date, integrating XGBoost with SHAP and partial dependence plots on a 169,344-scenario simulation dataset. Their study identified road gradient, initial SOC, HVAC usage, and road condition as primary determinants, revealed nonlinear interactions between gradient and stop density, and translated explainability outputs into a human-machine interface for operational decision support. However, none of these studies cross-validated SHAP attributions against LIME at the instance level to assess whether reported feature rankings are stable across explanation methods.

Xu et al. [61] demonstrated the wider applicability of XGBoost combined with SHAP for trip-level environmental performance analysis, showing that extended idling, frequent stop events, and low-speed congested operation are the most important contributors to poor eco-scores across 1113 real-world trips. Their study also highlighted a methodological caution relevant to EB energy research: SHAP importance values do not always imply direct causal mechanisms, because a feature may be important partly as a proxy for correlated but omitted variables. Domain knowledge must therefore accompany SHAP interpretation. Zhang et al. [62] combined LightGBM with Shapley value analysis across four feature categories for real-world EB operation, confirming that HVAC status, outside temperature, average speed, and congestion index are the most influential predictors in their Chinese urban context. Tian et al. [43] extended SHAP-based EB energy interpretation to cold-climate fleet data, identifying movement time, stop frequency, average speed, and regenerative braking as the most influential predictors, and explicitly noting that prior studies' failure to include road gradient, dwell time, and regenerative braking among their features limits their explanatory completeness.

A related and methodologically important strand of XAI literature extends beyond EB-specific contexts to examine how SHAP and LIME behave in complex nonlinear prediction tasks. Kanti et al. [63] compared XGBoost and deep neural networks with both SHAP and LIME for thermophysical property prediction, finding that XGBoost outperformed more complex DNN architectures, and that SHAP and LIME provided complementary insights: concentration dominated thermal conductivity effects while temperature dominated viscosity. This confirms a general finding across energy and physical systems: a simpler, well-calibrated ensemble model can outperform deeper architectures while remaining more interpretable. Devanathan et al. [64] applied a combined XGBoost, SHAP, LIME, permutation feature importance, and partial dependence plot framework for multi-scale household energy forecasting, showing that integrating multiple XAI methods provides richer and more reliable attributions than any single method alone.

The collective evidence establishes that SHAP is increasingly adopted as the standard post-hoc explanation tool for tree-based energy prediction models, but cross-validation of SHAP attributions against LIME at the instance level to confirm attribution stability remains rare [28,29]. Furthermore, pairwise SHAP interaction decomposition to characterise synergistic effects among speed, HVAC demand, and congestion has not been applied to EB energy modelling in tropical urban contexts. The present study addresses both gaps.

2.4 Electric Bus Energy Modelling in Tropical and Asian Urban Contexts

HVAC-driven energy demand is substantially larger in tropical cities than in temperate-climate contexts. Kambly and Bradley [31] quantified HVAC contributions of 15%–30% of total bus energy consumption in warm climates. Kammuang-lue and Boonjun [32] simulated EB energy consumption in Chiang Mai, Thailand, under international driving cycles, confirming that high ambient temperatures amplify HVAC-related energy penalties beyond what temperate-climate benchmarks would predict.

The energy penalty of hot-climate HVAC operation extends beyond simple additive effects. Corbet et al. [65] analysed real-world electric bus operational data combined with hourly weather records from a European coastal operator and found that a 1°C increase in temperature is associated with a 1.17% decrease in driving energy consumption in mild conditions, while freezing temperatures produce an additional 4.02% increase, underscoring the nonlinear and non-symmetric nature of thermal energy penalties. Critically, regenerative braking efficiency was found to decrease by 3.19% per 1°C reduction in temperature and by 0.29% per 1 km/h increase in wind speed. Idling energy consumption doubled under cold-weather conditions, from approximately 0.6 to 1.2 kWh/h. While this study focused on cold conditions, its methodological contributions are directly transferable to hot-climate contexts: weather variables produce nonlinear interactions with HVAC demand, regenerative braking, and auxiliary loads that additive regression models cannot fully capture, and that require nonlinear ML approaches with interaction-sensitive explainability tools. In Bangkok's tropical context, where ambient temperatures routinely exceed 35°C and air-conditioning operates continuously, analogous thermal interaction effects are expected to be substantially larger than any temperate-climate benchmark would suggest. Vepsäläinen et al. [14] further demonstrated that auxiliary thermal loads are the most variable component of total energy consumption across seasonal and climatic conditions, a finding consistent with Bangkok's near-constant HVAC demand throughout the operational year.

The Thai and regional policy context directly motivates the research priority. The ESCAP and KMUTT transition report [66] documents that as of 2021, Thailand's public bus fleet comprised 16,209 Category 1 buses, of which fewer than 1% were battery-electric. Bangkok's specific fleet included only 115 battery-electric buses out of 9326 total fixed-route Category 1 buses. Thailand's EV Development Plan 2022–2037 targets 4412 electric buses in the Bangkok Metropolitan Area by 2030, with the BMTA planning to replace 2511 conventional buses with electric alternatives. The plan projects annual energy savings of 492 ktoe and greenhouse gas reductions of 0.915 $MtCO_2eq$ by 2030. Well-to-wheel emission comparisons in the report indicate that electric buses produce 435 g/km CO_2 against 1105 g/km for diesel buses, with near-zero tailpipe particulate emissions. These targets situate Bangkok at the centre of Southeast Asia's most active bus electrification programme. Janjamraj et al. [67] further demonstrated that retrofitting existing internal combustion engine buses into electric buses is technically feasible in Thailand using dynamic force-based motor and battery sizing, with converted prototypes achieving approximately 1.056 kWh/km in city operation and passing safety tests including slope, emergency braking, and high-voltage insulation assessments. This retrofit pathway is relevant to the broader regional transition, where full new-bus replacement is often financially inaccessible for smaller operators.

Mohanty et al. [68] addressed the data-scarcity problem characteristic of developing urban contexts by demonstrating that synthetic EV operational datasets generated through temporal generative adversarial networks can support ensemble ML energy prediction with high accuracy, with LightGBM achieving 93.39% accuracy on synthetic charging data from Berhampur, India. The study illustrates both the potential and the limitations of synthetic data: while generative models can supplement sparse real-world records, validation against field data remains essential before operational deployment. This finding is directly relevant to Bangkok's context, where institutional constraints similarly restrict direct access to high-resolution

EB telemetry, and where physics-informed parametric datasets offer a validated alternative pathway that preserves operational realism through empirical parameter anchoring.

The preponderance of existing EB energy modelling studies has concentrated on temperate-climate European cities, cold-climate North American networks, and major Chinese urban corridors [20,38,39,43]. The tropical Southeast Asian urban EB context remains substantially underrepresented. Bangkok's combination of sustained high temperatures, continuous HVAC demand, severe peak-hour congestion, high stop and intersection densities, multilevel road infrastructure, and a heterogeneous three-model fleet creates an operational envelope qualitatively different from any context studied in the existing explainable EB energy literature.

2.5 Identified Research Gaps

Synthesising the evidence across all four streams of literature, four research gaps motivate the present study. First, the tropical Southeast Asian urban EB context is substantially underrepresented in the explainable machine learning literature, which has concentrated on temperate-climate European datasets, cold-climate North American networks, and northern Chinese city corridors [20,38,39,43]. The HVAC energy penalty, which is the most thermally variable component of total EB energy demand [14,31], has not been characterised through interaction-sensitive explainability methods in a sustained-heat tropical context. Second, SHAP attribution analysis in EB energy modelling has not been systematically cross-validated against LIME at the instance level, leaving attribution reliability unverified across the studies reviewed [28,29]. Third, companion ordinal efficiency classification models that enable practical trip-level operational screening have not been developed alongside continuous regression predictors in the EB energy literature [41,69]. While regression models predict consumption magnitudes, operators require actionable categories that can trigger routing, scheduling, or driver coaching decisions. Fourth, pairwise SHAP interaction analysis between speed, HVAC load, and congestion has not been applied in the EB energy literature to characterise the synergistic consumption dynamics specific to tropical urban transit, where the joint effect of low-speed congestion and high thermal demand produces compounding energy penalties that additive models cannot represent [29,65].

2.6 Summary of Reviewed Literature

Table 1 synthesises the reviewed studies across eight analytical dimensions relevant to the present study: study context and geographic region, dataset type, primary ML algorithm(s), explainability method employed, inclusion of HVAC or thermal load variables, consideration of feature interactions, cross-method XAI validation, and whether the study addresses a tropical or Asian urban context. This matrix highlights both the cumulative progress in the field and the specific methodological and geographic gaps that motivate the present work.

As Table 1 demonstrates, the present study is the only work that simultaneously addresses all four dimensions identified as research gaps: a tropical and Southeast Asian operational context, cross-method XAI validation combining SHAP and LIME, explicit pairwise feature interaction analysis, and full HVAC and thermal load integration. Studies operating in Asian and Chinese urban contexts have applied SHAP but have not cross-validated against LIME [28,29,62]. Studies that have compared SHAP and LIME in applied engineering contexts have done so outside the EB domain [60,63,64]. Physics-informed or simulation-based dataset studies exist in the EB literature [15,29,37] but none have been developed for a tropical Southeast Asian BRT network under Bangkok's specific climatic and operational conditions. This multi-dimensional gap defines the unique contribution of the present study.

Table 1: Summary matrix of reviewed literature on electric bus and electric vehicle energy consumption modelling. *Abbreviations*: GB = gradient boosting; RF = random forest; SVR = support vector regression; DL = deep learning (LSTM/CNN/MLP); LR = linear regression; GP = Gaussian process; Sim. = simulation/parametric dataset; Real = real-world operational data; ✓ = addressed; ○ = partially addressed; – = not addressed.

Study	Region	Dataset	ML Algorithm	XAI Method	HVAC/Thermal	Feature Interaction	Cross-Method XAI	Tropical/Asian
Lajunen (2014) [13]	Finland	Real	Physics/Tree	–	○	–	–	–
Vepsäläinen et al. (2018) [14]	Finland	Sim.	Physics	–	✓	–	–	–
Vepsäläinen et al. (2019) [34]	Finland	Sim.	Physics	–	✓	–	–	–
Fiori et al. (2021) [36]	Italy	Sim.	Physics	–	–	–	–	–
Hjelkrem et al. (2021) [20]	China/Norway	Real	Physics	–	○	–	–	○
Dabčević et al. (2024) [37]	Europe	Sim.	Physics + LR	–	✓	–	–	–
Chen et al. (2021) [38]	China	Real	DL (LSTM)	–	○	–	–	✓
Pamuła & Pamuła (2020) [39]	Poland	Real	DL	–	✓	–	–	–
Li et al. (2021a) [40]	China	Real	RF/GB	–	–	–	–	✓
Blades et al. (2024) [41]	UK/Ireland	Real	Multiple	–	–	✓	–	–
Liu & Liang (2022) [25]	Multiple	Real	GB	–	–	–	–	–
Kivëkäs et al. (2018) [70]	Finland	Real	Physics	–	✓	–	–	–
Zhao et al. (2025) [12]	China (BRT)	Real	LightGBM	SHAP	✓	○	–	✓
Hussain et al. (2025) [50]	USA	Real	XGB/LGB/ET	–	✓	–	–	–
Ullah et al. (2022) [51]	Japan	Real	XGB/LGB	–	✓	–	–	✓
Ma et al. (2025) [42]	China	Real	LightGBM	–	✓	–	–	✓
Zhang et al. (2024) [62]	China	Real	LightGBM	SHAP	✓	○	–	✓
Tian et al. (2025) [43]	Canada	Real	XGBoost	SHAP	✓	○	–	–
Dong et al. (2025) [45]	China	Real	XGB + LSTM	SHAP	✓	–	–	✓
Lyu et al. (2025) [44]	China	Real	XGB/SVM	–	–	✓	–	✓
Li et al. (2021b) [69]	Multiple	Sim.	Multiple	–	–	–	–	–
Abdelaty et al. (2021) [15]	Multiple	Sim.	Multiple	–	✓	–	–	–

(Continued)

Table 1 (continued)

Study	Region	Dataset	ML Algorithm	XAI Method	HVAC/Thermal	Feature Interaction	Cross-Method XAI	Tropical/Asian
Nan et al. (2023) [28]	China	Real	RF	SHAP/PDP	–	○	–	✓
Wang et al. (2026) [29]	China	Sim.	XGBoost	SHAP/PDP	✓	✓	–	✓
Xu et al. (2020) [61]	Canada	Real	XGBoost	SHAP	–	○	–	–
Kanti et al. (2024) [63]	India	Real	XGB/DNN	SHAP/LIME	✓	○	✓	✓
Devanathan et al. (2025) [64]	India	Real	XGBoost	SHAP/LIME/ PFI/PDP	–	✓	✓	✓
Givisis et al. (2025) [60]	Spain	Real	XGB/LGB	SHAP/LIME	–	–	✓	–
Mohanty et al. (2024) [68]	India	Syn.	LightGBM	SHAP	–	–	–	✓
Kambly & Bradley (2014) [31]	USA	Real	Physics	–	✓	–	–	–
Kammuang-lue & Boonjun (2021) [32]	Thailand	Sim.	Physics	–	✓	–	–	✓
Corbet et al. (2023) [65]	Europe	Real	OLS	–	✓	✓	–	–
ESCAP & KMUTT (2022) [66]	Thailand	Policy	–	–	–	–	–	✓
Janjamraj et al. (2025) [67]	Thailand	Real	Physics	–	✓	–	–	✓
Present Study	**Thailand**	**Sim. (PI)**	**XGB/LGB/ RF/SVR**	**SHAP + LIME**	✓	✓	✓	✓

3 Methodology

3.1 Study Context and Analytical Framework

This study examines electric bus (EB) operating scenarios representative of the Bangkok Metropolitan Region, with emphasis on the Bus Rapid Transit and Electric Vehicle (BRT/EV) corridor network and affiliated urban and suburban services. The modeled network covers five route categories: BRT corridor, urban inner, urban outer, airport link, and suburban routes. Three representative bus models were considered: BYD K9, Yutong E12, and King Long EV, with battery capacities of 324, 374, and 300 kWh, respectively.

Bangkok has a tropical savanna climate, with a hot season from March to May, a rainy season from June to October, and a cool season from November to February [30]. The city experiences high ambient temperatures and high relative humidity for much of the year. These conditions create sustained HVAC demand and make Bangkok's EB energy profile different from those of temperate-climate fleets [31].

The analytical workflow is shown in Fig. 1. The framework proceeds from physics-informed parametric dataset construction to preprocessing, model training, validation, uncertainty estimation, SHAP and LIME explainability, feature-group ablation, and final interpretation.

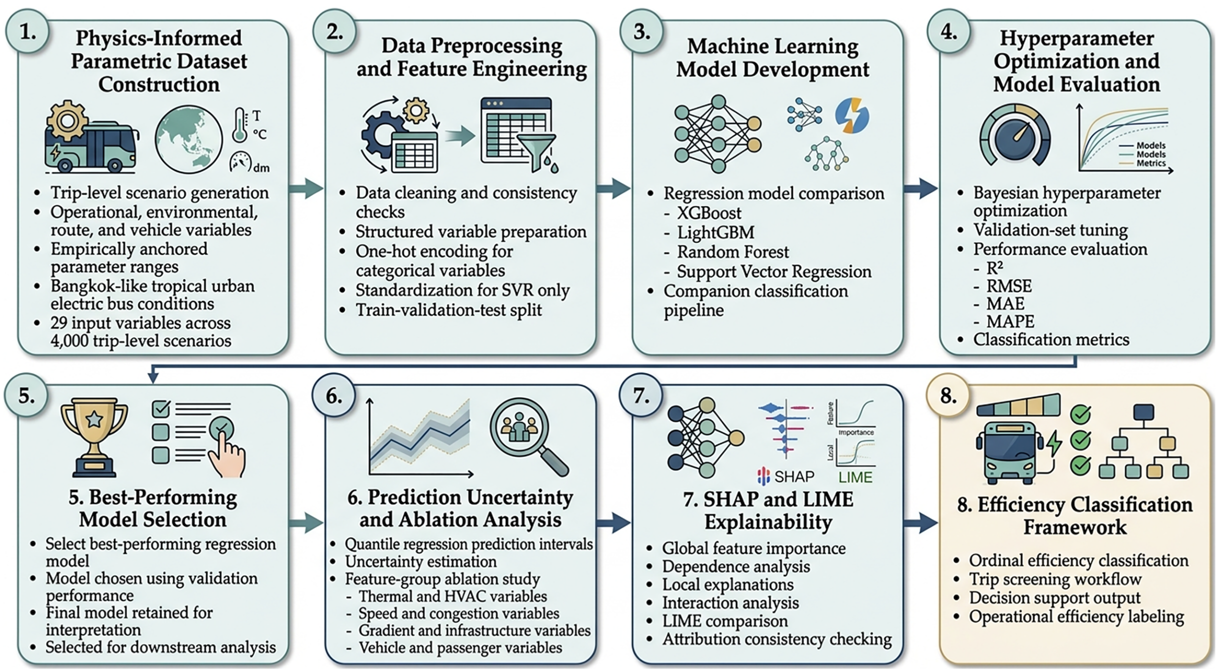

Figure 1: Overview of the analytical framework. The pipeline includes physics-informed parametric dataset construction, preprocessing, machine learning model training and validation, prediction uncertainty estimation, SHAP and LIME explainability, ablation analysis, and interpretation.

3.2 Physics-Informed Parametric Dataset Construction

This study uses a physics-informed parametric dataset rather than live high-resolution fleet telemetry. This framing is central to the study. The goal is to develop and evaluate an interpretable modeling framework for Bangkok-like EB operating scenarios, not to claim direct validation on operational Bangkok fleet records.

Direct access to trip-level EB telemetry from Bangkok transit agencies is institutionally constrained. Energy metering is not uniformly available across fleets, and detailed variables such as HVAC load, regenerative braking efficiency, and eco-driving behavior are rarely accessible from open administrative sources. Parametric and simulation-based data generation is widely used in EB energy studies when real-world telemetry is limited [15,29,34,36,69]. Accordingly, all results should be interpreted within the defined parametric scenario space.

The dataset contains 4000 complete trip-level scenarios. The input variables cover temporal and environmental conditions, route and infrastructure characteristics, traffic dynamics, vehicle specifications, and driver-operational factors. Variable ranges were anchored to published sources, including Thai Meteorological Department climate records, BMTA route information, Bangkok traffic data, manufacturer specifications, and international EB performance studies. Table 2 summarizes the feature definitions and empirical anchors.

Table 2: Input feature definitions, units, and empirical anchors for the Bangkok BRT/EV parametric dataset.

Variable	Description	Unit	Empirical Anchor
Temporal and environmental			
hour_of_day	Trip start hour	h	BMTA operational logs
day_of_week	Day of operation	Category	BMTA operating calendar
Month	Calendar month	1–12	Derived
Season	Hot, rainy, or cool season	Category	TMD classification [30]
ambient_temp	Ambient temperature	°C	TMD Bangkok records [30]
relative_humidity	Relative humidity	%	TMD Bangkok records [30]
Rainfall	Rainfall indicator	Binary	TMD precipitation records
Route and infrastructure			
route_type	BRT, urban, airport, or suburban route	Category	BMTA route structure
trip_distance	Total trip distance	km	BMTA route database
avg_road_gradient	Mean route gradient	%	Bangkok DEM
stop_density	Scheduled stops per kilometer	stops km^{-1}	BMTA stop database
intersection_density	Signalized junctions per kilometer	count km^{-1}	Bangkok traffic engineering data
peak_hour	AM or PM peak indicator	Binary	BMTA peak periods
Traffic dynamics			
congestion_index	Trip-level congestion severity	0–1	TomTom Bangkok Traffic Index [33]
avg_speed	Mean vehicle speed	km h^{-1}	BMTA GPS operating range
speed_variation	Within-trip speed variation	km h^{-1}	Li et al. [40]
stops_per_km	Actual stops per kilometer	Count km^{-1}	Lajunen [13]

(Continued)

Table 2 (continued)

Variable	Description	Unit	Empirical Anchor
Vehicle characteristics			
bus_model	BYD K9, Yutong E12, or King Long EV	Category	Manufacturer specifications
battery_capacity	Battery pack capacity	kWh	Manufacturer specifications
bus_age	Years since commissioning	Years	Ou [71]
soc_start	State of charge at trip start	%	Operating protocol
Operational and driver factors			
passenger_count	Number of passengers	Persons	BMTA ridership data
load_factor	Passenger load relative to capacity	%	BMTA operating data
hvac_load	Air-conditioning electrical load	kW	Kambly and Bradley [31]
aux_load	Auxiliary electrical load	kW	EB standard specifications
eco_driving_score	Driver eco-driving index	0–100	Nan et al. [28]
regen_efficiency	Regenerative braking efficiency	%	Fiori et al. [36]
trip_duration	Total trip duration	min	Derived from distance and speed

Each input variable was sampled from a distribution calibrated to its empirical anchor. Ambient temperature was generated from season-conditional truncated normal distributions. Relative humidity and rainfall probability were also season dependent. Trip distance was sampled by route type. Congestion index was constructed from time of day, rainfall, and random variation, then clipped to the observed Bangkok traffic range. Average speed was generated as an inverse function of congestion, adjusted by route type and stochastic variation.

The generator also imposed physically meaningful dependencies. For example, higher congestion reduced average speed, higher ambient temperature increased HVAC load, and passenger count affected load factor and vehicle mass. These dependencies were included to avoid unrealistic independent sampling of variables that are physically or operationally linked. Detailed descriptive statistics, distribution plots, and validation outcomes are reported in Section 4.

3.3 Energy Target Computation and Circularity Control

The primary regression target is energy consumption per kilometer, E_{km} in kWh km^{-1}. It was computed using a physics-informed composite model:

$$E_{\mathrm{km}} = \frac{E_{\mathrm{traction}} + E_{\mathrm{HVAC}} + E_{\mathrm{aux}} - E_{\mathrm{regen}}}{d}, \tag{1}$$

where d is trip distance. The traction component represents vehicle mass, road gradient, speed, stop frequency, battery age, and eco-driving behavior:

$$E_{\text{traction}} = \left[E_{\text{base}} + f_{\text{mass}} + f_{\text{gradient}} + f_{\text{speed}} + f_{\text{stops}} + f_{\text{age}} + f_{\text{eco}}\right]\varepsilon_{\text{traction}}. \tag{2}$$

The component functions were calibrated from published EB energy studies. Passenger mass was represented using a per-passenger load increment [15]. Gradient effects were calibrated using published traction-energy relationships [14]. Stop penalties represented repeated braking and acceleration events. Battery age captured capacity degradation [71], while eco-driving represented smoother driving behavior [28]. The speed function followed a U-shaped efficiency relationship, with penalties under very low speed due to stop-and-go operation and penalties at high speed due to aerodynamic drag [36,47].

HVAC and auxiliary energy were computed from power demand and trip duration:

$$E_{\text{HVAC}} = \frac{P_{\text{HVAC}} t_{\text{trip}}}{d}, \qquad E_{\text{aux}} = \frac{P_{\text{aux}} t_{\text{trip}}}{d}, \tag{3}$$

where $t_{\text{trip}} = d/v$ is trip duration in hours. This formulation captures the effect of slow travel on per-kilometer auxiliary and HVAC energy demand. Regenerative braking was represented as:

$$E_{\text{regen}} = \eta_{\text{regen}} E_{\text{traction}}, \tag{4}$$

where η_{regen} is the regenerative recovery fraction, calibrated from urban EB recovery ranges [36,70]. A stochastic residual term was added to represent measurement-like variation in real telematics streams.

Because the target was generated from a physics-informed equation whose inputs overlap with the machine learning features, circularity is a possible concern. The reported predictive performance should therefore be understood as the model's ability to recover the imposed physical structure and residual variation within the parametric scenario space. It should not be interpreted as equivalent to validation on independent live fleet telemetry. To address this issue, the study includes feature-group ablation experiments that quantify how much predictive information remains when major input domains are removed.

3.4 External Validation Procedure

External validation was conducted to assess whether the generated parametric scenarios were consistent with published EB measurements from Thailand, tropical climates, and international benchmarks. The validation checked energy consumption ranges, HVAC contribution, regenerative braking efficiency, gradient effects, speed-energy behavior, and passenger-load effects against published sources [13–15,31,32,36]. The purpose of this step was to evaluate whether the generated scenario space was physically realistic before machine learning model training. The validation results and figures are reported in Section 4.

3.5 Data Preprocessing and Partitioning

Categorical variables, including route type, bus model, day of week, and season, were one-hot encoded. Continuous predictors were standardized only for SVR because the RBF kernel is scale sensitive. Tree-based models received raw continuous variables and encoded categorical variables. CO_2-equivalent emissions were excluded from the regression feature set because they are a direct linear transformation of the energy target.

The ordinal efficiency category was derived from per-kilometer energy consumption thresholds. Trips below 1.65 kWh km^{-1} were labeled *Moderate Efficiency*. Trips from 1.65 to 2.20 kWh km^{-1} were labeled *Low Efficiency*. Trips above 2.20 kWh km^{-1} were labeled *Very Low Efficiency*. Class distributions and classification results are reported in Section 4.

The dataset was split into training, validation, and testing partitions using a stratified 70/10/20 split. Stratification preserved efficiency class proportions across partitions. The validation set was used for hyperparameter optimization and model selection. The test set was held out until final evaluation to reduce information leakage.

3.6 Machine Learning Models and Selection Rationale

Four regression algorithms were compared: XGBoost, LightGBM, Random Forest, and Support Vector Regression (SVR). A companion XGBoost classifier was trained for ordinal efficiency classification. The implementations used Python 3.11 with `scikit-learn` 1.3, `xgboost` 2.0, and `lightgbm` 4.1.

XGBoost and LightGBM were selected because gradient boosting methods are well suited to structured tabular prediction tasks and can capture nonlinear feature interactions [23,25]. They also support TreeSHAP-based interpretation [58]. Random Forest was included as a bagging baseline [48]. SVR with an RBF kernel was included as a nonlinear kernel-based baseline [49]. This model set therefore compares boosting, bagging, and kernel-based learning under the same dataset and evaluation design.

The final model for explainability was selected using validation-set RMSE as the primary criterion. The held-out test set was used only for final performance reporting.

3.7 Hyperparameter Optimization and Training

Hyperparameter optimization was conducted using the Tree-structured Parzen Estimator sampler in Optuna [72]. The optimization objective was validation RMSE. Gradient boosting models were optimized over 60 trials, while Random Forest and SVR were optimized over 40 trials. Early stopping was used for XGBoost and LightGBM when validation performance no longer improved.

After optimization, each model was retrained using its selected configuration and evaluated on the held-out test partition. The final search spaces and selected configurations are reported in Table 3 to support reproducibility.

Table 3: Hyperparameter search spaces and selected configurations.

Model	Hyperparameter	Search Range	Selected Value
XGBoost	n_estimators	200–1000	795
	max_depth	3–10	3
	learning_rate	0.01–0.30, log	0.024
	subsample	0.60–1.00	0.604
	colsample_bytree	0.50–1.00	0.726
	reg_alpha	10^{-4}–1.0, log	0.011
	reg_lambda	10^{-4}–5.0, log	0.994
LightGBM	n_estimators	200–1000	720
	num_leaves	20–150	20
	learning_rate	0.01–0.20, log	0.030
	min_child_samples	10–100	25
	feature_fraction	0.50–1.00	0.785
	bagging_fraction	0.50–1.00	0.607
Random Forest	n_estimators	100–500	365
	max_depth	5–30	15
	max_features	sqrt, log2	sqrt
	min_samples_split	2–15	3
	min_samples_leaf	1–10	1

(Continued)

Table 3 (continued)

Model	Hyperparameter	Search Range	Selected Value
SVR	C	0.1–100, log	40.88
	γ	10^{-4}–0.1, log	0.00108
	ε	0.001–0.5, log	0.010

3.8 Evaluation Metrics

Regression performance was evaluated using R^2, root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). These metrics jointly measure explained variance, large-error sensitivity, typical absolute error, and scale-free percentage error. RMSE was treated as the primary model-selection metric because large underpredictions are important for battery range planning [73]. MAE and MAPE were used as complementary measures. The RMSE-to-MAE ratio was also inspected as a residual diagnostic [74].

For the efficiency classification task, performance was assessed using overall accuracy, per-class precision, recall, F1 score, macro-F1, and a row-normalized confusion matrix. These metrics were selected because the Very Low Efficiency class is expected to be less frequent than the other classes.

3.9 Prediction Uncertainty Estimation

Prediction uncertainty was estimated using XGBoost quantile regression. Separate models were trained for the 0.05, 0.50, and 0.95 quantiles using the pinball loss objective. The resulting interval $[\hat{q}_{0.05}(\mathbf{x}), \hat{q}_{0.95}(\mathbf{x})]$ was treated as a 90% prediction interval for each observation. Empirical coverage and mean interval width were evaluated on the held-out test set and are reported in Section 4.

These quantile intervals provide uncertainty estimates for decision support, but they are not conformal intervals and do not guarantee finite-sample coverage. Conformal calibration is therefore identified as a future extension.

3.10 SHAP and LIME Explainability Workflow

SHAP decomposition was applied to the selected regression model using TreeSHAP [58]. Global importance was computed as the mean absolute SHAP value of each feature across the test set. Dependence plots, local waterfall explanations, and pairwise SHAP interaction plots were used to examine global, local, and interaction effects.

LIME explanations were generated for 50 systematically sampled test observations [27]. For each observation, the Spearman rank correlation between absolute SHAP values and LIME coefficient magnitudes was computed. The average correlation was used as the cross-method concordance measure. A value of $\rho \geq 0.70$ was treated as acceptable concordance, values from 0.60 to 0.70 were interpreted as near-threshold or partial concordance, and values below 0.60 were treated as weak concordance [59]. This interpretation avoids overstating attribution stability when the agreement is close to, but below, the acceptability threshold.

3.11 Ablation Study Design

An ablation study was conducted to assess the independent predictive contribution of major feature groups and to address the circularity concern in the physics-informed target. Four groups were removed one at a time:

1. *HVAC and thermal variables*: HVAC load, auxiliary load, ambient temperature, relative humidity, rainfall, and season indicators.
2. *Speed and congestion variables*: average speed, speed variation, congestion index, actual stops per kilometer, peak hour, trip duration, and intersection density.
3. *Gradient and infrastructure variables*: road gradient, stop density, and route type indicators.
4. *Vehicle and passenger variables*: passenger count, load factor, battery capacity, bus age, bus model indicators, and state of charge at trip start.

For each ablation condition, the selected model type was retrained from scratch using the reduced feature set and the same optimization protocol. Performance changes relative to the full-feature model were measured using ΔR^2, ΔRMSE, and ΔMAPE. Larger drops indicate that the removed feature group provides predictive information that cannot be recovered from the remaining variables.

4 Results

4.1 Exploratory Data Analysis

The distributional patterns of the target and selected predictors are summarized in Fig. 2. Panel (a) shows the energy consumption per kilometer target. The distribution is unimodal and moderately right-skewed (skewness = 0.84), with a modal peak near 1.55 kWh km^{-1}, a mean of 1.670 kWh km^{-1}, and an upper tail extending to 3.348 kWh km^{-1}. The interquartile range is 1.467–1.837 kWh km^{-1}. This asymmetry indicates that most trips fall under moderate conditions, while a smaller share combines severe congestion, elevated HVAC demand, road gradient, and heavy passenger loads.

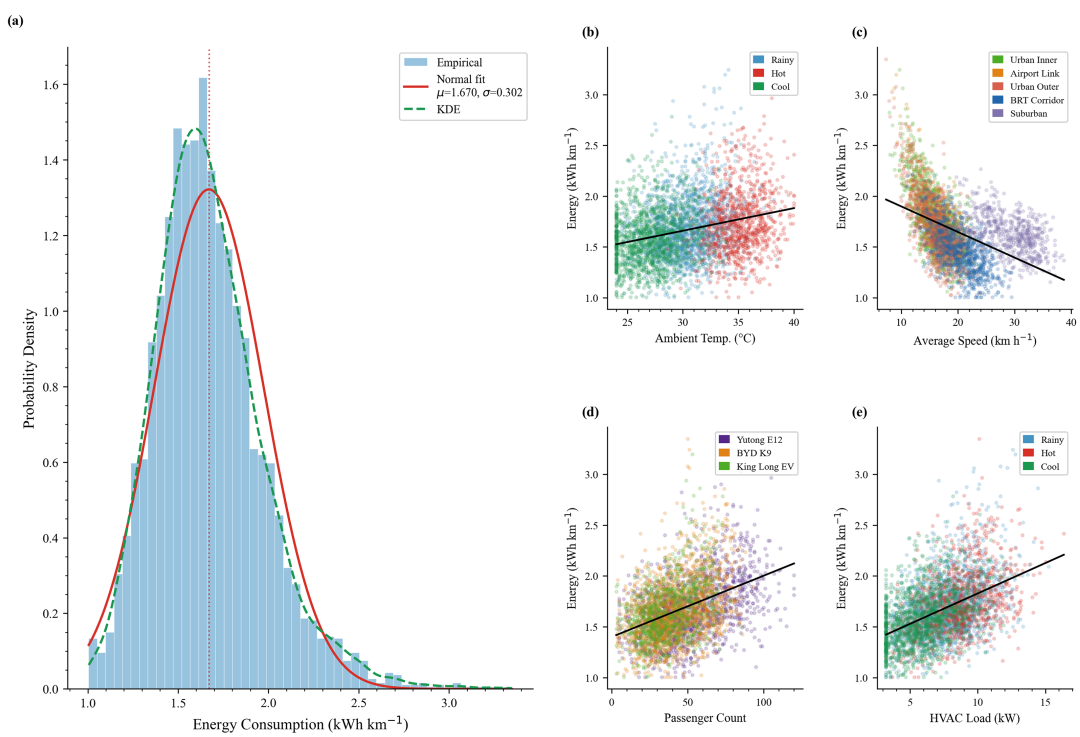

Figure 2: Distributional profiles of the primary regression target and selected input features across 4000 Bangkok BRT/EV bus trips. (**a**) Energy consumption per kilometer distribution with normal fit (μ = 1.670, σ = 0.302 kWh km^{-1}) and KDE overlay. (**b**) Scatter of energy vs. ambient temperature by season. (**c**) Scatter of energy vs. average speed by route type. (**d**) Scatter of energy vs. passenger count by bus model. (**e**) Scatter of energy vs. HVAC load by season.

Panel (b) shows a positive association between ambient temperature and energy consumption (r = +0.276), consistent with temperature-dependent HVAC demand. Panel (c) shows a negative association between average speed and energy consumption (r = −0.471), with many observations below 25 km h^{-1}. Panel (d) shows a positive payload-energy pattern across the three bus models (r = +0.412). Panel (e) shows a positive HVAC-energy association (r = +0.468) across seasons, supporting the importance of thermal load in Bangkok's tropical operating context. Before model training and interpretation, the distributional properties of the parametric scenario space were examined to verify that the generated operating conditions covered the expected Bangkok-like electric-bus operating envelope. Table 4 reports compact summary statistics for the primary target variable and selected operational, environmental, traffic, infrastructure, and vehicle-related predictors. These descriptive statistics provide the numerical basis for the subsequent histogram and seasonal-trend analyses.

Table 4: Summary statistics for selected continuous variables in the Bangkok-like BRT/EV parametric dataset (n = 4000 trips).

Variable	Unit	Mean	SD	Median [IQR]	Min–Max
Energy consumption	kWh km^{-1}	1.670	0.302	1.604 [1.467–1.837]	1.005–3.348
Trip distance	km	14.36	7.68	12.78 [8.75–18.27]	3.00–45.00
Trip duration	min	47.19	28.78	40.70 [29.80–54.10]	6.50–236.30
Average speed	km h^{-1}	19.08	5.66	17.80 [15.30–21.50]	7.30–38.70
Ambient temperature	°C	30.47	3.75	30.25 [27.60–33.30]	24.00–40.00
Relative humidity	%	72.27	7.69	72.00 [66.70–77.60]	51.10–97.90
HVAC load	kW	10.49	2.06	10.50 [9.10–11.88]	3.23–17.35
Passenger count	persons	45.25	20.64	43 [30–59]	3–120
Congestion index	0–1	0.482	0.165	0.438 [0.349–0.632]	0.124–0.937
Average road gradient	%	0.529	0.664	0.47 [0.08–0.91]	−1.52–2.93
Actual stops	stops km^{-1}	3.753	1.466	3.69 [2.59–4.77]	0.30–8.98
Regenerative braking efficiency	%	20.15	2.35	20.10 [18.60–21.70]	11.60–28.00

Route- and season-level variation in energy consumption is illustrated in Fig. 3. Urban inner routes show the highest median consumption, consistent with dense stop patterns and low average speeds under inner-city congestion. BRT corridor routes show lower median consumption, which is consistent with the speed advantage of dedicated bus lanes. Airport link and suburban routes show wider spreads, reflecting greater variability in gradient exposure and passenger load. Across seasons, hot-season trips show the highest median energy consumption, while the cool-season median is lower because of reduced cooling demand.

Fig. 4 decomposes total energy use into traction, HVAC, auxiliary, and regenerative components. Across all trips, HVAC energy accounts for 25.1% of net per-kilometer consumption, auxiliary load accounts for 11.9%, and regenerative braking recovers about 20% of traction energy. The mean components are E_{traction} = 1.319 kWh km^{-1}, E_{HVAC} = 0.418 kWh km^{-1}, E_{aux} = 0.198 kWh km^{-1}, and E_{regen} = 0.264 kWh km^{-1} recovered. HVAC share varies by season, with 28.0% in the hot season, 25.2% in the rainy season, and 20.0% in the cool season. These values fall within the 15%–30% warm-climate range reported by Kambly and Bradley [31]. The narrower seasonal variation compared with many temperate fleets is consistent with Bangkok's sustained cooling demand [14].

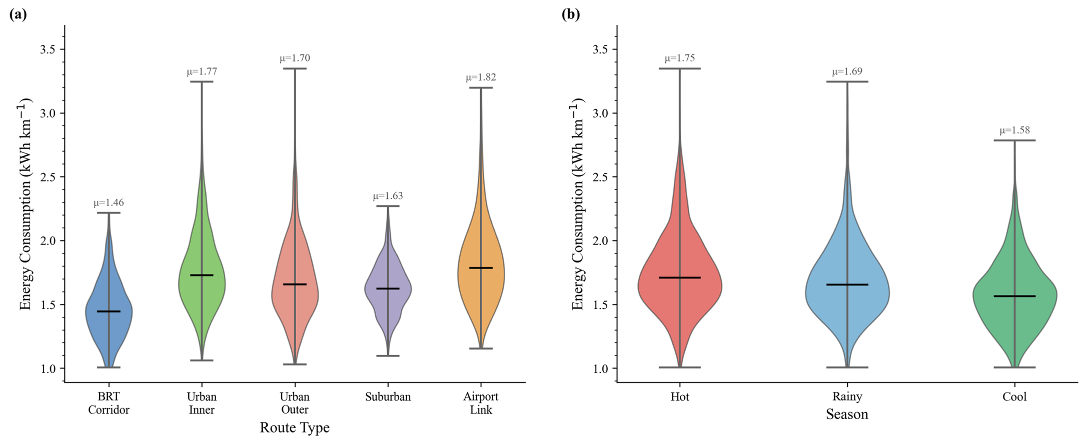

Figure 3: Violin plots of energy consumption (kWh km^{-1}) by (**a**) route type and (**b**) season. Horizontal marks indicate group medians.

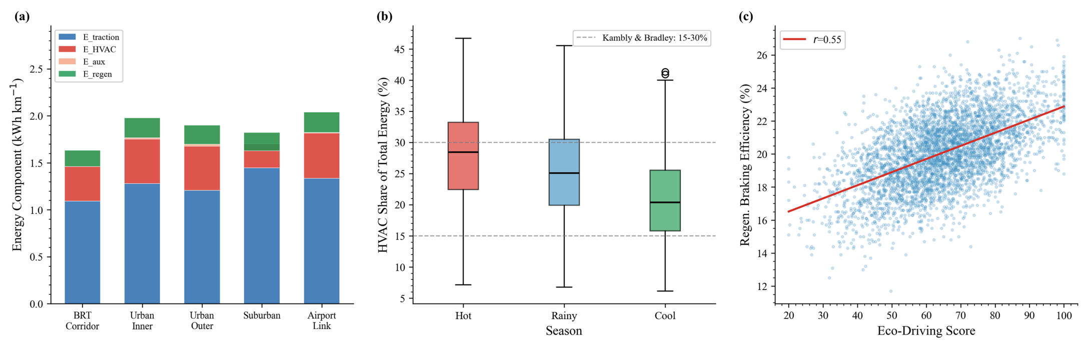

Figure 4: Energy component analysis. (**a**) Stacked bar chart of mean energy components (traction, HVAC, auxiliary) and regenerative recovery by route type. (**b**) HVAC share of total energy consumption by season; the shaded band marks the Kambly and Bradley [31] tropical reference range of 15%–30%. (**c**) Scatter of regenerative braking efficiency vs. eco-driving, indicating that smoother driving behavior is associated with improved energy recovery.

4.2 Model Performance Comparison

The held-out test-set performance of the four regression models is reported in Table 5 and visualized in Fig. 5. XGBoost gives the best result across all metrics: R^2 = 0.9439, RMSE = 0.0740 kWh km^{-1}, MAE = 0.0577 kWh km^{-1}, and MAPE = 3.51%. This indicates that the model explains approximately 94.4% of test-set variance. The RMSE-to-MAE ratio of 1.28 suggests broadly stable errors, with some influence from upper-tail observations.

Table 5: Regression model performance on the held-out test set (n = 800 trips). Bold values indicate the best result per metric. RMSE and MAE are in kWh km^{-1}.

Model	R^2	RMSE	MAE	MAPE (%)
XGBoost	**0.9439**	**0.0740**	**0.0577**	**3.51**
LightGBM	0.9322	0.0813	0.0618	3.73
SVR	0.9391	0.0771	0.0597	3.60
Random Forest	0.8132	0.1350	0.1020	6.18

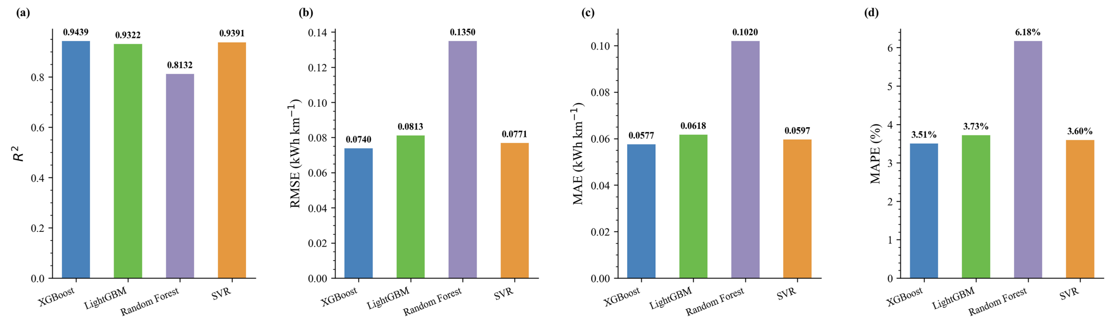

Figure 5: Performance comparison of all four regression models on the test set across (**a**) R^2, (**b**) RMSE, (**c**) MAE, and (**d**) MAPE. XGBoost achieves the best result on all four metrics. Random Forest is substantially weaker, consistent with the absence of sequential residual correction in bagging ensembles.

LightGBM also performs strongly, with R^2 = 0.9322 and MAPE = 3.73%. SVR achieves R^2 = 0.9391 and MAPE = 3.60%, showing that the tuned kernel baseline remains competitive. Random Forest performs more weakly, with R^2 = 0.8132 and MAPE = 6.18%. This pattern is consistent with the greater ability of boosting models to represent nonlinear interactions among speed, HVAC load, gradient, and congestion. XGBoost's MAPE of 3.51% is below the 5%–8% range reported in comparable EB energy prediction studies [38,39].

4.3 XGBoost Prediction Fidelity

Prediction fidelity for the selected XGBoost model is shown in Fig. 6 through a hexbin density scatter plot of predicted and observed energy consumption. The densest region aligns closely with the 1:1 reference line in the 1.4–2.0 kWh km^{-1} modal range. Most observations fall within the ±10% tolerance band, indicating small prediction errors across the core of the modeled operating distribution. Scatter increases above 2.5 kWh km^{-1}, which is consistent with the lower frequency of extreme-consumption scenarios in the training data.

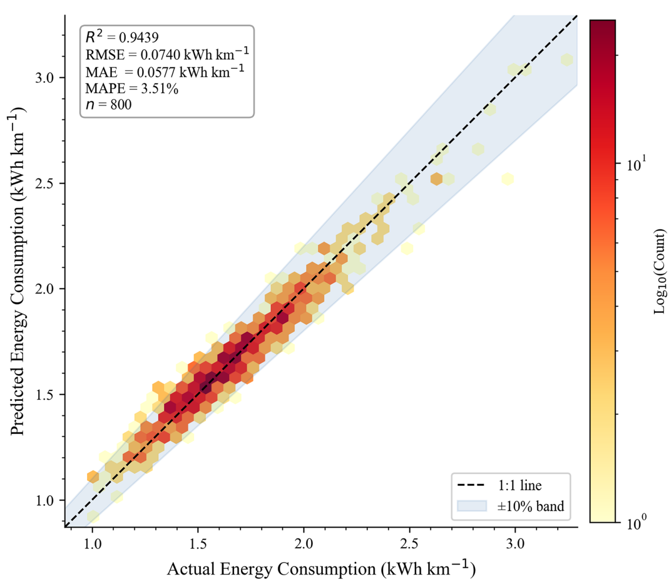

Figure 6: Hexbin density scatter of XGBoost-predicted vs. observed energy consumption (kWh km^{-1}) on the test set (n = 800). The dashed line is the 1:1 reference; the shaded band marks the ±10% tolerance. Color encodes $\log_{10}$ count per bin. Modest scatter above 2.5 kWh km^{-1} reflects the rarity of extreme-consumption events in the training data. R^2 = 0.9439, MAPE = 3.51%.

4.4 Residual Diagnostics

Residual behavior for the XGBoost model is evaluated in Fig. 7. Panel (a) shows residuals centered near zero, with a distribution center within 0.005 kWh km^{-1} of zero. Slightly heavier shoulders relative to the fitted normal density indicate a small number of larger residuals in the upper tail.

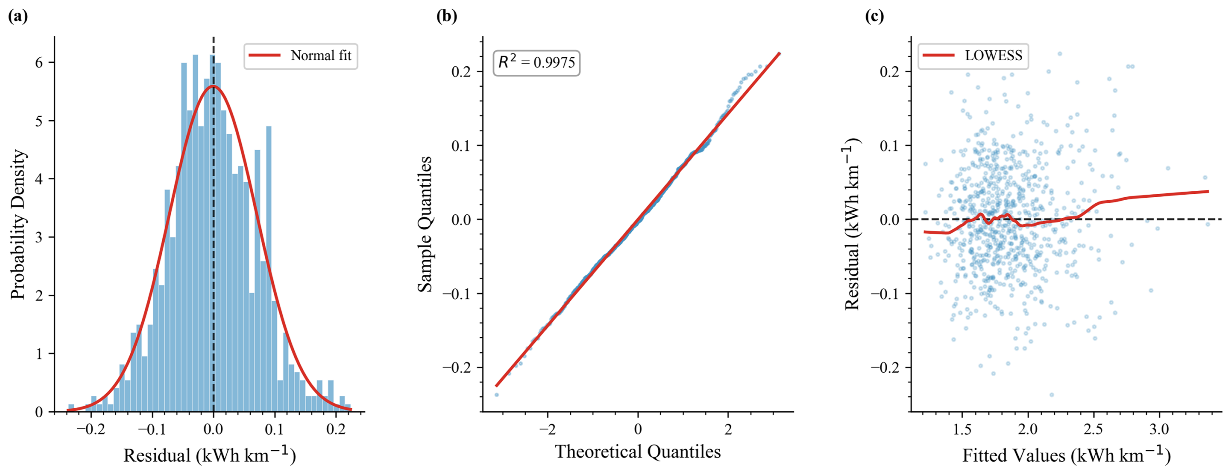

Figure 7: Residual diagnostic suite for the XGBoost model on the test set (n = 800). (**a**) Residual histogram with normal density overlay, showing near-zero mean and symmetric distribution. (**b**) Q-Q plot showing near-Gaussian residuals across the central range. (**c**) LOWESS-smoothed residuals vs. fitted values showing no clear heteroscedastic pattern or directional bias.

Panel (b) gives the Q-Q plot against theoretical normal quantiles. The residuals align closely with the reference pattern across the central range, with minor tail departures. Panel (c) shows the LOWESS-smoothed residuals as a function of fitted values. The trend remains close to zero and does not show clear curvature, fanning, or directional drift. These diagnostics indicate that the main systematic structure in the parametric test set is captured, while residual uncertainty remains largest for rarer tail cases.

4.5 Prediction Uncertainty Estimation

The uncertainty behavior of the XGBoost quantile regression model is summarized in Fig. 8. Panel (a) shows the 90% prediction intervals for test-set observations sorted by actual energy consumption. The empirical coverage is 79.4%, below the nominal 90% target. The mean interval width is 0.253 kWh km^{-1}.

The coverage shortfall of 10.6 percentage points suggests that the quantile intervals are under-calibrated. This may reflect the rarity of Very Low Efficiency cases in the upper tail, interaction complexity, and the fact that nonconformal quantile regression does not guarantee finite-sample coverage. Therefore, these intervals should be interpreted as exploratory uncertainty estimates rather than guaranteed operational bounds. Panel (b) shows that interval width increases with actual consumption, which is a useful qualitative pattern for identifying higher-uncertainty, high-consumption trips. Conformal calibration would be a suitable extension for stronger coverage control.

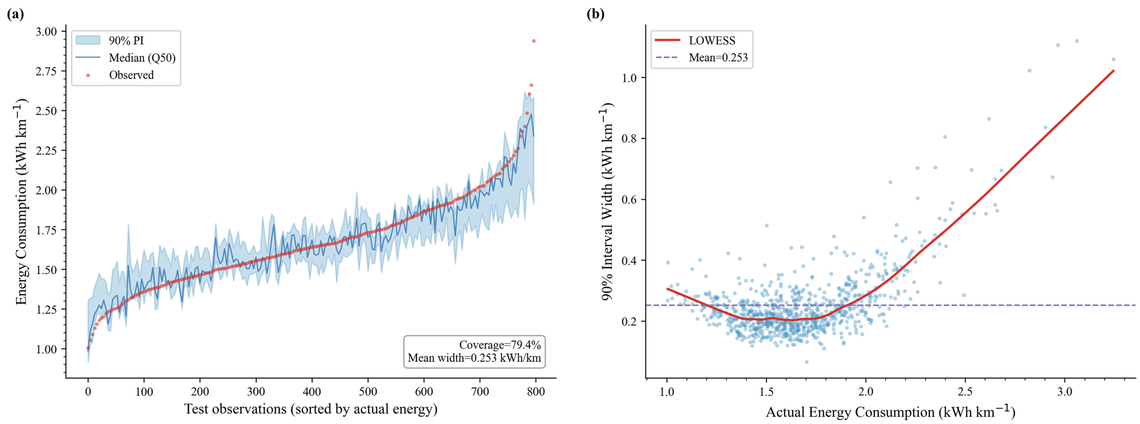

Figure 8: Prediction interval results from XGBoost quantile regression. (**a**) 90% prediction intervals for test-set observations sorted by actual energy consumption; empirical coverage = 79.4%, mean interval width = 0.253 kWh km^{-1}. (**b**) Interval width vs. actual energy consumption with LOWESS trend; width increases for high-consumption trips, indicating greater model uncertainty under extreme operating conditions.

4.6 Global SHAP Feature Importance

The global SHAP beeswarm plot in Fig. 9 shows how feature values influence the direction and magnitude of predicted energy consumption. Each point represents one trip, positioned by SHAP value and colored by feature value. The plot shows a three-tier importance structure.

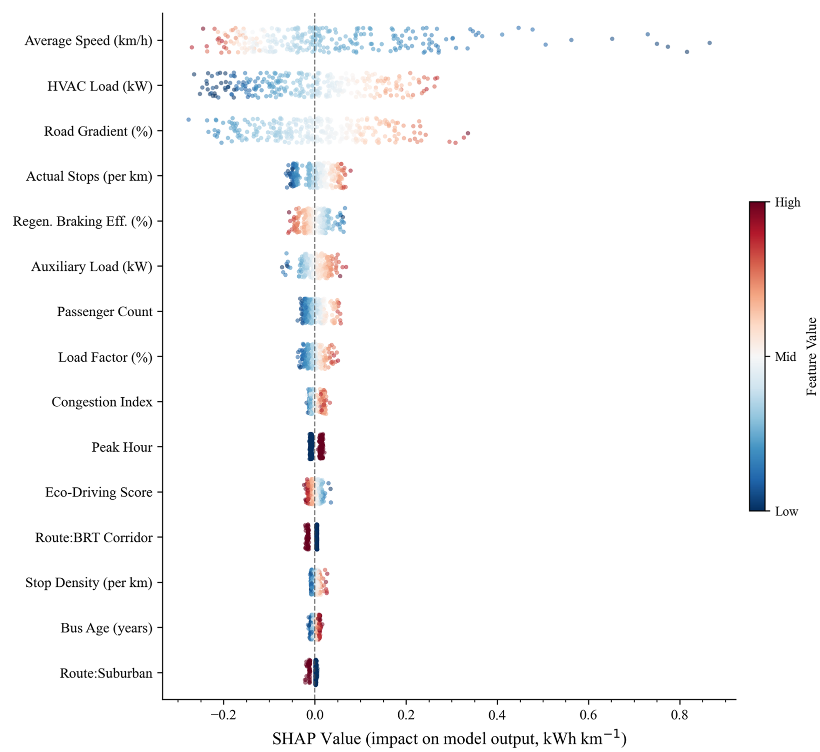

Figure 9: SHAP beeswarm plot for the XGBoost model on the test set. Each point represents one test observation. Points are colored by feature value (red: high, blue: low). Average speed is the dominant predictor (mean |SHAP| = 0.127 kWh km^{-1}), followed by HVAC load (0.105) and road gradient (0.096). The three-tier importance structure is evident.

The dominant tier includes average speed (mean |SHAP| = 0.127 kWh km^{-1}), HVAC load (0.105 kWh km^{-1}), and road gradient (0.096 kWh km^{-1}). The secondary tier includes actual stops per kilometer (0.029), regenerative braking efficiency (0.020), auxiliary load (0.019), and passenger count (0.017). The lower contextual tier includes load factor, congestion index, and peak hour indicator, each below 0.015 kWh km^{-1}.

A complementary bar-chart ranking of the top 15 SHAP features is given in Fig. 10.

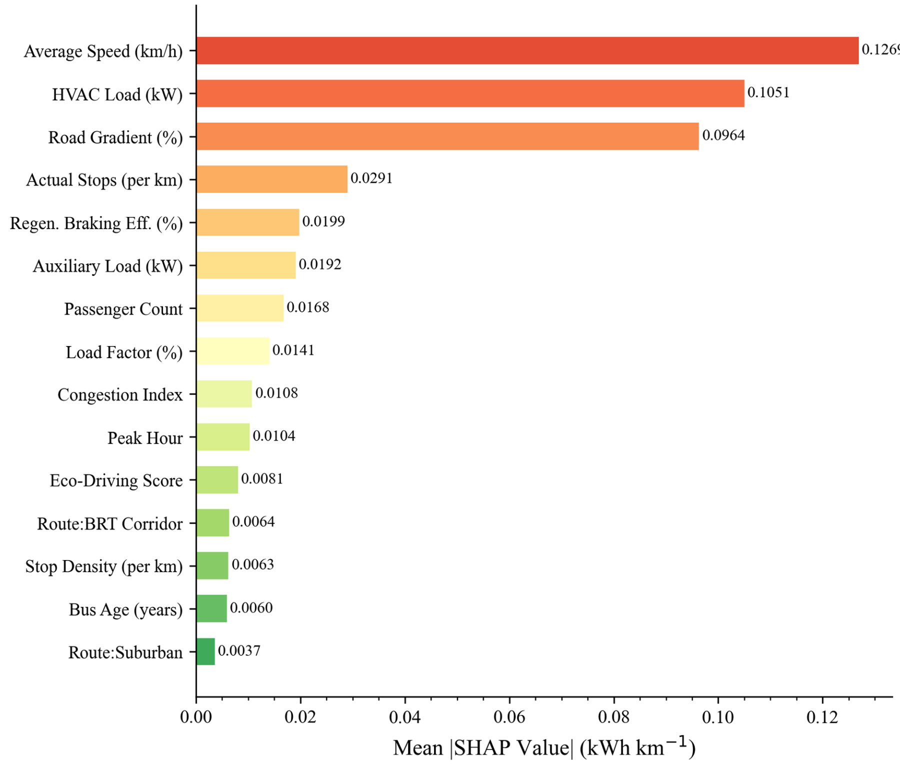

Figure 10: Global SHAP feature importance (mean |SHAP|, kWh km^{-1}) for the XGBoost model, top 15 features. Average speed (0.127), HVAC load (0.105), and road gradient (0.096) constitute the dominant tier. The remaining features form a secondary tier below 0.030 kWh km^{-1}.

Average speed ranks highest because it is linked to braking frequency, per-kilometer HVAC and auxiliary duration, and aerodynamic drag. Road gradient ranks highly despite Bangkok's generally flat terrain because even modest gradients affect heavy EB traction demand. HVAC load is also central because cooling demand remains persistent under tropical conditions. Lower-ranked features such as eco-driving score, congestion index, and bus age remain relevant, but their contributions are partly captured through correlated variables such as speed and stop frequency.

4.7 SHAP Dependence Analysis

4.7.1 Average Speed

The relationship between average speed and its SHAP contribution is examined in Fig. 11. At speeds below approximately 12 km h^{-1}, SHAP values become strongly positive, which is consistent with stop-and-go operation, longer per-kilometer auxiliary duration, frequent braking, and reduced regenerative recovery. As speed rises toward 20–25 km h^{-1}, SHAP contributions decline and become negative. The green band marks this efficiency zone, consistent with Fiori et al. [36] and Xu et al. [47]. The vertical dotted line at 19.1 km h^{-1} marks the fleet mean speed in the modeled dataset, placing many trips slightly below the most efficient range. Within this scenario space, measures that increase average speed toward this band may reduce per-kilometer consumption.

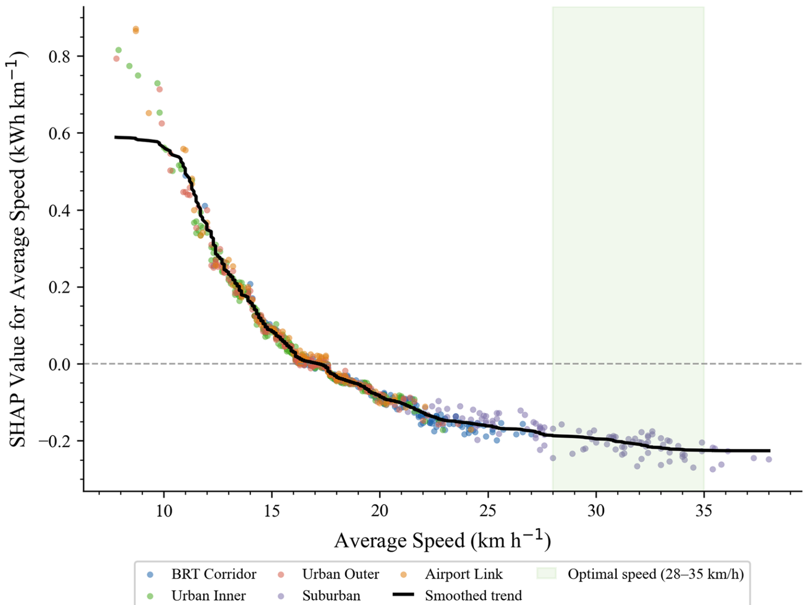

Figure 11: SHAP dependence plot for average speed (km h^{-1}), colored by route type. The green band marks the efficiency optimum at 20–25 km h^{-1}; the dotted line marks the fleet mean speed of 19.1 km h^{-1}. Points below the optimum experience increasing energy penalties from stop-and-go dynamics.

4.7.2 HVAC Load and Ambient Temperature

The temperature-dependent contribution of HVAC load is examined in Fig. 12, with relative humidity shown by color. SHAP values for HVAC load increase as ambient temperature rises and cross the zero-attribution line near 33°C. This indicates that HVAC load contributes positively to predicted energy consumption under hot conditions. The humidity color pattern is weaker than the temperature trend, suggesting that temperature is the main driver of HVAC attribution in this dataset. Because hot-season temperatures in Bangkok commonly exceed this range, HVAC-related penalties appear as a persistent feature of the modeled operating conditions.

4.7.3 Passenger Count

Passenger-load effects are visualized in Fig. 13. The positive, approximately linear pattern across the 3–120 passenger range is consistent with mass-traction coupling. Each additional passenger increases gross vehicle mass, which raises rolling resistance and acceleration energy. The vertical dotted line near 56 passengers marks the 70% capacity threshold for the BYD K9. The three bus models show similar patterns, suggesting that the payload-energy relationship is consistent across the modeled fleet.

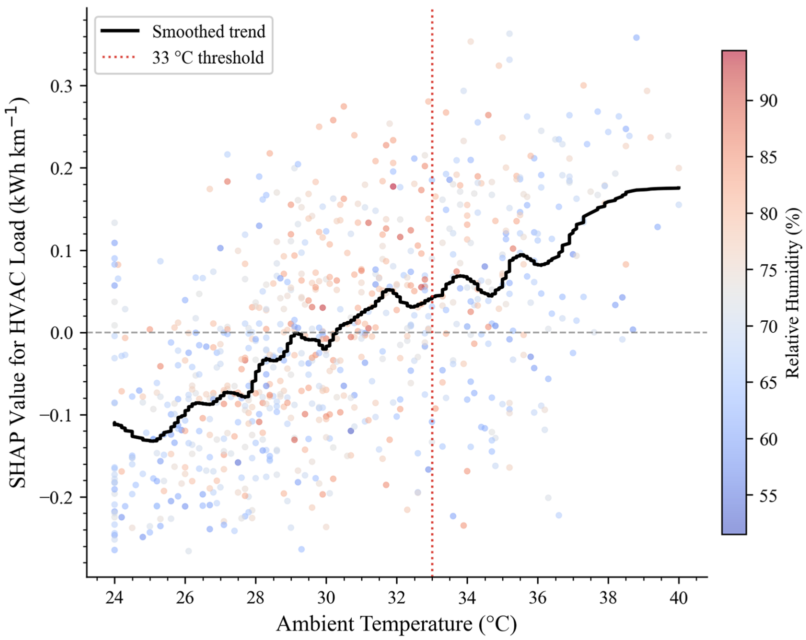

Figure 12: SHAP dependence plot for HVAC load as a function of ambient temperature (°C), with relative humidity (%) encoded as the color overlay. The red dotted line at 33°C marks the threshold above which the model assigns a net positive SHAP value to HVAC load. The smoothed trend indicates a monotonically increasing attribution with temperature.

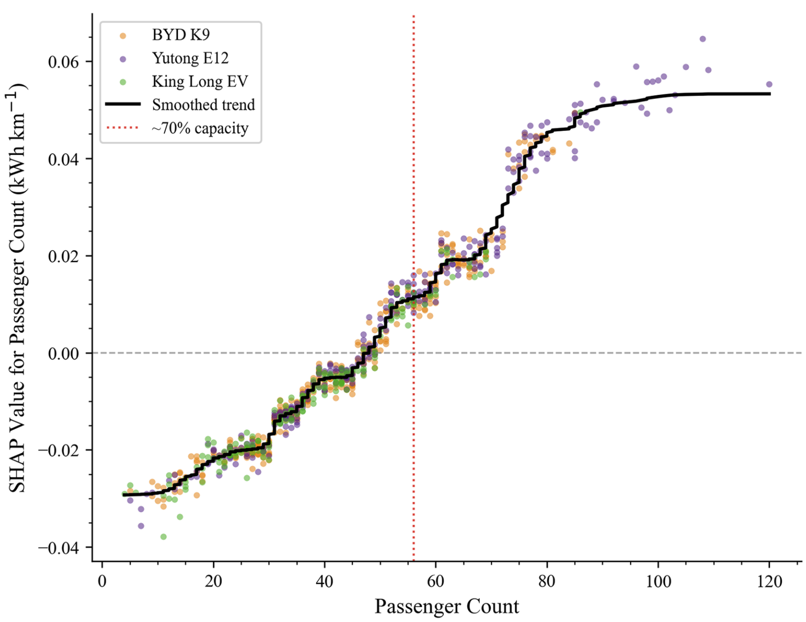

Figure 13: SHAP dependence plot for passenger count, colored by bus model (BYD K9, Yutong E12, King Long EV). The approximately linear positive relationship reflects the mass-traction coupling. The dotted line marks the 70% capacity threshold for the BYD K9. All three bus models show consistent behavior across the passenger count range.

4.8 Instance-Level Explanations: SHAP Waterfall Plots

4.8.1 High-Consumption Trip

The local explanation for the maximum-consumption test trip is shown through the SHAP waterfall plot in Fig. 14. The prediction is above the global mean baseline $\mathbb{E}[f(\mathbf{x})] \approx 1.670$ kWh km^{-1}. The largest positive contributor is very low average speed, consistent with stop-and-go congestion and longer per-kilometer auxiliary duration. Road gradient, elevated HVAC load, high stop frequency, and passenger payload further increase the prediction. Regenerative braking efficiency contributes slightly positively in this instance. The waterfall plot identifies how each factor contributes to the high predicted value within the modeled scenario.

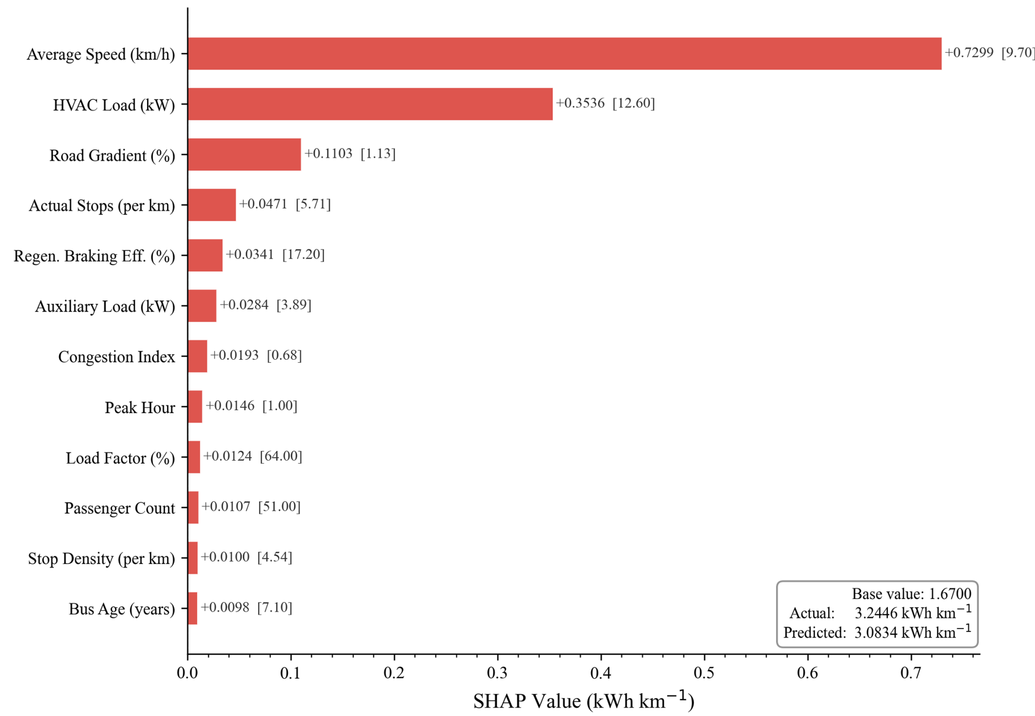

Figure 14: SHAP waterfall decomposition for the maximum-consumption trip in the test set.

4.8.2 Low-Consumption Trip

Fig 15 provides the corresponding SHAP waterfall explanation for the minimum-consumption test trip. This trip combines several favorable operating conditions. Average speed lies near the identified efficiency range and provides the largest negative SHAP offset. Near-zero or slightly favorable road gradient, lower HVAC load, lower auxiliary load, and fewer stops per kilometer further reduce the predicted energy consumption. In contrast, regenerative braking efficiency does not act as a reducing factor for this specific instance and contributes only slightly in the positive direction. The contrast between the two waterfall plots suggests that multiple favorable operating conditions can combine to produce much lower predicted energy consumption than any single factor alone.

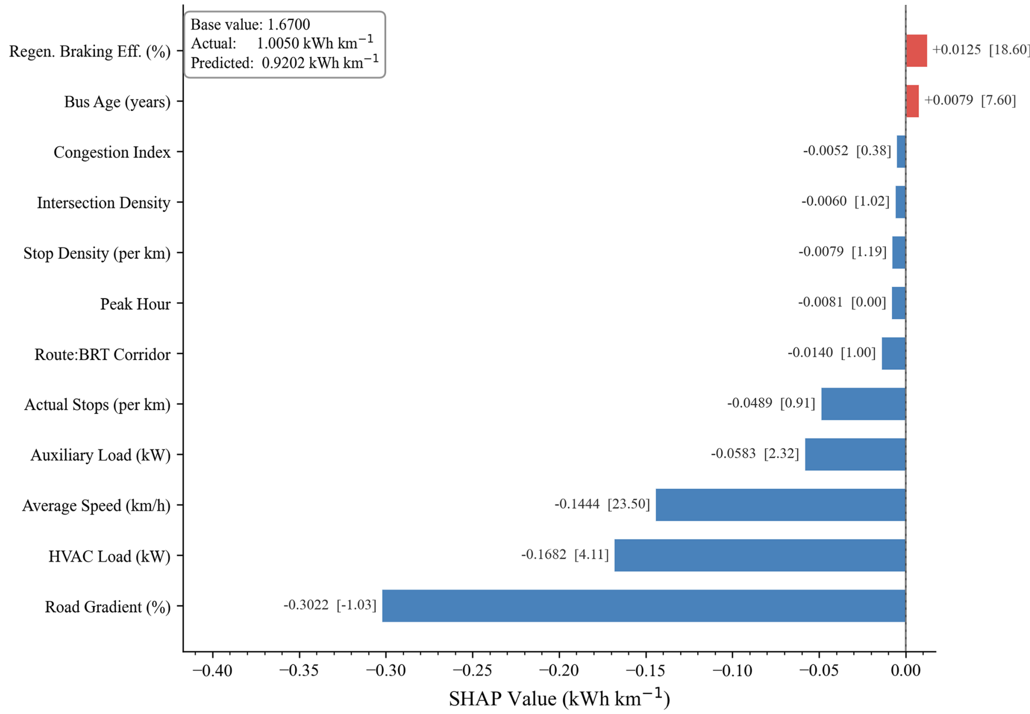

Figure 15: SHAP waterfall decomposition for the minimum-consumption trip in the test set.

4.9 SHAP Interaction Analysis

4.9.1 Pairwise Interaction Heatmap

Pairwise SHAP interaction strengths among the ten most important features are mapped in Fig. 16. Cell values represent mean absolute SHAP interaction values. The strongest interaction involves average speed and HVAC load. This is consistent with the mechanism that slower travel extends the per-kilometer duration of continuous HVAC draw. The second-strongest interaction involves average speed and road gradient, reflecting higher traction demand during low-speed uphill or elevated-section operation. The concentration of strong interactions around average speed indicates that speed is both a leading marginal predictor and a key interaction partner.

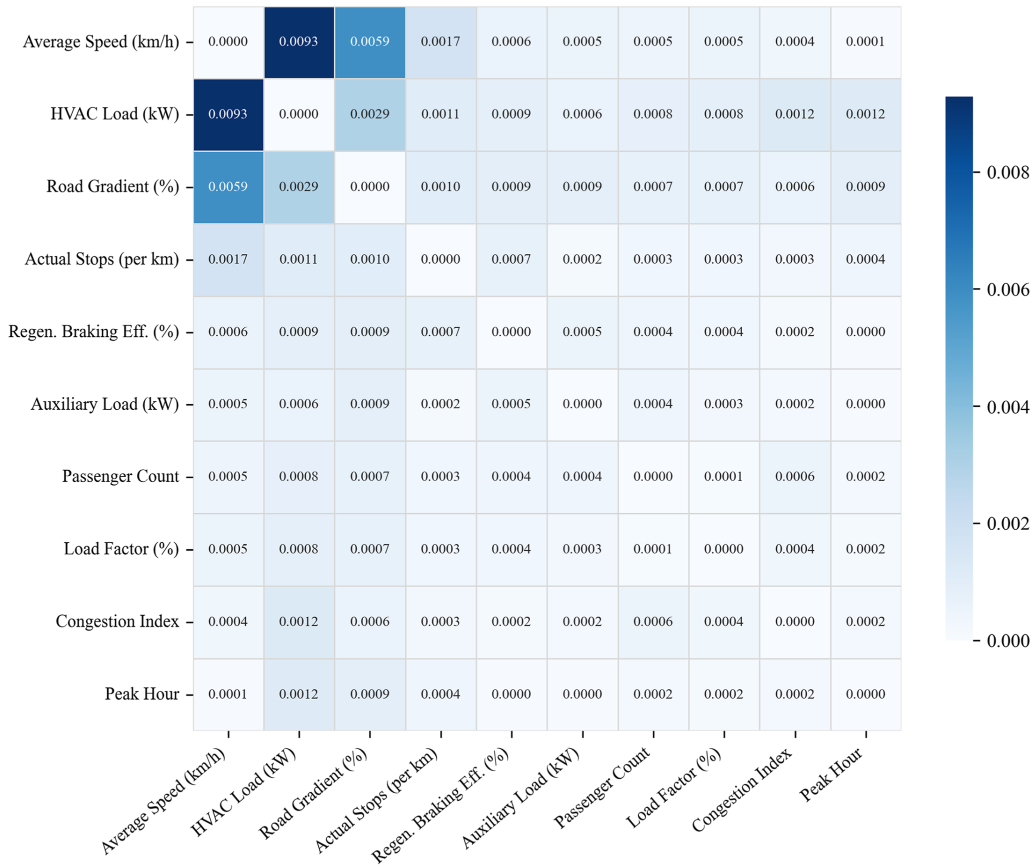

Figure 16: Pairwise SHAP interaction heatmap for the ten most important features (mean absolute SHAP interaction values, kWh km^{-1}). The speed-HVAC pair exhibits the strongest interaction, followed by speed-gradient. Average speed appears as the dominant interaction partner across multiple feature pairs.

4.9.2 Directional SHAP Interaction Plots

Directional interaction patterns are further explored in Fig. 17. Panel (a) shows the SHAP value for HVAC load against ambient temperature, colored by congestion level. SHAP values increase with temperature under both congestion conditions. High-congestion trips cluster at slightly higher SHAP values at similar temperatures, suggesting a modest amplification of HVAC effects during slow operation. Panel (b) shows the SHAP value for average speed against speed, stratified by HVAC load bands. The green band marks the 20–25 km h^{-1} efficiency range. Low speeds have positive SHAP values across all HVAC bands. At suboptimal speeds, high HVAC demand is associated with larger positive SHAP values than low HVAC demand. This pattern indicates that slow speed and high HVAC demand jointly increase predicted per-kilometer energy consumption. Panel (c) shows the SHAP value for congestion index against congestion level, grouped by speed category. Slow trips show higher congestion SHAP values than faster trips at similar congestion levels. This suggests that the energy effect of congestion is mainly mediated through average speed in the modeled scenario space.

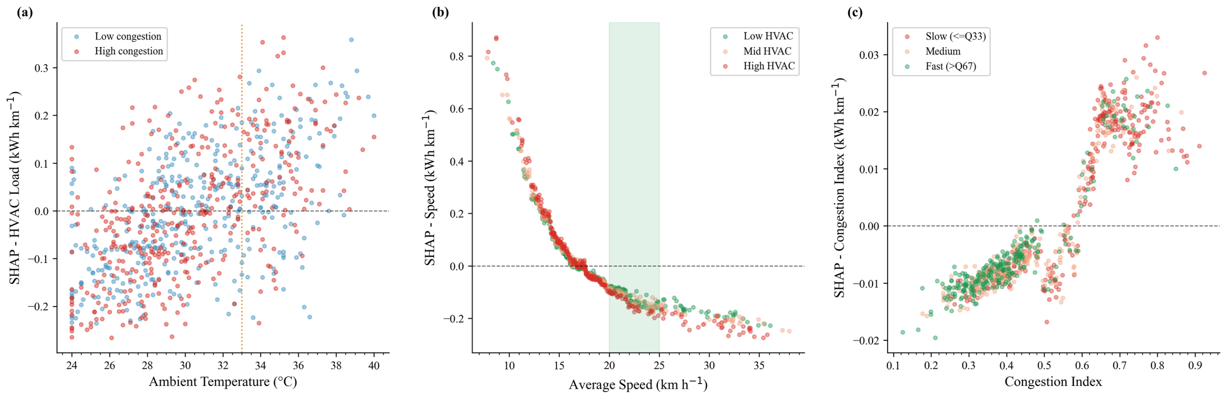

Figure 17: Directional SHAP interaction analysis. (**a**) SHAP value for HVAC load vs. ambient temperature, colored by congestion level; thermal energy penalties increase with temperature independently of congestion. (**b**) SHAP value for speed vs. speed, stratified by HVAC load band; high HVAC demand amplifies low-speed energy penalties. (**c**) SHAP value for congestion index vs. congestion, stratified by speed category; congestion penalties are mediated through speed.

4.9.3 Extended Interaction Surfaces

Additional two-variable response patterns are illustrated in Fig. 18. Panel (a) shows mean predicted energy consumption as a function of ambient temperature and passenger count. The surface rises as ambient temperature exceeds about 32°C and passenger count exceeds 60. Isoenergy contours shift from approximately 1.59–1.61 kWh km^{-1} in the cool, low-occupancy region to approximately 1.68–1.69 kWh km^{-1} in the hot, high-occupancy region. This pattern is consistent with joint thermal and payload loading during hot, crowded services.

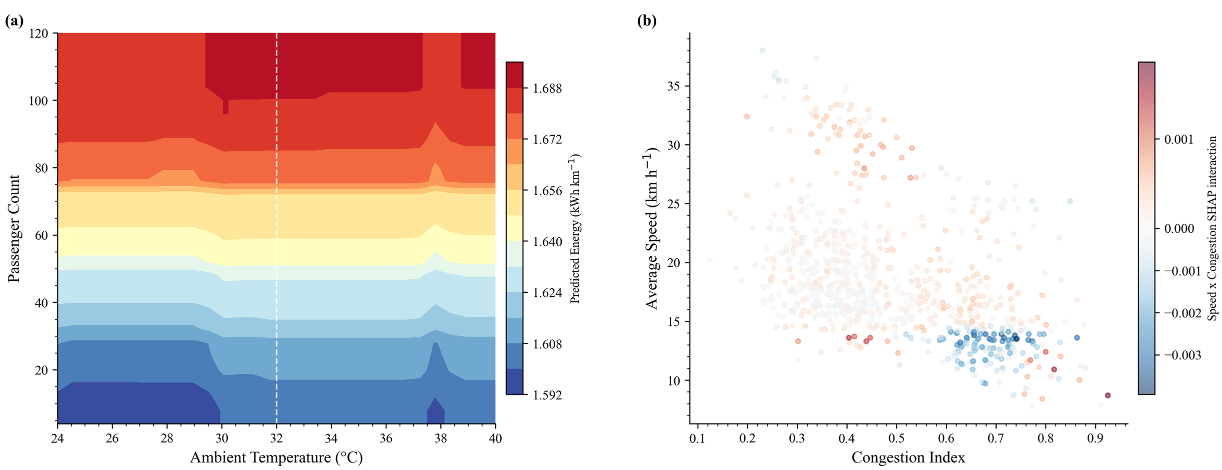

Figure 18: Extended interaction surfaces. (**a**) Mean predicted energy consumption across the ambient temperature and passenger count plane; isoenergy contours indicate joint thermal-loading amplification above 32°C and 60 passengers. (**b**) Speed-by-congestion SHAP interaction; negative values at high speed combined with low congestion indicate synergistic energy savings from uncongested operation.

Panel (b) shows the speed-by-congestion SHAP interaction. At speeds above about 28 km h^{-1} with congestion index below 0.4, interaction SHAP values are strongly negative. This suggests that higher-speed, low-congestion operation provides combined energy benefits within the modeled scenario space. The pattern is consistent with the potential benefit of operating conditions that both reduce congestion exposure and raise average speed.

4.10 Ablation Study Results

The feature-group ablation results are reported in Table 6. Removing HVAC and thermal variables reduces R^2 from 0.9439 to 0.7655 and increases MAPE from 3.51% to 7.34%. Removing speed and congestion variables produces the largest R^2 drop, with R^2 falling to 0.7380 and MAPE rising to 6.50%. Removing gradient and infrastructure variables also reduces performance (R^2 = 0.7991, MAPE = 6.80%). Removing vehicle and passenger variables causes only a modest change (R^2 = 0.9367, MAPE = 3.74%).

Table 6: Ablation study results. Each row reports XGBoost performance after removing one feature group. Δ values are relative to the full model baseline (R^2 = 0.9439, RMSE = 0.0740 kWh km^{-1}, MAPE = 3.51%).

Feature Group Removed	Feat.	R^2	MAPE (%)	ΔR^2	ΔMAPE (%)
Full model (baseline)	38	0.9439	3.51	—	—
Without HVAC variables	28	0.7655	7.34	−0.1784	+3.83
Without speed/congestion variables	28	0.7380	6.50	−0.2059	+2.99
Without gradient/infrastructure variables	30	0.7991	6.80	−0.1448	+3.29
Without vehicle/passenger variables	31	0.9367	3.74	−0.0072	+0.23

These results address the circularity concern associated with physics-informed target generation. The large performance drops after removing HVAC and speed/congestion variables show that these domains provide non-redundant predictive information within the parametric scenario space. This does not by itself establish live-fleet generalization, but it indicates that the full model is not relying only on easily recoverable information from the remaining correlated features.

4.11 SHAP-LIME Cross-Method Validation

Cross-method agreement between SHAP and LIME is summarized in Fig. 19. Panel (a) shows the distribution of Spearman rank correlations between SHAP absolute values and LIME coefficient magnitudes. The mean concordance is $\bar{\rho}$ = 0.686 (SD = 0.223), which is slightly below the pre-specified 0.70 threshold. This should therefore be interpreted as near-threshold agreement, not full attribution validation.

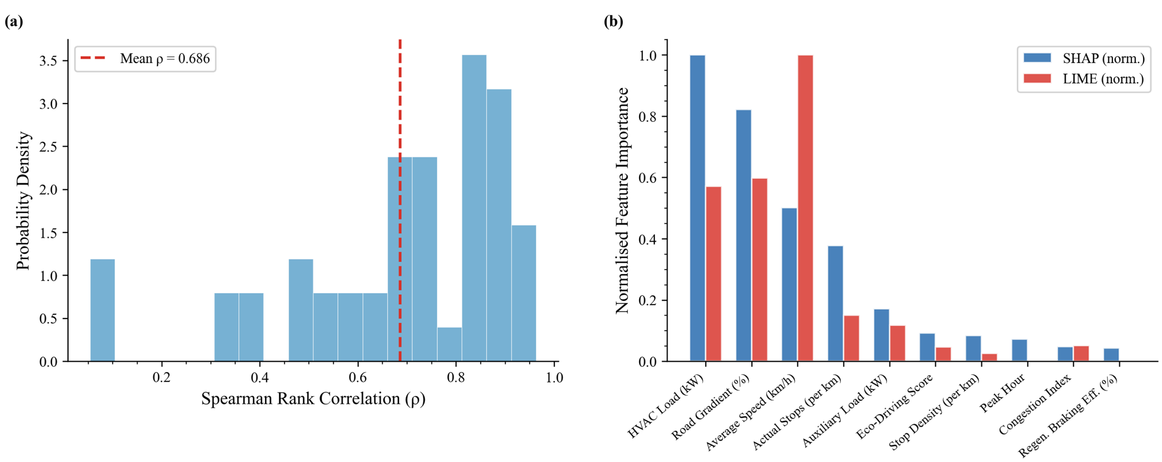

Figure 19: SHAP-LIME cross-method concordance across 50 test observations. (**a**) Distribution of per-observation Spearman rank correlations (ρ); mean $\bar{\rho}$ = 0.686 ± 0.223, marginally below the 0.70 threshold, with the shortfall attributed to interaction complexity and name-matching approximation. (**b**) Normalized global importance comparison.

The shortfall may reflect feature-name matching after one-hot encoding, the high interaction structure of the dataset, and the limits of LIME's local linear approximation. Panel (b) shows that both methods still agree on the top-three features: average speed, HVAC load, and road gradient. Thus, the SHAP-LIME comparison provides partial support for the dominant attribution hierarchy, while indicating that mid-ranked features should be interpreted with more caution.

4.12 Efficiency Classification Results

Classification performance is summarized in Table 7, with the row-normalized confusion matrix shown in Fig. 20 and per-class metrics displayed in Fig. 21.

Table 7: Per-class and macro-averaged classification performance of the XGBoost efficiency classifier on the held-out test set (n = 800). Overall accuracy = 88.25%, macro-F1 = 0.8423.

Class	Precision	Recall	F1 Score	Support
Moderate Efficiency	0.9089	0.9111	0.9100	416
Low Efficiency	0.8563	0.8713	0.8638	342
Very Low Efficiency	0.8286	0.6905	0.7532	42
Macro average	0.8646	0.8243	0.8423	800

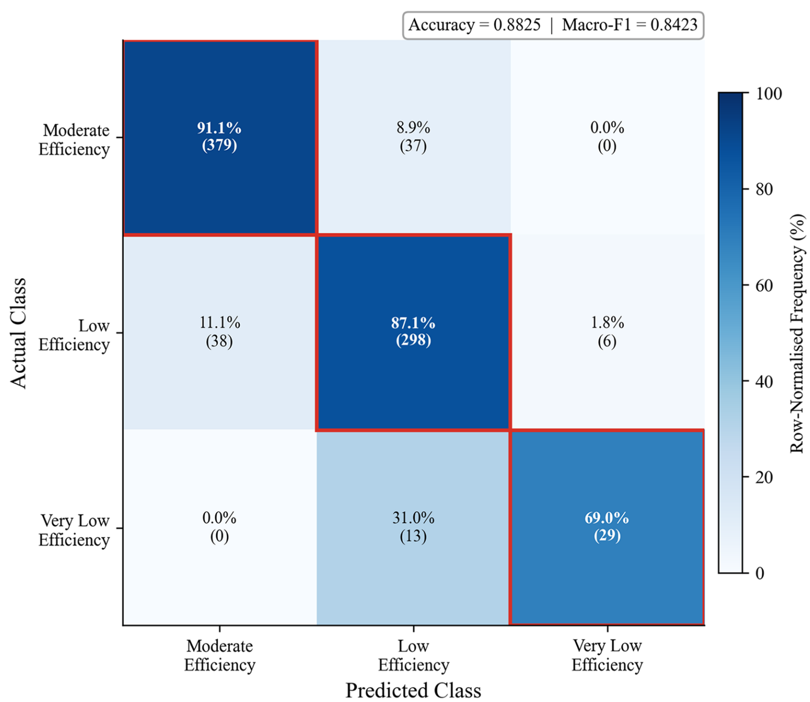

Figure 20: Row-normalized confusion matrix for the XGBoost efficiency classifier on the test set (n = 800). Overall accuracy is 88.25% and macro-F1 is 0.8423. Diagonal entries show correct classification rates per class. No misclassifications cross the two extreme class boundaries (Moderate vs. Very Low Efficiency), indicating that the most consequential category separation is preserved in the modeled test set.

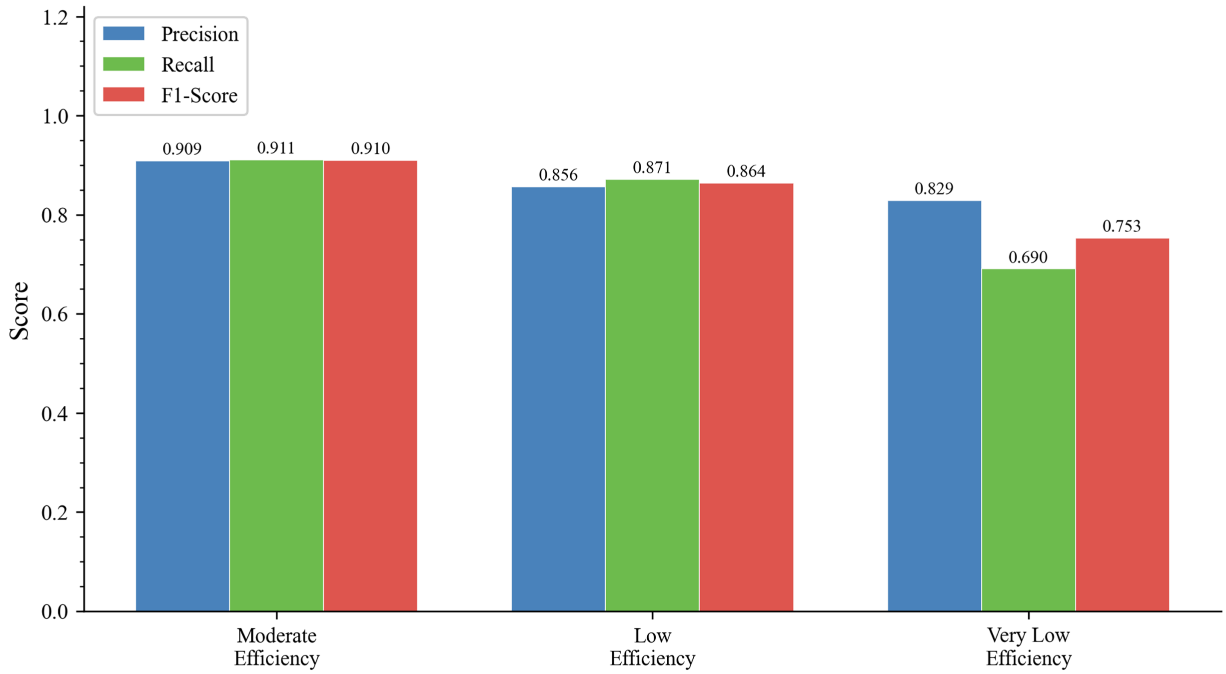

Figure 21: Per-class precision, recall, and F1 score for the XGBoost efficiency classifier. Moderate Efficiency achieves the highest performance on all three metrics. Very Low Efficiency achieves the lowest recall (0.6905), reflecting the class imbalance and the inherent boundary ambiguity for rare extreme-consumption events.

Overall accuracy is 88.25%, and macro-F1 is 0.8423. The Moderate Efficiency class achieves precision 0.9089, recall 0.9111, and F1 score 0.9100 across 416 test instances. The Low Efficiency class achieves precision 0.8563, recall 0.8713, and F1 score 0.8638 across 342 instances. Most errors occur between adjacent categories, which is expected near the 1.65 and 2.20 kWh km^{-1} class boundaries.

The Very Low Efficiency class achieves precision 0.8286, recall 0.6905, and F1 score 0.7532 across 42 instances. The lower recall reflects the small class size and the difficulty of separating rare high-consumption cases. No Moderate Efficiency trip is classified as Very Low Efficiency, and no Very Low Efficiency trip is classified as Moderate Efficiency. This suggests that the classifier preserves the most important separation between the two extreme categories in the modeled test set. Operational use would still require validation on live fleet data.

5 Discussion

5.1 Interpretation of the Main Findings

This study develops a physics-informed and interpretable machine learning framework for electric bus energy prediction under Bangkok-like tropical urban operating conditions. The findings should be interpreted within this framing. The dataset represents an empirically anchored parametric scenario space, not live high-resolution fleet telemetry from Bangkok transit operations.

Among the four regression models, XGBoost achieved the strongest performance, with R^2 = 0.9439, RMSE = 0.0740 kWh km^{-1}, MAE = 0.0577 kWh km^{-1}, and MAPE = 3.51%. The residual diagnostics suggest that the model captured the main systematic structure of the parametric energy distribution, with larger uncertainty mainly concentrated in rarer high-consumption cases. The weaker performance of Random

Forest is consistent with the need to model nonlinear interactions among speed, HVAC load, road gradient, and congestion.

The SHAP results identify average speed, HVAC load, and road gradient as the dominant predictors. This ranking is physically coherent. Average speed affects braking frequency, per-kilometer HVAC and auxiliary load duration, and aerodynamic drag. HVAC load is central because Bangkok's hot and humid climate creates sustained cooling demand. Road gradient is also important, even in a generally flat city, because flyovers, bridges, and elevated road sections impose additional traction demand on heavy electric buses.

The prediction interval analysis adds useful information but should be interpreted cautiously. The 90% quantile interval achieved 79.4% empirical coverage, which is below the nominal target. This indicates that the intervals are exploratory uncertainty estimates rather than guaranteed operational bounds. Their widening for high-consumption trips is still useful, because it highlights cases where model uncertainty is larger. Stronger calibration, such as conformal prediction, would be needed before using these intervals for guaranteed coverage [75].

5.2 Comparison with Previous Studies

The predictive accuracy of the proposed framework is consistent with recent EB energy prediction studies. Prior real-world and simulation studies commonly report MAPE values near 5%–8%, while more recent gradient boosting studies approach the 4%–5% range [29,38,39]. The MAPE of 3.51% obtained in this study compares favorably with those benchmarks. However, this comparison should be made carefully because the present dataset is physics-informed and parametric rather than directly measured from live fleet telemetry.

The findings also align with prior studies that identify speed, payload, route characteristics, and thermal load as major drivers of EB energy consumption [15,29]. The present study extends this literature by focusing on Bangkok's tropical operating conditions, where HVAC demand is not an occasional seasonal effect but a persistent energy burden. The HVAC contribution observed here is consistent with the 15%–30% warm-climate range reported by Kambly and Bradley [31] and with Thai evidence showing that high temperature increases EB cooling-related energy penalties [32].

5.3 Model Explainability and Circularity Assessment

A key concern in physics-informed parametric modeling is circularity. The target variable, E_{km}, is generated from a composite energy equation, and several predictors are also inputs to that equation. Therefore, the high predictive accuracy should not be interpreted as equivalent to validation on independent live telemetry. It reflects the model's ability to recover the imposed physical structure and residual variation within the defined scenario space.

The ablation study helps assess this issue. Removing HVAC and thermal variables, speed and congestion variables, or gradient and infrastructure variables caused clear performance degradation. This supports the view that these feature groups provide non-redundant predictive information within the parametric scenario space. The result does not prove live-fleet generalization, but it does show that the model is not relying only on easily recoverable information from the remaining correlated features. The modest effect of removing vehicle and passenger variables suggests that, in this dataset, much of their predictive contribution is already captured through operational and environmental variables.

The SHAP-LIME comparison provides partial support for the attribution structure. The mean Spearman concordance was $\bar{\rho} = 0.686$, which is slightly below the pre-specified 0.70 threshold. This should be described as near-threshold agreement rather than full attribution validation. Even so, both methods identified the same top-three features: average speed, HVAC load, and road gradient. This agreement strengthens confidence in the dominant feature hierarchy, while mid-ranked features should be interpreted more cautiously because LIME can be sensitive to local linear approximation, feature interactions, and one-hot encoded feature matching.

5.4 Practical Implications for Tropical Urban EB Operations

The findings point to several practical priorities for tropical urban EB systems, subject to validation on live fleet data. First, the strong role of average speed suggests that bus priority measures may reduce energy use by moving trips closer to the modeled 20–25 km h^{-1} efficiency range. Dedicated lanes, signal priority, and reduced stop-and-go delay could therefore provide energy benefits in addition to travel-time benefits.

Second, HVAC load should be treated as a core operational variable in warm-climate EB planning. In Bangkok-like conditions, cooling demand is structural rather than occasional. Depot pre-cooling, adaptive HVAC setpoints, improved insulation, and high-efficiency air-conditioning systems may reduce energy use, especially during hot-season peak periods. The speed-HVAC interaction further suggests that these measures are most valuable when paired with strategies that reduce slow crawling in congestion.

Third, the importance of road gradient indicates that digital elevation data should be included in route-level energy planning, even in cities that are generally flat. Flyovers, bridge approaches, and elevated sections can still create meaningful traction penalties. Route assignment and scheduling could account for these segments when battery margins are tight.

Finally, the companion efficiency classifier shows potential for screening trips into operational efficiency categories. Its strongest value is that no test case crossed the Moderate vs. Very Low Efficiency boundary. This suggests that the classifier could support trip-level review, driver coaching, or route investigation after validation on live fleet data. It should not yet be treated as a deployment-ready screening tool, but it provides a useful framework for turning continuous energy predictions into decision-oriented categories.

6 Conclusion

This study developed a physics-informed and interpretable machine learning framework for predicting electric bus energy consumption under Bangkok-like tropical urban operating conditions. A parametric dataset of 4000 trip-level scenarios was constructed using 29 input features across operational, environmental, infrastructural, and vehicular domains. All parameter ranges were anchored to published sources and externally benchmarked against Thai and tropical EB measurements. Therefore, the findings should be interpreted as evidence from a calibrated parametric scenario space, not as direct validation on live Bangkok fleet telemetry.

Among the four models evaluated, XGBoost achieved the strongest predictive performance, with $R^2 = 0.9439$, RMSE = 0.0740 kWh km^{-1}, and MAPE = 3.51% on the held-out test set. LightGBM and SVR also performed competitively, while Random Forest was weaker, consistent with the advantage of gradient boosting for interaction-rich tabular prediction tasks. Residual diagnostics indicated that the main systematic structure of the parametric energy distribution was captured, although higher uncertainty remained in rarer high-consumption cases.

SHAP analysis identified average speed, HVAC load, and road gradient as the dominant contributors to predicted energy consumption. Average speed was linked to stop-and-go losses, auxiliary load duration, and aerodynamic effects. HVAC load was important because Bangkok's tropical climate creates sustained cooling demand, and its interaction with speed showed that slow movement under hot conditions can compound energy use. Road gradient also contributed meaningfully, indicating that bridges, flyovers, and elevated sections should be considered in route-level energy planning even in a generally flat city.

The ablation study supported the non-redundant predictive contribution of the dominant feature groups within the parametric scenario space. Removing HVAC variables, speed and congestion variables, and gradient and infrastructure variables produced substantial performance losses. These findings suggest that the model was not relying only on easily recoverable information from correlated variables. However, because the target was generated from a physics-informed equation, these results should not be interpreted as proof of live-fleet generalization. SHAP-LIME comparison produced near-threshold concordance ($\bar{\rho}$ = 0.686), with both methods agreeing on the same top-three features. This provides partial support for the dominant attribution hierarchy, while mid-ranked features should be interpreted with more caution.

The companion XGBoost efficiency classifier achieved 88.25% accuracy and a macro-F1 score of 0.8423. No test case crossed the Moderate vs. Very Low Efficiency boundary, suggesting potential value for trip-level screening after validation on live fleet data. XGBoost quantile regression provided exploratory prediction intervals with 79.4% empirical coverage, below the nominal 90% target. These intervals are therefore useful for identifying higher-uncertainty cases, but conformal calibration is needed before they can be treated as guaranteed operational uncertainty bounds [75].

Several limitations should be considered. The dataset is parametric and does not fully capture live sensor noise, unobserved driver behavior, mechanical faults, long-term fleet degradation, or second-by-second acceleration and braking dynamics. Rainfall was also represented as a binary indicator rather than a continuous intensity measure. These constraints are common in physics-informed parametric EB modeling [15,29], but they should be considered when interpreting attribution magnitudes, classification performance, and transferability to real fleet operations.

Future work should validate the framework using GPS-matched high-resolution telemetry from Bangkok BRT/EV operations. Such data would allow calibration of the energy equation, external testing of the SHAP hierarchy, and stronger evaluation under real operational variability. Future extensions should also consider conformalized prediction intervals, partial dependence plots, and sequential models such as LSTM or Transformer architectures for within-trip driving dynamics. Applying the same framework to other tropical Asian cities, such as Jakarta, Kuala Lumpur, and Ho Chi Minh City, would help test transferability across different climates, fleet compositions, and congestion regimes.

Acknowledgement: The authors gratefully acknowledge the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, for supporting this research under grant number KFU261924.

Funding Statement: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under grant number KFU261924.

Author Contributions: Mohammed Almajed: Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing—original draft preparation. Mahbub Hassan: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing—original draft preparation, Writing—review and editing. Turjoy Das Turjo: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing—original draft preparation, Writing—review and editing. Md Ashequl Islam: Validation, Writing—review and editing. Md Ehtesamul Haque: Validation, Writing—review and editing. M. M. Hafizur Rahman: Supervision, Project administration, Writing—review and editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable, as this study used synthetic parametric data and did not involve human participants, animal subjects, clinical samples, or identifiable personal data.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Huang D, Li F, Yin J, Zhou Y, Bie Y, Wang H, et al. Evolution of electric bus systems: a review of models, methods, and future directions. Transp Res Part D Transp Environ. 2026;157:105387.
2. International Renewable Energy Agency. Electric buses: technology brief. Abu Dhabi, United Arab Emirates: IRENA; 2019 [cited 2026 Apr 20]. Available from: https://www.irena.org.
3. Mahmoud M, Garnett R, Ferguson M, Kanaroglou P. Electric buses: a review of alternative powertrains. Renew Sustain Energy Rev. 2016;62:673–84. doi:10.1016/j.rser.2016.05.019.
4. Zhou B, Wu Y, Zhou B, Wang R, Ke W, Zhang S, et al. Real-world performance of battery electric buses and their life-cycle benefits with respect to energy consumption and carbon dioxide emissions. Energy. 2016;96(2):603–13. doi:10.1016/j.energy.2015.12.041.
5. Suttakul P, Wongsapai W, Fongsamootr T, Mona Y, Poolsawat K. Total cost of ownership of internal combustion engine and electric vehicles: a real-world comparison for the case of Thailand. Energy Rep. 2022;8:545–53.
6. Basma H, Mansour C, Haddad M, Nemer M, Stabat P. Energy consumption and battery sizing for different types of electric bus service. Energy. 2022;239(May):122454. doi:10.1016/j.energy.2021.122454.
7. Zhou Y, Ong GP, Meng Q, Cui H. Electric bus charging facility planning with uncertainties: model formulation and algorithm design. Transp Res Part C Emerg Technol. 2023;150(2):104108. doi:10.1016/j.trc.2023.104108.
8. Qiu Z, Li S, Cui Y, Yang L, Kang Y, Ma X. Stochastic urban EV load forecasting: integrating markov-monte carlo simulation with spatio-temporal and traffic-weather dynamics. Intl J Electr Power Energy Syst. 2026;174(11):111510. doi:10.1016/j.ijepes.2025.111510.
9. Tian Z, Yan Q, Ma Y, Zhu C. Site and capacity planning of electric vehicle charging stations based on road–grid coupling. World Electr Veh J. 2026;17(2):101. doi:10.3390/wevj17020101.
10. Rahman MA, Holt D, Farajpour Y, Mammeri A, Khiabani H. Predictive modeling of energy demands for battery electric buses using real-world data. Energy Informatics. 2025;8(1):118. doi:10.1186/s42162-025-00564-y.
11. Wang Z, Zou L, Li P, Zhao L, Wu Z. Decadal analysis of spatial and temporal dynamics in energy consumption: insights from the Beijing Subway. Build Environ. 2025;282:113273. doi:10.1016/j.buildenv.2025.113273.
12. Zhao J, He J, Wang J, Liu K. Energy consumption prediction for electric buses based on traction modeling and LightGBM. World Electr Veh J. 2025;16(3):159. doi:10.3390/wevj16030159.
13. Lajunen A. Energy consumption and cost-benefit analysis of hybrid and electric city buses. Transp Res Part C Emerg Technol. 2014;38(3):1–15. doi:10.1016/j.trc.2013.10.008.
14. Vepsäläinen J, Ritari A, Lajunen A, Kivekäs K, Tammi K. Energy uncertainty analysis of electric buses. Energies. 2018;11(12):3267. doi:10.3390/en11123267.
15. Abdelaty H, Al-Obaidi AA, Mohamed M, Farag HEZ. Machine learning prediction models for battery-electric bus energy consumption in transit. Transp Res Part D Transp Environ. 2021;96(12):102868. doi:10.1016/j.trd.2021.102868.
16. Aziz A, Khan W, Yousaf MZ, Abdullah M, Khan RS, Farooq U, et al. Integrated energy scheduling for grid-connected microgrids using battery degradation-aware optimization and coordinated control strategies. Sci Rep. 2025;15(1):44033. doi:10.1038/s41598-025-28469-6.
17. Bracco S, Bianco G, Siri S, Barbagelata C, Casati C, Siri E. Simulation models for the evaluation of energy consumptions of electric buses in different urban traffic scenarios. In: 2021 Sixteenth International Conference on Ecological Vehicles and Renewable Energies (EVER). Piscataway, NJ, USA: IEEE; 2021. p. 1–6.

18. Belloni M, Tarsitano D, Sabbioni E. A comprehensive analysis of energy consumption in battery-electric buses using experimental data: impact of driver behavior, route characteristics, and environmental conditions. Electronics. 2025;14(4):735. doi:10.3390/electronics14040735.
19. Gallet M, Massier T, Hamacher T. Estimation of the energy demand of electric buses based on real-world data for large-scale public transport networks. Appl Energy. 2018;230:344–56. doi:10.1016/j.apenergy.2018.08.086.
20. Hjelkrem OA, Lervåg KY, Babri S, Lu C, Södersten CJ. A battery electric bus energy consumption model for strategic purposes: validation of a proposed model structure with data from bus fleets in China and Norway. Transp Res Part D Transp Environ. 2021;94:102804. doi:10.1016/j.trd.2021.102804.
21. Rajaperumal TA, Chinnappan CC. Integrating data-driven and physics-based approaches for robust wind power prediction: a comprehensive ML-PINN-Simulink framework. Sci Rep. 2025;15(1):29102. doi:10.1038/s41598-025-13306-7.
22. Meng C, Griesemer S, Cao D, Seo S, Liu Y. When physics meets machine learning: a survey of physics-informed machine learning. Mach Learn Comput Sci Eng. 2025;1(1):20.
23. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: a highly efficient gradient boosting decision tree. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 3149–57.
24. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189–232. doi:10.1214/aos/1013203451.
25. Liu Y, Liang H. A data-driven approach for electric bus energy consumption estimation. IEEE Trans Intell Transp Syst. 2022;23(10):17027–38. doi:10.1109/TITS.2022.3152679.
26. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 4765–74.
27. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM; 2016. p. 1135–44. doi:10.1145/2939672.2939778.
28. Nan S, Liao F, Li T, Chen H, Sun J. Identifying the electricity-saving driving behaviors of electric bus based on trip-level electricity consumption: a machine learning approach. Environ Sci Pollut Res. 2023;30(34):82743–59. doi:10.1007/s11356-023-28107-6.
29. Wang W, Mohamed M, Zhang B, Zhang Q, Abdelaty H. Electric bus energy prediction and factors interactions using explainable machine learning models. Eng Appl Artif Intell. 2026;163(26):112877. doi:10.1016/j.engappai.2025.112877.
30. Thai Meteorological Department. Climate data summary for Bangkok metropolitan area 2023. Bangkok, Thailand: TMD; 2023 [cited 2026 Apr 20]. Available from: https://www.tmd.go.th.
31. Kambly KR, Bradley TH. Estimating the HVAC energy consumption of plug-in electric vehicles. J Power Sources. 2014;259(1):117–24. doi:10.1016/j.jpowsour.2014.02.033.
32. Kammuang-lue N, Boonjun J. Energy consumption of battery electric bus simulated from international driving cycles compared to real-world driving cycle in Chiang Mai. Energy Rep. 2021;7:3267–72. doi:10.1016/j.egyr.2021.07.016.
33. TomTom. TomTom traffic index 2023. TomTom international BV; 2023 [cited 2026 Apr 20]. Available from: https://www.tomtom.com/traffic-index/.
34. Vepsäläinen J, Otto K, Lajunen A, Tammi K. Computationally efficient model for energy demand prediction of electric city bus in varying operating conditions. Energy. 2019;169(12):433–43. doi:10.1016/j.energy.2018.12.064.
35. Magnaval G, Boulay AM. Development of an analytical model of automobile energy consumption during use-phase for parametrized life cycle assessment. Renew Sustain Energ Rev. 2025;217(30):115716. doi:10.1016/j.rser.2025.115716.
36. Fiori C, Montanino M, Nielsen S, Seredynski M, Viti F. Microscopic energy consumption modelling of electric buses: model development, calibration, and validation. Transp Res Part D Transp Environ. 2021;98(3):102978. doi:10.1016/j.trd.2021.102978.

37. Dabčević Z, Škugor B, Cvok I, Deur J. A trip-based data-driven model for predicting battery energy consumption of electric city buses. Energies. 2024;17(4):911. doi:10.3390/en17040911.
38. Chen Y, Zhang Y, Sun R. Data-driven estimation of energy consumption for electric bus under real-world driving conditions. Transp Res Part D Transp Environ. 2021;98(4):102969. doi:10.1016/j.trd.2021.102969.
39. Pamuła T, Pamuła W. Estimation of the energy consumption of battery electric buses for public transport networks using real-world data and deep learning. Energies. 2020;13(9):2340. doi:10.3390/en13092340.
40. Li P, Zhang Y, Zhang Y, Zhang K, Jiang M. The effects of dynamic traffic conditions, route characteristics and environmental conditions on trip-based electricity consumption prediction of electric bus. Energy. 2021;218(6):119437. doi:10.1016/j.energy.2020.119437.
41. Blades LAW, Matthews T, McGrath TE, Early J, Cunningham G, Harris A. Predicting energy consumption of zero emission buses using route feature selection methods. Transp Res Part D Transp Environ. 2024;130:104158. doi:10.1016/j.trd.2024.104158.
42. Ma Y, Ye B, Wang S, Zhang Z. Accurate prediction of energy consumption of electric buses based on traffic condition, vehicle status, driving behavior, and environmental condition. IEEE Trans Transp Electrif. 2025;11(4):9778–92. doi:10.1109/TTE.2025.3556033.
43. Tian X, Wang B, Wang Z, Wan S, Peng H, An C. Unraveling energy demand in battery electric bus operations through an explainable machine learning approach using real-world cold-climate data. Energy. 2025;340:139256. doi:10.1016/j.energy.2025.139256.
44. Lyu A, Zhang H, Zhang Y, Li C, Chen Y. A study on energy consumption analysis and prediction of electric bus at intersections considering driving behavior. Sci Rep. 2025;15(1):44755. doi:10.1038/s41598-025-28835-4.
45. Dong C, Xiong Z, Li N, Yu X, Liang M, Zhang C, et al. A real-time prediction framework for energy consumption of electric buses using integrated machine learning algorithms. Transp Res Part E Logist Transp Rev. 2025;194:103884. doi:10.1016/j.tre.2024.103884.
46. Champahom T, Banyong C, Janhuaton T, Se C, Watcharamaisakul F, Ratanavaraha V, et al. Deep learning vs. Gradient boosting: optimizing transport energy forecasts in Thailand through LSTM and XGBoost. Energies. 2025;18(7):1685.
47. Xu N, Li X, Yue F, Jia Y, Liu Q, Zhao D. An eco-driving evaluation method for battery electric bus drivers using low-frequency big data. IEEE Trans Intell Transp Syst. 2023;24(9):9296–308. doi:10.1109/TITS.2023.3267187.
48. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. doi:10.1023/A:1010933404324.
49. Vapnik VN. The nature of statistical learning theory. New York, NY, USA: Springer; 1995.
50. Hussain I, Ching KB, Uttraphan C, Tay KG, Noor A. Evaluating machine learning algorithms for energy consumption prediction in electric vehicles: a comparative study. Sci Rep. 2025;15(1):16124. doi:10.1038/s41598-025-94946-7.
51. Ullah I, Liu K, Yamamoto T, Al Mamlook RE, Jamal A. A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: a path towards sustainability. Energy Environ. 2022;33(8):1583–612. doi:10.1177/0958305X211044998.
52. Kachalla IA, Ghiaus C, Baseer M. Comparative analysis of machine learning models for prediction and forecasting of electric water boilers energy consumption. Appl Therm Eng. 2025;267(2):125799. doi:10.1016/j.applthermaleng.2025.125799.
53. Kachalla IA, Ghiaus C, Ademuwagun A, Odeyinde OB, Baseer M. Data-driven hybrid SARIMAX-MLP framework for energy consumption prediction in residential micro-grid. Results Eng. 2025;26(2):105336. doi:10.1016/j.rineng.2025.105336.
54. Hai T, Ali AB, Agarwal D, Punia A, Jagga M, Anqi AE, et al. Predictive optimization using long short-term memory for solar PV and EV integration in relatively cold climate energy systems with a regional case study. Sci Rep. 2025;15(1):16414. doi:10.1038/s41598-025-01519-9.
55. Diouf B. Artificial intelligence in the analysis of energy consumption of electric vehicles. Energies. 2025;18(23):6338. doi:10.3390/en18236338.

56. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58(3):82–115. doi:10.1016/j.inffus.2019.12.012.
57. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access. 2018;6:52138–60. doi:10.1109/ACCESS.2018.2870052.
58. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56–67. doi:10.1038/s42256-019-0138-9.
59. Molnar C. Interpretable machine learning: a guide for making black box models explainable. 2019 [cited 2026 Apr 20]. Available from: https://christophm.github.io/interpretable-ml-book/.
60. Givisis I, Kalatzis D, Christakis C, Kiouvrekis Y. Comparing explainable AI models: SHAP, LIME, and their role in electric field strength prediction over urban areas. Electronics. 2025;14(23):4766. doi:10.3390/electronics14234766.
61. Xu J, Saleh M, Hatzopoulou M. A machine learning approach capturing the effects of driving behaviour and driver characteristics on trip-level emissions. Atmos Environ. 2020;224:117311. doi:10.1016/j.atmosenv.2020.117311.
62. Zhang Z, Ye B, Wang S, Ma Y. Analysis and estimation of energy consumption of electric buses using real-world data. Transp Res Part D Transp Environ. 2024;126:104017.
63. Kanti PK, Sharma P, Wanatasanappan VV, Said NM. Explainable machine learning techniques for hybrid nanofluids transport characteristics: an evaluation of SHAP and LIME. J Thermal Anal Calorimetry. 2024;149(21):11599–618. doi:10.1007/s10973-024-13639-x.
64. Devanathan B, Jnana Varshitha K, Pavan Kumar L, Lakshmanan SA, Krishna Prakash N. Explainable AI FRAMEWORK Using XGBoost with SHAP and LIME for multi-scale household energy forecasting. IEEE Access. 2025;13:149750–64. doi:10.1109/ACCESS.2025.3602673.
65. Corbet S, Larkin C, McCluskey J. The influence of inclement weather on electric bus efficiency: evidence from a developed european network. Case Stud Transp Policy. 2023;12(2):100971. doi:10.1016/j.cstp.2023.100971.
66. United Nations Economic and Social Commission for Asia and the Pacific. Study report on transitioning to electric public buses in Thailand. Bangkok, Thailand: United Nations Economic and Social Commission for Asia and the Pacific; 2023 [cited 2026 Apr 20]. Available from: https://www.unescap.org/sites/default/d8files/event-documents/ESCAP-2022-RP-Transitioning-to-Electric-Public-Buses-Thailand.
67. Janjamraj N, Changsarn C, Hiranvarodom S, Bhumkittipich K. The practical learning on electric bus conversion to support carbon neutrality policy in Thailand's transport sector. World Electric Veh J. 2025;16(3):181. doi:10.3390/wevj16030181.
68. Mohanty PK, Reddy KHK, Panigrahy SK, Sinha Roy D. Leveraging generative and explainable AI for electric vehicle energy toward sustainable, consumer-centric transportation. IEEE Access. 2024;12(12):143721–32. doi:10.1109/ACCESS.2024.3405959.
69. Li P, Zhang Y, Zhang Y, Zhang Y, Zhang K. Prediction of electric bus energy consumption with stochastic speed profile generation modelling and data driven method based on real-world big data. Appl Energy. 2021;298(17):117204. doi:10.1016/j.apenergy.2021.117204.
70. Kivekäs K, Lajunen A, Vepsäläinen J, Tammi K. City bus powertrain comparison: driving cycle variation and passenger load sensitivity analysis. Energies. 2018;11(7):1755. doi:10.3390/en11071755.
71. Ou S. Estimate long-term impact on battery degradation by considering electric vehicle real-world end-use factors. J Power Sources. 2023;573:233133.
72. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA: ACM; 2019. p. 2623–31.
73. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast. 2006;22(4):679–88. doi:10.1016/j.ijforecast.2006.03.001.
74. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res. 2005;30:79–82. doi:10.3354/cr030079.
75. Angelopoulos AN, Bates S. Conformal prediction: a gentle introduction. Found Trends Mach Learn. 2023;16(4):494–591.