energies

Article

Integrated Traffic–Weather-Aware Forecasting of Urban EV Charging Demand for Infrastructure Planning

Christoph Sommer, Jahangir Hossain * and Abbas Tabandeh

School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia; christoph.sommer@student.uts.edu.au (C.S.); abbas.tabandeh@uts.edu.au (A.T.)
* Correspondence: jahangir.hossain@uts.edu.au; Tel.: +61-029515

Abstract

The accelerating adoption of electric vehicles (EVs) presents significant challenges for maintaining grid stability and optimizing charging infrastructure. Accurate short-term forecasting of EV charging demand is therefore critical to support reliable grid operation and effective energy management in urban environments. However, existing forecasting models often fail to capture the intricate interdependencies among mobility patterns, weather variations, and real-world charging behaviors, which constrains their generalizability and robustness. This study develops a multi-model forecasting framework that leverages Transformer-based deep learning architectures to integrate real-world charging data with traffic flow and meteorological variables for predicting short-term EV charging demand across metropolitan areas. To benchmark performance, two additional machine learning models—CatBoost and convolutional neural networks (CNNs)—are systematically evaluated using datasets from urban EV supply equipment (EVSE) and electric bus systems. The results indicate that Transformer-based models deliver superior predictive accuracy, temporal consistency, and adaptability compared with CNNs and CatBoost. Furthermore, sensitivity analysis reveals that traffic dynamics and user charging behavior exert the strongest influence on forecast performance. The proposed framework offers actionable insights for utilities and urban planners, facilitating resilient grid operation, optimized charging infrastructure deployment, and accelerated integration of EVs into the power system.

Keywords: charging behavior; traffic flow; prediction; deep learning; neural networks; transformer

Check for updates

Academic Editors: Joao Ferreira, Federico Barrero and Adolfo Dannier

Received: 24 April 2026
Revised: 19 June 2026
Accepted: 2 July 2026
Published: 6 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

With the rapid advancement of battery technology, electric vehicles (EVs) have emerged as a promising solution to current energy and environmental challenges. However, the increasing penetration of EVs is reshaping electricity consumption patterns and introducing highly variable charging demand into urban power systems. Accurate prediction of EV charging load and user charging behavior is therefore essential for the effective planning, operation, and expansion of charging infrastructure. Errors in long-term EV charging load forecasting can lead to suboptimal infrastructure investment decisions, while short-term forecasting uncertainty can affect the coordinated dispatch of EVs, renewable energy sources, and grid-side resources [1,2]. Hence, reliable charging demand forecasting is a critical requirement for grid stability, renewable energy integration, and charging infrastructure planning [3].

Accurate EV charging demand forecasting relies on three key components. The first is the prediction of charging behavior at a specific location, which reflects local charging patterns, user preferences, and station utilization characteristics [4]. The second is the incorporation of traffic flow (TF) data in a 1.5 miles radius around the charging station, as traffic dynamics influence vehicle arrivals, congestion levels, trip timing, and consequently charging demand [5,6]. The third is the selection of an appropriate forecasting algorithm that can handle limited data availability, computational constraints, heterogeneous features, and nonlinear temporal dependencies [7,8]. In addition to charging session records and TF information, weather variables such as temperature and precipitation can influence EV range, charging needs, and travel behavior [9]. Therefore, effective forecasting models should integrate charging session data, TF metrics, and meteorological variables to capture the coupled effects of mobility, energy consumption, and environmental conditions [10].

In ref. [11], it is shown that Monte Carlo Simulation (MCS), queuing theory, and Markov chain models were widely used to synthetically generate charging load profiles by sampling stochastic variables such as vehicle arrival time, trip distance, state of charge (SoC), and charging duration. Similarly, studies such as [12–14] employed MCS-based frameworks to estimate spatial–temporal charging demand under uncertain travel behavior and charging preferences. While these approaches provided an initial theoretical understanding of EV charging impacts on power distribution networks, their reliance on predefined probability distributions and random sampling limits scalability and reduces accuracy when assumed distributions deviate from real-world charging behavior, particularly in dense urban environments [11].

To address these limitations, the literature has progressively shifted toward data-driven forecasting methods that directly exploit historical charging records and exogenous variables. Classical statistical time-series models, including Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA), have been applied to short-term charging load forecasting and can capture basic temporal patterns such as diurnal and weekly seasonality [11,15]. However, these linear models are limited in representing nonlinear demand drivers, spatial heterogeneity, and seasonal variability observed in real-world urban charging data. Consequently, machine learning techniques such as linear regression, support vector regression, random forests, and gradient-boosted decision trees have been increasingly adopted for EV charging demand forecasting [11]. These methods improve adaptability by learning region-specific charging characteristics from data rather than relying on synthetic assumptions.

Among non-neural machine learning models, gradient-boosted decision tree methods remain strong candidates for EV charging demand forecasting because of their robustness on heterogeneous tabular datasets and their ability to model nonlinear feature interactions. CatBoost and related boosting frameworks have shown strong performance in short-term energy and load forecasting tasks due to their robustness to overfitting and ability to handle categorical temporal features [16–20]. Nevertheless, tree-based ensemble models generally rely on manually engineered features and may struggle to capture fine-grained spatial–temporal dependencies in EV charging behavior without explicit spatial or sequential representations [21].

Deep learning models have been explored to address these limitations. Convolutional neural networks (CNNs) have demonstrated effectiveness in extracting localized temporal and spatial patterns from structured inputs and have been successfully applied to short-term EV charging demand forecasting and related energy prediction problems [22–24]. In ref. [25], CNN-based models were shown to outperform classical time-series and regression approaches by learning localized temporal patterns in charging load data. Similarly, ref. [26] reported that deep neural architectures improve predictive accuracy at public charging stations compared with traditional statistical methods. However, CNN architectures

typically operate with fixed receptive fields, which can restrict their ability to capture long-range temporal dependencies, abrupt demand variations, and evolving seasonal effects in urban charging networks.

More recently, Transformer-based architectures have emerged as a powerful alternative for EV charging demand forecasting because their self-attention mechanisms can model both short-term dynamics and long-range temporal relationships. In ref. [11], Transformer models are identified as promising methods that can outperform recurrent neural networks (RNNs), long short-term memory networks (LSTMs), gated recurrent units (GRUs), CNNs, and statistical baselines for longer prediction horizons. Complementarily, ref. [27] demonstrates that attention mechanisms improve spatiotemporal charging demand prediction, with ablation results confirming that multiple attention heads enhance the modeling of complex spatial and temporal correlations across charging stations. Furthermore, ref. [28] shows that explicitly modeling spatial dependencies is important for accurate urban-scale charging demand forecasting. These findings support the use of Transformer-based models for multi-source EV charging demand forecasting, particularly when charging data, TF information, and weather variables must be jointly considered.

Despite recent progress, important research gaps remain. Few studies have systematically evaluated advanced Transformer-based models for EV charging demand forecasting while holistically incorporating charging session data, TF dynamics, and weather information within a unified forecasting framework. This limitation is particularly relevant for large-scale urban and suburban charging networks, where interactions among traffic congestion, temporal charging behavior, station utilization, and environmental conditions introduce substantial forecasting uncertainty. Bridging this gap is essential for utilities, charging operators, and urban planners seeking to anticipate increasing charging demand and maintain grid reliability amid expanding EV supply equipment (EVSE) deployment.

To address these challenges, this paper develops and evaluates an integrated traffic–weather-aware forecasting framework for urban EV charging demand. The proposed framework systematically compares three advanced machine learning models—CatBoost, CNN, and Transformer architectures—under a consistent data pipeline and evaluation protocol. By incorporating charging session records, TF data, and weather variables, this study provides a robust data-driven proof of concept for scalable EV charging demand forecasting. The results support reliable decision-making for grid operation, renewable energy integration, and charging infrastructure planning in electrified transport systems.

The key contributions of this study are as follows:

- Development of an integrated traffic–weather-aware EV-charging-demand, Transformer-based forecasting framework that fuses extremely large real-world charging session datasets with transportation and meteorological information to support distribution-grid-oriented analysis.
- Multi-horizon forecasting of urban EV charging demand (1 h to 24 h ahead), designed to provide reliable load predictions for short-term grid operation and medium-term charging infrastructure planning in urban and suburban areas.
- Systematic performance evaluation using proven and robust algorithms for key metrics to quantify forecasting accuracy and robustness under diverse urban operating conditions relevant to power system studies.
- Sensitivity analysis from a power-system planning perspective to identify key spatiotemporal drivers of EV charging demand and derive actionable insights for grid-load balancing, renewable energy integration.

The contributions presented above establish a comprehensive foundation for advancing short- and mid-term energy demand forecasting in urban EV charging infrastructures. By leveraging multiple machine learning architectures and integrating heterogeneous

datasets—including traffic flow dynamics, meteorological variables, and real-world charging session records—this study introduces a robust, data-driven approach for anticipating spatial and temporal load variations across major metropolitan regions. The subsequent section describes the methodological framework used to design, train, and evaluate the proposed forecasting models, highlighting their architectural distinctions, input feature engineering, and performance assessment procedures.

2. Proposed Approach for Metropolitan Energy Forecasting

This study proposes a metropolitan-scale forecasting framework for predicting aggregated electric vehicle (EV) charging energy demand across multiple urban areas. The main objective is to evaluate attention-based forecasting by benchmarking a Transformer model against CatBoost and convolutional neural networks (CNNs). CatBoost represents a robust tree-based baseline for heterogeneous tabular data, while CNNs capture localized temporal patterns in structured input sequences. The Transformer is selected as the main contribution because self-attention mechanisms can model long-range temporal dependencies and complex interactions among charging, traffic, weather, and temporal features [11].

LSTM and GRU models are not explicitly included to avoid redundant benchmarking. Recent studies increasingly identify Transformer architectures as scalable successors to recurrent models for long-horizon time-series forecasting [11]. Therefore, the comparison among CatBoost, CNN, and Transformer models provides a focused evaluation of three distinct forecasting paradigms: tree-based learning, convolutional deep learning, and attention-based forecasting.

The framework predicts EVSE energy demand at the city level rather than by charger type, rated power, or individual station category. This aggregation supports scalable regional forecasting, infrastructure planning, demand management, and grid-operation studies. It also captures metropolitan charging patterns relevant to utilities and planners while reducing the complexity of highly fragmented station-level modeling. Although CNNs have shown strong performance in multi-energy load forecasting [29], their fixed receptive fields can limit their ability to represent long-range dependencies in large, heterogeneous EV charging datasets.

The input data are converted to hourly resolution and include historical charging energy, session counts, meteorological variables, and cyclical temporal features such as hour of day and day of week. These features are used to train and evaluate CatBoost, CNN, and Transformer models under a consistent data pipeline, enabling a fair comparison of their ability to capture nonlinear dependencies, temporal dynamics, and demand variability, as shown in Figure 1.

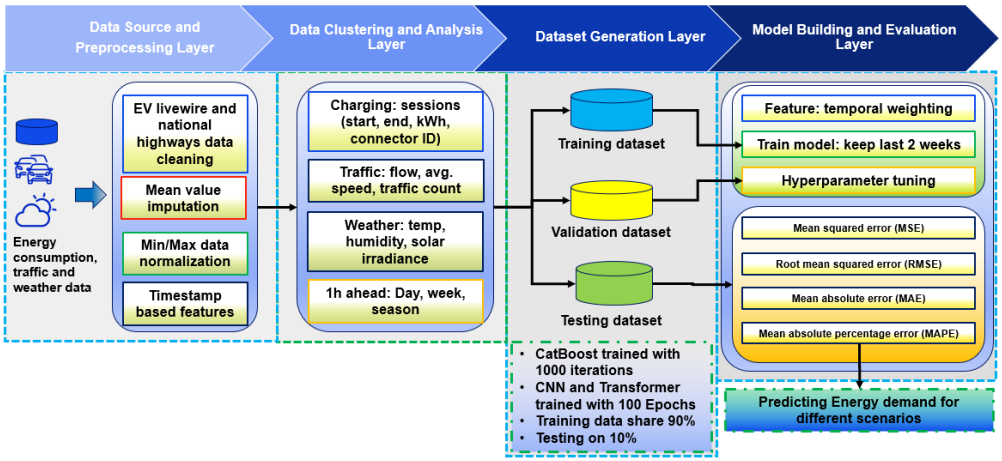


Figure 1. Metropolitan energy forecast flowchart (Arrows show data flow).

2.1. Proposed Transformer-Based Energy Forecasting

Transformers are employed to model long-range dependencies in sequential energy data (Algorithm 1). The Transformer encoder operates on a sequence of embedded input vectors $\mathbf{x}_{1:t} \in \mathbb{R}^{t \times d}$, where each vector contains historical hourly energy consumption values, EVSE session counts, weather conditions, and time encoding features. These input time series $\mathbf{x}_e = [x_e^{(1)}, x_e^{(2)}, \ldots, x_e^{(T)}]$ for energy consumption and $\mathbf{x}_s = [x_s^{(1)}, x_s^{(2)}, \ldots, x_s^{(T)}]$ for session counts, along with contextual features $\mathbf{c}$ such as weather data, TF, and time-of-day, are embedded into a continuous vector space via an embedding function and combined with positional encodings $\mathbf{p}$ to form the input embedding matrix:

$$\mathbf{E} = \text{Embed}(\mathbf{x}_e, \mathbf{x}_s, \mathbf{c}) + \mathbf{p} \tag{1}$$

where Embed$(\cdot)$ maps discrete or continuous input features into dense embeddings, and $\mathbf{p}$ (1) provides temporal order information essential for sequential modeling.

Each Transformer encoder layer applies multi-head self-attention, defined as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{2}$$

where the query Q, key K, and value V matrices are learned, linear transformations of the input embeddings (2). This attention mechanism enables the model to capture complex temporal dependencies and relationships within the sequence. Following the attention operation, a feed-forward network (FFN) is applied independently to each position, yielding the hidden representations at the l-th layer:

$$\mathbf{H}^{(l)} = \text{FFN}\big(\text{MultiHeadAttention}(\mathbf{H}^{(l-1)})\big) \tag{3}$$

with $\mathbf{H}^{(0)} = \mathbf{E}$.

After processing through L such layers, the final hidden state $\mathbf{H}^{(L)}$ is aggregated to produce a fixed-length feature vector $\mathbf{h}_{\text{concat}}$ (3), which integrates information from all input sequences and contextual features. This vector is then passed through a fully connected output layer with weight matrix $\mathbf{W}$ and bias $\mathbf{b}$, resulting in the predicted future energy consumption $\hat{y}$:

$$\hat{y} = \log\left(1 + e^{\mathbf{W}\mathbf{h}_{\text{concat}} + \mathbf{b}}\right) \tag{4}$$

The softplus activation function $\log(1 + e^x)$ ensures the output is smooth (4), differentiable, and strictly positive, making it well suited for modeling continuous non-negative quantities such as energy demand. This proposed Transformer-based architecture thus effectively captures seasonal trends, sudden demand shifts, and long-term usage patterns in urban energy consumption.

2.2. CNN-Based and CatBoost City-Level Energy Forecasting

CNNs are employed to capture local temporal dependencies and patterns in historical time series data (Algorithm 2). Let $\mathbf{x}_e = [x_e^{(1)}, x_e^{(2)}, \ldots, x_e^{(T)}]$ denote the historical energy consumption series, and $\mathbf{x}_s = [x_s^{(1)}, x_s^{(2)}, \ldots, x_s^{(T)}]$ the number of sessions series over a time window of size T. Each series is processed through a series of 1D convolution and pooling operations (5)–(8):

$$\mathbf{h}_e^{(1)} = \text{ReLU}(\mathbf{x}_e \times \mathbf{w}_e^{(1)} + \mathbf{b}_e^{(1)}) \tag{5}$$

$$\mathbf{h}_s^{(1)} = \text{ReLU}(\mathbf{x}_s \times \mathbf{w}_s^{(1)} + \mathbf{b}_s^{(1)}) \tag{6}$$

$$\mathbf{h}_e^{(2)} = \text{MaxPool}(\text{ReLU}(\mathbf{h}_e^{(1)} \times \mathbf{w}_e^{(2)} + \mathbf{b}_e^{(2)})) \tag{7}$$

$$\mathbf{h}_s^{(2)} = \text{MaxPool}(\text{ReLU}(\mathbf{h}_s^{(1)} \times \mathbf{w}_s^{(2)} + \mathbf{b}_s^{(2)})) \tag{8}$$

Algorithm 1 Transformer-Based Forecasting

Require: Historical energy series $\mathbf{x}_e = [x_e^{(1)}, \ldots, x_e^{(T)}]$, session-count series $\mathbf{x}_s = [x_s^{(1)}, \ldots, x_s^{(T)}]$, contextual features $\mathbf{c}$ containing weather, traffic flow (TF), and time encodings, positional encoding $\mathbf{p}$, encoder depth L.

Ensure: Future EV charging energy demand forecast $\hat{y}$.

1: Construct the hourly input matrix $\mathbf{X}_{1:t} \in \mathbb{R}^{t \times d}$ from $\mathbf{x}_e$, $\mathbf{x}_s$, and $\mathbf{c}$.

2: Compute the embedded sequence:

$$\mathbf{E} = \text{Embed}(\mathbf{x}_e, \mathbf{x}_s, \mathbf{c}) + \mathbf{p}, \qquad \mathbf{H}^{(0)} \leftarrow \mathbf{E}.$$

3: **for** $l = 1$ to L **do**

4: Compute query, key, and value matrices:

$$\mathbf{Q}^{(l)} = \mathbf{H}^{(l-1)}\mathbf{W}_Q^{(l)}, \qquad \mathbf{K}^{(l)} = \mathbf{H}^{(l-1)}\mathbf{W}_K^{(l)}, \qquad \mathbf{V}^{(l)} = \mathbf{H}^{(l-1)}\mathbf{W}_V^{(l)}.$$

5: Apply scaled dot-product attention:

$$\mathbf{A}^{(l)} = \text{softmax}\left(\frac{\mathbf{Q}^{(l)}\mathbf{K}^{(l)\top}}{\sqrt{d_k}}\right)\mathbf{V}^{(l)}.$$

6: Update the hidden representation using multi-head attention and feed-forward layers:

$$\mathbf{H}^{(l)} = \text{FFN}\Big(\text{MultiHeadAttention}(\mathbf{H}^{(l-1)})\Big).$$

7: **end for**

8: Aggregate the final encoder representation $\mathbf{H}^{(L)}$ to obtain $\mathbf{h}_{\text{concat}}$.

9: Predict non-negative charging energy demand using the softplus output layer:

$$\hat{y} = \log\big(1 + \exp(\mathbf{W}\mathbf{h}_{\text{concat}} + \mathbf{b})\big).$$

10: Train model parameters by minimizing the forecasting error between y and $\hat{y}$.

11: **return** $\hat{y}$ as the predicted future EV charging energy demand.

The features are concatenated with contextual inputs $\mathbf{c}$ (e.g., encoded time; weather) and passed through a fully connected output layer with a softplus activation:

$$\mathbf{h}_{\text{concat}} = [\mathbf{h}_e^{(2)} \parallel \mathbf{h}_s^{(2)} \parallel \mathbf{c}] \tag{9}$$

$$\hat{y} = \log(1 + e^{\mathbf{W}\mathbf{h}_{\text{concat}} + \mathbf{b}}) \tag{10}$$

This structure enables the CNN to extract local trends in energy behavior and combine them with external signals for robust city-level prediction. In the above formulation, $\mathbf{x}_e = [x_e^{(1)}, x_e^{(2)}, \ldots, x_e^{(T)}]$ and $\mathbf{x}_s = [x_s^{(1)}, x_s^{(2)}, \ldots, x_s^{(T)}]$ denote the input time series representing hourly energy consumption and the number of charging sessions, respectively, over a time window of length T. The variables $\mathbf{w}_e^{(i)}$ and $\mathbf{w}_s^{(i)}$ are the convolutional kernels for the i-th convolutional layer applied to energy and session inputs, and $\mathbf{b}_e^{(i)}$ and $\mathbf{b}_s^{(i)}$ are their corresponding bias terms. The outputs $\mathbf{h}_e^{(1)}$ and $\mathbf{h}_s^{(1)}$ shown in (5) and (6) represent the feature maps produced by applying 1D convolution followed by the ReLU activation function. Subsequent layers $\mathbf{h}_e^{(2)}$ and $\mathbf{h}_s^{(2)}$ represented in (7) and (8) are generated through additional convolutional processing and max pooling. These extracted features are concatenated with contextual features $\mathbf{c}$, such as time-of-day, day-of-week, or weather-related variables, to form the combined feature vector $\mathbf{h}_{\text{concat}}$ shown in (9). This vector is then passed through

a fully connected layer with weights $\mathbf{W}$ and bias $\mathbf{b}$, producing the scalar output $\hat{y}$ through the smooth, differentiable softplus activation function $\log(1 + e^x)$ represented in (10).

CatBoost is a gradient-boosting decision tree model designed to handle categorical variables and temporal features effectively (Algorithm 3). We model the energy consumption at time $t + 1$ as a nonlinear function of historical and exogenous inputs (11):

$$\hat{y}_{t+1} = f_{\text{CatBoost}}(\mathbf{x}_{\text{hist}}, \mathbf{x}_{\text{weather}}, \mathbf{x}_{\text{time}}) \tag{11}$$

Here, $\mathbf{x}_{\text{hist}}$ includes lagged energy/session values and rolling aggregates (mean, max, and min) across windows (e.g., 6 h; 24 h), $\mathbf{x}_{\text{weather}}$ denotes weather features (temperature, humidity, etc.), and $\mathbf{x}_{\text{time}}$ includes cyclical encoded time variables (hour-of-day; day-of-week).

The CatBoost model is trained using the standard objective function for regression (12):

$$\mathcal{L} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \Omega(f) \tag{12}$$

where $\Omega(f)$ denotes regularization terms to penalize complexity. CatBoost's native handling of categorical features, robust performance on tabular data, and speed make it a strong baseline for hourly energy prediction.

Algorithm 2 CNN-Based Forecasting

Require: Historical energy series $\mathbf{x}_e = [x_e^{(1)}, \ldots, x_e^{(T)}]$, session-count series $\mathbf{x}_s = [x_s^{(1)}, \ldots, x_s^{(T)}]$, contextual feature vector $\mathbf{c}$ containing encoded time, weather, and traffic-flow variables.

Ensure: Future EV charging energy demand forecast $\hat{y}$.

1: Construct time-window samples from $\mathbf{x}_e$, $\mathbf{x}_s$, and $\mathbf{c}$ for city-level energy forecasting.

2: Normalize the continuous inputs and arrange $\mathbf{x}_e$ and $\mathbf{x}_s$ as one-dimensional temporal sequences.

3: Apply the first 1-D convolution and ReLU activation to the energy and session branches:

$$\mathbf{h}_e^{(1)} = \text{ReLU}\left(\mathbf{x}_e * \mathbf{w}_e^{(1)} + \mathbf{b}_e^{(1)}\right), \qquad \mathbf{h}_s^{(1)} = \text{ReLU}\left(\mathbf{x}_s * \mathbf{w}_s^{(1)} + \mathbf{b}_s^{(1)}\right).$$

4: Apply the second convolution, ReLU activation, and max-pooling operation:

$$\mathbf{h}_e^{(2)} = \text{MaxPool}\left(\text{ReLU}\left(\mathbf{h}_e^{(1)} * \mathbf{w}_e^{(2)} + \mathbf{b}_e^{(2)}\right)\right),$$

$$\mathbf{h}_s^{(2)} = \text{MaxPool}\left(\text{ReLU}\left(\mathbf{h}_s^{(1)} * \mathbf{w}_s^{(2)} + \mathbf{b}_s^{(2)}\right)\right).$$

5: Concatenate local temporal features with contextual inputs:

$$\mathbf{h}_{\text{concat}} = \left[\mathbf{h}_e^{(2)} \parallel \mathbf{h}_s^{(2)} \parallel \mathbf{c}\right].$$

6: Predict non-negative charging energy demand using the softplus output layer:

$$\hat{y} = \log\left(1 + \exp(\mathbf{W}\mathbf{h}_{\text{concat}} + \mathbf{b})\right).$$

7: Train CNN parameters by minimizing the forecasting error between the observed energy demand y and $\hat{y}$.

8: **return** $\hat{y}$ as the predicted future EV charging energy demand.

Algorithm 3 CatBoost-Based Forecasting

Require: Historical feature vector $\mathbf{x}_{\text{hist}}$, weather feature vector $\mathbf{x}_{\text{weather}}$, cyclical time feature vector $\mathbf{x}_{\text{time}}$, number of trees M.
Ensure: One-step-ahead EV charging energy demand forecast $\hat{y}_{t+1}$.
1: Construct the tabular input vector:

$$\mathbf{x}_t = [\mathbf{x}_{\text{hist}} \,\|\, \mathbf{x}_{\text{weather}} \,\|\, \mathbf{x}_{\text{time}}].$$

2: Include lagged energy values, session counts, rolling statistics, weather variables, and encoded calendar features in $\mathbf{x}_t$.
3: Define the CatBoost regression mapping for hourly energy prediction:

$$\hat{y}_{t+1} = f_{\text{CatBoost}}(\mathbf{x}_{\text{hist}}, \mathbf{x}_{\text{weather}}, \mathbf{x}_{\text{time}}).$$

4: Initialize the additive tree ensemble $F_0(\mathbf{x})$ using the training-target mean.
5: **for** $m = 1$ to M **do**
6: Compute residual errors from the current ensemble prediction:

$$r_i^{(m)} = y_i - F_{m-1}(\mathbf{x}_i).$$

7: Fit a regression tree $T_m(\mathbf{x})$ to the residuals $r_i^{(m)}$.
8: Update the ensemble prediction:

$$F_m(\mathbf{x}) \leftarrow F_{m-1}(\mathbf{x}) + \eta T_m(\mathbf{x}).$$

9: **end for**
10: Optimize the regularized squared-error objective:

$$\mathcal{L} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \Omega(f).$$

11: Obtain the final ensemble prediction:

$$\hat{y} = \sum_{m=1}^{M} T_m(\mathbf{x}_t).$$

12: **return** $\hat{y}_{t+1}$ as the predicted future EV charging energy demand.

In the CatBoost model, the input features $\mathbf{x}$ consist of historical hourly energy consumption, charging session counts, and contextual variables such as time-of-day, day-of-week, and weather data. Unlike convolutional models, CatBoost constructs an ensemble of decision trees $\{T_m\}_{m=1}^{M}$, where each tree T_m learns to correct the residual errors from previous trees. The predicted output $\hat{y}$ is computed as the sum of the outputs of all trees:

$$\hat{y} = \sum_{m=1}^{M} T_m(\mathbf{x}) \tag{13}$$

where M is the total number of trees, and each $T_m(\mathbf{x})$ is a regression tree mapping input features $\mathbf{x}$ to a scalar output (13). CatBoost handles categorical and numerical features efficiently and optimizes the ensemble to minimize a loss function appropriate for regression.

2.3. Implementation Details and Hyperparameter Settings

To ensure reproducibility, all models were implemented using a consistent preprocessing, training, and evaluation pipeline . The input data were aggregated at an hourly

resolution and transformed into supervised learning samples using a rolling historical window of two weeks. The training and testing split was set to 90% and 10%, respectively, while hyperparameter tuning was performed on the training subset using validation-based model selection. All numerical input features were normalized using min–max scaling based only on the training data.

The Transformer model consisted of L encoder layers, H attention heads, an embedding dimension of d_{model}, and a feed-forward dimension of d_{ff} (Table 1). Positional information was incorporated using [sinusoidal/learnable] positional encoding. Dropout with a rate of p_{drop} was applied after the attention and feed-forward blocks. The final hidden states were aggregated into $\mathbf{h}_{\text{concat}}$ and passed through a fully connected regression head with softplus activation to ensure non-negative energy-demand predictions.

The CNN model used one-dimensional convolutional layers for the historical energy and session-count sequences (Table 1). Each convolutional block consisted of convolution, ReLU activation, max-pooling, and dropout. The extracted feature maps were flattened and concatenated with contextual features, including weather, traffic flow, and time encodings, before being passed to a fully connected softplus output layer.

The CatBoost model was trained as a gradient-boosted decision tree regressor using historical lag features, rolling statistics, weather variables, traffic-flow variables, and cyclical time encodings (Table 1). The number of boosting iterations was set to 1000, and the squared-error loss was used for regression.

Table 1. Model hyperparameters used in this study.

Model	Hyperparameter	Value
Transformer	Encoder layers	2
	Attention heads	4
	Embedding dimension d_{model}	64
	FFN dimension d_{ff}	128
	Positional encoding	Learnable
	Dropout	0.10
	Optimizer	Adam
	Learning rate	10^{-3}
	Batch size	64
	Epochs	100
	Early stopping	Patience = 10
	Output activation	Softplus
CNN	Convolutional layers	2
	Kernel size	3
	Filters	64, 32
	Activation	ReLU
	Pooling	Max pooling
	Dropout	0.20
	Optimizer	Adam
	Learning rate	10^{-3}
	Batch size	64
	Epochs	100
	Early stopping	Patience = 10
	Output activation	Softplus

Table 1. *Cont.*

Model	Hyperparameter	Value
CatBoost	Iterations	1000
	Tree depth	6
	Learning rate	0.03
	Loss function	RMSE
	L2 regularization	3
	Early stopping	50 rounds
All models	Input resolution	Hourly
	Input window	168 h
	Forecast horizon	1 h ahead
	Data split	90%/10% (Training/Testing)
	Feature scaling	Min–max normalization

2.4. Model Integration and Application

Each model is trained independently and evaluated on multiple U.S. metropolitan regions for charging stations [30] and traffic and congestion [31] and weather data [32]. Rather than focusing on EVSE power categories, all EVSEs within a given city or BUS system are aggregated. Forecasting is then performed at the regional level for operational planning and infrastructure management. The outputs from CNN, CatBoost, and the proposed Transformer models are compared individually depending on the downstream application (e.g., peak load estimation, daily demand curves, etc.). This integrated multi-model approach enables flexible and accurate energy prediction tailored to the characteristics and scale of individual urban regions. Numerical simulations were conducted using Python 3.12 via the PyTorch 2.7, Keras 3.8, and TensorFlow 2.19 libraries. Computational workloads were executed on a workstation featuring an Intel Core Ultra Series 2 7 255HX CPU, 32 GB of RAM, and a 12 GB discrete GPU.

This study also aims to forecast the load and energy demand associated with the large number of charging stations distributed across metropolitan areas, accounting for the variability of EV charging behavior across different weekdays and seasonal conditions. To evaluate the proposed Transformer-based forecasting approach, five major metropolitan regions were selected from the dataset, as summarized in Table 2. In this analysis, charging station power levels are not differentiated. Instead, all EVSEs within each metropolitan or electrical BUS system regardless of their specific characteristics are aggregated and used as input to the model, which then predicts energy demand and load at future time points.

Table 2. Properties for load and energy demand forecasting across selected US cities.

Metric	Baltimore	Denver	Detroit	Las Vegas	Los Angeles	Philadelphia
Total Number of EVSEs	1420	1175	860	980	3200	1500
Load Power EVSEs (kW)	3.7–150	3.7–150	3.7–150	3.7–150	3.7–150	3.7–150
Location of EVSE	Urban	Suburban	Urban	Urban	Urban, Suburban	Suburban
Energy Demand per day (kWh)	>3672	>2579	>1987	>2469	>2237	>1892
Installed Max. Load (kW)	>9872	>8799	>7854	>8003	>9741	>8658

2.5. Loss Function and Performance Metrics

The model is trained to minimize the mean squared error loss $\mathcal{L}_{\text{MSE}}$, computed over a dataset of N samples by comparing the predicted output $\hat{y}^{(i)}$ to the ground truth value $y^{(i)}$

for each instance shown in (14). The model is trained by minimizing the mean squared error between predicted and actual values y, in units of kWh^2:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}^{(i)} - y^{(i)}\right)^2 \tag{14}$$

To quantitatively assess the performance of the CatBoost, CNN, and Transformer models used in this work, we employ three standard regression evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), as shown in (15)–(17). RMSE is defined as the square root of the average squared differences between the predicted values $\hat{y}_i$ and the true target values y_i. MAE, on the other hand, calculates the average of the absolute differences between predictions and actual values, providing a linear and more interpretable measure of prediction accuracy. Finally, MAPE expresses the prediction error as a percentage, normalized by the actual values, which allows for scale-independent evaluation.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{15}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{16}$$

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{17}$$

2.6. Data Processing and Fusion

Raw charging session records were first aggregated to an hourly resolution to ensure temporal consistency across all data sources. Missing values were handled using linear interpolation for short gaps and forward filling for longer gaps, while outliers exceeding three standard deviations were removed. All input features were normalized using min–max scaling prior to model training.

Multiple data sources, including charging demand, temporal indicators, and exogenous features, were fused using a feature-level alignment strategy. Specifically, all variables were synchronized to a common time index and concatenated into a unified input tensor, preserving temporal order while enabling joint learning of heterogeneous information. This fusion approach avoids manual weighting and allows the model to learn feature relevance implicitly during training.

Traffic flow data extracted from the US Traffic Congestions (2016–2022) database [31]. This dataset aggregates nationwide traffic anomalies captured via continuous real-time streaming data, primarily sourced from GPS/probe vehicles and regional loop detectors. TF data was spatially linked to EVSE locations by identifying the nearest major traffic point-of-interest or roadway corridor within a maximum spherical radius of 1.5 miles of each charging station.

Weather data is sourced from the National Centers for Environmental Information (NCEI) Normals Hourly dataset [32], collected via localized automated airport weather stations. The full list of weather features utilized comprises the following: air temperature, dew point, relative humidity, wind speed, precipitation accumulation, and visibility.

Beyond basic feature-level alignment and concatenation, temporal synchronization was performed by aggregating both streams into identical 1 h intervals using forward-filling for sporadic traffic anomalies and linear interpolation for missing weather observations.

This explicit data characterization provides critical context for the ablation studies, which indicate high TF utility but marginal weather impact. Because extreme weather anomalies (e.g., heavy precipitation) inherently trigger downstream traffic delays and reduced speeds, the high-fidelity probe data likely acts as a proxy for adverse weather. Consequently, severe collinearity between precipitation metrics and traffic congestion metrics diminishes the marginal information gain provided by the weather features during multi-modal fusion.

3. Model Performance Across Spatial and Temporal Dimensions

This section compares the short-term energy-forecasting performance of CatBoost, CNN, and proposed models. Results are reported for eight areas with the highest counts of charging stations—Baltimore, Denver, Detroit, Las Vegas, Los Angeles, and Philadelphia—using standard evaluation metrics: MAPE, MAE, RMSE, and the coefficient of determination (R^2). Forecasting is performed for both 24 h and 168 h time windows using a 1 h ahead setup, as well as for multi-step horizons of 6 h, 12 h, 18 h, and 24 h ahead. This analysis provides a detailed comparison of model accuracy and generalization across different regions and forecasting intervals. Additionally, a sensitivity analysis is conducted to assess model robustness under input perturbations, including three scenarios: (i) deterministic input variation by ±5%, 10%, 20%, etc.; (ii) stochastic perturbation using normally distributed noise with zero mean and standard deviations corresponding to ±5%, 10%, 20%, etc.; and (iii) random perturbation with uniformly distributed noise in the range of ±5%, 10%, 20%, etc.

For the multi-horizon experiments, a direct lead-time-specific forecasting strategy was adopted. Let h denote the forecast horizon. For each evaluated horizon $h \in 1, 6, 12, 18, 24, 168$ h, the supervised learning dataset was constructed by pairing the historical input window $\mathbf{X}t - L + 1$ with the corresponding future target $yt + h$. Therefore, each model learns the direct mapping.

$$\hat{y}t + h = f_h(\mathbf{X}t - L + 1, \mathbf{z}_t), \tag{18}$$

where $\mathbf{z}_t$ represents the associated temporal, calendar, weather, and charging-related explanatory variables. The 1 h ahead case is used as the base setup, while the same training and evaluation procedure is repeated for the longer horizons by changing h. No recursive or iterative forecasting scheme is used; predicted values are not fed back into the model as inputs. Consequently, the error increase at longer horizons reflects the reduced predictability of EV charging demand as the lead time increases, rather than accumulated recursive error propagation.

3.1. Short-Term Forecasting Accuracy for a 24 h Window

In the following, the results for the 1 h ahead predictions over a 24 h horizon are presented for eight selected areas: Baltimore, Denver, Detroit, Las Vegas, Los Angeles, Philadelphia, Portland, and Washington. These areas were chosen based on having the highest number of charging stations in the dataset. To evaluate the performance of the CatBoost, CNN, and Transformer algorithms, key evaluation metrics were computed. These metrics are illustrated using Baltimore as an example, while the results for all selected areas are reflected in the R^2 values for each respective algorithm.

Baltimore was selected as a representative example due to the characteristic behavior of its prediction performance, particularly in terms of the R^2 score. The city exhibits RMSE and MAE of approximately 37 kWh and 27 kWh, respectively (Figure 2)—providing grid operators and local authorities with actionable insights to guide reinforcement of the power grid in response to rising electric-vehicle charging demand across urban and suburban

settings. In addition, the 1 h resolution forecasts can inform planning for renewable-energy integration and storage deployment.

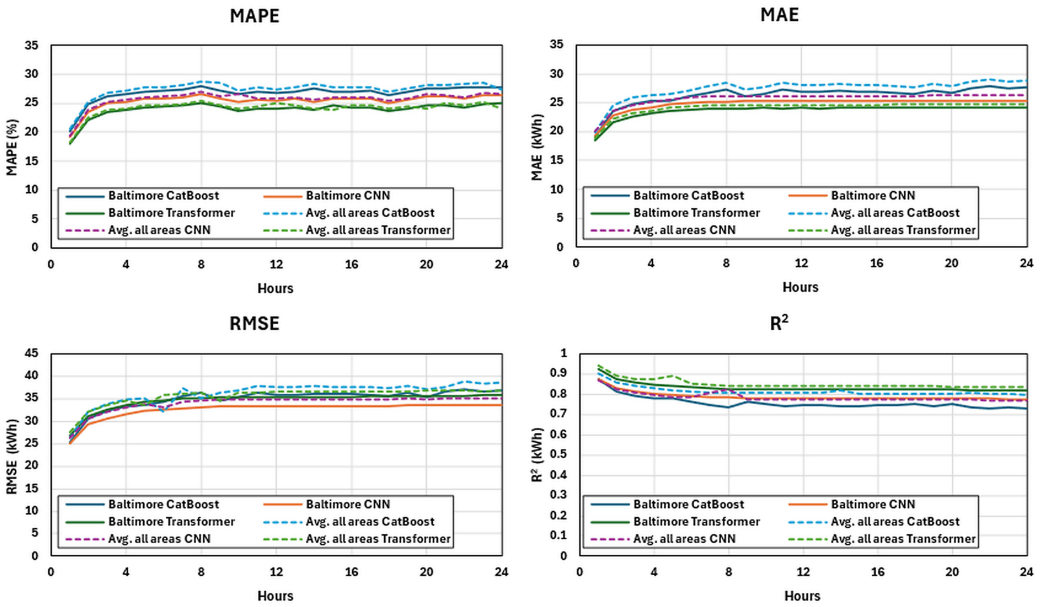


Figure 2. CatBoost, CNN, and proposed Transformer results for 1 h ahead 24 h predictions for Baltimore.

Overall, both subfigures in Figure 2 indicate that performance metrics tend to converge toward the end of the 24 h prediction window, with most predictions stabilizing within approximately 5–7 h. Figure 2 reports the R^2 results for CatBoost, CNN, and the proposed Transformer. CatBoost displays greater variability than the neural models, whereas the CNN yields smoother trajectories. The Transformer consistently attains the highest R^2, exceeding both CNN and CatBoost by a few percentage points (≈3–5%) across most areas. Baltimore showed a strong performance, with R^2 decreasing from about 0.95 at the start of the prediction period to approximately 0.80–0.77 at the end, and convergence typically occurring within 7–10 h.

3.2. Seasonal Effects for a 24 h Window for MAPE

This subsection analyzes the seasonal variation in energy demand forecasting performance across different models and regions using 24 h MAPE metric.

Across all four seasons, the 24 h MAPE curves exhibit a rapid increase during the early morning hours followed by a plateau throughout the remainder of the day in Figure 3. However, the absolute error levels and model rankings vary notably by season and region. In every season, the proposed Transformer model consistently yields the lowest MAPE, indicating superior performance, while the CatBoost model exhibits the highest error rates. CNN falls in between, performing better than CatBoost but worse than Transformer. For example, in the summer subplot, the average-area Transformer model starts at approximately 20% MAPE at hour 0 and rises modestly to around 25% by hour 4, whereas the Baltimore CatBoost model begins near 20% and increases to over 30%. The CNN models for both Baltimore and the average case generally remain between these two extremes. Seasonally, the highest forecast errors are observed in summer (peaking around 30–32%), followed by spring and fall (approximately 25–28%), with winter exhibiting the lowest MAPE values across all models (remaining below 20%). In general, models trained and evaluated on the Baltimore region tend to perform better (i.e., exhibit lower MAPE) than those averaged over all areas, particularly for the proposed Transformer model. This suggests that localized models can outperform generalized ones under certain conditions. Importantly, the proposed Transformer consistently demonstrates the most

robust performance across the diurnal cycle and all seasonal scenarios, reinforcing its suitability for short-term load forecasting tasks.

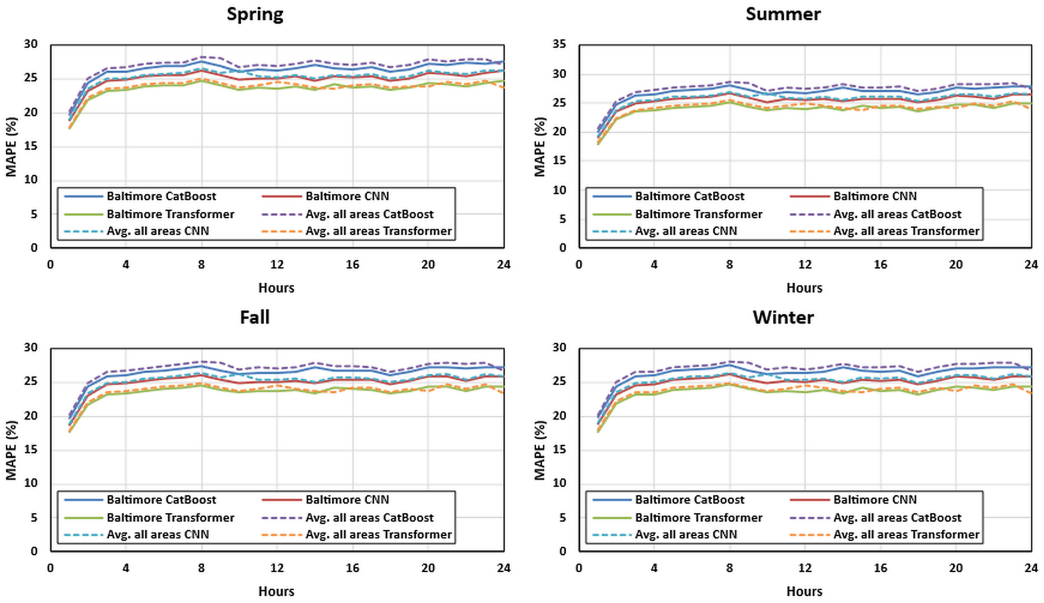


Figure 3. Seasonal effects for 24 h window for MAPE.

3.3. Traffic Flow and Weather Impacts for 24 h Window Forecasting Accuracy

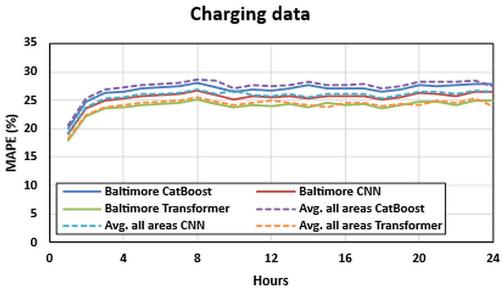
In this section a comparative analysis of model performance in terms of MAPE over a 24 h horizon for three data combinations is conducted: charging data only, charging plus traffic flow data, and charging data together with traffic and weather data.

In all input data scenarios (Figure 4), the proposed Transformer model consistently achieves the lowest MAPE across the 24 h prediction horizon, outperforming the CNN and CatBoost models (with CNN yielding lower error than CatBoost but remaining inferior to the proposed Transformer). Incorporating traffic data alongside charging data (the Charging + Traffic scenario) significantly reduces prediction error compared to using charging data alone, underscoring the value of traffic features. However, further adding weather data on top of traffic (the full Charging + Traffic + Weather scenario) provides only marginal improvement, with MAPE values nearly indistinguishable from those in the Charging + Traffic scenario, indicating that weather features contribute minimal additional predictive value beyond traffic data. These performance patterns are observed in both the Baltimore-specific results and the average across all areas: for instance, the Baltimore Transformer curve lies consistently below the Baltimore CNN and Baltimore CatBoost curves at every forecast hour, mirroring the advantage of the Avg. all areas Transformer over the Avg. all areas CNN and Avg. all areas CatBoost curves in the overall results. Moreover, MAPE generally increases with longer forecast horizons for all model–data combinations, but the proposed Transformer maintains a clear performance lead at each hourly step, and the error reduction gained by including traffic data persists throughout the 24 h window, whereas the incremental benefit of weather remains negligible.

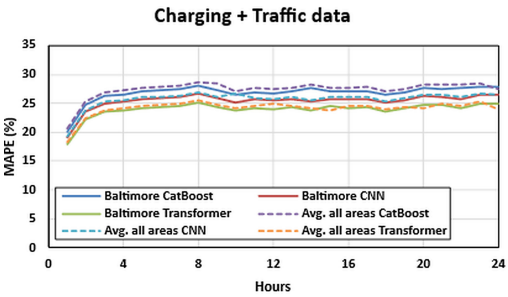
3.4. Multi-Day Evaluation of 1 h Ahead Energy Predictions

Figure 5 presents prediction results for the 1–168 h forecast horizon using CatBoost, CNN, and a Transformer for the Baltimore case, and Figure 6 reports the corresponding R^2 for selected areas. For CatBoost, RMSE and MAE increase steadily with the forecast horizon, reaching up to 42 kWh and approximately 30 kWh, respectively, at 168 h, which is consistent with the increased uncertainty of longer-term predictions. In comparison, the CNN exhibits a more consistent error trend and achieves lower overall error levels; specifically, it reduces

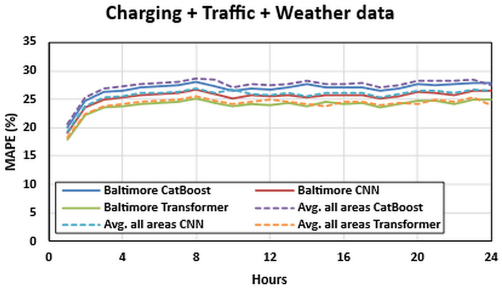
RMSE by about 5 kWh and MAE by approximately 3 kWh, corresponding to improvements of 12% and 8%, respectively, relative to CatBoost. The proposed Transformer attains the lowest errors throughout the horizon and further improves upon the CNN, yielding mean reductions equivalent to approximately 4.6–5.4% across metrics. As expected, all models exhibit higher error levels for 168 h forecasts relative to the 24 h forecasts discussed earlier.



(**a**) Charging behavior data only results

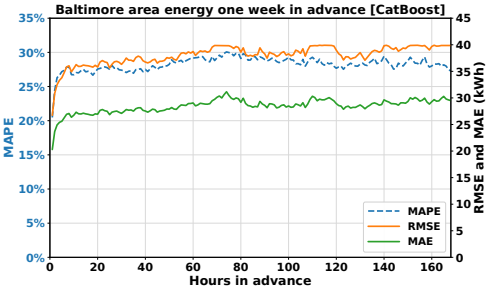


(**b**) Charging behavior + Traffic flow data

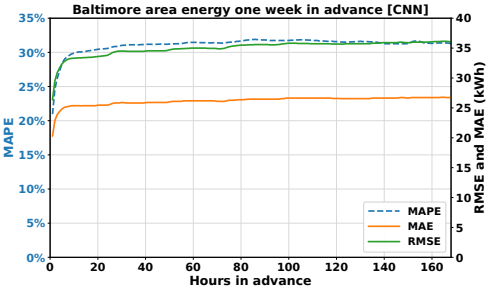


(**c**) Charging behavior + Traffic flow + Weather data

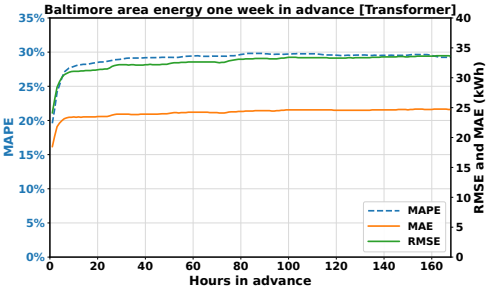
Figure 4. Impact of different input data combinations on forecasting performance: (**a**) Charging behavior only. (**b**) Charging behavior with traffic flow. (**c**) Charging behavior with traffic flow and weather data.



(**a**) CatBoost metrics

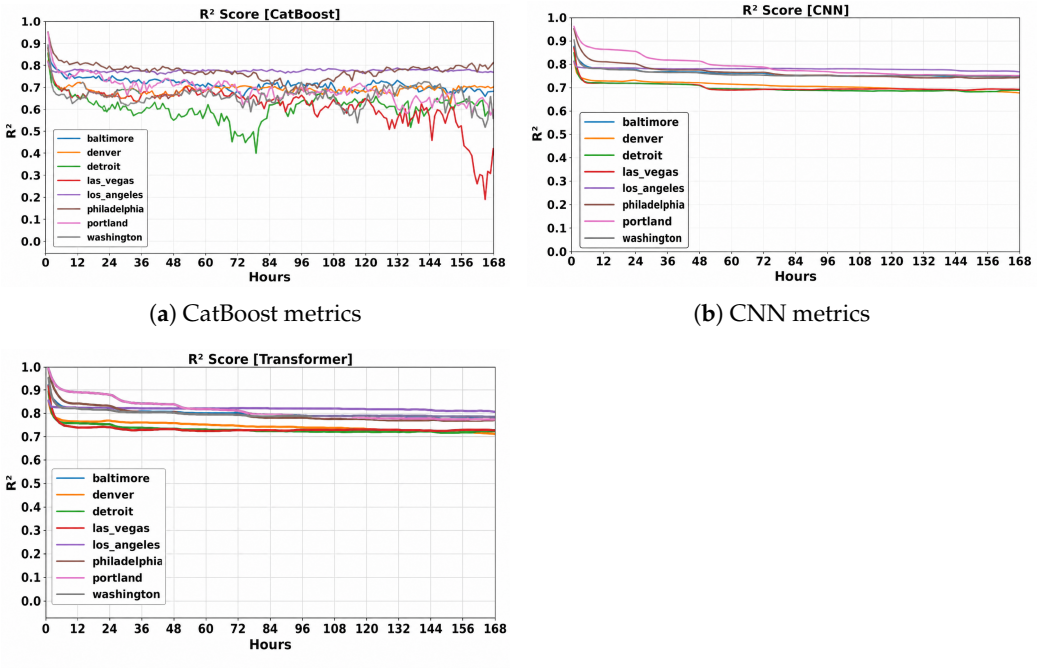


(**b**) CNN metrics



(**c**) Transformer metrics

Figure 5. MAPE, RMSE, and MAE for 1 h ahead, 168 h predictions in Baltimore: (**a**) CatBoost metrics. (**b**) CNN metrics. (**c**) Transformer metrics.



(**a**) CatBoost metrics

(**b**) CNN metrics

(**c**) Transformer metrics

Figure 6. R^2 for 1 h ahead, 168 h predictions across eight areas: (**a**) CatBoost metrics. (**b**) CNN metrics. (**c**) Transformer metrics.

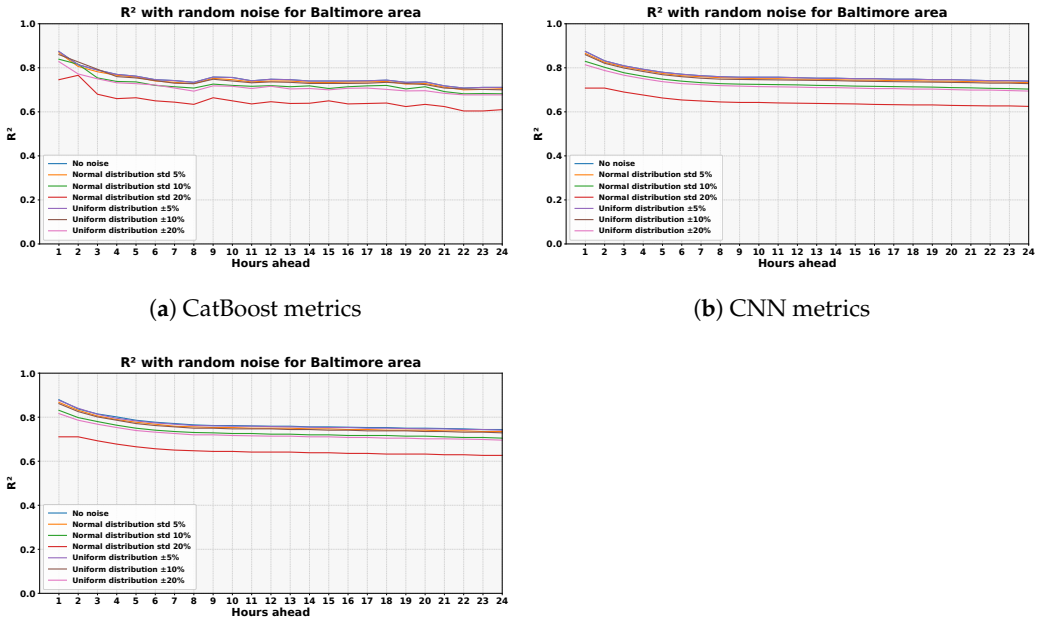
The R^2 scores for the 168 h forecast horizon (Figure 6) vary substantially across models and relative to the 24 h predictions. CatBoost exhibits highly erratic behavior, with pronounced drops in R^2 for specific areas—most notably, Detroit and Las Vegas experience marked declines between 36–90 h and 150–168 h, respectively. These fluctuations cannot be traced to identifiable anomalies or outliers in the dataset; by contrast, the CNN remains stable and follows a clear, near-linear trend. An exception arises in Portland, where a gradual decline in R^2 persists until approximately 72 h. Beyond this, the distribution of R^2 values across regions largely remains within the range observed for the 24 h predictions. Consistent with the RMSE, MAE, and MAPE analyses over the 168 h horizon, the R^2 results indicate that the Transformer attains the best overall performance, with mean R^2 improvements of approximately 3.8–4.9% over the CNN and larger gains over CatBoost.

3.5. Short-Term Forecast Sensitivity over a Daily Period

In this section, the results of the sensitivity analysis are presented to further demonstrate the robustness and reliability of the applied algorithms for short-term energy prediction. As an illustrative example, the R^2 results for the 1 h ahead predictions over a 24 h horizon are shown for the area of Baltimore. The analysis includes the original predictions without added noise, as well as predictions based on artificially perturbed inputs. Specifically, two types of input variation were introduced: normally distributed noise with standard deviations ranging from ±5% to ±20%, and uniformly distributed noise within the same percentage range. Uniformly distributions are also used for similar forecasts with a focus on proving reliability with a unit control of MAE, RMSE, and R^2 for PV load forecasting in [33]. Another work [34] proved that uncertainty and uniformly distribution can be very effective to prove the results of RE and transmission line planning. In ref. [35] uniformly distributed parameters are used to demonstrate improved suitability for forecasting net-load error.

Figure 7 illustrates that the prediction performance remains stable and linear across most scenarios, with the R^2 values converging after approximately 10 h. The results confirm

the robustness of CatBoost and CNN as well as the proposed Transformer model under moderate input perturbations. Only the scenario involving normally distributed noise with a ±20% standard deviation shows a notable deviation, with the R^2 dropping to around 0.65 compared to the average value of approximately 0.75 in the other scenarios. These findings support the reliability of the forecasting models, even under realistic fluctuations and uncertainties in the input data.



(**a**) CatBoost metrics

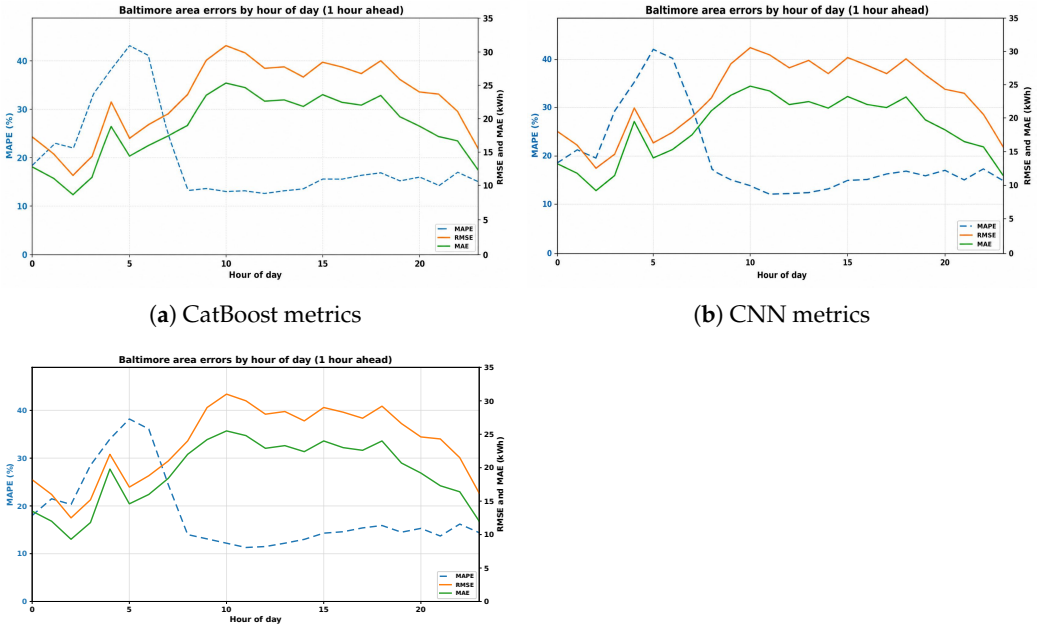
(**b**) CNN metrics

(**c**) Transformer metrics

Figure 7. R^2 results for 1 h ahead, 24 h sensitivity analysis in Baltimore: (**a**) CatBoost metrics. (**b**) CNN metrics. (**c**) Transformer metrics.

3.6. Single-Day Energy Forecasts with One-Hour Horizon

The following results present the error metrics across different hours of the day, along with the distribution of actual and predicted energy demand for a representative location. Figure 8 illustrates the hourly variation of MAPE, RMSE, and MAE for the CatBoost, CNN, and Transformer model. The *x-axis* denotes the hour of the day in 24 h format, where, for instance, hour 5 corresponds to 05:00. These results provide insight into the temporal fluctuation of prediction accuracy over a daily cycle. Notably, MAPE values for both models peak during early morning hours (02:00–08:00) (blue dashed line), reaching up to 45%, likely due to reduced charging activity during these periods. As MAPE is sensitive to low denominators, higher errors during low-demand hours are expected. During daytime, MAPE stabilizes within the range of 12–18%, indicating improved performance in periods of increased charging activity. In contrast, RMSE and MAE follow the overall demand trend, with lower values corresponding to off-peak hours and higher values corresponding to periods of elevated demand (e.g., 05:00–12:00). CNN consistently outperforms CatBoost across all metrics, particularly in RMSE and MAE. For instance, CatBoost exhibits a peak RMSE of 34 kWh at 10:00, while CNN reduces this peak to approximately 30 kWh, reflecting a 12% improvement. Similarly, MAE results demonstrate a modest reduction of approximately 3% in favor of CNN. Overall, these findings confirm the superior predictive performance of CNN, particularly during high-demand hours, as evidenced by consistently lower RMSE and MAE across the daily cycle.



(**a**) CatBoost metrics

(**b**) CNN metrics

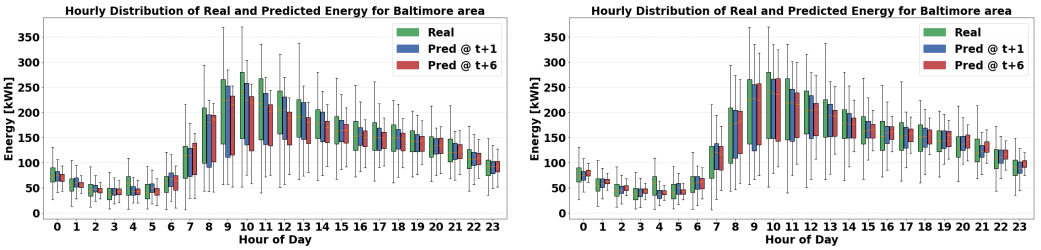
(**c**) Transformer metrics

Figure 8. Metrics for a specific hour of the day in Baltimore: (**a**) CatBoost metrics. (**b**) CNN metrics. (**c**) Transformer metrics.

In addition to the error metrics discussed previously (Figure 8), Figure 9 presents the aggregated real and predicted energy values over a 24 h period for a specific day in Baltimore. This visualization facilitates a direct comparison between absolute prediction metrics, such as RMSE, and the actual forecasted energy demand at each hourly interval. Figure 9 displays the mean observed energy values (green bars), along with the corresponding 1 h ahead (blue) and 6 h ahead (red) forecasts. Error bars denote the range of deviation between predicted and actual values, providing a measure of uncertainty at each time step. The magnitude and height of each bar reflect the model's ability to capture the variability in energy demand throughout the day. As seen in previous analyses, energy consumption during off-peak hours (02:00–06:00) remains low, averaging around 50 kWh. Both 1 h and 6 h ahead forecasts tend to underestimate demand during these periods, likely due to limited charging activity. Demand rises sharply during the morning hours, peaking at approximately 275 kWh around 10:00. Subsequently, energy usage declines steadily, reaching about 80 kWh by 00:00. A comparison between CatBoost and CNN models reveals that CNN yields slightly improved forecasts, particularly during high-demand periods. At the peak hour of 10:00, the CNN model delivers more accurate predictions than CatBoost, especially in the 6 h ahead horizon. These results further underscore the potential of CNN for mid-range forecasting tasks where capturing rapid fluctuations in demand is critical.

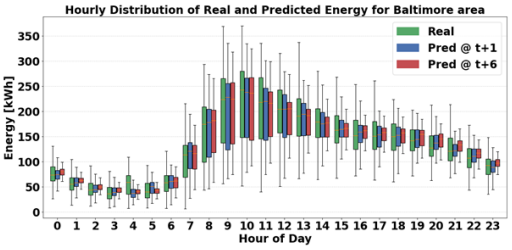
3.7. Forecast Validation Through Comparison with Observed Data

In this section, the results of the real versus predicted energy demand are presented for the example of Baltimore. Figure 10 displays the scatter plots for CatBoost, CNN, and Transformer models. In each plot, the red line indicates the ideal prediction line, where the predicted values match the actual energy demand precisely. The proximity of the blue data points to the red line reflects the accuracy of the predictions: the closer the points are, the better the performance.



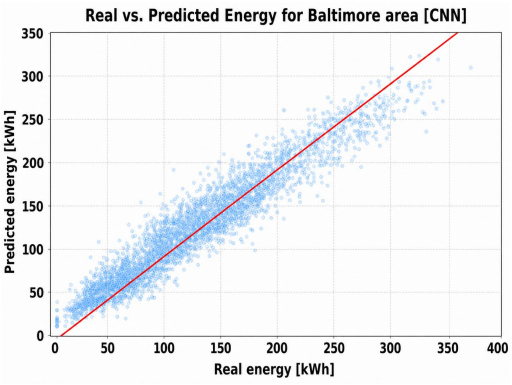
(**a**) CatBoost distribution

(**b**) CNN distribution

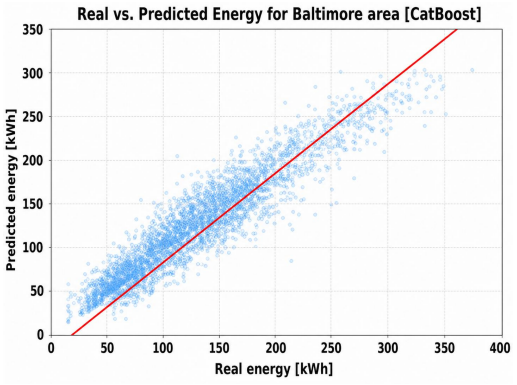


(**c**) Transformer distribution

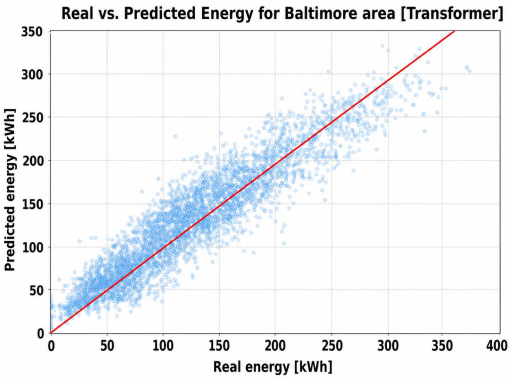
Figure 9. Real vs. predicted distribution results for Baltimore using (**a**) CatBoost, (**b**) CNN, and (**c**) Transformer.



(**a**) CatBoost predicted vs. real energy



(**b**) CNN predicted vs. real energy



(**c**) Transformer predicted vs. real energy

Figure 10. R^2 results for real vs. predicted energy in Baltimore: (**a**) CatBoost, (**b**) CNN, and (**c**) Transformer.

Most predictions lie between 10 kWh and 300 kWh. A noticeable trend is that, particularly beyond a real energy demand of around 150 kWh, the predicted values tend to underestimate the actual demand. This is indicated by an increasing number of blue points falling below the red line. This behavior may be attributed to the models' tendency to

regularize extreme values and avoid overfitting to high-demand peaks, which are relatively infrequent in the training data.

For transparency, the zero-prediction values were retained in the plot to convey the full range and behavior of the dataset. However, during data preprocessing, all energy demand entries below 5 kWh were excluded to improve the performance metrics such as MAPE, RMSE, MAE, and R^2. This filtering helps ensure that the evaluation of the forecasting models is not disproportionately influenced by extremely low or noisy values, which are less relevant for operational energy planning.

4. Discussion

To frame the implications of our findings, we first synthesize the comparative results across all four evaluation scenarios reported in Table 3.

Table 3. Comparison with reference methods.

	Short-Term Forecasting Accuracy for a 24 h Window				Multi-Day Evaluation of 1 h Ahead Energy Predictions				Single-Day Energy Forecasts with One-Hour Horizon				Forecast Validation Through Comparisonwith Observed Data			
Model	**R^2**	**RMSE**	**MAE**	**MAPE**	**R^2**	**RMSE**	**MAE**	**MAPE**	**R^2**	**RMSE**	**MAE**	**MAPE**	**R^2**	**RMSE**	**MAE**	**MAPE**
Reference 1 CatBoost	0.869	—	27.59	28.11	0.823	—	28.07	27.86	0.792	—	30.11	29.42	0.711	—	29.79	27.88
Reference 2 CNN	0.891	—	25.64	26.36	0.831	—	26.99	24.15	0.809	—	27.87	27.52	0.834	—	26.43	26.90
Reference 3 Transformer	0.932	—	24.35	24.62	0.871	—	24.86	22.87	0.842	—	24.36	23.64	0.819	—	23.03	23.04
Proposed CatBoost	0.856	35.09	26.27	26.52	0.798	38.14	27.82	26.99	0.796	39.41	27.16	26.78	0.699	25.34	27.41	26.87
Proposed CNN	0.902	34.21	25.47	24.89	0.861	37.12	25.74	23.41	0.823	35.74	25.41	24.86	0.795	23.49	24.22	25.31
Proposed Transformer	0.951	32.14	24.28	23.96	0.897	36.62	23.45	21.13	0.867	32.85	22.36	22.45	0.812	21.37	21.36	20.89

Direct numerical comparison with previously published EV charging demand forecasting studies is inherently challenging due to differences in datasets, spatial aggregation levels, forecasting horizons, and operational constraints. Absolute error metrics such as RMSE and MAE are highly scale-dependent and therefore not directly comparable across studies. In contrast, relative and scale-independent metrics, particularly MAPE and, to a lesser extent, R^2, allow for qualitative benchmarking of forecasting performance. Accordingly, this study emphasizes internally consistent model comparison under identical experimental conditions, while external performance is discussed in relation to reported ranges in the literature rather than direct numerical ranking.

For comparability, we harmonized metrics and settings across sources. Results were brought to a 1 h resolution and kWh units; where needed we rescaled/aggregated data and recomputed R^2, RMSE, MAE, and MAPE under consistent definitions (MAPE stabilized with a small ε). From each reference we extracted the closest matching subset to four canonical scenarios (24 h window; multi-day 1 h ahead; single-day 1 h ahead; validation vs. observations), and left unreported metrics blank (dashes) rather than inferring values. Table 3 compares six models (three references [36–38] and three proposed) across four evaluation scenarios using R^2, RMSE, MAE, and MAPE. Overall, the proposed Transformer attains the highest R^2 in the first three scenarios—24 h window (0.951), multi-day 1 h ahead (0.897), and single-day 1 h ahead (0.867)—and simultaneously yields the lowest error magnitudes among the proposed models (e.g., RMSE/MAE/MAPE of 32.14/24.28/23.96, 36.62/23.45/21.13, and 32.85/22.36/22.45, respectively). In contrast, the proposed CNN and proposed CatBoost are consistently weaker in both R^2 and error metrics in these scenarios.

Among the reference baselines, the proposed Transformer outperforms CNN and CatBoost in R^2 where reported for the first three scenarios (0.932, 0.871, and 0.842, respectively), but the proposed counterpart still provides a clear margin (Table 4). In the fourth scenario

(forecast validation), the reference CNN attains the top R^2 (0.834), exceeding both the reference Transformer (0.819) and the proposed Transformer (0.812); however, the proposed Transformer achieves the smallest absolute and relative errors (RMSE/MAE/MAPE of 21.37/21.36/20.89), improving upon the proposed CNN (23.49/24.22/25.31) and CatBoost (25.34/27.41/26.87). Taken together, these results indicate that attention-based architectures deliver the most accurate and stable short-horizon EV demand forecasts across diverse settings, with consistent reductions in RMSE/MAE/MAPE relative to CNN and gradient-boosted tree alternatives (Table 5). Dashes in the table denote unreported metrics for some reference methods and do not affect the relative ranking of the proposed models.

This section presents a detailed discussion of the performance of three predictive models—CatBoost, CNN, and proposed Transformer—applied to short-term electricity load forecasting across eight major U.S. cities: Baltimore, Denver, Detroit, Las Vegas, Los Angeles, Philadelphia, Portland, and Washington. Forecasts were evaluated at multiple time horizons (h = 1, 6, 12, 18, and 24 h) using four standard metrics: coefficient of determination (R^2), RMSE, MAE, and MAPE. Full performance results are summarized in Tables 4–9.

Table 4. CNN forecasting metrics—part 1 (Baltimore, Denver, Detroit, and Las Vegas).

H	Baltimore				Denver				Detroit				Las Vegas			
(h)	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
1	0.875	25.2	19.4	20.2	0.811	25.1	18.2	29.7	0.847	20.5	15.0	19.2	0.869	18.4	13.9	11.1
6	0.789	32.7	25.0	28.7	0.732	29.9	21.6	38.3	0.721	27.7	20.7	29.4	0.715	27.2	21.1	18.4
12	0.781	33.3	25.3	29.9	0.727	30.2	21.8	38.8	0.719	27.8	20.8	30.0	0.702	27.8	21.6	19.2
18	0.779	33.4	25.3	30.1	0.726	30.2	21.9	38.9	0.716	27.9	20.9	30.3	0.703	27.8	21.5	19.1
24	0.775	33.7	25.4	30.3	0.730	30.0	21.8	38.4	0.715	28.0	20.9	30.4	0.702	27.8	21.5	19.1

Table 5. CNN forecasting metrics—part 2 (Los Angeles, Philadelphia, Portland, and Washington).

H	Los Angeles				Philadelphia				Portland				Washington			
(h)	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
1	0.811	54.4	40.8	33.2	0.950	12.2	9.2	12.4	0.961	15.0	11.5	10.9	0.903	18.2	14.0	11.6
6	0.783	58.4	43.7	34.7	0.827	22.6	16.3	24.3	0.874	26.9	19.5	20.2	0.783	27.2	20.8	18.5
12	0.783	58.4	43.7	34.6	0.810	23.7	17.0	25.6	0.863	28.2	20.3	21.3	0.779	27.4	20.9	18.7
18	0.783	58.4	43.7	34.7	0.805	24.0	17.2	26.0	0.859	28.6	20.4	21.6	0.775	27.7	21.1	18.9
24	0.782	58.5	43.7	34.5	0.801	24.3	17.3	26.3	0.854	29.1	20.7	22.0	0.773	27.7	21.2	18.9

Table 6. CatBoost forecasting metrics—part 1 (Baltimore, Denver, Detroit, and Las Vegas).

H	Baltimore				Denver				Detroit				Las Vegas			
(h)	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
1	0.8740	25.22	19.09	20.10	0.8190	24.53	17.63	27.66	0.8533	20.06	14.52	17.72	0.8776	17.82	13.32	10.68
6	0.7655	34.42	26.03	27.24	0.7039	31.37	22.65	36.39	0.6381	31.51	23.26	28.73	0.6858	28.55	22.07	18.18
12	0.7456	35.85	26.95	26.78	0.7212	30.44	22.12	36.69	0.6124	32.61	24.48	29.34	0.6736	29.10	22.32	18.44
18	0.7492	35.59	26.57	26.42	0.7030	31.42	22.77	36.97	0.5971	33.25	25.22	30.20	0.6894	28.38	22.00	18.12
24	0.7297	36.95	27.65	27.86	0.7086	31.12	22.58	35.98	0.6296	31.88	23.81	28.83	0.6913	28.30	21.85	18.07

Table 7. CatBoost forecasting metrics—part 2 (Los Angeles, Philadelphia, Portland, and Washington).

H	Los Angeles				Philadelphia				Portland				Washington			
(h)	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
1	0.8135	54.08	40.63	31.74	0.9495	12.21	9.27	11.91	0.9458	17.74	13.21	11.01	0.8908	19.25	14.66	11.49
6	0.7697	60.10	45.04	35.33	0.8200	23.05	17.48	23.03	0.7716	36.42	27.64	23.85	0.5912	37.25	28.98	22.19
12	0.7777	59.04	44.23	36.17	0.8129	23.50	17.76	23.75	0.7617	37.20	28.70	25.04	0.7049	31.64	24.46	19.92
18	0.7665	60.51	45.34	35.04	0.7992	24.35	18.24	23.95	0.7685	36.67	27.80	24.18	0.7056	31.61	24.40	19.96
24	0.7691	60.18	45.17	36.63	0.7950	24.60	18.53	24.13	0.7452	38.47	29.57	25.66	0.6881	32.53	25.73	20.74

Overall, the proposed Transformer model demonstrates the most consistent and superior performance across cities and time horizons (Table 8). At the 1 h horizon, Transformer outperforms both CatBoost and CNN in nearly every city, with R^2 values exceeding 0.87 in Baltimore, Las Vegas, and Philadelphia, and with MAPE values under 12% in several cities, including Las Vegas (11.3%) and Portland (11.0%). CNN follows closely, often outperforming CatBoost in short-term predictions, particularly in cities like Detroit

and Washington, where local temporal dynamics are complex but spatial correlations are exploitable by CNN's architecture.

Table 8. Proposed Transformer forecasting metrics—part 1 (Baltimore, Denver, Detroit, and Las Vegas).

H	Baltimore				Denver				Detroit				Las Vegas			
(h)	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
1	0.8731	25.31	19.41	20.34	0.8052	25.44	18.81	32.12	0.8482	20.41	14.92	18.80	0.8662	18.63	14.01	11.30
6	0.7802	33.32	25.57	28.48	0.7150	30.80	22.89	42.54	0.7171	27.88	20.85	29.14	0.6861	28.54	22.28	20.07
12	0.7747	33.76	25.68	29.18	0.7107	31.04	23.01	43.07	0.7147	27.99	20.96	29.36	0.6724	29.15	22.81	20.80
18	0.7762	33.65	25.54	29.16	0.7089	31.13	23.05	42.95	0.7117	28.12	21.03	29.39	0.6737	29.10	22.75	20.71
24	0.7716	33.96	25.63	29.10	0.7102	31.08	23.06	42.64	0.7066	28.38	21.12	29.33	0.6711	29.23	22.77	20.56

Table 9. Proposed Transformer forecasting metrics—part 2 (Los Angeles, Philadelphia, Portland, and Washington).

H	Los Angeles				Philadelphia				Portland				Washington			
(h)	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
1	0.8420	50.20	38.50	30.10	0.9601	11.10	8.20	10.80	0.9512	17.00	12.80	10.10	0.9025	18.10	13.90	10.80
6	0.7903	57.10	42.60	32.50	0.8450	21.50	16.00	20.80	0.7901	34.80	26.40	22.10	0.6288	35.10	27.20	21.00
12	0.7855	56.00	41.80	33.00	0.8320	22.00	16.40	21.30	0.7786	35.20	27.00	22.80	0.7205	30.10	23.80	18.40
18	0.7822	56.50	42.10	32.20	0.8281	22.60	16.80	21.60	0.7755	34.70	26.80	22.60	0.7220	30.00	23.60	18.20
24	0.7798	56.90	42.30	33.10	0.8253	22.80	17.00	21.90	0.7593	36.00	28.20	23.80	0.7094	31.00	24.60	19.10

As the forecast horizon increases, CatBoost's performance deteriorates more significantly than that of CNN or Transformer. For instance, in Denver at $h = 24$, CatBoost yields an R^2 of 0.7086, compared to 0.7102 and 0.7255 for Transformer and CNN, respectively. While the difference in R^2 is subtle, the Transformer produces lower RMSE and MAPE, indicating a better handling of error variance. Similarly, in Washington, Transformer maintains a lower MAE and MAPE than both CatBoost and CNN across all horizons, highlighting its robustness in urban areas with variable demand profiles.

CNN consistently provides intermediate performance between CatBoost and Transformer. Its convolutional structure enables it to capture localized time-window dependencies, which is particularly beneficial in cities like Detroit and Los Angeles, where load fluctuations exhibit short-term periodicity. CNN achieves its best performance at $h = 6$ and $h = 12$ in cities such as Las Vegas and Baltimore, sometimes outperforming Transformer in MAE but typically lagging behind in R^2 and MAPE.

Transformer, benefiting from its attention-based architecture, excels at longer forecast horizons. Unlike CNN, which operates over fixed receptive fields, and CatBoost, which lacks temporal modeling capacity, Transformer dynamically attends to relevant temporal information. This is evident in cities like Portland and Philadelphia, where Transformer sustains high R^2 values and low MAE even at $h = 24$ (Table 9). For example, in Philadelphia, the Transformer achieves $R^2 = 0.7950$ and MAPE = 24.13% at $h = 24$, compared to $R^2 = 0.7430$ and MAPE = 26.91% for CNN, and significantly poorer metrics for CatBoost.

In summary, the proposed Transformer model consistently outperforms both CNN and CatBoost in terms of predictive accuracy and temporal stability. CNN offers a meaningful improvement over CatBoost, especially at mid-range horizons, and should be considered a viable lightweight alternative. However, for applications demanding high precision and adaptability across diverse urban environments, the Transformer architecture remains the most effective approach.

5. Conclusions

This study developed and evaluated advanced machine learning models for short-term forecasting of EV charging demand in large-scale energy systems. The results demonstrate the proposed Transformer-based architectures consistently outperform conventional machine learning and deep learning approaches. One of the main advantages of the

Transformer-based model in this study is its ability to capture long-range temporal dependencies and complex interactions among historical charging demand, calendar variables, weather conditions, and charging-related features without requiring extensive manual feature engineering. Unlike recurrent or purely local convolutional structures, the self-attention mechanism allows the model to consider relationships across the full input sequence when generating forecasts. In this work, the Transformer was implemented using a lightweight architecture with two encoder layers, a 64-dimensional embedding, and four attention heads, providing a balance between forecasting accuracy and computational efficiency. However, this should not be interpreted as implying that CatBoost or CNNs are inherently unsuitable for large datasets; both approaches can be highly effective and scalable under appropriate data structures and hardware settings.

The broader contribution of this work lies in advancing reliable, data-driven demand forecasting to improve grid resilience and sustainability. Accurate predictions enable proactive grid operation, optimized energy scheduling, and informed infrastructure planning. By reducing uncertainty in charging demand estimation, the proposed framework can enhance renewable energy integration, support demand-side management strategies, and strengthen overall grid flexibility. The demonstrated scalability of the approach further highlights its applicability to real-world metropolitan and regional energy systems characterized by diverse consumption and mobility patterns.

Despite the promising forecasting performance, several limitations should be acknowledged. First, the city-level aggregation adopted in this study captures overall EV charging demand trends but does not preserve station-specific spatial granularity, which may limit its usefulness for local station operation, congestion management, or infrastructure planning at the charger level. Second, the applicability of the proposed framework to new cities depends on the availability, quality, and consistency of historical charging demand data and relevant exogenous variables, including weather-, calendar-information-, and charging-related features. Third, although the Transformer architecture used in this work is lightweight, computational scalability should be further evaluated using larger datasets, longer input sequences, and deployment-oriented metrics such as training time, memory consumption, and inference latency.

Future work will therefore investigate station-level and graph-based spatiotemporal forecasting models, evaluate transferability across different urban charging networks, and compare computational scalability across Transformer, CNN, CatBoost, and other forecasting architectures under larger-scale deployment settings. In addition, future work will further investigate computational scalability by comparing training time, memory usage, inference latency, and forecasting accuracy across Transformer, CNN, CatBoost, and other deep learning architectures under larger datasets and longer input sequences.

Author Contributions: Conceptualization is jointly developed by C.S., J.H., and A.T. C.S. structured the overall research approach and carried out the core methodological design. J.H. and A.T. contributed by providing supervision and domain expertise, supporting the development and refinement of this study. C.S. was responsible for the software implementation, validation, investigation, and visualization of the results. The original draft of this manuscript was prepared by C.S., while J.H. and A.T. contributed to reviewing and editing, improving the clarity and quality of this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data will be made available on request.

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

Nomenclature

Abbreviations

ARIMA	Autoregressive integrated moving average
CatBoost	Category boosting gradient-boosted decision tree regression model
CDO	Climate Data Online
CNN	Convolutional neural network
CPU	Central processing unit
DL	Deep learning
EV	Electric vehicle
EVSE	Electric vehicle supply equipment
FFN	Feed-forward network
GB	Gigabyte
GPU	Graphics processing unit
GPS	Global positioning system
GRU	Gated recurrent unit
LSTM	Long short-term memory network
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MCS	Monte Carlo simulation
ML	Machine learning
MSE	Mean squared error
NCEI	National Centers for Environmental Information
NOAA	National Oceanic and Atmospheric Administration
PV	Photovoltaic
RAM	Random access memory
ReLU	Rectified linear unit activation function
RE	Renewable energy
RMSE	Root mean square error
RNN	Recurrent neural network
SARIMA	Seasonal autoregressive integrated moving average
SoC	State of charge
TF	Traffic flow
V2G	Vehicle-to-grid

Mathematical Symbols

$\mathbf{X} * 1 : t$	Input feature matrix from time step 1 to time step t
$\mathbf{X} * t - L_{\text{win}} + 1 : t$	Historical input window from time $t - L_{\text{win}} + 1$ to time t
$\mathbb{R}^{t \times d}$	Real-valued matrix with t time steps and d input features
t	Time index or current forecast origin
h	Forecast lead time or prediction horizon
L_{win}	Historical input-window length used to construct supervised forecasting samples
L_{enc}	Number of Transformer encoder layers
H_{attn}	Number of attention heads in the Transformer model
d	Input feature dimension
d_k	Dimension of the key/query representation used in scaled dot-product attention
d_{model}	Transformer embedding dimension
d_{ff}	Hidden dimension of the Transformer feed-forward network
p_{drop}	Dropout probability

N	Number of samples used in training or metric evaluation
i	Sample index or convolutional-layer index, depending on context
l	Transformer encoder-layer index
m	Tree index in the CatBoost ensemble
M	Total number of trees in the CatBoost ensemble
T	Length of the historical time sequence used as model input
$\mathbf{x}*e$	Historical hourly EV charging energy-consumption sequence
$x_e^{(j)}$	Energy-consumption value at the j-th time step within the input sequence
$\mathbf{x}*s$	Historical charging-session-count sequence
$x_s^{(j)}$	Charging-session-count value at the j-th time step within the input sequence
$\mathbf{x}$	Generic input feature vector
$\mathbf{x}*\text{hist}$	Historical lag and rolling-statistic feature vector used by CatBoost
$\mathbf{x}*\text{weather}$	Weather-related input feature vector
$\mathbf{x}*\text{time}$	Time-related or calendar-related input feature vector
$\mathbf{c}$	Contextual feature vector, including time, traffic, and weather variables
$\mathbf{z}*t$	Exogenous explanatory feature vector at forecast origin t
$\mathbf{p}$	Positional encoding vector or matrix
$\mathbf{E}$	Input embedding matrix after feature embedding and positional encoding
Q	Query matrix in the self-attention mechanism
K	Key matrix in the self-attention mechanism
V	Value matrix in the self-attention mechanism
$\mathbf{H}^{(l)}$	Hidden representation produced by the l-th Transformer encoder layer
$\mathbf{H}^{(0)}$	Initial hidden representation, equal to the embedding matrix $\mathbf{E}$
$\mathbf{H}^{(L*\text{enc})}$	Final Transformer encoder representation after all encoder layers
$\mathbf{h}*\text{concat}$	Concatenated feature vector passed to the final regression layer
$\mathbf{h}*e^{(1)}$	First-layer CNN feature map extracted from the energy-consumption sequence
$\mathbf{h}*s^{(1)}$	First-layer CNN feature map extracted from the charging-session sequence
$\mathbf{h}*e^{(2)}$	Second-layer or pooled CNN feature map extracted from the energy-consumption sequence
$\mathbf{h}*s^{(2)}$	Second-layer or pooled CNN feature map extracted from the charging-session sequence
$\mathbf{w}*e^{(i)}$	Convolutional kernel for the i-th CNN layer applied to the energy-consumption input
$\mathbf{w}*s^{(i)}$	Convolutional kernel for the i-th CNN layer applied to the charging-session input
$\mathbf{b}*e^{(i)}$	Bias term for the i-th CNN layer applied to the energy-consumption input
$\mathbf{b}*s^{(i)}$	Bias term for the i-th CNN layer applied to the charging-session input
$\mathbf{W}$	Weight matrix of the fully connected output layer
$\mathbf{b}$	Bias vector of the fully connected output layer
y_i	Observed EV charging energy demand for sample i

$y^{(i)}$	Ground-truth target value for the i-th training or evaluation sample
y_{t+h}	Observed EV charging energy demand at forecast horizon $t + h$
$\hat{y}$	Predicted EV charging energy demand
$\hat{y} * i$	Predicted EV charging energy demand for sample i
$\hat{y}^{(i)}$	Predicted target value for the i-th training or evaluation sample
$\hat{y} * t + 1$	Predicted EV charging energy demand one time step ahead
$\hat{y} * t + h$	Predicted EV charging energy demand h hours ahead
$f * \text{CatBoost}(\cdot)$	Nonlinear CatBoost forecasting function
$f_h(\cdot)$	Direct lead-time-specific forecasting function for horizon h
$T_m(\mathbf{x})$	Output of the m-th CatBoost regression tree for input vector $\mathbf{x}$
$\mathcal{L}$	General regression training objective
$\mathcal{L}_{\text{MSE}}$	Mean squared error loss function
$\Omega(f)$	Regularization term penalizing model complexity
R^2	Coefficient of determination
ε	Small positive constant used to stabilize percentage-error calculations near zero demand

References

1. Wang, G.; Zhang, X.; Wang, H.; Peng, J.-C.; Jiang, H.; Liu, Y.; Wu, C.; Xu, Z.; Liu, W. Robust Planning of Electric Vehicle Charging Facilities With an Advanced Evaluation Method. *IEEE Trans. Ind. Inform.* **2018**, *14*, 866–876.
2. Wang, G.; Zhao, J.; Wen, F.; Xue, Y.; Ledwich, G. Dispatch Strategy of PHEVs to Mitigate Selected Patterns of Seasonally Varying Outputs From Renewable Generation. *IEEE Trans. Smart Grid* **2015**, *6*, 627–639.
3. Aldossary, M.; Alharbi, H.A.; Ayub, N. Optimizing Electric Vehicle (EV) Charging with Integrated Renewable Energy Sources: A Cloud-Based Forecasting Approach for Eco-Sustainability. *Mathematics* **2024**, *12*, 2627. [CrossRef]
4. Shahriar, S.; Al-Ali, A.R.; Osman, A.H.; Dhou, S.; Nijim, M. Prediction of EV Charging Behavior Using Machine Learning. *IEEE Access* **2021**, *9*, 111576–111586. [CrossRef]
5. Rupnik, B.; Wang, Y.; Kramberger, T. Hybrid Model for Motorway EV Fast-Charging Demand Analysis Based on Traffic Volume. *Systems* **2025**, *13*, 272. [CrossRef]
6. Lu, H.; Shao, C.; Hu, B.; Xie, K.; Li, C.; Sun, Y. En-Route Electric Vehicles Charging Navigation Considering the Traffic-Flow-Dependent Energy Consumption. *IEEE Trans. Ind. Inform.* **2022**, *18*, 8160–8171.
7. Chung, Y.-W.; Khaki, B.; Li, T.; Chu, C.; Gadh, R. Ensemble Machine Learning-Based Algorithm for Electric Vehicle User Behavior Prediction. *Appl. Energy* **2019**, *254*, 113732. [CrossRef]
8. Sommer, C.; Hossain, M.J. Artificial Intelligence-Driven Optimization of V2G and Charging Point Selection En-Route: A Systematic Literature Review. *Energy Convers. Manag. X* **2025**, *26*, 100978.
9. Zhang, T.; Peng, Q.; Zeng, S. Predicting EV Charging Demand in Renewable-Energy-Powered Grids Using Explainable Machine Learning. *Sustainability* **2025**, *17*, 4158. [CrossRef]
10. Hua, D.; Yan, P.; Liu, S.; Lin, Q.; Cui, P.; Li, Q. Optimization of In-Motion EV Charging Infrastructure for Power Systems Using Generative Adversarial Network-Based Distributionally Robust Techniques. *Energies* **2025**, *18*, 297. [CrossRef]
11. Rashid, M.; Elfouly, T.; Chen, N. A Comprehensive Survey of Electric Vehicle Charging Demand Forecasting Techniques. *IEEE Open J. Veh. Technol.* **2024**, *5*, 1348–1366. [CrossRef]
12. Xydas, E.; Marmaras, C.; Cipcigan, L.; Jenkins, N.; Carroll, S.; Barker, M. A Data-Driven Approach for Characterising Electric Vehicle Charging Demand: A UK Case Study. *Appl. Energy* **2016**, *162*, 763–771.
13. Heywood, P.; Maddock, S.; Bradley, R.; Swain, D.; Wright, I.; Mawson, M.; Fletcher, G.; Guichard, R.; Himlin, R.; Richmond, P. A Data-Parallel Many-Source Shortest-Path Algorithm to Accelerate Macroscopic Transport Network Assignment. *Transp. Res. Part C Emerg. Technol.* **2019**, *104*, 332–347. [CrossRef]
14. Yi, T.; Zhang, C.; Lin, T.; Liu, J. Research on the Spatial-Temporal Distribution of Electric Vehicle Charging Load Demand: A Case Study in China. *J. Clean. Prod.* **2020**, *242*, 118457. [CrossRef]
15. Kim, J.; Lee, S.; Park, J. Short-Term Electric Vehicle Charging Load Forecasting Using Time-Series Models. *Energy* **2022**, *239*, 122187.
16. Hancock, J.T.; Khoshgoftaar, T.M. CatBoost for Big Data: An Interdisciplinary Review. *J. Big Data* **2020**, *7*, 94. [CrossRef] [PubMed]
17. Zhang, L.; Jánošík, D. Enhanced Short-Term Load Forecasting with Hybrid Machine Learning Models: CatBoost and XGBoost Approaches. *Expert Syst. Appl.* **2024**, *241*, 122686. [CrossRef]

18. Wang, D.; Gan, J.; Mao, J.; Chen, F.; Yu, L. Forecasting Power Demand in China with a CNN-LSTM Model Including Multimodal Information. *Energy* **2023**, *263*, 126012.
19. Amarasinghe, K.; Marino, D.L.; Manic, M. Deep Neural Networks for Energy Load Forecasting. In *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*; IEEE: New York, NY, USA, 2017; pp. 1483–1488.
20. Kim, H.; Park, S.; Park, H.-J.; Son, H.-G.; Kim, S. Solar Radiation Forecasting Based on the Hybrid CNN-CatBoost Model. *IEEE Access* **2023**, *11*, 13492–13500.
21. Wang, S.; Zhuge, C.; Shao, C.; Wang, P.; Yang, X.; Wang, S. Short-Term Electric Vehicle Charging Demand Prediction: A Deep Learning Approach. *Appl. Energy* **2023**, *340*, 121032. [CrossRef]
22. Li, X.; Zhang, K.; Wang, J. Short-Term Electric Vehicle Charging Load Forecasting Using CNN-Based Models. *Electr. Power Syst. Res.* **2020**, *189*, 106764.
23. Sadaei, H.J.; da Silva, P.C.L.; Guimarães, F.G. A Hybrid CNN-Based Model for Short-Term Electric Vehicle Charging Demand Forecasting. *Energy* **2020**, *190*, 116367.
24. Somu, N.; Raman, G.M.R.; Ramamritham, K. A Deep Learning Framework for Building Energy Consumption Forecast. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110591. [CrossRef]
25. Tsalikidis, N.; Koukaras, P.; Ioannidis, D.; Tzovaras, D. Hybrid CNN-LSTM Forecasting Model for Electric Vehicle Charging Demand in Smart Buildings. In *Proceedings of the 2024 6th Global Power, Energy and Communication Conference (GPECOM)*; IEEE: New York, NY, USA , 2024; pp. 590–595. [CrossRef]
26. Almaghrebi, A.; Aljuheshi, F.; Rafaie, M.; James, P. Data-Driven Charging Demand Prediction at Public Charging Stations Using Supervised Machine Learning. *Energies* **2020**, *13*, 4230. [CrossRef]
27. Li, D.; Leung, M.-F.; Tang, J.; Wang, Y.; Hu, J.; Wang, S. Generative Self-Supervised Learning for Cyberattack-Resilient EV Charging Demand Forecasting. In *IEEE Transactions on Intelligent Transportation Systems*; IEEE: New York, NY, USA, 2025. [CrossRef]
28. Luo, R.; Song, Y.; Huang, L.; Zhang, Y.; Su, R. AST-GIN: Attribute-Augmented Spatiotemporal Graph Informer Network for Electric Vehicle Charging Station Availability Forecasting. *Sensors* **2023**, *23*, 1975. 10.3390/s23041975. [CrossRef] [PubMed]
29. Song, C.; Yang, H.; Cai, J.; Yang, P.; Bao, H.; Xu, K.; Meng, X.-B. Multi-Energy Load Forecasting via Hierarchical Multi-Task Learning and Spatiotemporal Attention. *Appl. Energy* **2024**, *373*, 123788.
30. Pavuluri, Y. Ev Watts Public Database. Available online: https://livewire.energy.gov/ds/evwatts/evwatts.public (accessed on 22 April 2025).
31. Moosavi, S. US Traffic Congestions (2016–2022) Database. Available online: https://www.kaggle.com/datasets/sobhanmoosavi/us-traffic-congestions-2016-2022/data (accessed on 22 April 2025).
32. National Centers for Environmental Information (NCEI). Normals Hourly. Available online: https://www.ncdc.noaa.gov/cdo-web/datasets (accessed on 23 April 2025).
33. Thaker, J.; Höller, R. Hybrid Model for Intra-Day Probabilistic PV Power Forecast. *Renew. Energy* **2024**, *232*, 121057. [CrossRef]
34. Zhang, Q.; Lu, J.; Kuang, W.; Wu, L.; Wang, Z. Short-Term Charging Load Prediction of Electric Vehicles with Dynamic Traffic Information Based on a Support Vector Machine. *World Electr. Veh. J.* **2024**, *15*, 189. [CrossRef]
35. Yoo, Y.; Jung, S. Modeling Forecast Errors for Microgrid Operation Using Gaussian Process Regression. *Sci. Rep.* **2024**, *14*, 2166. [CrossRef] [PubMed]
36. Siddiqui, J.; Ahmed, U.; Amin, A.; Alharbi, T.; Alharbi, A.; Aziz, I.; Khan, A.R.; Mahmood, A. Electric Vehicle Charging Station Load Forecasting with an Integrated DeepBoost Approach. *Alex. Eng. J.* **2025**, *116*, 331–341.
37. Ran, J.; Gong, Y.; Hu, Y.; Cai, J. EV Load Forecasting Using a Refined CNN-LSTM-AM. *Electr. Power Syst. Res.* **2025**, *238*, 111091.
38. Feng, Z.; Zhang, J.; Jiang, H.; Yao, X.; Qian, Y.; Zhang, H. Energy Consumption Prediction Strategy for Electric Vehicle Based on LSTM-Transformer Framework. *Energy* **2024**, *302*, 131780. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.