Prediction of Installed Power Capacity in EV Charging Stations Through Urban Mobility and Infrastructure Indicators via Machine Learning

Abdullah Haşim Atınç[1], Nuğman Sağır[2], and Zeynep Hasırcı Tuğcu[3]

[1]Karadeniz Technical University, Department of Electronics and Automation, Trabzon, Türkiye
abdullahhasimatinc@ktu.edu.tr
[2]Karadeniz Technical University, Department of Electronics and Communication Engineering, Trabzon, Türkiye
nugmansagir@ktu.edu.tr
[3] Karadeniz Technical University, Department of Electrical and Electronics Engineering, Trabzon, Türkiye
zhasirci@ktu.edu.tr

Abstract

This study proposes a machine learning-based framework for forecasting cumulative installed power demand based on urban mobility and infrastructure indicators. The dataset includes features such as population density, number of motorvehicles (MVs), electric vehicles (EVs), electric vehicle charging stations (EVCSs) and location information normalized using a min-max scaling technique to ensure fair contribution during learning. An extensive hyperparameter optimization is performed to fine-tune each model for improved accuracy and generalizability. GPR showed the highest prediction performance by reaching the highest R^2 values and the lowest RMSE, MAE, and MSE values, emphasizing continuity. of the regularity of nonlinear relationships. ANN closely follows the model, providing a well-balanced generalization. The comparative analysis of the model outputs and input feature distributions reveal that the proposed approach can provide robust and accurate power demand forecasts and support efficient planning and decision-making processes in sustainable urban development and smart energy systems.

1. Introduction

With the development of urbanization and industry, environmental problems such as climate change and air pollution are becoming more important on a global scale. The increase in air pollutants in urban areas, harmful emissions from vehicles running on fossil fuels, and noise pollution have necessitated the transition to sustainable solutions in the transportation sector [1]. In this context, electric vehicles emerge as an important component of sustainable transportation with their low emissions and environmentally friendly features. International and national climate targets such as the Paris Climate Agreement and Net Zero 2050 also play an important role in the spread of electric vehicless [2].

Although electric vehicles are beneficial to the environment and society, there are various difficulties in their widespread use. These include insufficient fast-charging infrastructure, high purchasing costs of batteries, limitations of lithium-ion batteries, and lack of appropriate legal and regulatory frameworks. Insufficient support for the development of the EV ecosystem in many countries also leads to low adoption rates. Despite the known benefits of electric vehicles, these structural and technical problems need to be addressed to promote their widespread use[3, 4, 5]. In conclusion, the growth of electric vehicles depends on more than just technological advances. It also depends on factors such as social awareness, government regulations, and investments in infrastructure.

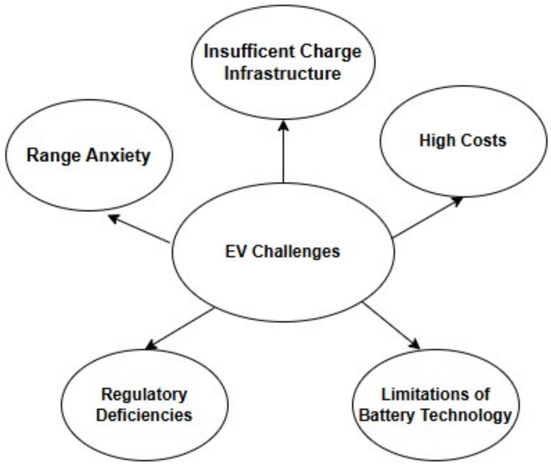

Fig. 1. EV Challenges

1.1. Development of Public Charging Infrastructure and Trends in the EV Market

By 2024, the number of electric vehicles reached 58 million and accounted for four per cent of all passenger cars. Global EV sales vary depending on factors such as regional government incentives and carbon targets. The establishment of domestic car brands by developing countries such as Vietnam, Turkey, and Argentina has also contributed positively to the local and international EV markets [6].

In recent years, the number of publicly available charging stations worldwide has increased significantly. China and the European Union (EU) have demonstrated a significant increase in the ratio of charging stations to EVs in recent years. In the United States and the United Kingdom, the ratio of vehicles to public fast charging stations exceeds the optimal level, indicating that the charging infrastructure is inadequate to satisfy demand [6].

Public charging stations in urban areas offer a viable alternative for people who do not have access to charging facilities at home or at work. However, high-traffic corridors such as highways require the establishment of fast-charging stations. In this context, over seventy-five percent of highways in Europe possess fast-charging infrastructure at intervals of no more than 50 km. Conversely, such infrastructure is present on fewer than half of the highways in the United States [6].

To meet the projected increase in global EV sales requires charging capacity to be nine times larger than current. The grid infrastructure must be renewed to meet this increase in charging capacity, and the future regional charging load must be accurately predicted by grid managers. Investments in grid infrastructure such as capacity increase, transformer reinforcement, and charging systems are of great importance in preventing supply-demand imbalances. Failure to upgrade the electricity grid to manage the demand will have a disruptive effect on grid stability and energy quality. Using data analytics, machine learning, and other artificial intelligence methods in grid planning is vital to ensure a sustainable and secure energy supply in the face of evolving EV charging demands.

1.2. Related Works

Several studies have proposed various methodologies for the planning and forecasting of public charging infrastructure. Sreekumar and Lekshmi created a demand forecasting model for EV charging stations, employing multiple scenarios in their analysis. This model was developed using a dataset sourced from the electric vehicle charging stations operated by Perth and Kinross Council. The data was first analyzed to determine the average demand across 1, 2, 3, and 4-hour intervals. The resultant dataset was subsequently employed in multiple models, including Random Forest, Categorical Boosting, Extreme Gradient Boosting (GB), and Light Gradient Boosting (LGB). The Categorical Boosting Regression (CBR) model was chosen for its exceptional performance, evidenced by a reduced mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) [7]. Sears et al. estimated the number of EV charging stations in Vermont using a statistical model and Geographic Information System (GIS) infrastructure developed based on the number of EVs, EV range, and daily travel demand [8]. Roy and Law analyzed the spatial heterogeneity in the distribution of electric vehicle charging stations (EVCS) employing machine learning techniques. At first, they looked at how demographic and socio-economic factors affected how easy it was to arrive to EVCS. Thereafter, the variables were combined with information about current EVCS locations to predict how many EVCS there would be in the future at different spatial resolutions. We used a number of machine learning algorithms and looked at how well they predicted things to determine the best spatial granularity for accurate prediction. The Random Forest (RF) algorithm was the most accurate of the models that were tested[9].

Mohammad et al. introduced two encoder-decoder models that use convolutional long short-term memory networks (ConvLSTM) and bidirectional ConvLSTM (BiConvLSTM), in addition to a standard long short-term memory (LSTM) network, to forecast the energy demand of electric vehicle charging stations in both temporal and spatial dimensions. The models use data from four different cities' EVCS energy demand to make predictions. We used three error metrics—RMSE, MSE, and MAPE—to see how well these forecasting systems worked. The results derived from the proposed method are ultimately contrasted with standard deep learning models and traditional machine learning approaches [11].

Akıl et al. are investigating a methodology to quantify the total electric vehicle charging demand by utilizing historical charging demand datasets alongside mobility statistics for electric vehicle operators. This model employs a Monte Carlo Simulation (MCS) to examine the charging behaviors of EV users and generate EV charging durations. To make the model more useful, the dataset on electric vehicle demand, which was made up of time series data, was first split up using empirical mode decomposition (EMD). Subsequently, each unique signal was forecasted independently using a Bayes-optimized long short-term memory (LSTM) network. The model's effectiveness was assessed via an examination of the IEEE 33-bus test system, concentrating on power system performance, distribution network power losses, bus voltage reductions, and transformer loading conditions [13].

Mystakidis et al. use real-world data from a residential complex in Norway, including the start and end times of charging sessions, the energy charged, hourly traffic flow, and weather data. The methodology consists of a series of steps that gradually integrate predicted values, such as traffic flow and average charging loads per private and shared user (PWALU and SWALU), rather than a single-step approach. This step-by-step approach demonstrated superior performance compared to the traditional model. [15].

Table 1. Overview of Recent Studies on EV Charging Demand Forecasting

Ref.	Year	Models	Proposed	Objective
7	2024	RFR, CBR, XGBR, LGBMR	CBR	EV charge demand prediction
8	2014	Statistical model	Statistical	EVCS number estimation
9	2022	SVM Linear, RF, MLR	RF	Spatial heterogeneity in EVCS placement
10	2024	SARIMA, ARIMA	SARIMA	Demand & revenue forecasting of EVCS
11	2023	LSTM, CNN, CNN-LSTM, BiLSTM, ConvLSTM	ConvLSTM-BiLSTM	Energy load demand forecasting
12	2022	EMD, AOA, DLSTM	EMD–AOA–DLSTM	Forecasting using decomposed time series
13	2022	MCS, EMD, LSTM (BO)	EMD–BO–LSTM	Bulk EV charging demand & grid impact
14	2024	AdaBoost, RF, Bagging, LR, GB, LSTM, CNN, NN	Ensemble trees	15-min EVCS demand forecasting
15	2025	ETR, HGBR, LGBMR, CBR, LSTM, CNN	LogZI	Hourly EV charging forecasting
This Study	2025	RF, SVM, ANN, GPR	GPR	City-scale EVCS installed power forecasting

1.3. Motivation of This Study

The rapid adoption of EVs is revolutionizing distribution grid planning. However, most capacity planning analyses for charging infrastructure either (i) focus on national totals, overlooking city-level differences, or (ii) depend on short-term station-level data that lack sufficient socio-economic context. In Turkey, local distribution companies are responsible for implementing policies and investment decisions regarding electric vehicle infrastructure. Thus, forecasts regarding the capacity of EVCS at the municipal level are more realistic. This research fills this gap by creating a monthly framework for six cities from August 2022 to April 2025. It combines charging data that has been approved by the Energy Market Regulatory Authority (EMRA) with data on urban mobility and socio-economic factors. We looked at different ways of learning and made clear, policy-compliant predictions that support phased implementations that work for site selection, transformer sizing, and budget limitations. The outcomes of this study will not only enhance the electric vehicle charging infrastructure but also promote sustainable urban development. By tailoring distribution strategies to local needs and financial constraints, we aspire to create a replicable model that other distribution companies can adopt to establish their own electric vehicle charging networks.

In this study, a dataset consisting of population density, motor vehicles (MCV), EVs, EVCS, and location information, normalized by min-max scaling to ensure fair contribution of each feature during training, was trained using RF, SVM, ANN, and GPR models to predict cumulative installed power demand based on urban mobility and infrastructure indicators. The model performance was evaluated using a five-fold cross-validation strategy.

Our research makes the following contributions:

We offer a city-scale, policy-ready dataset, which we see as one of the inaugural monthly city-level panels in Turkey. This dataset delineates the correlation between the installed capacity of EMRA-approved EVCS and diverse urban mobility, infrastructure, and socioeconomic metrics across six cities in the region from August 2022 to April 2025.

Ultimately, we guarantee the reproducibility of our results. We are cataloging our preprocessing and hyperparameter optimization methodologies and establishing a replicable framework to promote adoption in other areas.

2. Theory

Machine learning has become a powerful way to model complex systems when it's hard to make analytical or physical models. It lets systems learn patterns from data and make accurate predictions or classifications in a wide range of real-world situations. Supervised learning models like decision trees, support vector machines, neural networks, and probabilistic methods have become very popular among the many different types of machine learning techniques. This is because they are flexible and work well for both classification and regression tasks. Numerous comparative studies in the literature, including [16], elucidate the practical advantages and disadvantages of various learning algorithms across diverse conditions. Building upon these foundations, this section presents the theoretical backgrounds of the models employed in this study.

2.1. Random Forest

One of the most popular ensemble learning methods is Random Forest (RF), which was first suggested by Breiman [17]. It combines multiple decision trees trained on different bootstrap samples and random feature subsets to improve the predictive performance.

The final prediction is obtained by majority voting (classification) or averaging (regression):

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x) \tag{1}$$

where T is the total number of trees, and $h_t(x)$ denotes the prediction of the t-th tree.

This method effectively reduces variance and mitigates overfitting by aggregating diverse weak learners.

2.2. Support Vector Machine (SVM)

Support Vector Machines offer a powerful supervised learning framework for both classification and regression tasks. The core idea is to find the optimal hyperplane that maximally separates the data points from different classes in the feature space [18].

For a linearly separable binary classification problem, the decision function is defined as:

$$f(x) = \text{sign}(w^T x + b) \tag{2}$$

where w is the weight vector, x is the input vector, and b is the bias term.

The optimal separating hyperplane is obtained by solving the following convex optimization problem:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{subject to} \quad y_i(w^T x_i + b) \geq 1 \tag{3}$$

where $y_i \in \{-1, 1\}$ are the class labels.

When data is not linearly separable, kernel functions (e.g., radial basis function, polynomial) are used to transform it into a higher-dimensional space.

2.3. Artificial Neural Networks (ANN)

Inspired by the biological structure of the human brain, Artificial Neural Networks (ANN) consist of layers of interconnected neurons that can model complex non-linear relationships through learning from data [16], .

Each neuron computes a weighted sum of its inputs followed by a non-linear activation function:

$$a_j = \phi\left(\sum_i w_{ij} x_i + b_j\right) \tag{4}$$

where x_i are the input signals, w_{ij} is the weight from input i to neuron j, b_j is the bias term, and $\phi(\cdot)$ is the activation function (e.g., sigmoid, ReLU).

The network is trained using the backpropagation algorithm to minimize a loss function (e.g., mean squared error), updating weights via gradient descent.

2.4. Gaussian Process Regression (GPR)

In contrast to parametric models, Gaussian Process Regression (GPR) provides a flexible, probabilistic approach to regression by modeling distributions over functions using a mean and covariance function [20].

A Gaussian process is formally defined as:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')) \quad (5)$$

where

$$m(x) = E[f(x)], \quad k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] \quad (6)$$

Given training data $\mathcal{D} = \{X, \mathbf{y}\}$, the predictive mean and variance for a new test point x_* are:

$$\mu(x_*) = k(x_*, X)[K(X, X) + \sigma_n^2 I]^{-1}\mathbf{y} \quad (7)$$

$$\sigma^2(x_*) = k(x_*, x_*) - k(x_*, X)[K(X, X) + \sigma_n^2 I]^{-1}k(X, x_*) \quad (8)$$

where $K(X, X)$ is the kernel matrix and σ_n^2 is the noise variance.

GPR not only provides point estimates but also quantifies the uncertainty of predictions.

3. Methods

3.1. EV Charging Infrastructure Dataset

The dataset created for training and testing purposes in the machine learning models evaluated within the scope of the study was created to analyze the electric vehicle charging infrastructure in Turkey using technical data from the charging station, such as the cumulative number of EVCS and the cumulative installed power of EVCS, as well as the location of the station and socioeconomic and demographic factors such as population density, total number of motor vehicles, and number of EVs on a monthly basis between August 2022 and April 2025.

In this context, technical data for charging stations located in six different cities in Turkey were provided by Aras Distribution System Operator's Smart Grids and R&D Directorate, which is responsible for the distribution of electricity in the relevant cities, with the approval of the Charging Services Group of the Energy Transformation Department of the EMRA[21].

In addition to technical data, to reflect the socioeconomic and demographic structure of cities, variables such as population information, the number of vehicles, the total number of motorized land vehicles, and the number of electric vehicles were obtained using current data published by the Turkish Statistical Institute (TSI) [22, 23, 24].

The data obtained from all these sources were consolidated into a multivariate dataset suitable for subsequent analysis.

3.2. Data Processing

All simulations and machine-learning analyses were conducted in MATLAB R2024b on a 64-bit Windows 11 Pro system (Intel® Core™ i9-14900HX, 2.20 GHz; 64 GB RAM). To ensure comparable feature contributions and mitigate scale disparities, all input variables were normalized to the range [-1, 1] using min–max scaling (see Eq. (9)). All preprocessing steps (e.g., feature scaling, categorical encoding) were implemented within a single, reproducible pipeline. The scaler was fitted exclusively on the training data—and within each cross-validation fold, on that fold's training partition—and only applied to the corresponding validation and test partitions, thereby preventing data leakage. When the target variable was normalized for model fitting, predictions and performance metrics were transformed back to the original units before being reported.

$$x_n = \frac{2(x - x_{\min})}{x_{\max} - x_{\min}} - 1 \quad (9)$$

where x_n is the normalized value, x is the original input, and $x_{\min}$, $x_{\max}$ represent the minimum and maximum values of each feature, respectively. TThis scaling process made the models learn faster and more stable by speeding up convergence and stopping high-magnitude features from taking over during optimization. After normalization, the dataset was split into 85% for training and 15% for testing to thoroughly test how well the model can generalize. To make sure it was strong and to avoid overfitting, a five-fold cross-validation scheme was used.

Figure 2 shows the overall flow of data that the proposed machine-learning framework uses to predict installed power. Some of the important input parameters were population density, the number of internal combustion engine vehicles (ICEVs), the number of EVs, the number of EVCSs, and the geographical location identifier. These features were chosen because they could affect the total installed power demand for planning smart mobility and energy infrastructure. The machine learning model takes these inputs and uses them to find complex nonlinear relationships. It then uses these relationships to figure out the total installed power (kW) as the output. This modeling structure lets us use data to learn about the energy needs of cities, which helps us make smart choices about how to integrate electric vehicles and plan cities in a way that is good for the environment.

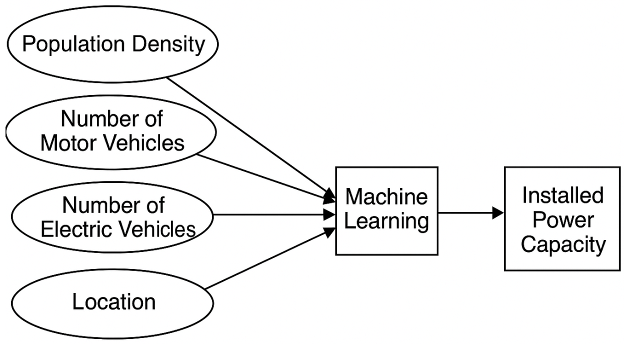

Fig. 2. Installed Power Capacity Prediction.

Because the panel size was limited (monthly data from six cities), we intentionally used a simple specification to avoid overfitting and keep the values the same. Instead of filling the design with complicated time-related terms, we focused on a clear and short feature set during cross-validation. We did a lot of hyperparameter optimization on all of the models to improve their ability to predict, as shown in Table 2. In Random Forest, augmenting the number of trees diminishes the model variance, whereas the leaf size regulates the complexity of individual trees. A leaf size of 10 achieved an ideal balance between bias and variance, preventing the underfitting and overfitting of the model. The ensemble of 100 trees achieved variance reduction while maintaining manageable computational costs. The Gaussian (RBF) kernel was chosen for the SVM because of its capacity to transform nonlinearly separable data into a higher-dimensional space, facilitating the identification of intricate nonlinear relationships. The optimized box constraint value (C = 1) facilitated a balanced margin that accommodated specific errors while preserving the generalization. An epsilon of 0.05 in the epsilon-insensitive loss function established a limited tolerance zone, thereby augmenting the model resilience to minor variations in the target values. The kernel scale was set to "Auto," which let the algorithm change the feature space on its own. Different architectures in artificial neural networks were evaluated to control model capacity and mitigate overfitting. A shallow network with only one hidden layer with 10 neurons and a ReLU activation function showed better performance. The ReLU function helps with vanishing gradient problems and makes sparse activation easier, which speeds up learning. An epoch count of 500 enabled sufficient convergence without causing overfitting, thereby guaranteeing both depth of learning and generalization. The squared exponential kernel was chosen for Gaussian Process Regression (GPR) because it is smooth and can show continuous nonlinearities based on the Euclidean distance. The noise parameter (σ = 0.1) combined the stochastic variance and reduced overfitting by taking into account the noise in the measurements. The constant basis function helps center the Gaussian Process prior, which makes the kernel function mostly responsible for capturing the data structure. We used 5-fold cross-validation on the training dataset to test all possible combinations of hyperparameters. The configurations that produced the minimal average validation error were chosen, guaranteeing a balanced bias–variance framework for each model. This systematic tuning process markedly enhanced the generalization capability, as evidenced by the final predictive performance.

Table 2. Hyperparameter optimization results for the machine learning models

Model	**Parameter**	**Range Tested**	**Optimized Value**
2***RF**	Number of Trees	10, 50, 100, 200	100
	Minimum Leaf Size	1, 5, 10, 20, 50	10
4***SVM**	Kernel Function	Linear, Gaussian, Polynomial	Gaussian
	Box Constraint (C)	0.1, 1, 10, 100	1
	Epsilon	0.01, 0.05, 0.1, 0.2	0.05
	Kernel Scale	Auto, 1, 10, 100	Auto
4***ANN**	Number of Layers	1, 2, 3	1
	Number of Neurons	5, 10, 20, 50	10
	Activation Function	ReLU, Sigmoid, Tanh	ReLU
	Epochs	100, 200, 500, 1000	500
3***GPR**	Kernel Function	SquaredExponential, RationalQuadratic	Squared Exponential
	Sigma (Noise)	0.01, 0.1, 1	0.1
	Basis Function	Constant, Linear, None	Constant

Figure 3 shows the distribution of the input features used in the machine learning model. As shown, the majority of features—namely, population density, number of motor vehicles (MKT_Cumulative), number of electric vehicles (EA_Cumulative), and number of EV charging stations (EVCS_Cumulative)—are heavily skewed toward the lower end of the normalized range. This indicates a concentration of data points in regions with relatively low population and transportation infrastructure, which may reflect real-world imbalances in urban and rural distribution. In contrast, the *Location* parameter exhibited a uniform categorical distribution, suggesting a balanced representation across all spatial regions in the dataset. These visual insights are crucial for understanding the model's input structure and ensuring that the learning algorithms are not biased toward over-represented conditions.

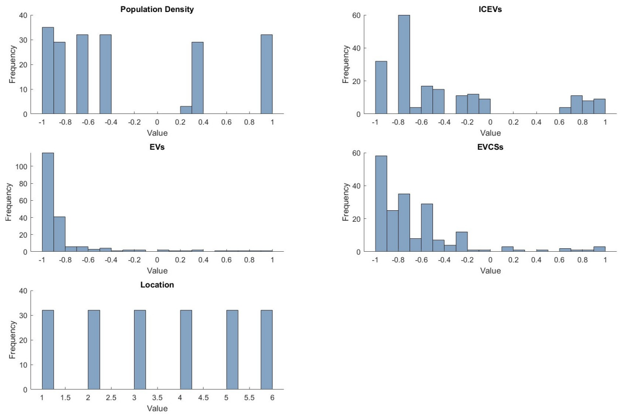

Fig. 3. Distribution of input parameters

4. Results and Discussion

In this section, the performance of the employed machine learning models is comprehensively evaluated using both quantitative metrics and graphical interpretation. The regression models were trained to predict the normalized installed power capacity in EV Charging Stations, and their effectiveness was assessed based on four standard statistical performance indicators: coefficient of determination (R^2), RMSE, MAE, and MSE. These metrics offer complementary perspectives, with R^2 reflecting the proportion of variance explained by the model and RMSE/MAE/MSE capturing the absolute and squared deviations from the ground truth. Furthermore, actual versus predicted value plots are provided for each model to visually inspect the alignment of the predictions with the real measurements, allowing for an intuitive assessment of the bias, variance, and distributional behavior. By combining numerical and graphical evaluations, a more holistic understanding of the generalization ability and learning behavior of each model is achieved, highlighting their relative strengths and limitations in the context of cumulative load prediction. As shown in Fig. As shown in Figure 4, the comparative evaluation of the machine learning models demonstrated notable differences in predictive accuracy. While the GPR model achieved the lowest RMSE, MAE, and MSE values with the highest R^2 score, indicating superior performance, the ANN model also provided competitive results with a strong generalization ability. In contrast, the SVM exhibited relatively higher error values and a lower R^2, suggesting a weaker predictive capability. Table 3 quantitatively supports these visual findings. GPR achieved the highest R^2 (0.99619) and the lowest RMSE, MAE, and MSE values among all models, confirming its superior performance in learning complex nonlinear mappings and capturing uncertainties. The ANN ranks second with an R^2 of 0.93486 and consistently low error metrics, which highlights its capability to generalize well under a compact neural architecture. RF follows with a decent R^2 of 0.91999, but slightly elevated RMSE and MSE values reflect its reduced model complexity owing to the intentionally low tree count. Finally, SVM recorded the poorest performance, with an R^2 of only 0.69362 and the highest error rates, verifying its lack of capacity to generalize under limited training data and non-optimal parameter settings. The overall ranking suggests that nonparametric, flexible models, such as GPR and ANN, are more suitable for capturing the underlying structure of the data in this context, whereas RF and SVM may require further tuning or richer data representations to achieve comparable accuracy.

Table 3. Performance comparison of ML models

Model	R^2	RMSE	MAE	MSE
RF	0.9200	0.0745	0.0456	0.00555
SVM	0.6936	0.1458	0.1110	0.02121
ANN	0.9349	0.0672	0.0506	0.00452
GPR	**0.9962**	**0.0163**	**0.0107**	**0.00026**

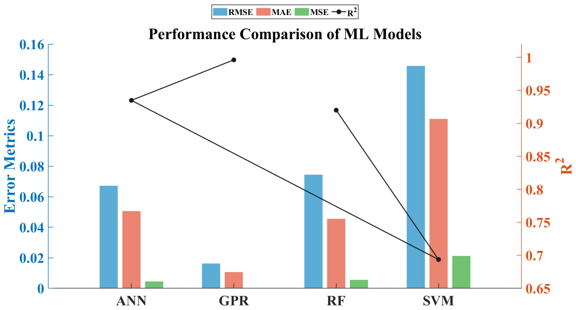

Fig. 4. Error Metrics and R² Comparison of ML Models

The actual versus predicted scatter plots for the four ML models are shown in Figure 5. The GPR showed almost perfect alignment along the diagonal, which meant that the predictions were very accurate and had very few errors. This is because it is probabilistic and smooth, thanks to the squared exponential kernel, which effectively captures continuous nonlinearities. Adding a noise parameter ($\sigma = 0.1$) makes the model even more regular by taking into account the measurement noise, which lowers the risk of overfitting. The ANN also showed that it could make good predictions, with a lot of points clustered around the ideal line. Its shallow architecture with ReLU activation allows for an effective nonlinear approximation without making the model too complicated. This keeps a balance between learning capacity and generalization. RF, while giving moderate accuracy, had more spread around the diagonal, especially at higher values. This is mostly because there aren't many trees (10), which makes it harder for the ensemble to reduce variance and makes it more likely to underfit. SVM had the worst alignment with the diagonal, with big mistakes all over the place. This can be explained by both the reduced training set and suboptimal kernel parameter tuning, which hinder its ability to form flexible regression boundaries despite using an RBF kernel.

5. Conclusions

This study presents a machine learning-based predictive framework to estimate the cumulative installed power capacity using urban mobility and infrastructure indicators, such as population density, number of ICEVs, EVs, EVCSs, and location identifiers. Among the four models evaluated—RF, SVM, ANN, and GPR—the GPR model demonstrated the best performance, achieving an R² of 0.99619 and an RMSE of 0.01625. Compared to

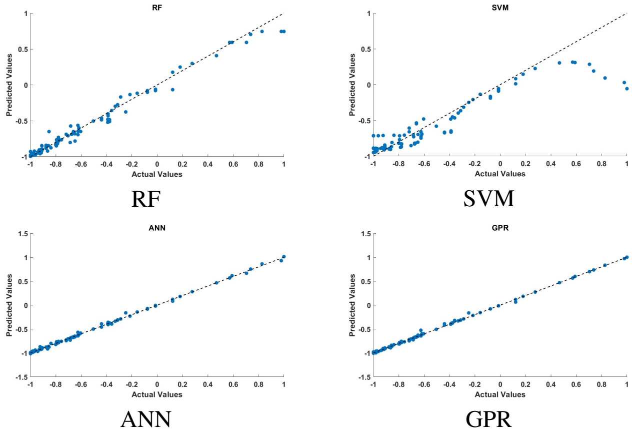

Fig. 5. Predicted performance visualizations of the four machine learning models based on actual vs. predicted values for normalized power estimation

the weakest model, SVM, GPR reduced the RMSE by approximately 88.8% and improved R² by more than 43%. Similarly, compared with RF and ANN, GPR provided 78.2% and 75.8% lower RMSE, respectively. These improvements show that the GPR is better at finding complex nonlinear patterns and measuring predictive uncertainty. The ANN model also did well (R² = 0.93486), which means it can generalize well even with a small structure. The SVM, on the other hand, had trouble modeling nonlinearities well and got the lowest R² (0.69362). In general, the results indicate that probabilistic and neural-based models, especially GPR, work better for planning the power needs of EV infrastructure. This approach not only enhances prediction accuracy but also provides actionable insights for energy-aware planning and decision-making in smart cities.

6. Acknowledgement

We are grateful for the datasets that EMRA and the Aras Distribution System Operator have made available to us. The authors' opinions, findings, and conclusions in this material do not necessarily reflect the views of the data providers.

7. References

References

[1] S. N. Toolib, W. N. W. Hanafi, S. Daud, and H. A. Afsarizal, "Factors influencing electric vehicle adoption: A conceptual paper," *European Proceedings of Finance and Economics*, vol. EpFE 23081.77, 2023, doi: 10.15405/epfe.23081.77.

[2] A. A. Ghani, N. Abdullah, K. Balasubramaniam, R. Saidur, and A. K. Pandey, "Assessing technological-driven challenges and policies associated with electric vehicle (EV) adoption," *Transport Policy*, vol. 169, pp. 167–177, 2025, doi: 10.1016/j.tranpol.2025.04.030.

[3] A. Pamidimukkala, S. Kermanshachi, J. M. Rosenberger, and G. Hladik, "Barriers and motivators to the adoption of electric vehicles: A global review," *Green Energy and Intelligent Transportation*, vol. 3, p. 100153, 2024, doi: 10.1016/j.geits.2024.100153.

[4] F. Alanazi, “Electric Vehicles: Benefits, Challenges, and Potential Solutions for Widespread Adaptation,” *Applied Sciences*, vol. 13, no. 10, Art. no. 6016, 2023, doi: 10.3390/app13106016.

[5] M. Koniak, P. Jaskowski, and K. Tomczuk, “Review of Economic, Technical and Environmental Aspects of Electric Vehicles,” *Sustainability*, vol. 16, no. 22, Art. no. 9849, 2024, doi: 10.3390/su16229849.

[6] International Energy Agency (IEA), *Global EV Outlook 2025: Catching Up with Climate Ambitions*, IEA Publications, Paris, France, 2025. [Online]. Available: www.iea.org/reports/global-ev-outlook-2025

[7] A. Menon, M. Hosseinian-Far, and S. Velu, “Assessment of electric vehicle charging infrastructure using multi-criteria decision-making: A review and case study,” *Green Energy and Intelligent Transportation*, vol. 3, p. 100185, 2024, doi: 10.1016/j.geits.2024.100185.

[8] J. Sears, K. Glitman, and D. Roberts, “Forecasting demand of public electric vehicle charging infrastructure,” in *Proc. IEEE Conf. Technologies for Sustainability (SusTech)*, 2014, pp. 250–254, doi: 10.1109/SusTech.2014.7046225.

[9] S. Roy and S. Law, “A machine learning approach for forecasting energy demand of electric vehicle charging stations,” *Sustainable Cities and Society*, vol. 86, p. 104129, 2022.

[10] K. C. Akshay, G. H. Grace, K. Gunasekaran, and R. Samikannu, “Power consumption prediction for electric vehicle charging stations and forecasting income,” *Scientific Reports*, vol. 14, no. 1, pp. 1–11, 2024, doi: 10.1038/s41598-024-56507-2.

[11] F. Mohammad, D.-K. Kang, M. A. Ahmed, and Y.-C. Kim, “Energy Demand Load Forecasting for Electric Vehicle Charging Stations Network Based on ConvLSTM and BiConvLSTM Architectures,” IEEE Access, vol. 11, pp. 67350–67366, May 9, 2023, doi: 10.1109/ACCESS.2023.3274657. :contentReference[oaicite:0]index=0

[12] J. Shanmuganathan, A. A. Victoire, G. Balraj, and A. Victoire, *Deep Learning LSTM Recurrent Neural Network Model for Prediction of Electric Vehicle Charging Demand*, *Sustainability*, vol. 14, no. 16, p. 10207, Aug. 2022, doi: 10.3390/su141610207.

[13] M. Akıl, E. Dokur, and R. Bayındır, “Analysis of Electric Vehicle Charging Demand Forecasting Model based on Monte Carlo Simulation and EMD-BO-LSTM,” in *Proc. 10th IEEE Int. Conf. on Smart Grid (icSmartGrid 2022)*, Istanbul, Turkey, Jun. 27–29, 2022.

[14] A. Ostermann and T. Haug, *Probabilistic forecast of electric vehicle charging demand: analysis of different aggregation levels and energy procurement*, *Energy Informatics*, vol. 7, art. 13, 2024, doi: 10.1186/s42162-024-00319-1.

[15] A. Mystakidis, N. Tsalikidis, P. Koukaras, G. Skaltsis, D. Ioannidis, C. Tjortjis, and D. Tzovaras, *EV charging forecasting exploiting traffic, weather and user information*, *International Journal of Machine Learning and Cybernetics*, vol. 16, pp. 6737–6763, 2025, doi: 10.1007/s13042-025-02643-8.

[16] M. Pal and P. M. Mather, “A comparative study of artificial neural network and support vector machine classifiers for land cover classification,” *Pattern Recognition Letters*, vol. 27, no. 5, pp. 622–628, Apr. 2006, doi: 10.1016/j.patrec.2005.11.018.

[17] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[18] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[19] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., Upper Saddle River, NJ, USA: Prentice Hall, 1999.

[20] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Cambridge, MA, USA: MIT Press, 2006.

[21] Aras Electricity Distribution Company, “Technical data for charging stations in six Turkish cities, monthly data August 2022–April 2025,” Data set on CD-ROM, May 24, 2025. [Unpublished].

[22] Turkish Statistical Institute, “Motor Vehicle Statistics,” TurkStat Online Database, 2025. [Online]. Available: https://data.tuik.gov.tr/MotorluKaraTasitIstatistikleri. [Accessed: Jun. 15, 2025]

[23] Turkish Statistical Institute, “Address-Based Population Registration Results,” TurkStat Online Database, 2025. [Online]. Available: https://data.tuik.gov.tr/AdreseDayaliNufusKayitSistemiSonuclari. [Accessed: Jun. 15, 2025].

[24] Turkish Statistical Institute, “Fuel Distribution of Road-Registered Vehicles, 2004– 2025 (March),” Unpublished raw data, provided by request, Jun. 18, 2025.