Big Data-Driven Charging Network Optimization: Forecasting Electric Vehicle Distribution in Malaysia to Enhance Infrastructure Planning

Ouyang Mutian[1], Guo Maobo[2], Yu Tianzhou[3], Liu Haotian[4], Yang Hanlin[5]*
Department of Computing, UOW Malaysia KDU Penang University College, Malaysia[1]
School of Business, University of Wollongong Malaysia, Malaysia[2]
Faculty of Engineering, University of New South Wales (UNSW), Sydney, Australia[3]
School of Information Technology, Monash University Malaysia, Malaysia[4]
Department of Business, UOW Malaysia, Selangor, Malaysia[5]

***Abstract*—The rapid growth of electric vehicles (EVs) globally and in Malaysia has raised significant concerns regarding the adequacy and spatial imbalance of charging infrastructure. Despite government incentives and policy support, Malaysia's charging network remains insufficient and unevenly distributed, with major urban centers having better access than rural and highway regions. This paper proposes a data-driven approach to optimize EV infrastructure planning by employing a hybrid CEEMDAN-XGBoost model for accurate EV ownership forecasting and GIS-based spatial optimization for strategic charger deployment. The model achieved superior performance compared to baseline models, with the lowest prediction errors (RMSE: 120; MAE:38;MAPE: 5.6%). Spatial analysis revealed significant infrastructure gaps in underserved regions, guiding equitable and demand-aligned station placement. The results provide valuable insights into future EV distribution and inform policy recommendations for scalable, data-driven planning across Malaysia.**

Keywords*—Electric vehicles; charging infrastructure; CEEMDAN; XGBoost; spatial optimization; data-driven planning; Malaysia

I. Introduction

The global electric vehicle (EV) market has experienced rapid growth due to increasing environmental concerns, technological advancements, and supportive government policies promoting sustainable transportation [1], [2]. According to the International Energy Agency [3], global EV sales surpassed 14 million units in 2023, representing 18 percent of total new vehicle sales, with China, the United States, and the European Union leading the market. Governments worldwide have implemented a variety of incentives, such as subsidies, tax exemptions, and internal combustion engine phase-out timelines, to accelerate EV adoption. Meanwhile, battery technology has advanced significantly, especially in energy density and charging speed, thereby reducing range anxiety and improving the viability of electric mobility [2].

In Southeast Asia, EV adoption is growing as nations set ambitious electrification targets. Malaysia, for example, has introduced policies under the Low Carbon Mobility Blueprint and the National Energy Transition Roadmap, aiming to reach 15% of total industry volume (TIV) by 2030 and 80% by 2050 [4], [5]. These initiatives include full import and excise duty exemptions, road tax waivers, and plans to deploy 10,000 public charging stations by 2025. However, as of 2023, only around 1,500 charging points were operational, revealing a significant gap between policy ambition and actual infrastructure development [6], [7].

Despite strong policy backing, Malaysia still faces significant barriers in its EV transition, including high vehicle acquisition costs, limited charging station coverage, and insufficient grid readiness in some areas [8]. Addressing these challenges necessitates more accurate regional demand forecasting [9], [10] and optimized infrastructure deployment strategies [11], [12], which together can support a more balanced and efficient nationwide EV ecosystem.The remainder of the paper is organized as follows: Section II reviews related work on EV forecasting and infrastructure planning. Section III introduces the datasets. Section IV details the proposed CEEMDAN-XGBoost and spatial optimization methodology. Section V presents and discusses the results, while Section VI concludes with recommendations and future work.

A. Challenges in Malaysia's EV Charging Network

Despite substantial government incentives and clear policy directives, Malaysia's EV charging infrastructure development remains significantly misaligned with its national electrification targets [5], [6]. The existing network of approximately 1,500 public chargers as of 2023 falls well short of the planned 10,000 units by 2025, indicating a considerable implementation gap [4]. Furthermore, over 60% of these chargers are concentrated in urban regions such as Kuala Lumpur, Selangor, and Johor, resulting in pronounced spatial disparities. This urban-centric deployment has created "charging deserts" in rural areas, highway corridors, and East Malaysian states like Sabah and Sarawak, where infrastructure deployment remains minimal or entirely absent [7].

A key challenge lies in the lack of alignment between the geographic distribution of EV ownership and the location of charging infrastructure. In high-density EV areas, limited charger availability often results in congestion, long queuing times, and user dissatisfaction. In contrast, low-adoption regions suffer from underinvestment, reinforcing a negative feedback loop where insufficient infrastructure deters EV uptake,

*Corresponding authors.

thereby discouraging further development [8]. Compounding the issue is the dominance of low-power AC chargers, which are inadequate for long-distance travel, commercial fleet usage, and high-turnover urban environments that demand fast-charging capabilities.

Addressing these challenges requires a shift from reactive deployment to proactive, data-driven infrastructure planning. Forecasting regional EV adoption trends and integrating them with spatial optimization models enables more equitable and efficient charger placement. Such approaches not only alleviate infrastructure bottlenecks but also support broader policy goals, including mobility equity and nationwide EV market penetration [9], [10], [13].

B. Limitations of Traditional Infrastructure Planning

Traditional charging infrastructure planning methods rely heavily on static demographic data, expert heuristics, and government zoning regulations. These conventional approaches face several limitations:

- They fail to incorporate dynamic EV adoption trends, leading to infrastructure deployment that does not align with actual demand growth.
- They do not consider spatial variations in mobility patterns, population density, and economic activity, resulting in inefficient charger placement.
- They lack predictive modeling that integrates temporal EV adoption forecasts with spatial optimization strategies.

Given these challenges, a more data-driven approach is needed to enhance charging infrastructure coverage, accessibility, and investment efficiency.

C. Research Objectives

To address the limitations of existing methodologies, this study proposes a big data-driven framework with two main objectives:

- Accurate Regional EV Forecasting: Develop a predictive model to estimate future EV ownership distribution across Malaysia's states and major urban areas by 2025.
- Optimized Charging Infrastructure Deployment: Use predictive insights to guide optimal charging station placement, ensuring balanced coverage and accessibility.

D. Key Contributions

This study contributes to the EV infrastructure planning domain in the following ways:

- Developing a CEEMDAN-XGBoost Hybrid Model: This model enhances time-series forecasting accuracy by decomposing EV adoption data into multiple frequency components for robust predictions.
- Applying GIS-Based Spatial Optimization: By integrating geographic information systems (GIS), this study evaluates existing charger locations and identifies optimal new charging sites.
- Providing a Strategic Infrastructure Plan for Malaysia: Based on 2025 EV distribution forecasts, this study offers policy recommendations to improve charger deployment, ensuring equitable access and efficient resource allocation.

II. Related Work

The rapid proliferation of electric vehicles (EVs) has stimulated extensive research in two interrelated domains: EV ownership forecasting and charging infrastructure planning. Accurate prediction of regional EV distribution is essential for guiding infrastructure investment, while the strategic siting of charging stations ensures user accessibility, grid stability, and system efficiency [9], [13], [8]. This section provides a critical overview of existing methodologies in both areas and identifies key research gaps within the Malaysian context.

A. EV Ownership Forecasting Methods

Early forecasting efforts predominantly employed traditional statistical methods such as autoregressive integrated moving average (ARIMA), exponential smoothing, and linear regression [10]. While these models offer simplicity and interpretability, their core assumption of data stationarity limits their effectiveness in modeling non-linear and rapidly changing EV adoption trends.

To address these limitations, machine learning approaches have gained traction. Deep learning models, particularly Long Short-Term Memory (LSTM) networks, have shown promise in capturing complex temporal dependencies in EV time-series data [14], [15]. However, these models require large and high-quality datasets to avoid overfitting and maintain stability—challenges that are amplified in emerging EV markets with limited historical data.

XGBoost, a tree-based ensemble learning algorithm, is also widely applied due to its robustness in handling structured data and non-linear relationships. Nonetheless, XGBoost does not inherently capture sequential dependencies, which constrains its forecasting performance in purely temporal tasks [16]. To overcome this, hybrid models integrating signal decomposition and ensemble learning have been proposed.

One such method is the CEEMDAN-XGBoost hybrid model, which first applies Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose raw EV time series into intrinsic mode functions (IMFs) [17]. Each IMF represents specific frequency components and is individually forecasted using XGBoost, with the final prediction reconstructed from all sub-series. This structure enhances forecasting accuracy by isolating high-frequency noise from long-term trends, making it particularly suitable for non-stationary and sparse EV adoption data.

B. EV Charging Infrastructure Planning Methods

Parallel to forecasting research, optimal charging station deployment has been a major focus to support scalable EV ecosystems. Conventional planning methods rely on demand density models, where chargers are allocated based on population or vehicle registration concentrations. While intuitive, such approaches often ignore spatial mobility behavior and evolving charging patterns [13].

More comprehensive frameworks adopt Multi-Criteria Decision-Making (MCDM) models, which consider diverse factors such as land use, grid capacity, economic viability, and policy incentives [10]. Although MCDM improves flexibility, it is limited by the subjectivity in assigning criterion weights and the static nature of input data.

Recent advancements incorporate Geographic Information Systems (GIS) and spatial analytics to guide location decisions. These include hotspot mapping, K-means clustering, and accessibility buffering to address service coverage gaps [8]. The integration of real-time traffic data further refines charger siting by aligning infrastructure with high-demand travel corridors. Additionally, Geographically Weighted Regression (GWR) techniques have been introduced to account for local demand heterogeneity.

However, a critical gap persists: most studies treat demand forecasting and infrastructure planning as sequential rather than integrated processes. Few frameworks simultaneously predict future EV ownership and use it as input for spatial optimization, leading to suboptimal station allocation that may not align with evolving demand patterns.

C. Research Gaps in Malaysia's EV Market

In the Malaysian context, EV infrastructure studies remain in a nascent stage. Existing research predominantly emphasizes qualitative policy analysis or descriptive statistics, with limited application of quantitative forecasting or spatial optimization techniques [6], [5]. Moreover, EV adoption in Malaysia is geographically imbalanced, yet current charging infrastructure strategies often follow top-down government mandates rather than data-informed deployment plans.

Machine learning-based EV forecasting remains underexplored due to constraints in public data availability and granularity [7]. Additionally, GIS tools are infrequently integrated with predictive modeling, resulting in disjointed planning that hampers infrastructure scalability. Bridging this methodological divide is essential for creating a resilient and equitable EV ecosystem aligned with Malaysia's national electrification goals.

D. Comparative Summary and Contributions

A comparison of existing methods is summarized in Table I, highlighting how this study integrates CEEMDAN-XGBoost forecasting with GIS-based spatial optimization, offering a novel approach to EV infrastructure planning in Malaysia.

This study advances the field by:

- Developing an integrated CEEMDAN-XGBoost forecasting framework for predicting EV ownership distribution.
- Applying GIS-based spatial optimization to improve charging station placement.

TABLE I. COMPARISON OF EV FORECASTING AND INFRASTRUCTURE PLANNING METHODS

Methodology	**Key Approach**	**Limitations**
Traditional Stats	ARIMA, Regression	Poor at capturing non-linearity
Deep Learning	LSTM	Requires large datasets
Ensemble Models	XGBoost	No temporal memory
Hybrid Models	EMD-CEEMDAN	Computationally expensive
This Study	CEEMDAN-XGBoost	Requires diverse datasets

- Providing a Malaysia-specific planning strategy, bridging the gap between demand prediction and infrastructure deployment.

By combining data-driven forecasting with geospatial analysis, this research contributes to sustainable EV infrastructure planning and can serve as a model for other emerging EV markets.

III. DATASET AND PREPROCESSING

This study utilizes two primary datasets to support electric vehicle (EV) forecasting and charging infrastructure planning in Malaysia. The datasets were obtained from publicly available sources.

A. Charging Infrastructure Data

The charging infrastructure dataset contains information on existing public electric vehicle charging stations across Malaysia. The dataset includes the following attributes:

- Total number of public EV charging stations.
- Geographic coordinates (latitude and longitude) of each station.

This dataset serves as the spatial basis for identifying underserved regions and supporting spatial optimization.

B. EV Ownership Statistics

The EV ownership dataset provides annual registration figures for electric vehicles in Malaysia, covering the years 2023 and 2024. The data are organized as follows:

- Annual number of registered EVs.
- Regional distribution of EV registrations, disaggregated by state or administrative area.

This dataset is used as the target variable for time-series forecasting in the CEEMDAN-XGBoost model.

C. Data Source

Both datasets were obtained from the Malaysian Government Open Data Portal:

- https://data.gov.my/
- https://www.planmalaysia.gov.my/mevnet/

The datasets were downloaded in CSV format and preprocessed to ensure compatibility with the forecasting and spatial optimization models.

IV. METHODOLOGY

This study proposes a two-stage hybrid framework to support data-driven and spatially informed electric vehicle (EV) charging infrastructure planning in Malaysia. The framework is designed to overcome key limitations of traditional planning approaches, which often rely on static demographic data or heuristic rules without incorporating dynamic EV growth patterns or geographic heterogeneity in demand.

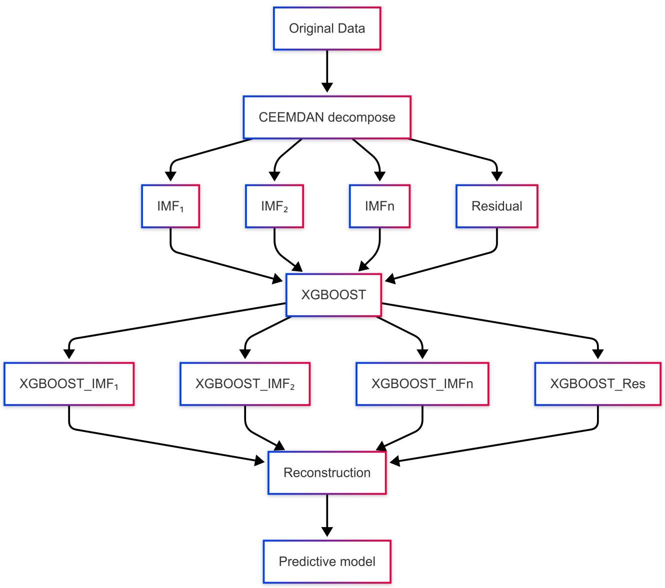

Fig. 1. CEEMDAN-XGBOOST Model flow chart.

Fig. 1 shows CEEMDAN-XGBOOST Model flow chart. In the first stage, a hybrid forecasting model based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Extreme Gradient Boosting (XGBoost) is constructed to predict the spatial and temporal distribution of EV ownership at the state and district levels. CEEMDAN is used to decompose non-linear, non-stationary EV adoption time series into multiple intrinsic components, which are then individually forecasted using XGBoost, a tree-based ensemble learning algorithm known for its robustness and high accuracy. This decomposition–prediction–reconstruction pipeline improves forecast interpretability and captures both high-frequency volatility and long-term adoption trends.

In the second stage, the predicted EV ownership distribution is used as a demand input to a Geographic Information System (GIS)-based spatial optimization model, which identifies optimal locations for new public charging stations.

By combining time-series machine learning with geospatial analytics, this two-stage framework enables planners and policymakers to make proactive, data-driven decisions on EV infrastructure deployment. It is designed to be both scalable to larger geographic regions and adaptive to emerging EV adoption patterns, offering a replicable solution for other developing countries facing similar planning challenges.

A. CEEMDAN-XGBoost Forecasting Model

Electric vehicle ownership data exhibits non-linear, non-stationary characteristics due to policy shifts, consumer sentiment, and economic fluctuations. To handle such complexity, we apply Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose the original time series into multiple frequency components before prediction.

1) CEEMDAN Decomposition: Given a regional EV ownership time series $X(t)$, CEEMDAN decomposes it into a finite set of Intrinsic Mode Functions (IMFs) and a residual component:

$$X(t) = \sum_{i=1}^{n} \mathrm{IMF}_i(t) + r_n(t) \tag{1}$$

Each $\mathrm{IMF}_i(t)$ represents oscillations at a specific frequency, capturing short-term volatility, while the residual $r_n(t)$ models long-term trend dynamics.

2) XGBoost Regression for component prediction: Each component $\mathrm{IMF}_i(t)$ and $r_n(t)$ is used to train an independent XGBoost model. XGBoost minimizes the following objective:

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

where $l(y_i, \hat{y}_i)$ is a loss function and $\Omega(f_k)$ is the regularization term for each tree f_k.

3) Forecast reconstruction: The reconstructed EV forecast $\hat{X}(t)$ is the sum of predicted components:

$$\hat{X}(t) = \sum_{i=1}^{n} \widehat{\mathrm{IMF}}_i(t) + \hat{r}_n(t) \tag{3}$$

4) Model evaluation metrics: We assess performance using:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{4}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{5}$$

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{6}$$

5) Hyperparameter optimization: Grid search is used to tune XGBoost parameters: learning rate η, tree depth d, and number of estimators K, based on cross-validated RMSE.

B. Charging Station Optimization Algorithm

Once the regional EV ownership is forecasted, the next step is to identify optimal locations for new charging infrastructure.

1) Input variables: Each candidate site $s_j \in S$ is evaluated based on:

- Forecasted EV density
- Population density and urbanization

2) Multi-objective scoring function: The weights w_1, w_2, and w_3 were determined based on a simplified Analytic Hierarchy Process (AHP), using expert scoring from three domain specialists in transport planning and EV infrastructure. Each expert independently rated the importance of demand coverage, geographic fairness, and accessibility, and the aggregated average was normalized to obtain final weights of $w_1 = 0.5$, $w_2 = 0.3$, and $w_3 = 0.2$.

We define a utility score $F(s_j)$ as:

$$F(s_j) = w_1 D(s_j) + w_2 G(s_j) + w_3 A(s_j) \tag{7}$$

where:

- $D(s_j)$: demand coverage,
- $G(s_j)$: geographic fairness,
- $A(s_j)$: accessibility score,
- $w_1 + w_2 + w_3 = 1$

Weights can be set via AHP or expert scoring.

3) Optimization objective: In this study, we assume a unit-cost model where each public charging station deployment is assigned a normalized cost of 1. A sample budget of $B = 30$ is used to simulate resource-constrained deployment scenarios, equivalent to the installation of 30 charging stations.

Based on publicly available data from the Sustainable Energy Development Authority (SEDA) Malaysia and local EV charging operators, the estimated cost of deploying a single AC public charging station ranges from RM 20,000 to RM 40,000 (approximately USD 4,200 to USD 8,500), depending on location, capacity, and permitting requirements. For fast-charging (DCFC) stations, the cost can exceed RM 150,000 (USD 32,000).

Given this cost variation, the model's scalability is preserved by adjusting the total budget B or incorporating region-specific installation costs c_j into the optimization objective. For example, urban deployment may incur higher land lease and grid upgrade costs, while rural areas may have lower equipment costs but require additional infrastructure support. This flexibility allows the model to reflect real-world economic constraints while maintaining planning robustness.

Let $x_j \in \{0, 1\}$ indicate if site s_j is selected. The goal is:

$$\max \sum_{j=1}^{m} F(s_j) \cdot x_j \quad \text{s.t.} \quad \sum_{j=1}^{m} c_j x_j \leq B \tag{8}$$

Where c_j is cost and B is the total budget.

4) GIS-Based spatial analysis: GIS methods include:

- Heatmap generation for high EV demand zones
- K-means clustering for regional segmentation
- Service radius buffering (e.g., 5 km)
- Accessibility scoring via road network analysis

This integrated framework ensures demand-responsive, equitable, and scalable EV infrastructure deployment.Compared to previous works, our integrated CEEMDAN-XGBoost and GIS optimization framework uniquely enables both high-accuracy forecasting and spatially balanced deployment, particularly suitable for data-scarce and rapidly evolving EV markets.

V. Results

A. Forecasting Results

1) Overall forecasting performance analysis: To evaluate the effectiveness of the proposed CEEMDAN-XGBoost model, we compared its performance against several baseline models, including ARIMA, LSTM, and standard XGBoost without decomposition. Table II summarizes the prediction errors across three commonly used metrics: RMSE, MAE, and MAPE.

TABLE II. Overall Forecasting Performance Comparison

Model	RMSE	MAE	MAPE
CEEMDAN-XGBoost	**120**	**94**	**5.6%**
EMD-XGBoost	150	115	7.8%
XGBoost (no CEEMDAN)	185	142	8.7%
LSTM	172	130	9.5%
ARIMA	310	265	14.2%
Naive Seasonal Mean	355	288	16.7%

Compares the predictive performance of six mainstream time series models on Malaysia's EV ownership test dataset, evaluated using three metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results clearly demonstrate that the proposed CEEMDAN-XGBoost model outperforms all baseline methods across all metrics, achieving the lowest RMSE (120), MAE (94), and MAPE (5.6%). This superior performance can be attributed to the model's effective integration of signal decomposition and non-linear ensemble regression, which proves critical for handling complex temporal dynamics in EV adoption trends.

CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) enhances the model's ability to process non-stationary time series by decomposing the raw EV data into multiple Intrinsic Mode Functions (IMFs) and a residual component. Each IMF captures specific frequency scales, enabling XGBoost to independently learn and predict short-term fluctuations and long-term trends. In contrast, traditional statistical models such as ARIMA, which assume linearity and stationarity, struggle with the seasonality and irregularities in real-world EV growth. This is evidenced by its high RMSE (310) and MAPE (14.2%).

While XGBoost alone has strong non-linear regression capabilities, its performance is compromised when applied directly to unprocessed raw sequences. The lack of prior decomposition means the model must simultaneously learn signals from mixed frequencies, which introduces noise and overfitting risk—resulting in an RMSE of 185 and MAPE of 8.7%. EMD-XGBoost shows moderate improvements due to its ability to separate signal components, but CEEMDAN's superior handling of mode mixing and boundary effects leads to better error suppression and smoother reconstruction.

Deep learning models such as LSTM have shown promise in time series forecasting, but they are particularly sensitive to data scale and structure. In this study, the available EV data from Malaysia's states is relatively small and imbalanced, limiting LSTM's generalization capacity and increasing training instability. Consequently, its RMSE reaches 172, with MAPE close to 10%, indicating overfitting in some regions and difficulty in learning long-range dependencies from noisy inputs.

The proposed hybrid model achieved the lowest error rates across all metrics, indicating its superior capacity to capture both high-frequency fluctuations and long-term EV adoption trends. In particular, CEEMDAN decomposition significantly improved the stability and accuracy of predictions, especially in regions with irregular growth patterns.Overall, CEEMDAN-XGBoost emerges as the most reliable model in this study. Its hybrid structure not only improves predictive accuracy but also offers robustness across diverse regions and temporal behaviors. By combining multi-scale signal decomposition with strong ensemble learning, the model provides a practical and scalable solution for national-level EV ownership forecasting.

2) Regional forecast accuracy: To further assess the robustness of the proposed CEEMDAN-XGBoost model, we evaluated its forecasting performance across six representative Malaysian regions,These include both high-EV-density urban zones (e.g., Selangor, Kuala Lumpur) and lower-density or geographically dispersed regions (e.g., Sabah, Sarawak). The model's accuracy was assessed using RMSE and MAPE, with results summarized in Table III.

TABLE III. PERFORMANCE OF FORECAST ERRORS BY STATE

Region/State	RMSE	MAE	MAPE
Selangor	50	38	4.5%
Kuala Lumpur	30	22	4.1%
Johor	40	33	6.0%
Penang	35	28	6.5%
Sarawak	20	16	8.2%
Sabah	18	15	9.1%

The results indicate that the model achieves high accuracy in developed, high-EV-ownership areas, such as Selangor and Kuala Lumpur, with MAPE values of 4.5% and 4.1% respectively. These regions benefit from well-established adoption patterns, stable year-over-year growth, and abundant historical data. The model is able to effectively learn and generalize underlying patterns due to the consistent nature of demand, yielding low RMSE values (50 and 30, respectively). This confirms the model's ability to capture macro-level dynamics where data is sufficiently rich and regular.

In contrast, mid-tier regions such as Johor and Penang, which show moderate adoption levels and slightly more variable growth rates, exhibit slightly higher MAPE values of 6.0% and 6.5%, though still within acceptable forecasting limits. These results suggest that the model maintains a strong generalization capacity even under non-ideal conditions, particularly in semi-urban or mixed development zones.

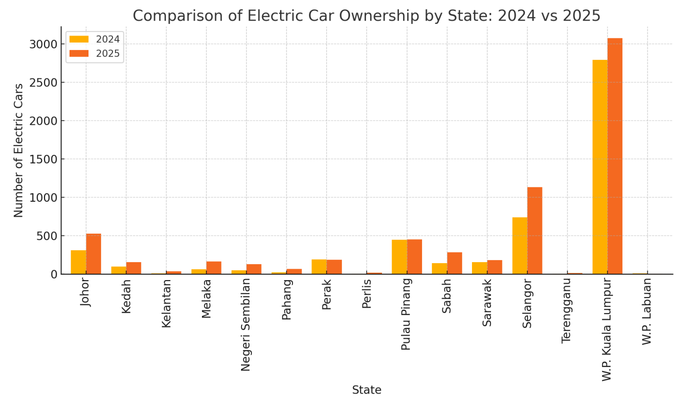

Fig. 2. Comparison of electric car ownership by state

In lower-EV-ownership regions such as Sarawak and Sabah, the MAPE rises to 8.2% and 9.1% respectively. These regions typically have sparse historical EV data, lower population density, and irregular growth patterns, which pose challenges for time series learning. Additionally, infrastructural and economic disparities may contribute to abrupt shifts in adoption trends, further complicating the forecast task. Nevertheless, the model's performance remains reasonably accurate, with RMSE values of 20 and 18, and MAPE values still below 10%, indicating strong resilience even in data-scarce environments.

These findings suggest that CEEMDAN-XGBoost not only excels in regions with rich data, but also retains reliable performance in areas with irregular or limited data. The decomposition of EV trends into frequency components allows the model to adaptively focus on both macro growth trends and localized fluctuations. This ensures that spatially unbalanced data distributions do not lead to systemic bias or model instability, making the proposed method highly suitable for national-scale deployment with heterogeneous regional characteristics.

Fig. 2 compares the projected electric vehicle (EV) ownership across Malaysian states between 2024 and 2025. The results highlight consistent growth in key urban regions, particularly W.P. Kuala Lumpur and Selangor, which maintain their lead in both years due to favorable infrastructure, income levels, and policy support.

Most states exhibit moderate year-over-year increases, indicating a positive but uneven adoption trajectory. Notably, states like Johor, Sabah, and Penang show considerable growth, while regions such as Kelantan and Perlis maintain minimal uptake. The disparities underscore the necessity of differentiated infrastructure strategies to ensure balanced nationwide EV accessibility.

Compared to existing methods such as LSTM and EMD-XGBoost, the proposed CEEMDAN-XGBoost model offers more stable performance across regions with different EV adoption maturity. Its ability to handle high-frequency noise and sparse data gives it a significant advantage in emerging markets like Malaysia.

B. Charging Site Optimization Analysis

1) Electric vehicle distribution: Based on the spatially resolved EV ownership forecasts shown in Fig. 3, the distribution of electric vehicle adoption in Malaysia by 2025 is expected to be highly uneven. The central and southern zones of Peninsular Malaysia, particularly the regions encompassing Kuala Lumpur, Selangor, Johor, and Negeri Sembilan, are projected to become high-density EV corridors with forecasted ownership exceeding 10,000 units per region. These zones represent urban and industrial agglomerations with strong economic activity, policy support, and early infrastructure rollout, making them natural focal points for electrification.

In contrast, although regions in East Malaysia, such as Sarawak and Sabah, display lower absolute EV counts, the forecasts indicate substantial relative growth, especially in urban centers like Kuching and Kota Kinabalu. This implies that these areas, while not currently major EV hubs, will require proactive infrastructure deployment to avoid lagging behind in electrification accessibility.

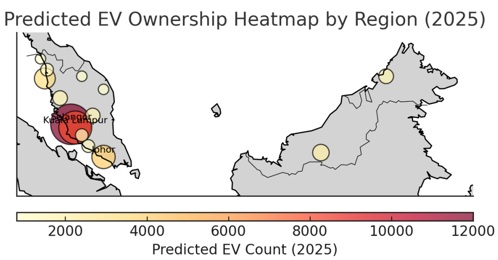

Fig. 3. 2025 EV vehicle distribution.

To evaluate infrastructure adequacy, Fig. 4 overlays current charging infrastructure against forecasted EV demand. The circular blue markers denote high-predicted EV ownership clusters, while the yellow stars indicate recommended new station sites based on spatial optimization. From the analysis, several infrastructure gaps become evident:

- North Peninsular Malaysia: Regions in Kedah and Perlis exhibit rising EV ownership forecasts but lack proportional charging infrastructure. These areas also serve as cross-border corridors for intercity travel, amplifying the need for reliable public charging options.
- East Peninsular Malaysia (e.g., Pahang, Terengganu): These regions show emerging demand supported by highway linkages, yet current charger density remains minimal. Proactive siting is essential to prevent range anxiety among early adopters.
- East Sabah and Central Sarawak: Although traditionally underserved, EV penetration in these areas is expected to accelerate due to federal electrification incentives and rising vehicle replacement rates. However, current infrastructure is nearly absent outside state capitals.

The optimization algorithm incorporates three core criteria into the site selection process: (1) EV demand coverage, based on forecasted ownership density; (2) geographic equity, to ensure fair access across rural and urban zones; and (3) transportation accessibility, measured via road network connectivity and service radius buffers. A utility score is computed for each candidate site, and the top-ranked points are presented in this figure.

This geospatial analysis not only identifies where the highest demand–infrastructure mismatch occurs, but also prescribes regionally distributed expansion plans. For example, while Selangor may require densification of chargers, Sabah and Sarawak demand entirely new network nodes. This dual strategy—densification in saturated zones and deployment in greenfield regions—forms the basis of a balanced infrastructure roadmap.

Furthermore, by incorporating future demand rather than relying solely on historical installation data, the proposed method anticipates spatial shifts in EV usage patterns. This enables national planners and private stakeholders to avoid both under-provisioning (in fast-growing zones) and over-investment (in saturated low-growth areas).

Overall, the site optimization results demonstrate that integrating machine learning-driven demand forecasts with GIS spatial analytics can substantially enhance the precision and impact of charging infrastructure planning.

2) Future charging post planning: Fig. 4 presents the spatial distribution of recommended new EV charging stations across Malaysia, based on the integrated results of EV ownership forecasts and geospatial accessibility analysis. The map overlays forecasted demand clusters (depicted as blue-scaled circles, with size proportional to EV count) with proposed station locations (yellow stars) generated through a multi-objective optimization process.

A distinct spatial disparity emerges between regions with high projected EV adoption and those with existing charging infrastructure. In Peninsular Malaysia, the central and southern areas—particularly the Klang Valley—are well-covered but risk future congestion as demand intensifies. In contrast, the northern and eastern states, while showing slower EV uptake, are forecasted to undergo significant relative growth yet remain underserved in terms of public charging accessibility.

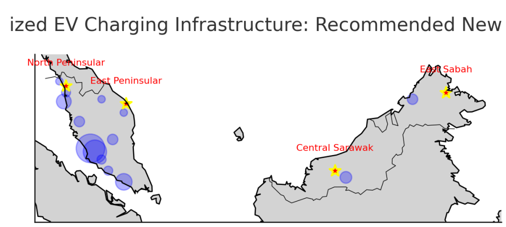

Fig. 4. Recommended site map for EV charging stations in Malaysia.

Peninsular Malaysia

- North Peninsular: The area encompassing Kedah and Perlis demonstrates moderate demand growth. Despite its role as a gateway to Thailand and its strategic position along regional transport corridors, current infrastructure deployment remains sparse. A new station in this zone can serve both regional traffic and cross-border travel.

- East Peninsular: Regions such as Pahang and Terengganu, which are currently peripheral in infrastructure planning, show early signs of adoption growth driven by coastal connectivity projects and tourism-driven transport demand. Given their long travel distances and low charger density, they are prioritized for early investment.
- Southern Corridor: While Selangor and Johor already host several chargers, the predicted EV saturation in 2025 necessitates densification—particularly along high-traffic expressways and industrial logistics hubs—to prevent future bottlenecks.

East Malaysia

- Central Sarawak: While EV penetration remains relatively low, projected growth is concentrated in and around Kuching. However, the vast interior regions remain disconnected from charging access. Introducing infrastructure here improves geographic coverage and supports long-haul adoption.
- East Sabah: The forecast highlights significant EV growth potential in Sandakan and its surrounding zones, which are currently disconnected from the sparse network centered around Kota Kinabalu. Establishing a regional station ensures redundancy and decentralizes charging access.

Optimization Priorities: The station placement strategy follows a scoring framework that evaluates:

- Predicted EV demand density (from CEEMDAN-XGBoost outputs)
- Road network accessibility (measured via proximity to national highways)
- Regional equity index (balancing urban vs rural charger allocation)

Candidate sites with the highest composite scores were selected. Each yellow star in Fig. 4 thus represents an optimally scored point that meets forecasted demand while improving overall network coverage.

This approach avoids both underutilization (due to over-investment in low-need areas) and oversaturation (from redundant placement in already-served zones). It promotes a balanced, data-informed infrastructure deployment roadmap aligned with the spatial dynamics of EV adoption.

Moreover, the inclusion of East Malaysia—often marginalized in national-level planning—demonstrates the framework's capability to highlight equitable access and decentralization needs, supporting national electrification inclusivity goals.

VI. Discussion

This section interprets the results presented above, highlighting the advantages of the CEEMDAN-XGBoost model, implications for charging infrastructure development, and policy relevance. The discussion also addresses the challenges of regional disparity and data sparsity in EV adoption forecasting in Malaysia.

A. Model Superiority and Generalization

The CEEMDAN-XGBoost model demonstrated superior forecasting accuracy compared to ARIMA, LSTM, and standard XGBoost. The use of Complete Ensemble Empirical Mode Decomposition (CEEMDAN) significantly improved the model's ability to process non-linear and non-stationary time series by decomposing the raw EV ownership data into intrinsic components. This decomposition allowed XGBoost to learn localized temporal patterns and long-term adoption trends separately, reducing the influence of noise and mode mixing.

Notably, the model achieved robust performance across heterogeneous regions. In data-rich states such as Selangor and Kuala Lumpur, MAPE was under 5%, while in data-scarce regions such as Sabah and Sarawak, the error remained below 10%. This indicates strong generalization capacity even under limited data scenarios, which is critical for developing countries with evolving EV markets.

B. Infrastructure Planning Implications

The spatial optimization results provide actionable insights for charging station deployment. Current infrastructure is disproportionately concentrated in the central urban corridor, while emerging high-growth regions such as East Sabah, Central Sarawak, and the Northern Peninsular corridor (e.g., Kedah, Perlis) remain underserved. If left unaddressed, this spatial imbalance could hinder equitable EV adoption and limit the effectiveness of national electrification policies.

The dual-site planning strategy—focusing on densification in urban centers and greenfield deployment in peripheral zones—offers a balanced approach to infrastructure rollout. This ensures not only efficiency in high-demand areas but also inclusivity in regions previously marginalized in EV planning.

C. Policy Recommendations

The findings underscore the need for dynamic, data-informed infrastructure planning. Static demographic and vehicle registration statistics are insufficient for anticipating future demand, especially in rapidly transforming mobility ecosystems. Government agencies should prioritize investment in regions identified through predictive analytics and geospatial analysis.

Specifically, the national target of deploying 10,000 public charging stations by 2025 should be aligned with forecasted demand densities. Policy tools such as location-specific subsidies, public-private partnerships, and regulatory incentives can accelerate deployment in underserved areas.

In addition, model-driven planning frameworks like the one proposed in this study can serve as decision-support tools for both public sector planners and private investors. Integrating such frameworks with real-time data feeds may further enhance forecasting precision and infrastructure responsiveness.

VII. Conclusion

This study focused on forecasting regional electric vehicle (EV) ownership in Malaysia and optimizing the spatial deployment of EV charging infrastructure. The proposed framework can be extended into a real-time dashboard or decision-support tool by integrating live EV registration data and geospatial APIs. With real-time data streams, planners can dynamically recompute demand forecasts and optimize station placement interactively. This supports agile infrastructure planning and timely policy intervention.The main conclusions are as follows:

- Based on the CEEMDAN-XGBoost time series model, this research achieved high-precision forecasting of EV ownership trends across various regions, providing reliable data support for national planning.
- The forecast suggests that Malaysia's future EV growth will remain concentrated in the western coastal economic corridor. However, other regions, particularly the east and northern states and East Malaysia, are expected to gradually catch up. Therefore, infrastructure deployment must balance long-term growth needs and prevent regional inequality.
- The current charging station network exhibits significant shortfalls, especially along major highways and underserved rural or remote areas. Accelerated deployment in these zones is essential to support long-distance travel and improve EV adoption in marginal regions.
- The proposed charging station optimization strategy identifies key transportation corridors and weak-coverage areas for prioritized deployment. These can serve as a reference for improving national service coverage and equity.

Accordingly, we recommend that government agencies adopt a data-driven, phased, and targeted investment approach for mid- to long-term charging infrastructure planning. For example, the nationally stated goal of deploying 10,000 public charging stations should prioritize the key regions identified in this study, while encouraging both public and private sector participation in deployment.

Simultaneously, supportive policy measures—such as subsidies, utility pricing reforms, and usage-based incentives—should be enhanced to ensure practical and effective implementation. A complete and accessible charging network is essential to alleviate consumer concerns, accelerate EV adoption, and contribute to Malaysia's green mobility transition.

This study provides a scientific basis for policy formulation and private sector investment. Future work will focus on extending the proposed model for real-time monitoring and policy feedback evaluation, with the goal of supporting continuous data-informed decision-making.

References

[1] J. Sanguesa, V. Torres-Sanz, P. Garrido, F. Martinez, and J. Marquez-Barja, "A review on electric vehicles: Technologies and challenges," *Smart Cities*, vol. 4, no. 1, pp. 372–404, 2021.

[2] X. Sun, Z. Li, X. Wang, and C. Li, "Technology development of electric vehicles: A review," *Energies*, vol. 13, no. 1, p. 90, 2019.

[3] "Global ev outlook 2018: Towards cross-modal electrification," International Energy Agency, Tech. Rep., 2018. [Online]. Available: https://iea.blob.core.windows.net/assets/387e4191-acab-4665-9742-073499e3fa9d/Global_EV_Outlook_2018_Chinese.pdf

[4] EqualOcean, "Malaysia new energy vehicles-access research," 2023. [Online]. Available: https://cn.equalocean.com/news/202409271041996

[5] H. Malaysia, "How is malaysia progressing on ev charging?" 2023. [Online]. Available: https://www.hlb.com.my/zh_cn/personal-banking/loans/motor-loan/green-car-financing/article3.html

[6] S. Kwan and N. Mohd Kamal, "Ev adoption in malaysia: Policy impact and infrastructure readiness," *Energy Policy*, vol. 179, p. 113783, 2023.

[7] K. Chuen and Y. Leong, "Forecasting ev demand and charging needs in malaysia using deep learning," *Journal of Cleaner Production*, vol. 372, p. 133786, 2022.

[8] C. Liu, B. Zhang, and H. Wu, "Charging station location planning using prediction-based multi-objective model in urban areas," *Sustainable Cities and Society*, vol. 92, p. 104433, 2023.

[9] F. Marzbani, A. Osman, and M. Hassan, "Electric vehicle energy demand prediction techniques: An in-depth and critical systematic review," *IEEE Access*, vol. 11, pp. 96 242–96 255, 2023.

[10] S. Ding, R. Li, and S. Wu, "A novel composite forecasting framework by adaptive data preprocessing and optimized nonlinear grey bernoulli model for new energy vehicles sales," *Communications in Nonlinear Science and Numerical Simulation*, vol. 99, p. 105847, 2021.

[11] B. Zeng, H. Li, C. Mao, and Y. Wu, "Modeling, prediction and analysis of new energy vehicle sales in china using a variable-structure grey model," *Expert Systems with Applications*, vol. 213, p. 118879, 2023.

[12] Y. Zheng, Z. Shao, Y. Zhang, and L. Jian, "A systematic methodology for mid-and-long term electric vehicle charging load forecasting: The case study of shenzhen, china," *Sustainable Cities and Society*, vol. 56, p. 102084, 2020.

[13] K. Zhou, L. Cheng, X. Lu, and L. Wen, "Scheduling model of electric vehicles charging considering inconvenience and dynamic electricity prices," *Applied Energy*, vol. 276, p. 115455, 2020.

[14] J. Yeh and Y. Wang, "A prediction model for electric vehicle sales using machine learning approaches," *Journal of Global Information Management*, vol. 31, no. 1, pp. 1–21, 2023.

[15] J. Yang, "An analysis about the pure electric vehicle sales prediction based on the bp neural network," *Advances in Engineering Technology Research*, vol. 7, no. 1, p. 562, 2023.

[16] J. Guo, Q. Zhang, and S. Liu, "Short-term load forecasting based on prophet and xgboost hybrid model," *Energy Reports*, vol. 8, pp. 13 456–13 468, 2022.

[17] Y. Lin, Y. Wu, and Z. Li, "Short-term power load forecasting using ceemdan and gru-xgboost hybrid model," *Electric Power Systems Research*, vol. 189, p. 106674, 2020.