scientific reports

OPEN

Charging stations demand forecasting using LSTM based hybrid transformer model

Adil Hussain[1]✉, Vishwanath Eswarakrishnan[2]✉, Ayesha Aslam[3] & Sajib Tripura[1,4]

Accurate forecasting of energy demand for electric vehicles (EVs) is crucial for maintaining the stability and reliability of power systems. By predicting demand over various periods, charging station owners can ensure a continuous energy supply. Medium-term and long-term demand predictions, which extend from a few weeks to several days, help analyze charging demand across different periods based on historical trends. This study proposes a Transformer model that utilizes an LSTM-based encoder-decoder for forecasting the demand at electric vehicle charging stations (EVCS). The proposed model is compared with traditional deep learning-based LSTM and Transformer models. The research employs open datasets from ACN, including charging data from Caltech and JPL. Both datasets are used to train and test the models. Predictions are made for 30 days, 120 days, and 240 days ahead, with results compared to actual demand. Performance is evaluated using Mean Absolute Error (MAE) and Mean Squared Error (MSE). Compared to baseline models, the proposed LSTM-Transformer model for Caltech data shows a significant improvement at the 30-day horizon, lowering MAE by up to 17.27% and MSE by 19.79%. The accuracy improvement are minor but consistent in longer horizons (120 and 240 days), with an MAE and MSE improvement of up to 5.71% and 4.85%, respectively. The LSTM-Transformer model also shows better accuracy across all horizons for JPL data by reducing MAE and MSE by up to 24.91% and 23.17% at 30 days, 5.00% and 5.17% at 120 days, and 3.90% and 4.86% at 240 days. The results indicate that the Hybrid Transformer model outperforms the baseline models for both datasets in medium-term and long-term predictions. The 30, 120, and 240-day predictions demonstrate lower error rates with the proposed model when utilizing Caltech and JPL charging data for these time frames.

Keywords Electric vehicles, Demand prediction, Hybrid transformer, Charging demand, Deep learning

Using fossil fuels has significantly contributed to global warming in recent decades. The Paris Agreement in 2015 aims to restrict the rise of global temperature to a maximum of two degrees Celsius by the end of this century[1]. Electricity generating and transport systems are major contributors to carbon emissions, with transport systems representing around 25% of global energy consumption[2]. Electric Vehicles (EVs) yield a 45% decrease in carbon emissions compared to conventional vehicles equipped with Internal Combustion Engines (ICEs)[3]. As such, transitioning to an electrified transportation system and evolving conventional transport into smart, intelligent networks is crucial for moving towards a sustainable, low-carbon energy future and minimizing greenhouse gas emissions. Furthermore, adopting smart city concepts and intelligent transportation systems could enhance the efficiency of existing infrastructure and services.

In smart cities, smart transportation, as a primary pillar, has recently progressed to the development stage following the successful completion of the conceptual phase. In this context, conventional transportation systems have recently been implemented to enhance vehicle electrification, hence adhering to pollution regulations and facilitating the attainment of smart city objectives. Consequently, a quick increase in the quantity of EVs has been noted since 2015. EVs on the road have increased twice from 2017 to 2021. These numbers are anticipated to rise to 145 million by 2030 from approximately 16.5 million by 2021[4]. The expected extensive integration of EVs into power systems presents significant challenges and possibilities for power systems management.

The integration of EVs into the power grid may result in various issues for the current electric power system despite the economic and environmental benefits. The principal adverse implications of such deployment

[1]School of Electronics and Control Engineering, Chang'an University, Xi'an 710000, China. [2]Meta Platforms Inc., 1 Hacker Way, Menlo Park, CA 94025, USA. [3]School of Information Engineering, Chang'an University, Xi'an, China. [4]Department of Computer Science, Rangamati University of Science and Technology, Rangamati 4500, Bangladesh. ✉email: 2022032907@chd.edu.cn; vishwanath.ek@ieee.org

encompass power quality and reliability issues, considerable bottlenecks in the distribution network, and increased peak load demand[5–7]. Accurate demand forecasting can significantly lower carbon emissions. By adjusting charging times, EVs can be better integrated into the power grid, thereby diminishing their environmental footprint. This is accomplished by encouraging cleaner energy sources and reducing peak-hour consumption. It's crucial to consider these effects in detail to foster a transportation system that is more eco-friendly and sustainable.

An accurate projection of EV charging demand prediction will improve power system managers' decision-making capabilities for smart city energy management. Efficient EV charging station network management necessitates accurate forecasting of charging demand[8]. Moreover, this approach can boost the efficiency of charging stations while lowering maintenance expenses. Accurately predicting charging demand is crucial for designing, developing, and expanding EV charging infrastructure[9]. Therefore, it's crucial to incorporate modelling and forecasting of EV charging demand into energy system planning and scheduling frameworks. Most of the existing researches are focused on using traditional statistical and machine learning based models, which are unable to capture the demand trends over time for different horizons, resulting in lower accuracy. The usage of individual models, such as LSTM or the Transformer model, can also result in low accuracy, as capturing the charging demand trends over time, especially for medium and long-term predictions, is difficult. The recent studies are focused on combining the LSTM with CNN, or GRU-based models; however, the combination of LSTM and Transformer model is mainly ignored for EV charging demand predictions. The primary goal of this study is to enhance the precision of medium-term and long-term forecasts. The research proposes a Hybrid Transformer model with LSTM-based Encoder-Decoder. The proposed model is implemented using the ACN dataset, which includes Caltech and JPL charging stations. The proposed model and the baselines are trained and tested using both datasets. The performance comparison is performed with the baselines, including the LSTM, and Transformer models. The main contributions of this research work are as follows:

- To the best of our knowledge, this is the first study to propose an LSTM encoder-decoder-based Transformer model for EV Charging Demand Forecasting.
- The proposed model is used for medium-term, i.e., 30 days, and Long-Term, i.e., 120 and 240-day Charging Demand Predictions.
- The study performs a comparative analysis with the deep learning models, including LSTM and Transformer Models.
- The study evaluates the performance of the proposed model for medium and long-term predictions for the Caltech and JPL charging locations from the ACN-Data using the deep learning models, including LSTM and Transformer Model as baselines.

This research is organized as follows: Section II contains the literature, including machine learning, deep learning, and hybrid models. Section III is the methodology and provides an overview of the dataset, evaluation metrics, and baseline models. Section IV is the design and implementation of the proposed models and the training process for the proposed and baseline models. Section V is the results and analysis, briefly analyzing the performance. Section VI concludes the study.

Literature review

The EV industry is undergoing substantial growth. EVs are available in hybrid and fully electric configurations, which combine an internal combustion engine with an electric powertrain. Hybrid Electric Vehicles (HEVs), Plug-in Hybrid Electric Vehicles (PHEVs), and Battery Electric Vehicles (BEVs) are the three primary categories of EVs. Two distinct energy sources, an internal combustion engine (ICE) and an electric motor, propel HEVs. In comparison to ICEs, electric motors optimize fuel efficiency.

Unlike HEVs, BEVs are primarily propelled by a single source and consist of three primary technical components: a rechargeable battery cell, a power controller, and an electric motor. The third category of electric vehicles is PHEVs. They use batteries to power an electric motor and may recharge from an external source[10,11]. As reported[4], the BEV is the most frequently employed type of EV in various markets.

Forecasting the charging demand for EVs is difficult because of several uncertainties. The irregularities arise from the unpredictable nature of electric vehicle drivers' behaviours. Thus, accounting for the uncertainties affecting the system in modelling and management is essential for the seamless integration of electric vehicles into the energy grid. Numerous research efforts have been undertaken in the literature to address the issue of delivering accurate EVEC predictions in power networks.

The initial classification of charging demand prediction models identifies linear techniques that employ linear functions for modeling and forecasting time series data. The most commonly applied linear methods for predicting charging demand include Bayesian inference models, Principal Component Analysis (PCA), Autoregressive Moving Average (ARMA) (along with its derivatives), and multiple linear regression (MLR).

Deep learning

Deep Learning (DL) began in 1940[12]. This approach has only been effectively utilized in numerous domains in recent years. The main reasons are the rapid rise of powerful computers that can train complicated mathematical models and the availability of larger data sets for DL training[13]. Deep learning algorithms have been widely used to predict EV charging demand[14–21]. A few DL-based studies are discussed below.

The study in[22] forecasts the charging demand of Caltech and JPL charging stations one day in advance. Variational inference is employed for the posterior distribution, while LSTM parameters are employed for the prior distribution. The proposed strategy is more effective than SVR and MLR. In[14] the authors conducted a comparative analysis of deep learning techniques for the ultrashort-term (minute-level) prediction of EV charging demand. The feasibility of six machine learning algorithms—ANN, LSTM, RNN, GRU, SAEs, and Bi-LSTM—was evaluated using historical EV data from Shenzhen, southern China, from July 1, 2017, to June 30, 2018. This dataset comprises power charging consumption and session start and finish times. The efficacy of the above approaches is evaluated using Root Mean Square Error (MSE) and Mean Absolute Error (MAE). The six models predicted super short-term events well, while LSTM outperformed the others. In[23], Sequence to Sequence, a DL-based forecasting approach, predicted charging demand one to five months ahead. Data from 1200 Los Angeles charging stations was used to evaluate this strategy. Standard DL models, including ARIMA, LSTM, and XGBoost, were utilized to test the Sequence-to-Sequence technique. Sequence to Sequence outperformed the other models for both short- and long-term forecasts, showing considerable accuracy gains.

A novel probabilistic framework for charging demand prediction, based on deep learning, was introduced in[18]. Three phases make up this structure. It has two primary steps. First, wavelet decomposition and normalization were used to normalize and partition traffic flow data by frequency. Second, a CNN predicted traffic flow data, followed by a mixture model-based EV arrival rate prediction. Finally, the charging demand prediction was calculated using a Queuing-based probabilistic model and projected arrival rates. The research used U.K. data from January to December 2014. The proposed model predicted charging demand with better accuracy according to the results. A novel DL-based approach is developed by the authors[24] for charging demand forecasting. Using Empirical Mode Decomposition (EMD), the approach divides the input time series into subtime series and forecasts charging demand values using Deep LSTM (DLSTM). Arithmetic Optimization Algorithm optimizes DLSTM parameters. Case study data came from Georgia Tech EV charging stations. Comparing the suggested model to other DL-based approaches showed it was more successful. While[24] suggesting a unique charging demand prediction algorithm[25], compared four well-established DL-based techniques (ANN, LSTM, GRU, and RNNs) to estimate future values of Moroccan charging demand time series. The 2000 data points were obtained from 1793 charging transactions at two three-phase charging sites (22KW and 11KW). All four methods worked, although the single-layer GRU model predicted best. In their study, Vishnu, et al.[26] used three learning frameworks to real-time ACN data and examined their effectiveness to estimate short-term EV demand. LSTM, support vector regression (SVR), and auto-regressive (AR) forecasting frameworks all performed well in predicting EV charging demand. LSTM performed best, with an RMSE of 5.9 kW and an MAE of 4 kW.

The utilization of DL techniques has contributed to their widespread adoption in the field of time series prediction due to their various advantages. They are known for their capacity to manage and analyze massive volumes of complicated and nonlinear data. Additionally, they can automatically select features. Nevertheless, DL models have certain drawbacks, including their limited interpretability, tendency to overfit, and difficulty in diagnosing and rectifying errors.

Hybrid models

Hybrid modelling approaches integrate different modelling strategies to form a comprehensive and unified method, where every model adds its unique perspective on the data. This hybrid modelling technique is an effective tool for analyzing time series, merging multiple methods to enhance prediction accuracy and robustness by taking advantage of the strengths of various techniques[27]. The literature contains numerous studies that have combined the primary techniques discussed in the previous Section to construct hybrid models for charging demand prediction[28–38]. Some of these studies are discussed in the subsequent paragraphs.

A recent study[39] combined SARIMA and DL techniques to create a hybrid model for predicting charging demand. SARIMA captures linear features and seasonality, while LSTM captures nonlinear characteristics. The model outperformed other methods in predicting charging demand based on real-world data from a charging station in Spain during 2015–2016. The research findings were compared to other techniques, such as SARIMA, LSTM, Extreme Learning Machine (ELM), SVR, SARIMA-ELM, and SARIMA-SVR. The results of the proposed hybrid model are better than those of the baseline methods.

Similarly, the authors[40] combined the multivariate residual correction Grey Model (GM) and the LSTM network into a cohesive model for predicting charging demand. The GM was utilized to evaluate how various elements, such as the quantity of electric vehicles, the cost of electricity, and weather conditions, influence charging demand. Subsequently, the LSTM was implemented to accurately identify influential factors, thereby decreasing the probability of prediction errors. The efficacy of the devised technique was affirmed by the study's concentration on charging demand data recorded in China in 2017. The authors[41] introduce a hybrid LSTM-TED model to enhance EV charging station prediction. The encoder reduces input data dimensions to preserve and choose meaningful information, while LSTM determines input variable temporal relationships. Two sets of charging demand predictions from Chinese and US charging stations are utilized to test this strategy. The hybrid model is assessed using conventional prediction models, including ARIMA, LSTM, and Historical Average (HA). The proposed methodology surpasses conventional methodologies. ANNs, RNNs, and Q-learning are introduced in[42] for prediction. The input of the Q-learning model is derived from the outputs of ANN and RNN. The investigation investigated the charge of EVs in three modes: coordinated, uncoordinated, and intelligent. Keras was employed to acquire the case study data. The results indicate that the hybrid model enhances the accuracy by 50% for ANN and RNN-based predictions.

Hybrid approaches are often more precise and dependable than singular techniques, as they blend the advantages of various models, leading to enhanced performance. Furthermore, they exhibit enhanced resilience to outliers and disturbances. However, hybrid approaches in forecasting time series also have their disadvantages. One major downside is the high computational expense involved in integrating different methodologies. Additionally, finding the optimal mix of models and assigning the correct weight poses a significant challenge.

Transformer-based models

Koohfar, et al.[43] utilized the Transformer model to predict EV charging demand. Predictions for three time steps are considered: 7 days, 30 days, and 90 days to cover both short-term and long-term forecasting of EV charging capacity. The study employed ARIMA, SARIMA, RNN, and LSTM as baseline models and compared their performance using RMSE and MAE. The findings show that the Transformer beats the other models in terms of short-term and long-term forecasts, indicating its capacity to solve time series challenges. Manzoor, et al.[44] used transformer-based models to forecast EV charging demand. They also provide a comparison evaluation of two extensively utilized attention mechanisms, Self-attention and ProbSparse attention, with the goal of determining the best solution for the EV charging demand forecasting problem. Experiments using real-world ACN dataset reveal that the computationally efficient, ProbSparse attention based model beats the classic self-attention model, offering a new solution to the EV charging demand forecasting. Yılmaz, et al.[45] compared the Transformer model to traditional LSTM, Bi-directional LSTM (Bi-LSTM), and Support Vector Regression (SVR) models for predicting Battery State of Charge (SoC) using NASA, BMW i3, Stanford University Battery Datasets, and real-world battery data from the Musoshi brand's L5 EV. Bouhamed, et al.[46] present an encoder–decoder model that leverages potential of Transformer-based encoders to generate probabilistic forecasts, i.e., a distribution over future predictions, using two real-world datasets: hourly load data from the city of Johor in Malaysia's power supply company and hourly load consumption data from one of the buildings of Grenoble Institute of Technology. The predictions are made for 24 h, 1 week, and 1 month. Accuracy improved from 84.7% to 89.7% for Malaysian data and from 45.5% to 57.2% for Grenoble data. Ke and Wang[47] used historical smart meter data, and developed a home charging prediction method inspired by the concept of non-intrusive load monitoring (NILM). Unlike NILM, which detects EV charging that has already occurred, our method provides predictive information of future EV charging occurrences, thus increasing its utility for charging management. Specifically, our method uses a transformer model based on a self-attention mechanism, employing a "divide-conquer" strategy, to process historical meter data effectively and learn EV charging representation for charging occurrence prediction, enabling prediction at one-minute intervals. Experimental results show that our method is effective, consistently achieving high accuracy of over 96.81% across various prediction time spans.

The current research shows that load prediction in smart solar microgrids can be enhanced remarkably using hybrid deep learning models. Compared to standalone models, hybrid CNN-LSTM model along with feature selection methods are useful in capturing temporal patterns and significant features in hourly load forecasting[48]. In a similar vain, the combination of CNN with bidirection directional gated recurrent units (Bi-GRU) based on advanced metering infrastructure data improves short-term three-phase loads prediction due to its capacity to model more complicated temporal and spatial trends[49]. Production of metaheuristic algorithms and two steps deep learning architecture enhances fault classification within a diesel generator and makes microgrid operation reliable[50]. The methods provide an example of how convolutional, recurrent and optimization can be combined to achieve high levels of accuracy of the prediction and robustness of the system during the management of renewable energy.

Transformer based models have improved the prediction accuracy as proven in the literature, especially for EV charging demand, however, to investigate further the medium and long term predictions, like 30 days, 120 days, and 240 days, the variations in the charging demand can differ based on the days and the charging patterns. Therefore, combining the Transformer model with the LSTM model can help achieve better accuracy, as both LSTM and Transformer are efficient in predictions. Therefore, the main objective of this study is to design an LSTM-based Transformer model where the LSTM layers are used in the encoder and decoder of the Transformer model. The main contribution of the study is to present a hybrid model by comining the LSTM and Transformer models for the EV charging demand predictions for the medium and long-term periods.

Methodology

With the increase in EV adoption, there is a growing need for accurate power consumption estimates to optimize charging station management[33]. The effective planning and development of EV infrastructure can be significantly enhanced through precise predictions of power needs. This study conducts demand forecasting for EV charging stations using long short-term memory (LSTM), and transformer models. The dataset used for this work includes charging sites from Caltech and JPL. The charging need is anticipated for thirty days in the middle term and for longer durations such as one hundred twenty days and two hundred forty days. The following activities are carried out: data collection, preparation, and exploratory data analysis. The training and test sets are divided from the overall datasets. Data from Caltech and JPL are employed to train and test the models. These datasets are used to assess the performance of the models being studied. Performance is evaluated using two metrics: the Mean Absolute Error (MAE) and the Mean Square Error (MSE). The methodology is shown in Fig. 1.

Baseline models

LSTM model

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) designed to recognize patterns in data sequences over long periods. They manage information flow through memory cells and three types of gates: input, output, and forget gates. These features help overcome the vanishing gradient problem common in standard RNNs. LSTMs excel in maintaining context over long sequences, making them highly effective for applications in natural language processing, voice recognition, and time series analysis. Their ability to retain important information across extended sequences makes them a popular choice in deep-learning projects that require handling temporal dependencies. However, LSTMs may encounter difficulties in capturing distant temporal dependencies due to vanishing gradient issues observed with lengthy sequences. Sparse EV charging data in specific regions can degrade performance, and LSTMs necessitate large datasets to prevent

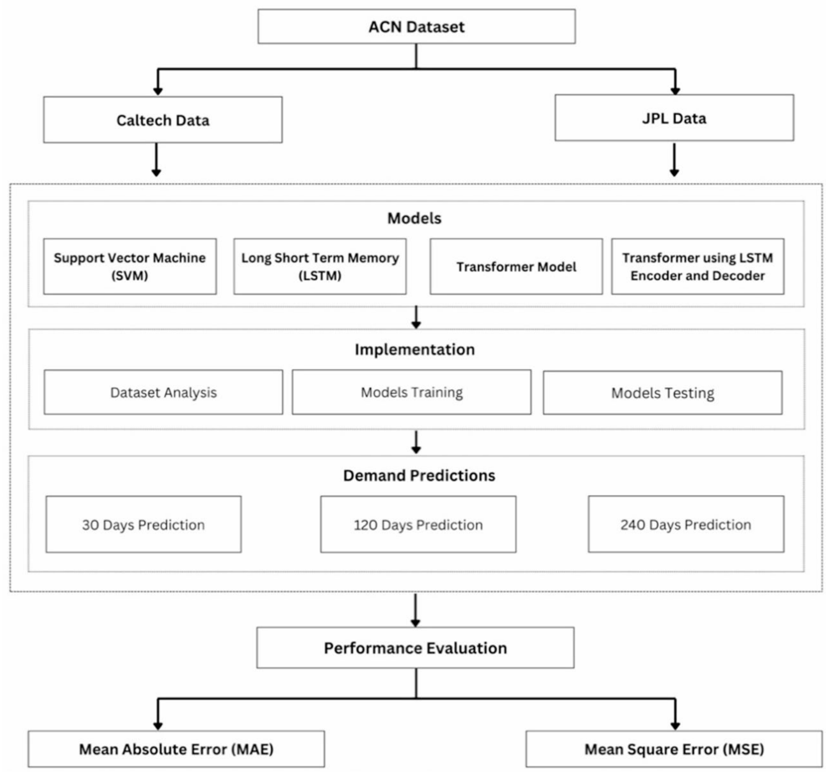

Fig. 1. An overview of the methodology.

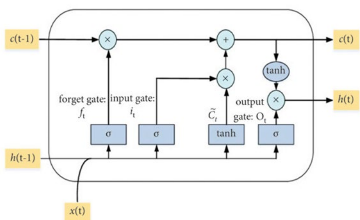

Fig. 2. An overview of LSTM architecture.

overfitting. Moreover, their computationally intensive step-by-step processing approach and their black-box character render them less interpretable for stakeholders. Figure 2 shows the architecture of an LSTM model.

The LSTM model predicts EV charging demand $\widehat{y}_t$ by learning temporal dependencies from historical demand data. The mathematical formulation for LSTM, including input, forget, cell state update and output gates, are represented as follows.

Forget gate The forget gate decides which information from the previous cell state C_{t-1} should be retained or discarded. It uses a sigmoid function σ to assign a value between 0 as completely forget and 1 as completely retain for each piece of information.

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{1}$$

Where W_f is the weight matrix for the forget gate, b_f is the vector for the forget gate, h_{t-1} is the hidden state from the previous timestep, and x_t is an input feature vector at time t, representing factors influencing charging demand.

Input gate The input gate determines how much of the new input x_t should be added to the cell state. It works alongside the candidate cell state $\widetilde{C}_t$ to update the memory with new, relevant information.

$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

Where i_t decides the importance of the current input.

The candidate cell state is a set of potential updates for the cell memory, created based on the input x_t and previous hidden state h_{t-1}. It uses the $\tanh$ function to constrain values between − 1 and 1.

$$\widetilde{C}_t = \tanh(W_c\,[h_{t-1}, x_t] + b_C) \tag{3}$$

Where $\widetilde{C}_t$ are the candidate values to update the cell state.

<u>Cell-state update gate</u> The cell state combines retained past information that is controlled by f_t and new information that is controlled by i_t and $\widetilde{C}_t$ to update the memory.

$$C_t = f_t \odot C_{t-1} + i_t \odot \underset{t}{\widetilde{C}} \tag{4}$$

Where combines retained past information $f_t \odot C_{t-1}$ and new information $i_t \odot \underset{t}{\widetilde{C}}$ to update the cell state.

<u>Output gate</u> The output gate controls how much of the updated cell state C_t is passed through *tanh* to produce the new hidden state h_t, which represents the output of the LSTM for this timestep.

$$o_t = \sigma\left(W_o\,[h_{t-1}, x_t] + b_o\right) \tag{5}$$

$$h_t = o_t \odot \tanh(C_t) \tag{6}$$

Where o_t controls how much of the cell state is passed to the output, and h_t is the updated hidden state.

The predicted EV charging demand $\widehat{y}_t$ is obtained by applying a linear transformation to the hidden state h_t.

$$\widehat{y}_t = W_y h_t + b_y \tag{7}$$

<u>Transformer model</u> Regarding parallel processing, the Transformer neural network architecture shines since it can manage sequential data without depending on recursion. The model can grasp connections between faraway items in a sequence due to the self-attention mechanism introduced in this version. Because of their scalability and capacity to manage long-range relationships, transformers have completely transformed jobs such as machine translation, time series forecasting, and natural language processing (NLP). Newer versions of the model, such as BERT and GPT, which use an encoder-decoder structure, provide the basis for powerful language models. While transformers are effective in capturing long-term dependencies, their computational requirements are high due to their quadratic complexity with sequence length, rendering them resource-intensive. Limited data can result in underfitting, and they operate most effectively with extensive datasets. If not adequately regularized, transformers may overfit to medium- and long-term trends and may struggle to adapt to non-stationary patterns, such as abrupt shifts in demand or policy changes. Furthermore, the process of designing Transformers to integrate a variety of contextual features, such as weather or socioeconomic factors, is intricate and susceptible to implementation.

The architecture of this model contains a MultiHeadAttention layer, which separates the input into several heads. The model analyzes the separate parts of the input sequences simultaneously. The attention mechanism uses scaled dot-product attention to determine queries, keys, and value representations. Once attention is applied, a GlobalAveragePooling1D layer compresses the sequence. A thick layer is used for regression, flattens the output and predicts a single output value. Figure 3 shows the Transformer model overview, as used in this study.

The key components in the Transformer model tailored for EV charging demand prediction are provided mathematically.

<u>Dense layer to encode input</u> The dense layer at the encoder's input transforms raw input features *x* into a higher-dimensional representation. The activation function *ReLU* introduces non-linearity, allowing the model to capture complex relationships in the data.

$$i_e = ReLU(W_{input}x + b_{input}) \tag{8}$$

<u>Multi head attention</u> The Query (Q), Key (K), and Value (V) are derived from the encoded input features, enabling the model to learn dependencies between different timesteps. The objective is to identify the importance of past EV charging demand features for predicting future demand.

$$Attention\,(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{9}$$

<u>Global average pooling</u> The global average pooling aggregates information across all timesteps into a single vector, summarizing the global context of the input data. This step reduces dimensionality while preserving essential features.

$$h_{pool} = \frac{1}{n}\sum{}_{i=1}^{n} h_i \tag{10}$$

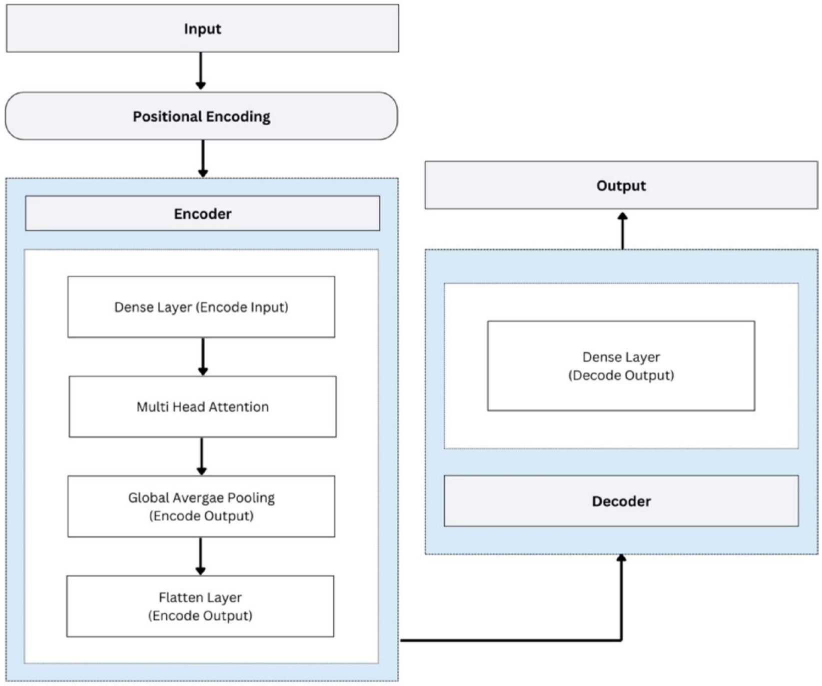

Fig. 3. An overview of the transformer model architecture.

Where h_{pool} is the pooled output.

<u>Flatten layer</u> The flattened layer in the encoder converts the pooled output into a 1D vector, preparing it for further processing in the decoder.

$$h_{flat} = Reshape\left(h_{pool}\right) \tag{11}$$

Where h_{pool} is the pooled output received from global average pooling, and h_{flat} is the flattened representation of h_{pool}.

<u>Dense layer in decoder</u> The decoder contains only one dense layer used to decode the output received from the flattening layer of the encoder, which was encoded by reshaping the pooled output. The objective is to map the encoded representation h directly to the EV charging demand prediction $\widehat{y_t}$. The dense layer acts as the final regression step, outputting the predicted demand values for future timesteps.

$$\widehat{y_t} = W_{decode}h + b_{decode} \tag{12}$$

Proposed hybrid transformer model

The Transformer and LSTM architectures are combined in the hybrid model. By combining LSTMs and Transformers for demand prediction, the strengths of both models are leveraged to overcome their limitations. While Transformers effectively model long-range dependencies using self-attention, LSTMs excel at capturing local and short-term temporal patterns. The computational complexity of Transformers is reduced by this hybrid approach, which enables LSTMs to preprocess and compress sequential data. Subsequently, Transformers became more focused on global trends. Furthermore, the superiority of LSTMs in managing smaller datasets reduces Transformers' requirement for a substantial amount of data. Collectively, they establish a robust framework for capturing both fine-grained and broad temporal patterns, thereby improving the accuracy of predictions for medium- and long-term horizons.

A multi-head attention mechanism handles context and long-range dependencies, while an LSTM-based encoder and decoder store temporal dependencies in the system. After an LSTM encoder processes the input sequence, positional encoding is used to keep track of the order in which events occur. The encoded representation is improved using multi-head attention layers, followed by layer normalization and dropout for regularization. After processing the encoder's output, the decoder uses an extra LSTM layer to examine the target sequence. Positional encoding and attention layers then follow. A flattened layer is used to predict the output sequence. Figure 4 shows the architecture of the proposed Hybrid Transformer model using LSTM encoder-decoder.

The encoder in this hybrid Transformer-LSTM model is created a single LSTM layer that processes the input sequence. This LSTM layer comprises 128 cells, as indicated by the model dimensionality (d_model). The LSTM layer detects temporal connections in the input data and produces a sequence of hidden states that reflect the

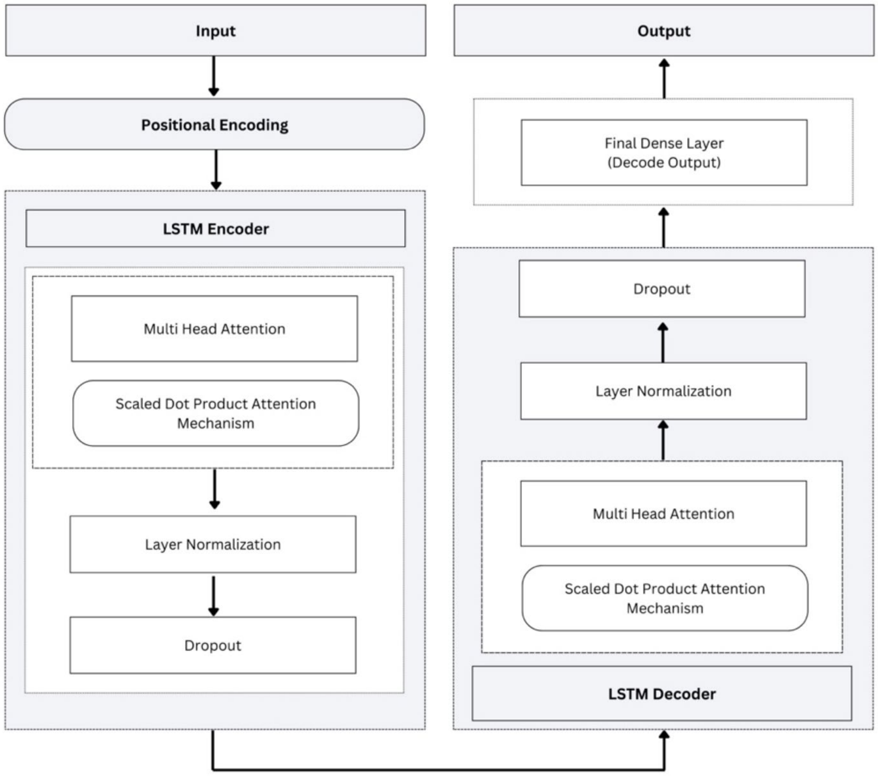

Fig. 4. Proposed hybrid transformer model using LSTM-based encoder–decoder.

encoded input. These hidden states are then transmitted via a multi-head attention mechanism, which captures relationships throughout the whole sequence and improves the model's capacity to focus on key areas of the input. The encoder's function is to convert the input sequence to a format that the decoder can use to make predictions. The decoder in this hybrid model, like the encoder, has a single LSTM layer with 128 cells. This layer takes the target sequence (such as previously anticipated or ground truth values) and mixes it with the encoded data from the encoder. After the LSTM processes the target sequence, the output is routed through a multi-head attention mechanism that attends to the relevant encoded input, allowing the decoder to make predictions by focusing on critical temporal relationships. The decoder constructs the output sequence by taking advantage of both the LSTM's sequential processing capacity and the attention mechanism's ability to simulate long-range relationships in the data.

The input of the LSTM at the Encoder is a sequence of historical charging demand data, which is 30 days, 120 days, or 240 days based on the prediction period as "time steps, where the look-back window is similar to the prediction period. Also, the model takes batch size and features as input. So, the input is (batch size, time_steps and features). The LSTM encoder outputs a sequence of encoded representations with shape (batch_size, time_steps, d_model), and captures the temporal dependencies in the inut. Further, the LSTM in decoder takes input which was the output of the LSTM encoder, and provides a sequence of decoded features shaped as (batch_size, time_steps, d_model), which is given to the attention layers, and then provided to the dense layer for the final prediction output for ev demand as (batch_size, time_steps, 1).

Dataset

Data on Adaptive Charging Networks (ACN) was collected from two locations in California. The CAN, located at the Caltech campus, is within a parking garage, featuring 54 Electric Vehicle Supply Equipment (EVSE) units or charging stations and a 50 kW DC fast charging station. This Caltech ACN is open to the general public and is often used by drivers not affiliated with Caltech. The proximity of the parking garage to the campus gym enables many drivers to charge EVs while they work out, typically in the mornings or evenings. Conversely, the ACN at JPL comprises 52 EVSEs housed in a parking facility. Access to the JPL campus is restricted, with charging facilities exclusively available to employees, marking it as a workplace charging location. In contrast, the Caltech site serves as a workplace and a public charging destination. JPL experiences significant EV usage, leading to high demand for EVSEs and a spontaneous system where drivers relocate their EVs after charging to make room for others.

Evaluation metrics

The most widely used metric for evaluating forecast accuracy in research on charging demand prediction is the Mean Absolute Error (MAE). It calculates the average of the absolute differences between predicted and actual values, as shown in Eq. 13.

$$MAE = \frac{1}{N}\sum{}_{i=1}^{n}\left|\widehat{y}_i - y_i\right| \tag{13}$$

The average of the squared differences between actual and predicted values in a dataset is called the Mean Squared Error (MSE). MSE calculation is performed using Eq. 14.

$$MSE = \frac{1}{N}\sum{}_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2 \tag{14}$$

Implementation

This section provides an overview of the ACN dataset, including Caltech and JPL charging locations. It also briefly outlines the models used in this study and provides an overview of the implementation using the hyperparameters configuration used for the models during the implementation.

Caltech dataset analysis

Figure 5 illustrates energy use in kilowatt-hours (kWh) over time, with the x-axis labelled “Timestamp” spanning from April 2018 to mid-2021 and the y-axis denoting energy values from 0 to 1000 kWh. The data reveals a clear cyclical or seasonal trend, marked by intervals of increased energy use succeeded by substantial declines. Peak energy usage transpired from mid-2018 to late 2018, subsequently diminishing throughout 2019 and attaining a low point by the conclusion of 2020. Consequently, energy usage will increase by mid-2021. The chart indicates fluctuations potentially associated with seasonal or operational factors affecting energy use over time. Figure 5 shows the energy consumption of the Caltech charging station.

Figure 6 shows the energy consumption from the Caltech Dataset, categorized into training and testing sets. The blue line represents the training data, covering the period from early 2018 to early 2021, whilst the orange line indicates the testing data, which spans from mid-2021 onwards. Both datasets measure energy in kWh on the y-axis. The training data indicates significant seasonal or operational fluctuations, with peak energy use occurring in mid-2018, followed by a general decrease until 2020 when energy usage stabilizes. The testing data reveals a comparable pattern of increasing energy use commencing in mid-2021. The clear distinction between training and testing data signifies that the data is utilized for predictive analysis or model evaluation.

JPL dataset analysis

Figure 7 illustrates the trends of energy consumption statistics, with the y-axis denoting energy provided in kilowatt-hours (kWh) and the x-axis displaying timestamps from September 2018 to roughly mid-2021. The energy data exhibits significant fluctuation, with a pronounced peak occurring between late 2018 and mid-2020, during which the values consistently remain elevated. A notable decrease in energy use occurred in mid-2020, succeeded by relatively low energy usage. Commencing in late 2020, energy consumption started an upward trend, persisting in its slow ascent throughout 2021. The graph depicts patterns indicating operational changes, seasonal effects, or other external factors affecting energy distribution throughout this period.

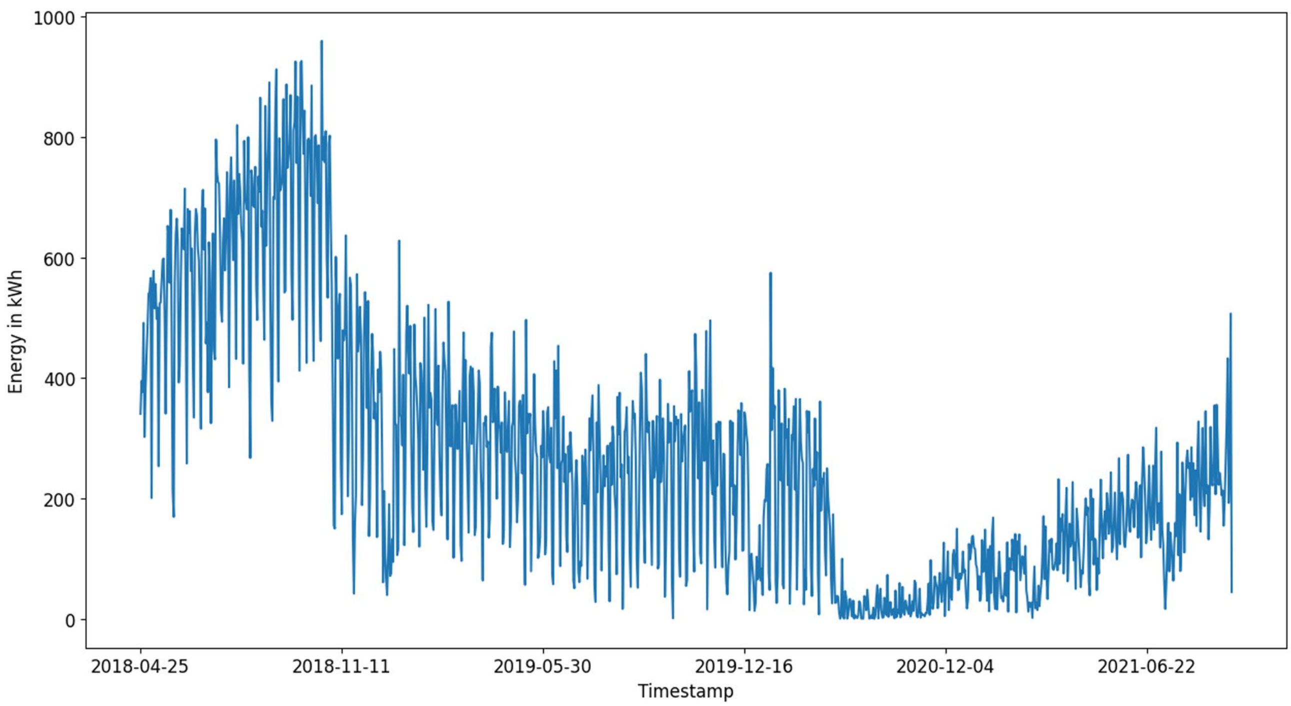

Fig. 5. Energy consumption analysis of Caltech dataset.

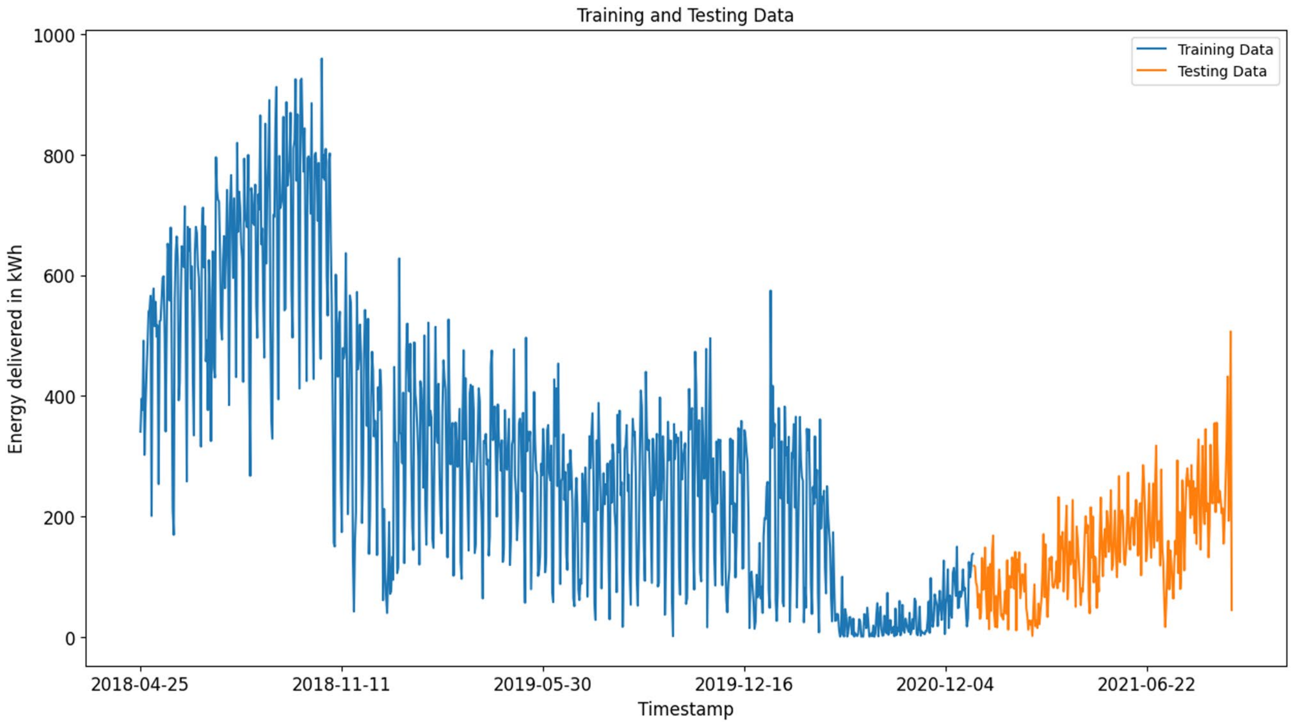

Fig. 6. Caltech dataset split into training and testing sets.

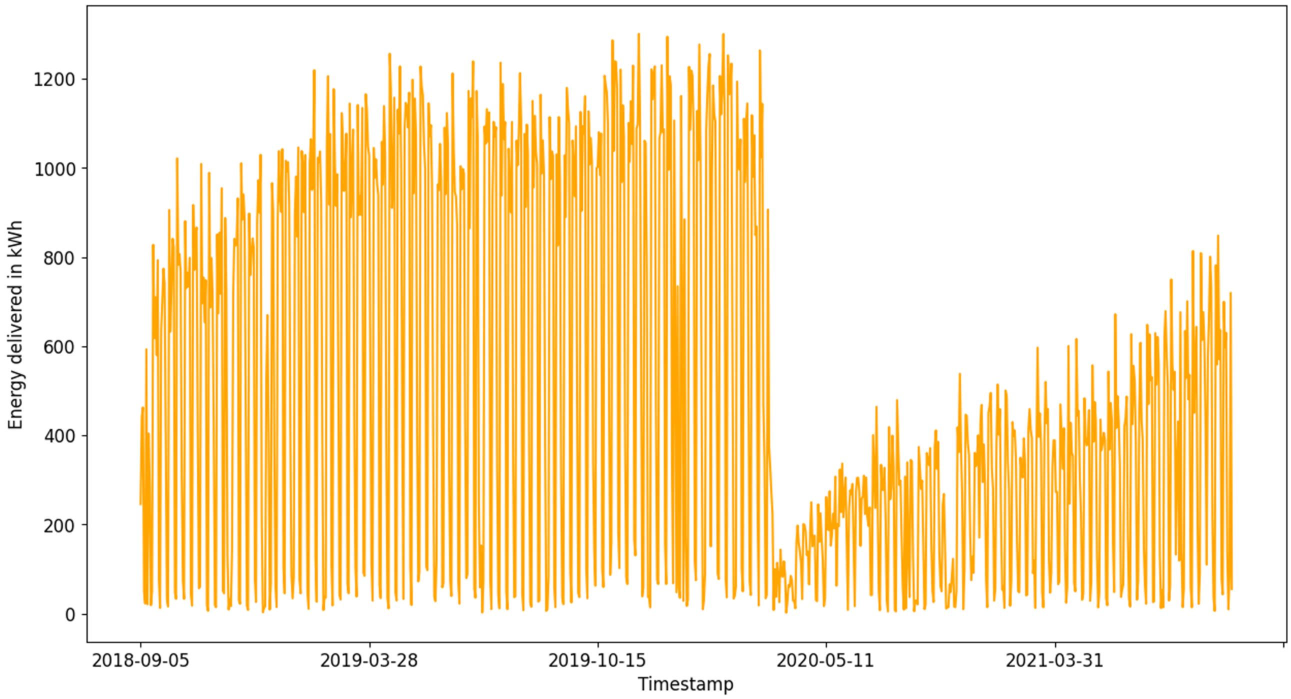

Fig. 7. Energy consumption analysis of JPL dataset.

The energy consumption data from the JPL dataset is segmented into training and testing sets, as seen in Fig. 8, and is represented graphically over time. The blue line represents the training data from September 2018 to mid-2020, while the orange line depicts the testing data covering mid-2020 to 2021. Both datasets measure kWh energy consumption, ranging from 0 to 1200 kWh. The training data indicates significant fluctuations, peaking around 1000–1200 kWh from late 2018 to mid-2020, followed by a sharp decline in energy use. The testing data reveal a steadily increasing energy consumption phase beginning in late 2020, suggesting upward

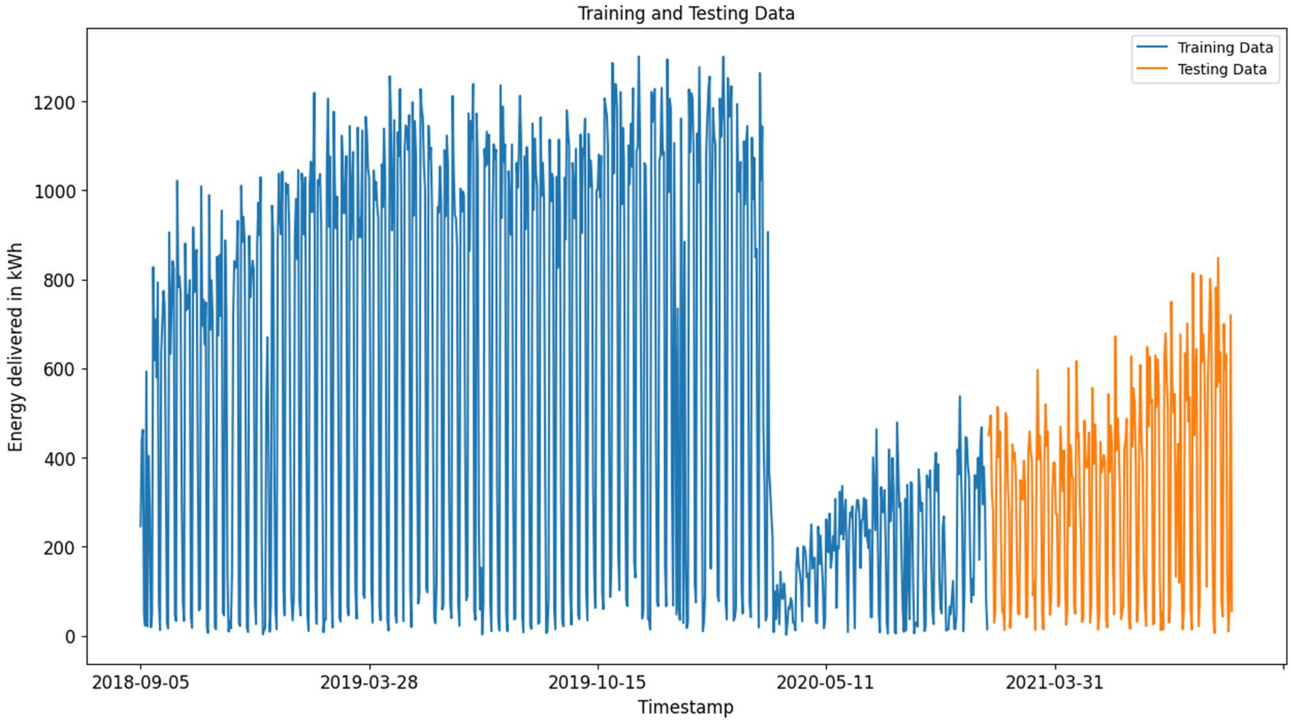

Fig. 8. JPL dataset split into training and testing sets.

Variables	Values
Scaler	MinMaxScaler
Optimizer	Adam
Error metrics	MAE, MSE
Epochs	100
Batch size	32
Learning rate	0.001
Activation functions	ReLu
Dropout rates	0.2

Table 1. An overview of hyperparameters.

trends. The differentiation between training and testing data indicates that the dataset is utilized for model prediction or evaluation based on the JPL dataset. The overall number of training timesteps is 739, whereas the number of testing timesteps is 210, as shown in Fig. 8.

The energy consumption analysis of Caltech and JPL data shows that the charging demand contains the variation, where the energy consumption has decreased from 2020 and increases gradually till the end of 2021, however, the increase in energy consumption is still lower as compared to the increased consumption from the September 2018 till December 2019.

Implementation

Table 1 shows the setup for the model's implementation process. Data preprocessing is handled by MinMaxScaler, which normalizes feature values to a specified range, enhancing model performance and convergence. The training utilizes the Adam optimizer, known for its adaptive learning rate capabilities, set here at 0.001 to ensure stable and effective updates to the model's weights. A batch size of 32 ensures that the model's accuracy and computational efficiency are balanced during the 100 training epochs. Performance is assessed by calculating the MAE and MSE, which offer valuable information regarding the accuracy of predictions and the extent of errors. The model's capacity to learn intricate patterns is improved by employing ReLU (Rectified Linear Unit)activation functions, which introduce non-linearity. A dropout rate of 0.2 is implemented to mitigate overfitting, which involves the random deactivation of 40% of neurons during the training process. This approach is intended to enhance robustness and generalization.

An input size of 1 was set in the proposed LSTM model to correspond to the nature of the dataset that is univariate time series. The size 64 was chosen as a compromise between time-window size, comparable to the temporal dependencies involving complex temporal dependencies and avoiding excessive computational and overfitting expenses. Regularization and enhancement of generalization were included by the setting of dropout rate to 0.2. The fully-connected layers (64 32 1) consecutively decrease dimension and the network learns the non-linear mappings through them, and then gives the final prediction. The Adam optimizer for its strength and effectiveness with sparse gradients was used at the initial learning rate set at 0.001. To ensure that the training did not involve overfitting and that the learning was dynamic enough, simple mechanisms were used to automatically control the learning rate while training by using a learning rate scheduler (ReduceLROnPlateau) and early stopping. The set of values was founded on the general suggestions in the literature and justified by pre-experiments on tuning in order to obtain stable training and satisfactory performance of the predictive model.

The models are trained using the training sets, and the predictions are performed using the test sets for each period, including 30 days, 120 days, and 240 days, for Caltech and JPL charging stations. The 30-day period is appropriate for medium-term predictions, allowing the model to account for recent trends and provide timely forecasts for daily or weekly operations. The 120-day period, on the other hand, provides a medium-term perspective, capturing seasonal variations, such as changes in charging patterns. The 240-day period covers long-term forecasting demands, which are critical for making strategic choices, developing infrastructure, and predicting future charging requirements. While longer periods may present more uncertainty, they give a more comprehensive perspective, which aids in future-proofing EV charging infrastructure and managing resources over time. This mix of periods enables a balanced approach, delivering insights across multiple timelines and matching with the changing demands of EV charging demand forecasts.

Results and analysis

This section provides an overview of the model's performance for the test sets using Caltech and JPL charging stations of the ACN data for the 30, 120, and 240 days, and compares with the actual data. This section provides a brief analysis of the results.

Caltech dataset

The baseline models, including the LSTM and Transformer models, and the proposed Hybrid Transformer model, are implemented using the Caltech Dataset for the predictions and the test set. The performance of the test set predictions is compared with the actual demand over the specific periods, and the evaluation is performed using MSE and MAE.

LSTM model

With an MAE of 38.683 and an MSE of 47.658, the 30-day test set prediction results using LSTM on the Caltech data show promise. Figure 9 shows the predicted and actual energy demand over 30 days.

Using LSTM on the Caltech data, the prediction results for the 120-day test set show an MAE of 91.940 and an MSE of 97.749. Figure 10 shows the actual and predicted values over 120 days using LSTM for Caltech data.

The 240-day test set LSTM prediction results using the Caltech data show an MAE of 109.754 and an MSE of 114.567. Figure 11 compares the predicted and actual values for 240 days.

Transformer

The 30-day test set prediction results for the Transformer model on the Caltech data show an MAE of 39.183 and an MSE of 47.507. Figure 12 shows the actual and predicted values for 30 days using Caltech data for the Transformer.

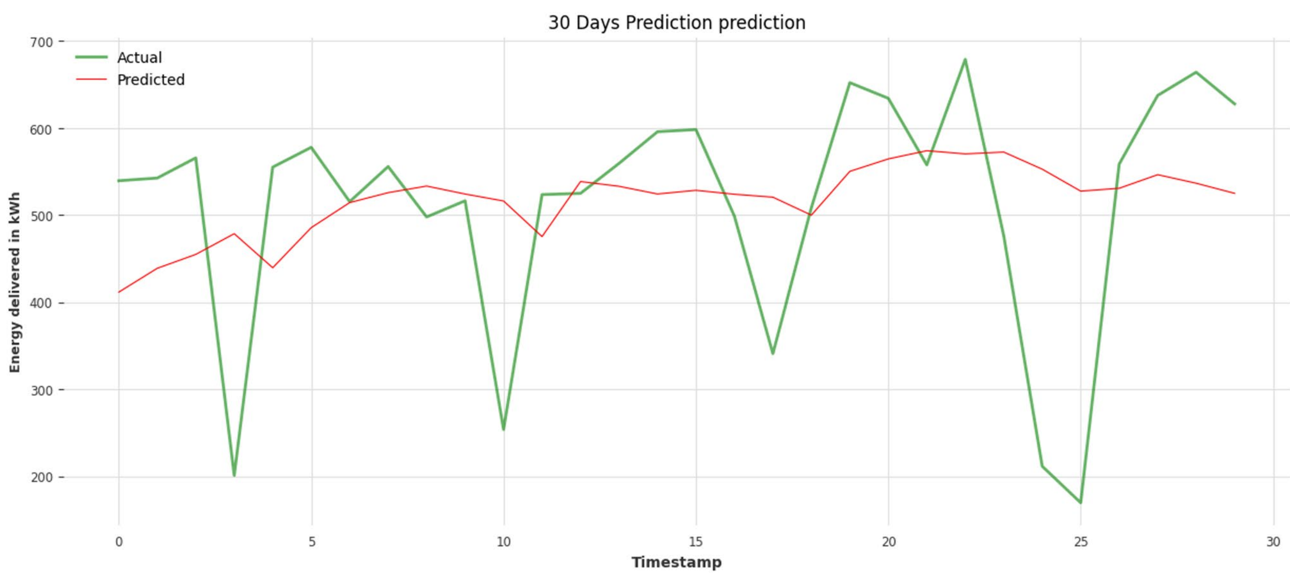

Fig. 9. 30 days prediction using LSTM for test set of Caltech data.

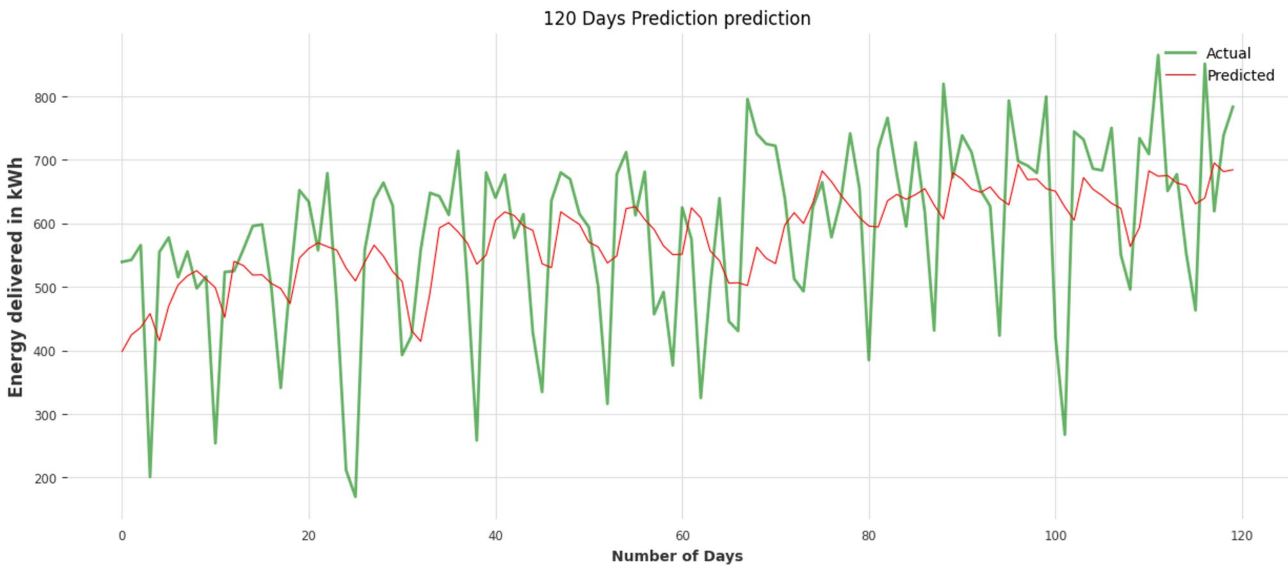

Fig. 10. 120 days prediction using LSTM for test set of Caltech data.

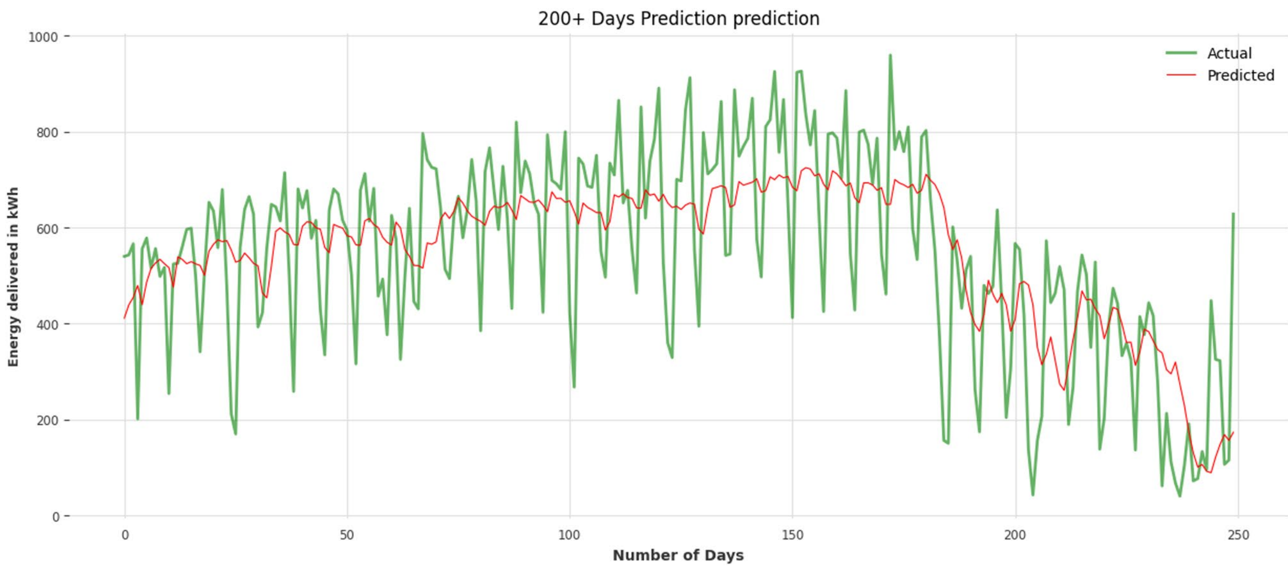

Fig. 11. 240 days prediction using LSTM for test set of Caltech data.

Using a Transformer model with the Caltech data, the predictions for the 120-day test set yielded an MAE of 93.461 and an MSE of 98.835. Figure 13 shows the actual and predicted values for 120 days using Caltech data for the Transformer.

The results of the 240-day test set using the Transformer model on the Caltech data show an MAE of 113.956 and an MSE of 117.653 for the predictions. Figure 14 shows the actual and predicted values for 240 days using Caltech data for the Transformer.

Hybrid transformer model

The proposed Hybrid Transformer model, utilizing the Caltech data and an LSTM encoder-decoder, produced an MAE of 32.416 and an MSE of 38.221 for the 30-day test set. With very few error metrics, the hybrid model captures the data's accurate patterns and makes reliable short-term forecasts. Combining the Transformer and LSTM components improves the model's ability to handle sequential dependencies, leading to far better predictive performance than using either design alone. According to the findings, this hybrid model outperforms baselines for the 30-day forecast. Figure 15 compares the actual and predicted energy for Caltech data using the proposed model for 30 days.

In the 120-day test set, the suggested Hybrid Transformer model using Caltech data produced predictions with an MAE of 89.256 and an MSE of 94.040. These error metrics show that the hybrid model keeps getting

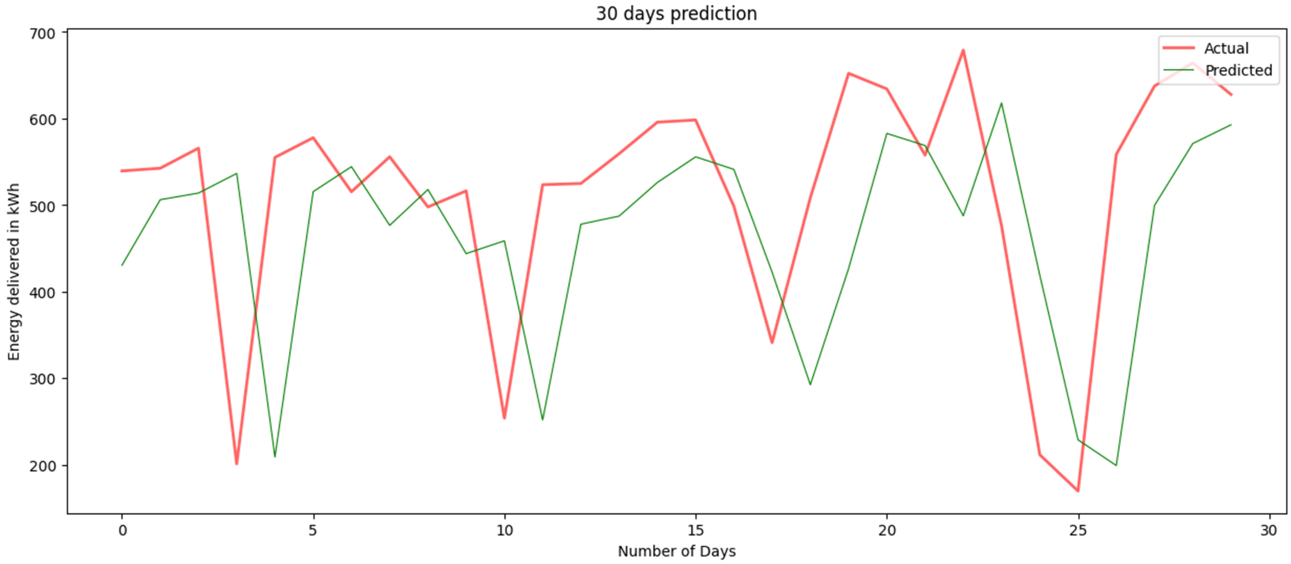

Fig. 12. 30 days prediction using transformer for test set of Caltech data.

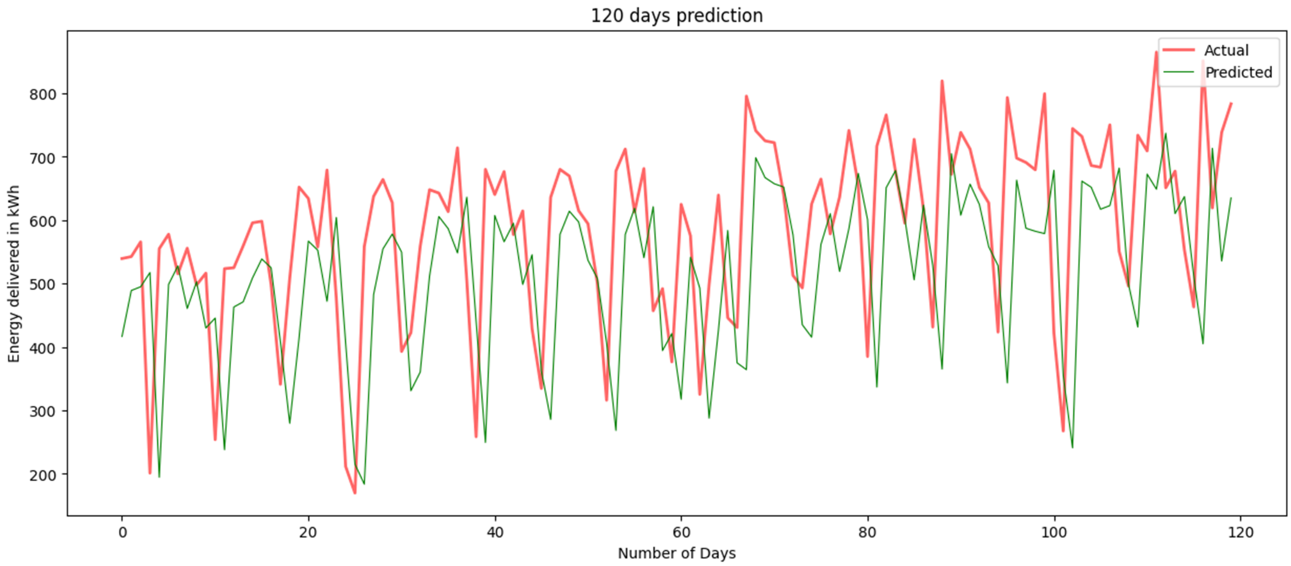

Fig. 13. 120 days prediction using transformer for test set of Caltech data.

rather good predictions over a long period. Figure 16 compares the actual and predicted values for the following prediction.

The 240-day test set prediction results using the proposed Hybrid Transformer model on the Caltech data were 107.452 MAE and 112.040 MSE. Figure 17 compares the actual and predicted demand.

JPL dataset

The test set is used to test and predict the demand using the baseline models, such as LSTM, and Transformer. Predictions for 30, 120, and 240 days are performed. We use MAE and MSE to evaluate the performance.

LSTM

The 30-day test set prediction results for LSTM on the JPL data show an MAE of 82.813 and an MSE of 90.528. Figure 18 compares actual and predicted values across 30 days using LSTM with JPL data.

For the 120-day test set, the predictions made using LSTM on the JPL data show an MAE of 94.231 and an MSE of 98.424. Figure 19 compares actual and predicted values across 120 days using LSTM with JPL data.

The 240-day test set predictions using LSTM on JPL data have an MAE of 105.238 and an MSE of 112.687. Figure 20 compares actual and predicted values across 240 days using LSTM with JPL data.

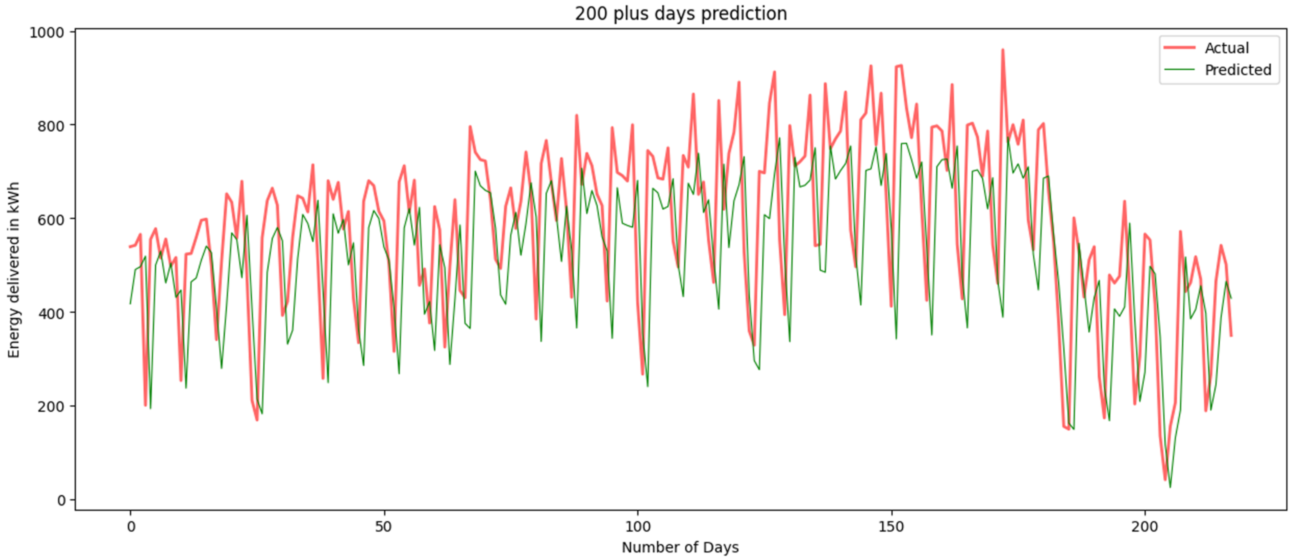

Fig. 14. 240 days prediction using transformer for test set of Caltech data.

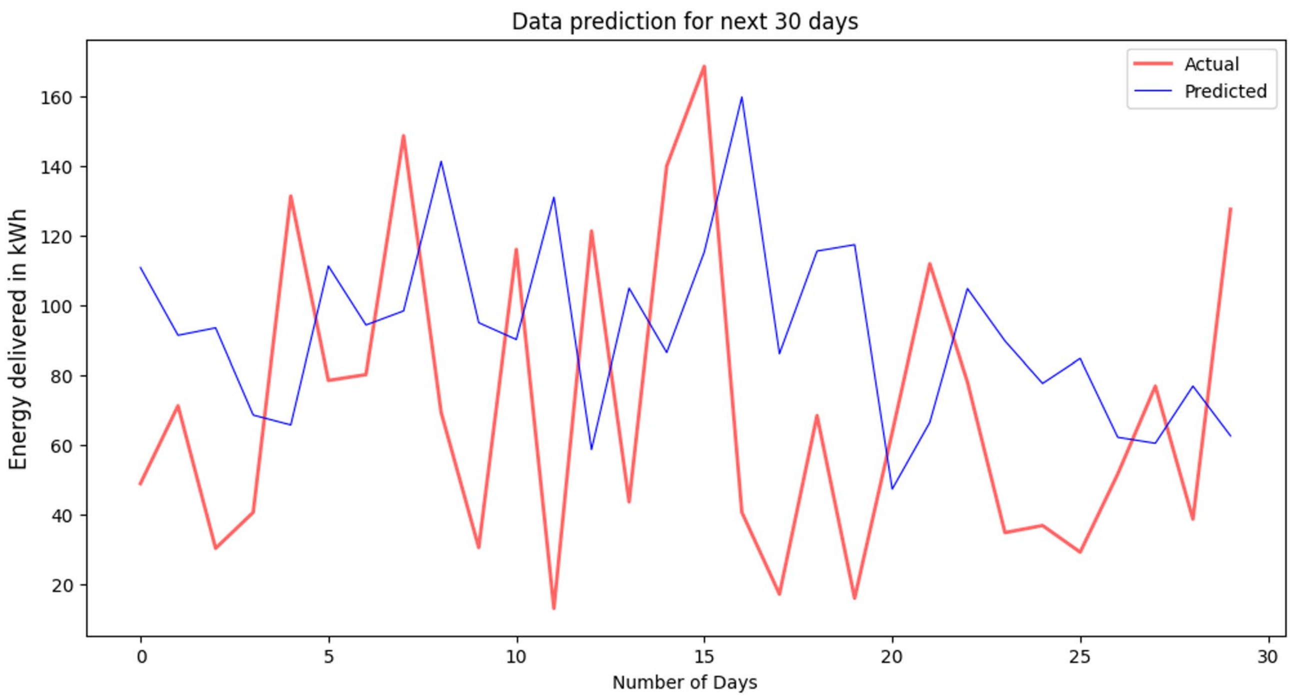

Fig. 15. 30 days prediction using proposed hybrid transformer for test set of Caltech data.

Transformer
The Transformer model was utilized to predict the 30-day test set using JPL data. The results revealed an MAE of 82.813 and an MSE of 87.098. Figure 21 compares actual and predicted values across 30 days using LSTM with JPL data.

The Transformer model's performance on the 120-day test set using JPL data yielded an MAE of 92.247 and an MSE of 97.392. Figure 22 compares actual and predicted values across 30 days using LSTM with JPL data.

The Transformer model was employed to generate predictions for the 240-day trial set. The results yielded an MSE of 107.266 and an MAE of 114.581. These substantial error values indicate that the Transformer model struggles significantly with making accurate long-term predictions using JPL data. Figure 23 compares actual and predicted values across 30 days using LSTM with JPL data.

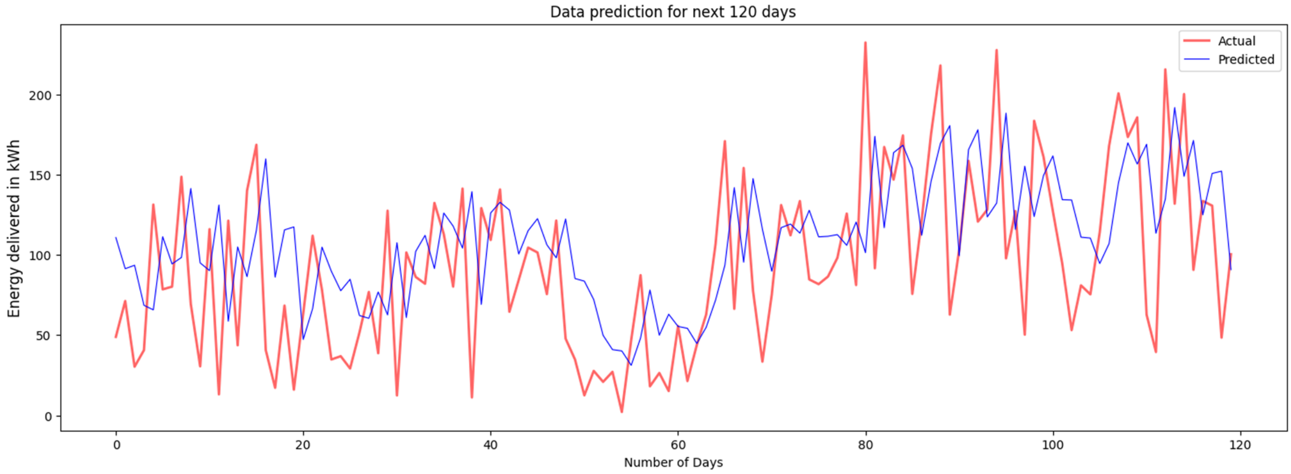

Fig. 16. 120 days prediction using proposed hybrid transformer for test set of Caltech data.

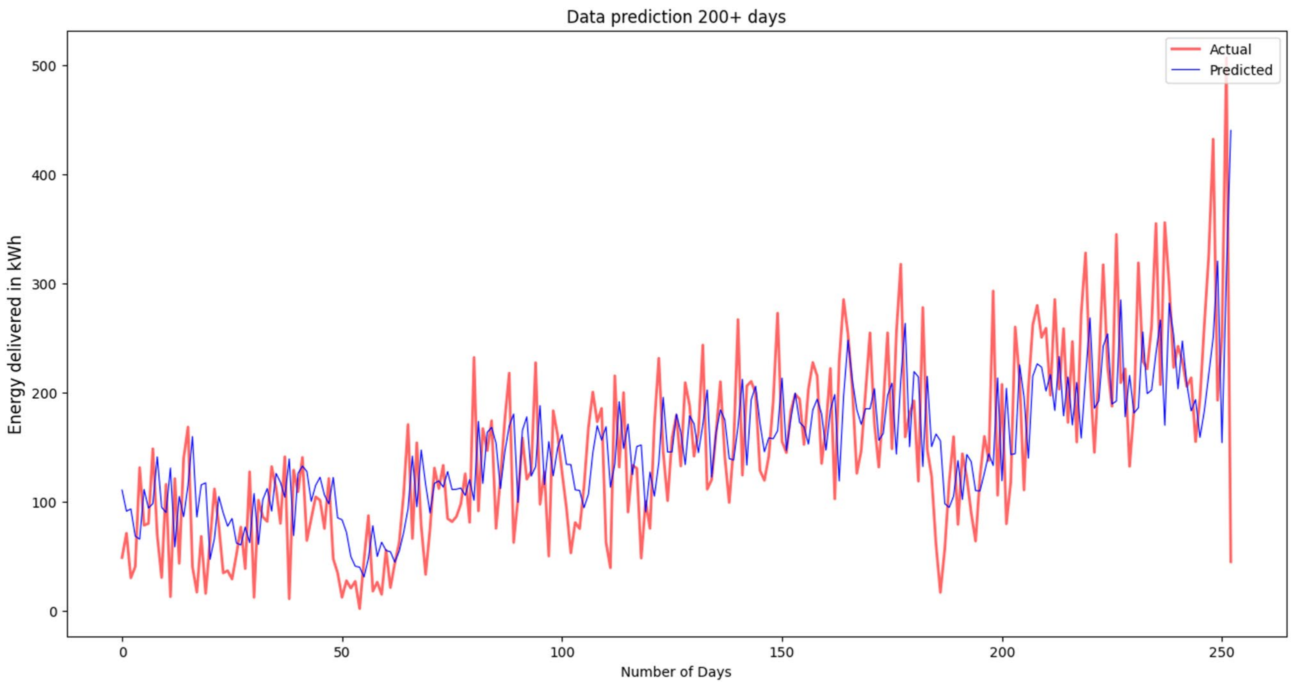

Fig. 17. 240 days prediction using proposed hybrid transformer for test set of Caltech data.

Hybrid transformer
Using the JPL data, the suggested Hybrid Transformer model produced 30-day test set prediction results with an MAE of 62.186 and an MSE of 69.536. Based on these measures, the hybrid model performs better for the 30-day prediction. Figure 24 compares actual and predicted values across 30 days using the proposed hybrid Transformer using LSTM with JPL data.

Using the JPL data, the suggested Hybrid Transformer model produced MAE and MSE forecasts of 89.526 and 93.328, respectively, for the 120-day test set. As these error metrics show, the hybrid model can capture the more significant trends over a longer time horizon. Figure 25 compares actual and predicted values across 120 days using the Transformer with JPL data.

Using the proposed Hybrid Transformer model with the JPL data, the results for the 240-day test set of predictions reveal an MAE of 103.085 and an MSE of 109.010. Figure 26 compares actual and predicted values across 240 days using the proposed hybrid LSTM-based Transformer with JPL data.

Performance analysis
The proposed and baseline models are implemented using Caltech and JPL charging stations. The predictions are made for 30, 120, and 240 days for all the models using these datasets. The results show that the Hybrid

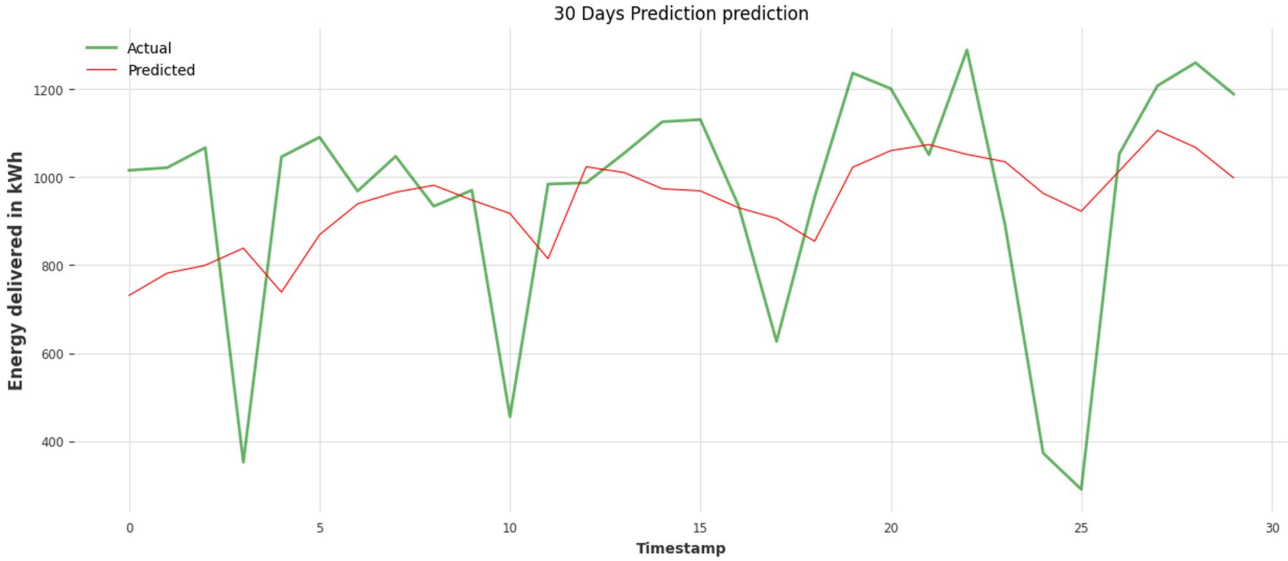

Fig. 18. 30 days prediction using LSTM for test set of JPL data.

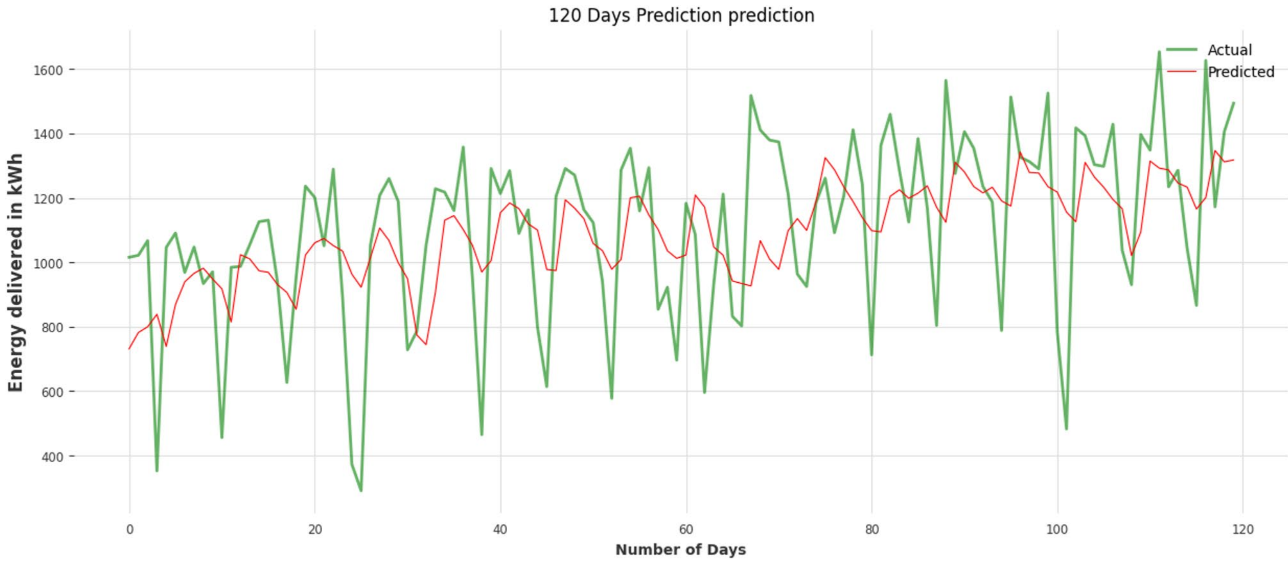

Fig. 19. 120 days prediction using LSTM for test set of JPL data.

Transformer model performs better than the baseline models. Furthermore, the comparison of the MAE and MSE is provided for each period. In the 30 days, the LSTM-Transformer model performed better than LSTM and Transformer, achieving the lowest MAE of 32.416 and MSE of 38.221. Therefore, it performed the best across both metrics. For 120 days, the LSTM-Transformer model sustained its accuracy, achieving the lowest MAE of 89.256 and MSE of 94.040, signifying enhanced prediction accuracy relative to other models. During the 240-day prediction, the Transformer model demonstrated better performance, achieving the lowest MAE of 107.452 and MSE of 112.040, continuously surpassing other methods in error minimization during all periods. The comparison of the results using MAE and MSE, as well as predictions for 30, 120, and 240 days ahead using Caltech test data, is provided in Table 2. The comparison is also illustrated in Fig. 27.

The comparative performance analysis across different time horizons as shown in Table 3 reveals that the LSTM-Transformer model consistently outperforms both standalone LSTM and Transformer models in terms of MAE and MSE. At the 30-day horizon, the LSTM-Transformer shows the most substantial improvement, reducing MAE by up to 17.27% and MSE by 19.79%, indicating its strength in short-term prediction. The improvements, while more moderate, remain consistent for longer horizons—120 and 240 days—with the LSTM-Transformer achieving up to 5.71% better MAE and 4.85% lower MSE. These results highlight the LSTM-Transformer's effectiveness in capturing both sequential and contextual patterns for enhanced forecasting

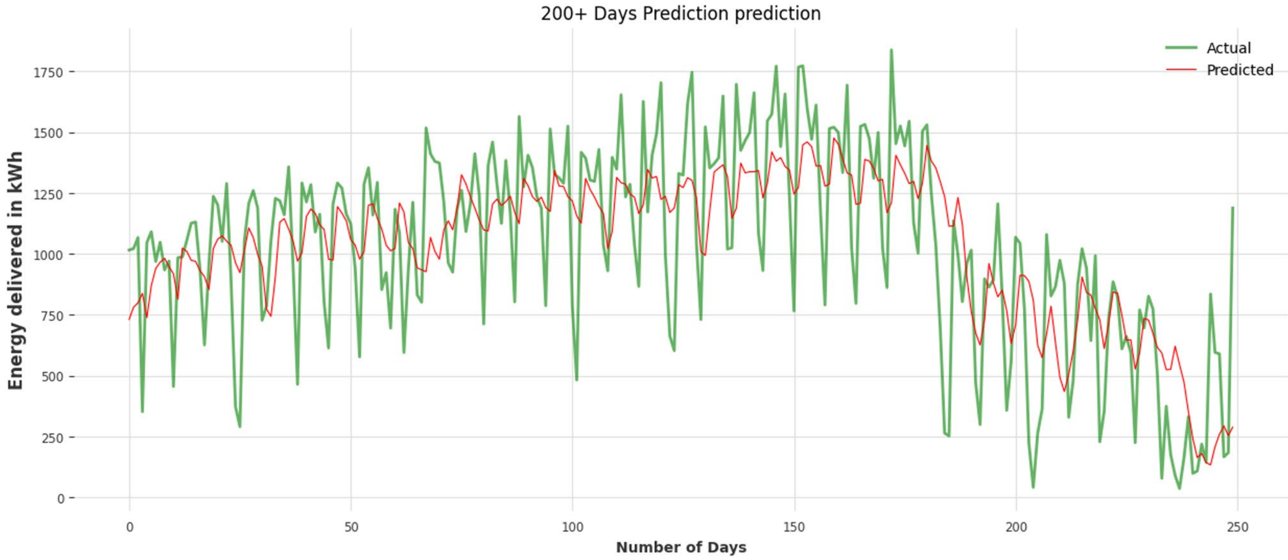

Fig. 20. 240 days prediction using LSTM for test set of JPL data.

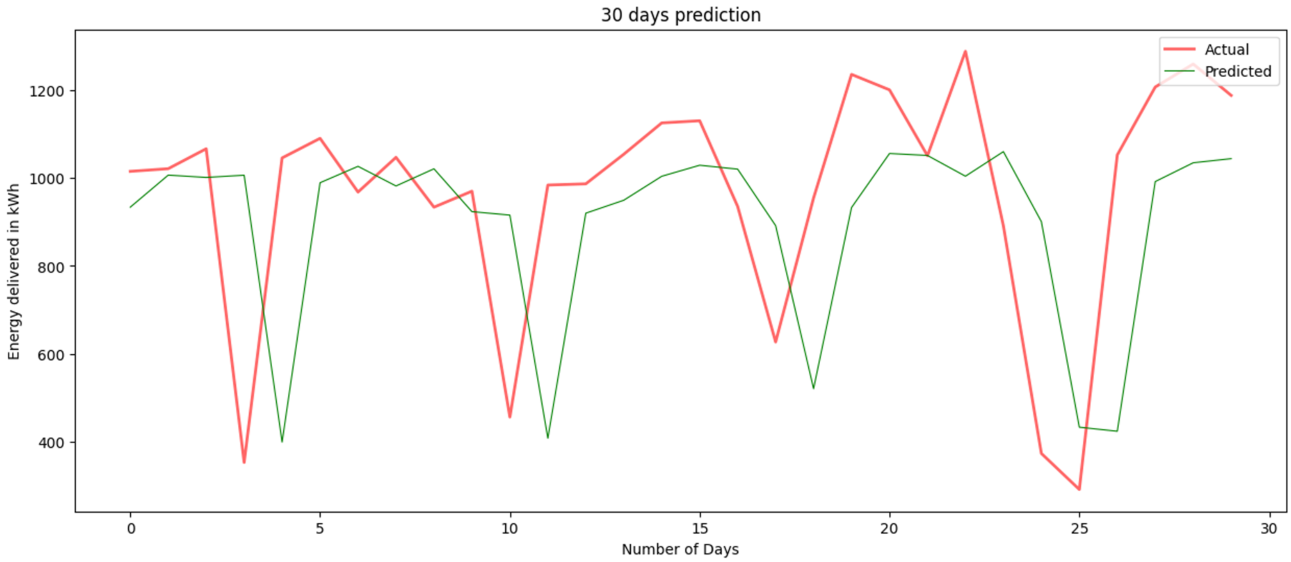

Fig. 21. 30 days prediction using transformer for test set of JPL data.

accuracy. Figure 28 illustrates the comparison of relative improvements using the radar map for the proposed model compared with the baselines using caltech data.

For the JPL data, during the 30-day interval, the LSTM-Transformer model showed outstanding results, with the lowest MAE of 62.186 and MSE of 69.536, surpassing the LSTM and Transformer models. Over 120 days, the LSTM-Transformer also demonstrated exceptional performance by achieving the lowest MAE of 89.516 and MSE of 93.328, with the Transformer trailing closely behind. During the 240-day interval, LSTM-Transformer sustained its dominance, reaching the lowest MAE (103.085) and MSE (109.010), continuously surpassing the other models in error minimization across all durations. The performance of the models using JPL data is provided using MAE and MSE for the test sets, including 30-, 120- and 240-day periods, in Table 4. Also, the comparison is illustrated in Fig. 29.

The LSTM-Transformer model shows significant improvements in predictive accuracy throughout all periods when compared to the separate LSTM and Transformer models, as seen in Table 5. With MSE down 23.17% over LSTM and 9.82% over Transformer, and MAE down 24.91% over both baselines, it provides the most gains over the 30-day horizon. Throughout the 120-day period, the LSTM-Transformer raises MAE by 2.96% (in comparison to Transformer) and 5.00% (in comparison to LSTM), and MSE by 5.17% and 4.17%. While the advantages are less evident throughout the 240-day period, they remain stable, with MSE improvements of up to

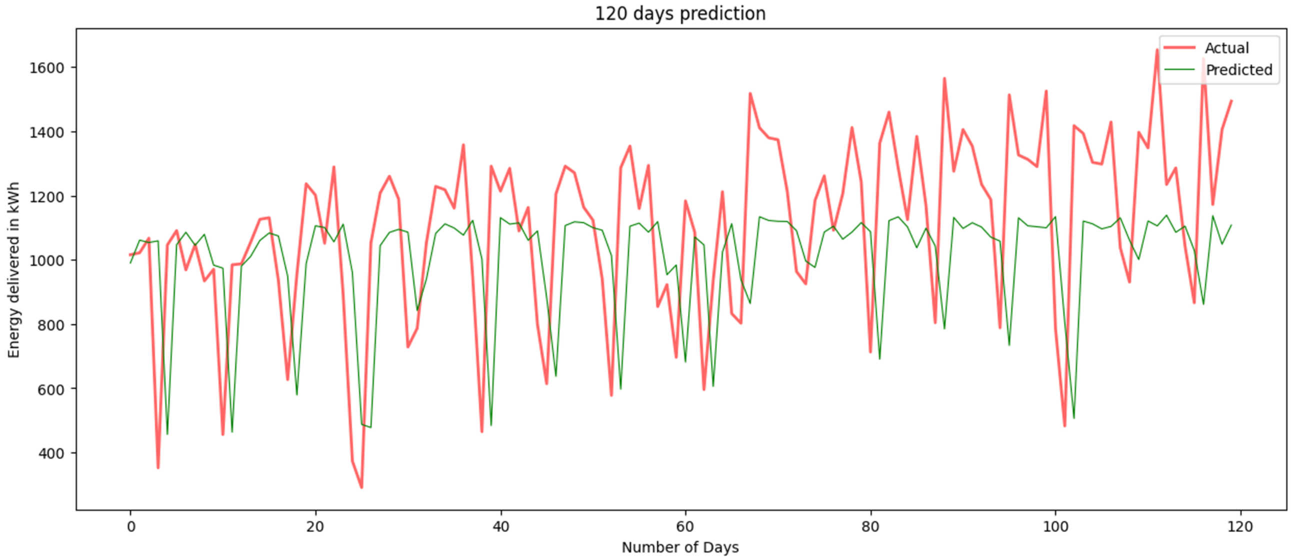

Fig. 22. 120 days prediction using transformer for test set of JPL data.

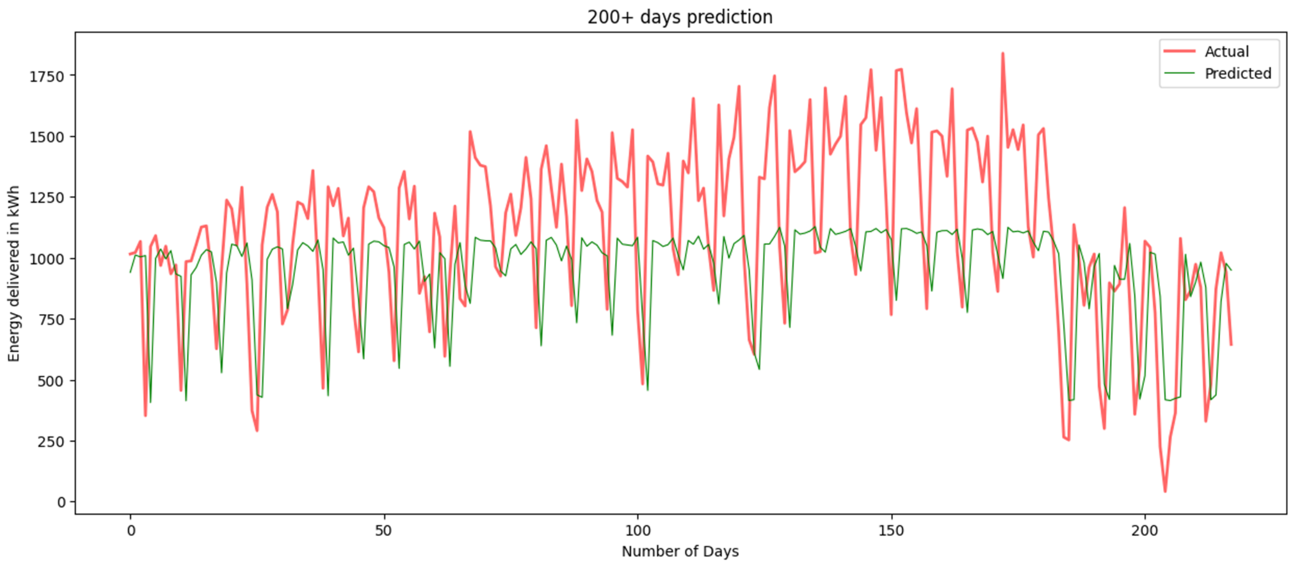

Fig. 23. 240 days prediction using transformer for test set of JPL data.

4.86% and MAE reductions of up to 3.90%. Figure 30 illustrates the comparison of relative improvements using the radar map for the proposed model compared with the baselines using JPL data.

Discussion

The combination of LSTM-Transformer, which includes an LSTM-based encoder and decoder, is also useful to infer both short-term time series and long-term contextual dependencies due to its effectiveness in predicting medium and long-term EV charging demand. Transformer self-attention increases the capability of the model to identify its higher level trends and complicated time dependence in larger areas, whereas LSTM layers are better suited to model sequential dependence and local fluctuations in the charging actions. The result of this synergy is more precise and solid predictability, which takes into account the dynamism and heterogeneity of the EV charging requirement.

The results shows that the proposed model performs better in the 30-day prediction because it can capture recent and steadier patterns in EV charging demand, where short-term variations are more predictable and less influenced by external variables such as seasonal changes. The 120-day and 240-day predictions exhibit lesser gains since these longer horizons are intrinsically more difficult to predict due to greater uncertainty and impact from seasonal fluctuations, changes in user behavior, and external situations such as weather or public events. The model's ability to accurately predict EV charging demand reduces over time due to complicated dynamics

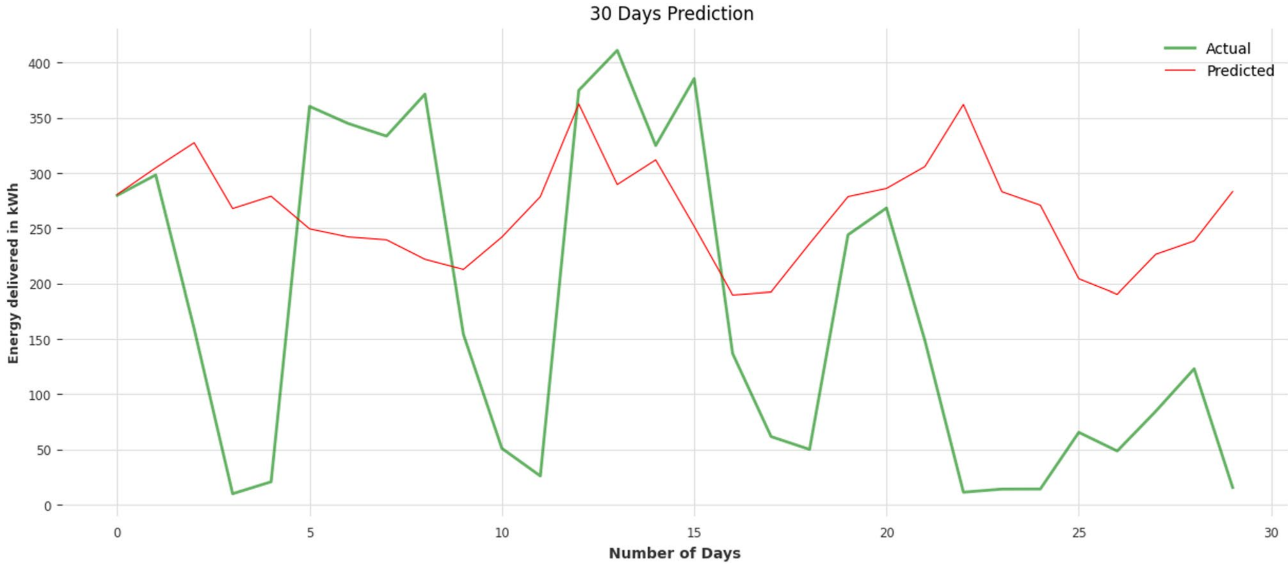

Fig. 24. 30 days prediction using proposed hybrid transformer for test set of JPL data.

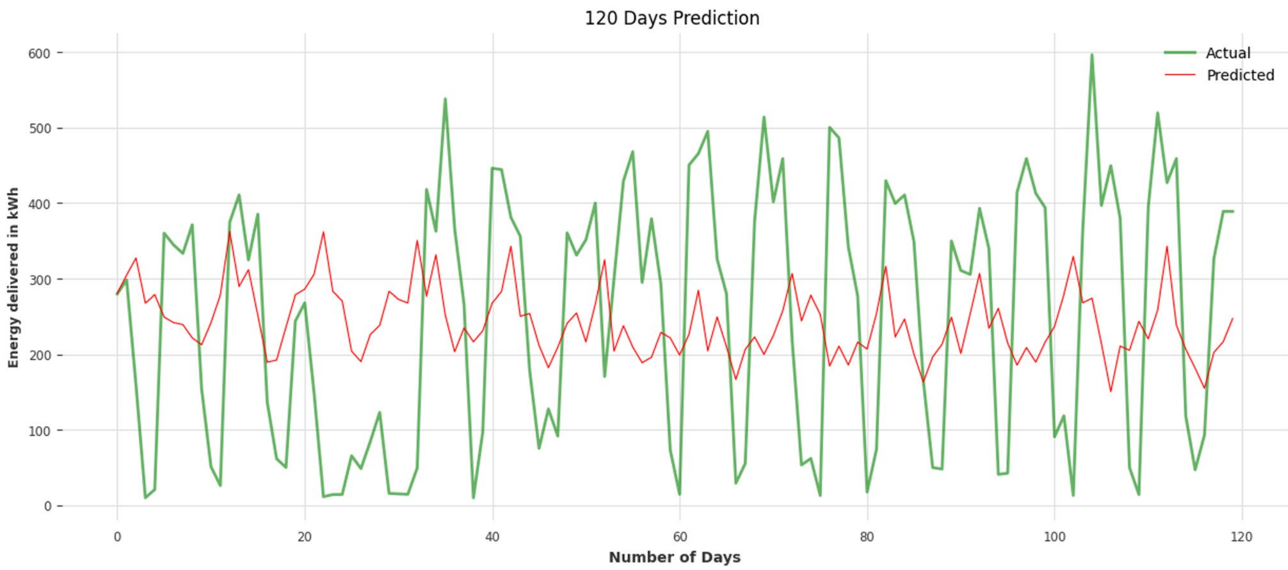

Fig. 25. 120 days prediction using proposed hybrid transformer for test set of JPL data.

and longer-term trends that are more difficult to capture. Furthermore, the increasing variability in the data over longer periods makes it more difficult for the model to maintain high accuracy, resulting in lesser gains in the 120-day and 240-day forecasts.

However, despite the improved results of the proposed hybrid LSTM-Transformer architecture in terms of the ability to accomplish the tasks of medium- and long-term EV charging demand prediction, several drawbacks should be noted. The complexity of the model makes it more demanding and the training time can be longer which can be restrictive in the usage of real-time applications in other environments that are resource-limited. Also, the model is now based on historical finding of charging and it is possible that it can be more accurate by considering the external factors like weather, traffic patterns or user behavior. In future, the work could be aimed at the integration of multi-modal data sources, the exploration of model compression methods that would minimise the computational requirements, and the development of adaptive solutions that would dynamically update the prediction based on the altered charging patterns and external circumstances.

Conclusion

This study introduces a Hybrid Transformer model that combines an LSTM-based encoder and decoder to predict medium-term (30 days) and long-term (120 and 240 days) outcomes. The model is developed and

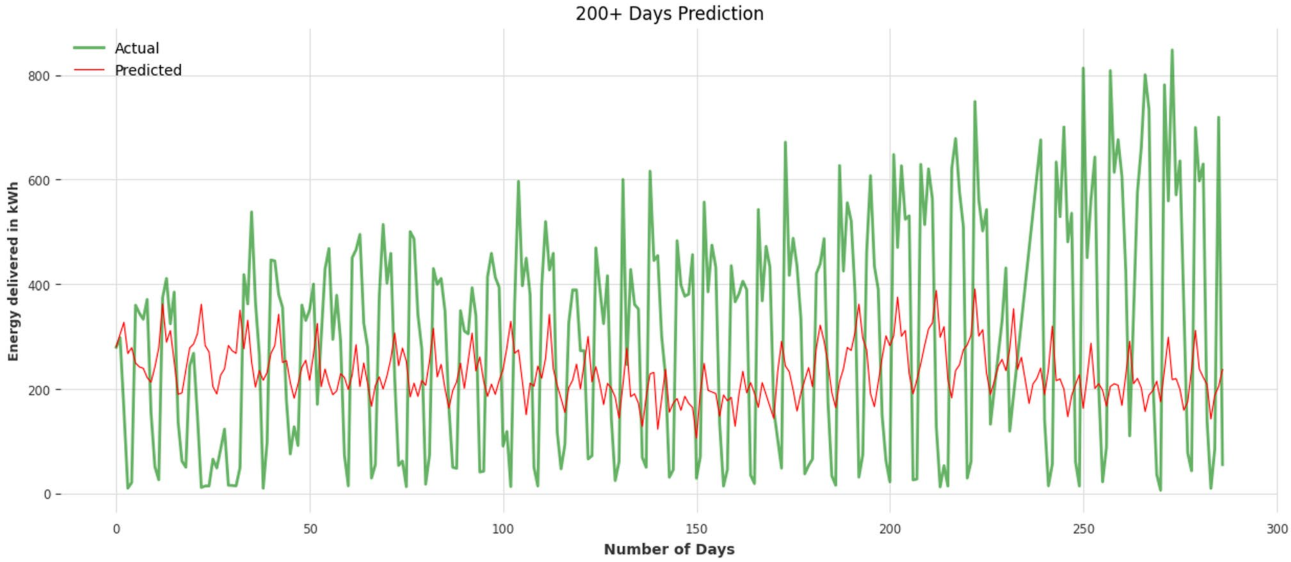

Fig. 26. 240 days prediction using proposed hybrid transformer for test set of JPL data.

Period	C	LSTM	Transformer	LSTM-transformer
30 days	MAE	38.683	39.183	32.416
	MSE	47.658	47.507	38.221
120 days	MAE	91.940	93.461	89.256
	MSE	97.749	98.835	94.040
240 days	MAE	109.754	113.956	107.452
	MSE	114.567	117.653	112.040

Table 2. Performance comparison for 30 days, 120 days, and 240 days using Caltech test Set.

evaluated using ACN data from Caltech and JPL. Baseline models, including LSTM and a standard Transformer, are implemented for comparison. Each dataset is divided into training and testing sets, with the latter used for making predictions. The forecasts are then compared to actual demand across 30, 120, and 240 days. For Caltech data, the proposed LSTM-Transformer beats LSTM across all periods, with a 16.18% increase in MAE and 19.79% in MSE at 30 days. The gains are more moderate, at 2.92% in MAE and 3.79% in MSE for 120 days, and 2.10% in MAE and 2.20% in MSE for 240 days, demonstrating persistent superior performance. Furthermore, with JPL data, the LSTM-Transformer outperforms the Transformer, particularly in MAE. After 30 days, it boosts MAE by 17.27% and MSE by 19.56%. For 120 days, it improves MAE by 4.50% and MSE by 4.85%, while at 240 days, it outperforms by 5.71% and 4.77%, respectively. The results show that the Hybrid Transformer using the LSTM encoder-decoder model performs better than the baselines, including LSTM and the Transformer model, for all the periods using both datasets. The predictions performed for 30, 120, and 240 days ahead also show that the error for the following period predictions using the proposed hybrid Transformer model is better than all the baselines. Furthermore, the LSTM model also performs better than the Transformer model.

This research performs medium—and long-term predictions only; however, the hybrid Transformer model can be used for short-term predictions, including a few hours to a few days ahead, i.e., 1 h ahead, 24 h ahead, or 7 days ahead. The proposed model can be further fine-tuned for better results, especially for the long term. External factors, including Weather data and seasonal factors, can also be included in the future.

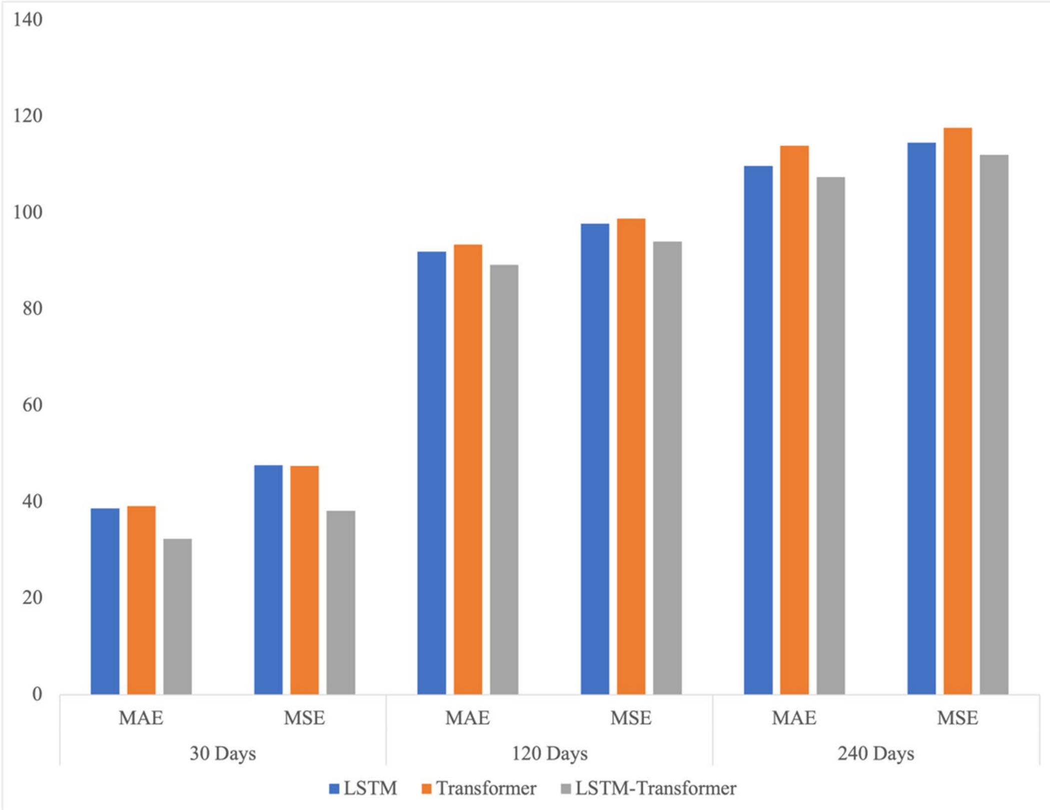

Fig. 27. Performance comparison using Caltech data.

Periods	Metrics	vs. LSTM (%)	vs. transformer (%)
30 days	MAE	16.18	17.27
	MSE	19.79	19.56
120 days	MAE	2.92	4.50
	MSE	3.79	4.85
240 days	MAE	2.10	5.71
	MSE	2.20	4.77

Table 3. Relative improvement of the proposed LSTM-Transformer using Caltech Data.

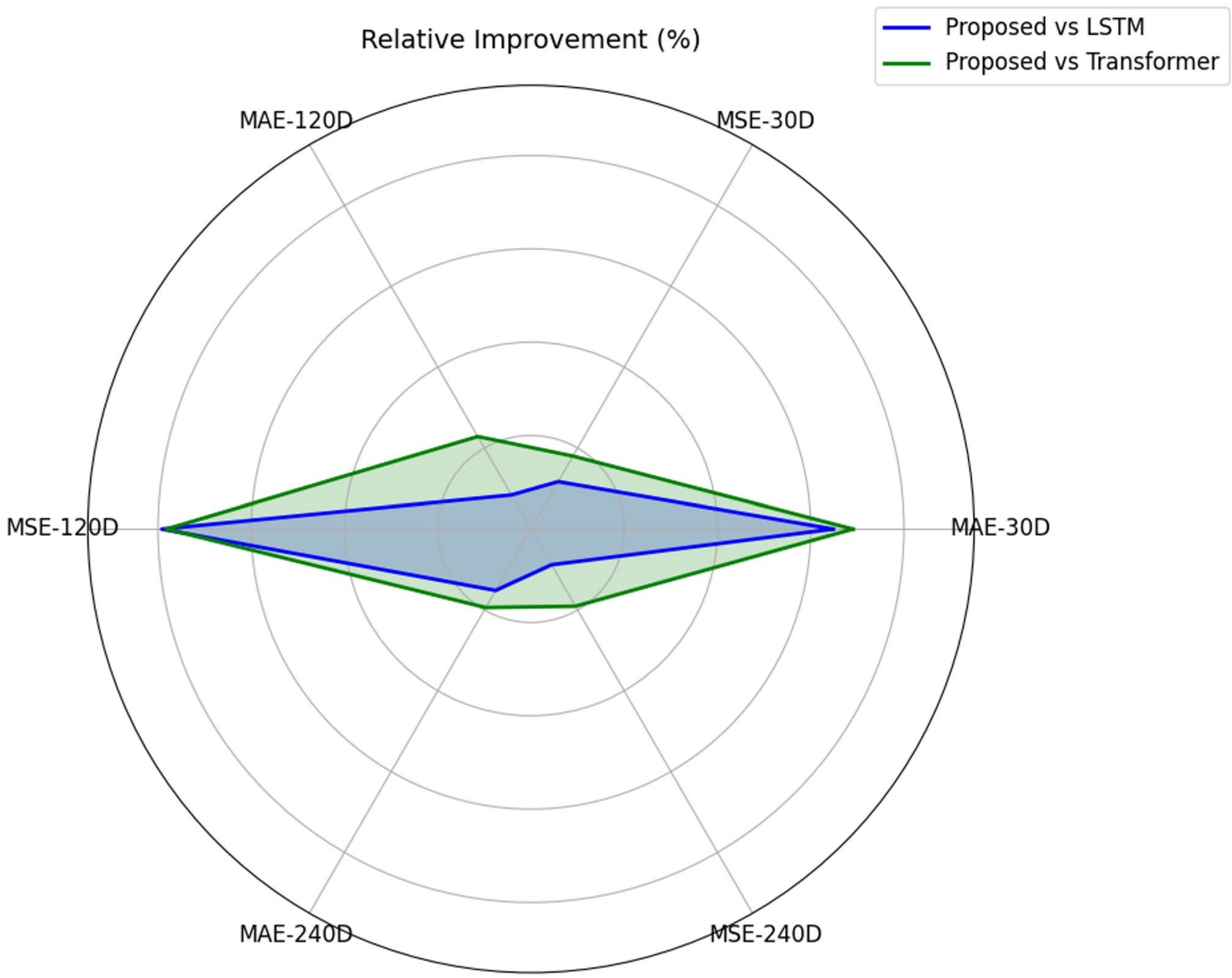

Fig. 28. Radar map of relative improvements for the proposed model compared with baselines using Caltech data.

Period	Models	LSTM	Transformer	LSTM-transformer
30 days	MAE	82.813	82.813	62.186
	MSE	90.528	77.098	69.536
120 days	MAE	94.231	92.247	89.516
	MSE	98.424	97.392	93.328
240 days	MAE	105.238	107.266	103.085
	MSE	112.687	114.581	109.010

Table 4. Performance comparison for 30 days, 120 days and 240 days using JPL test Set.

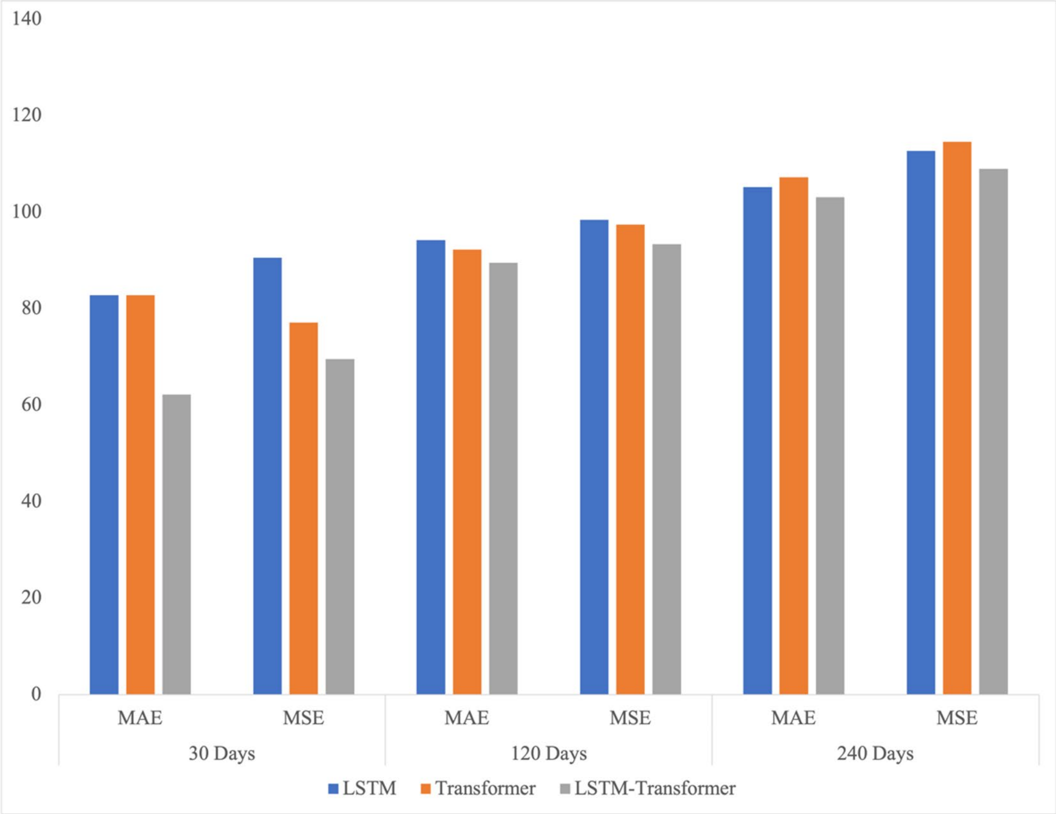

Fig. 29. Performance comparison using JPL data.

Periods	Metrics	vs. LSTM (%)	vs. transformer (%)
30 days	MAE	24.91	24.91
	MSE	23.17	9.82
120 days	MAE	5.00	2.96
	MSE	5.17	4.17
240 days	MAE	2.05	3.90
	MSE	3.26	4.86

Table 5. Relative improvement of the proposed LSTM-Transformer using JPL data.

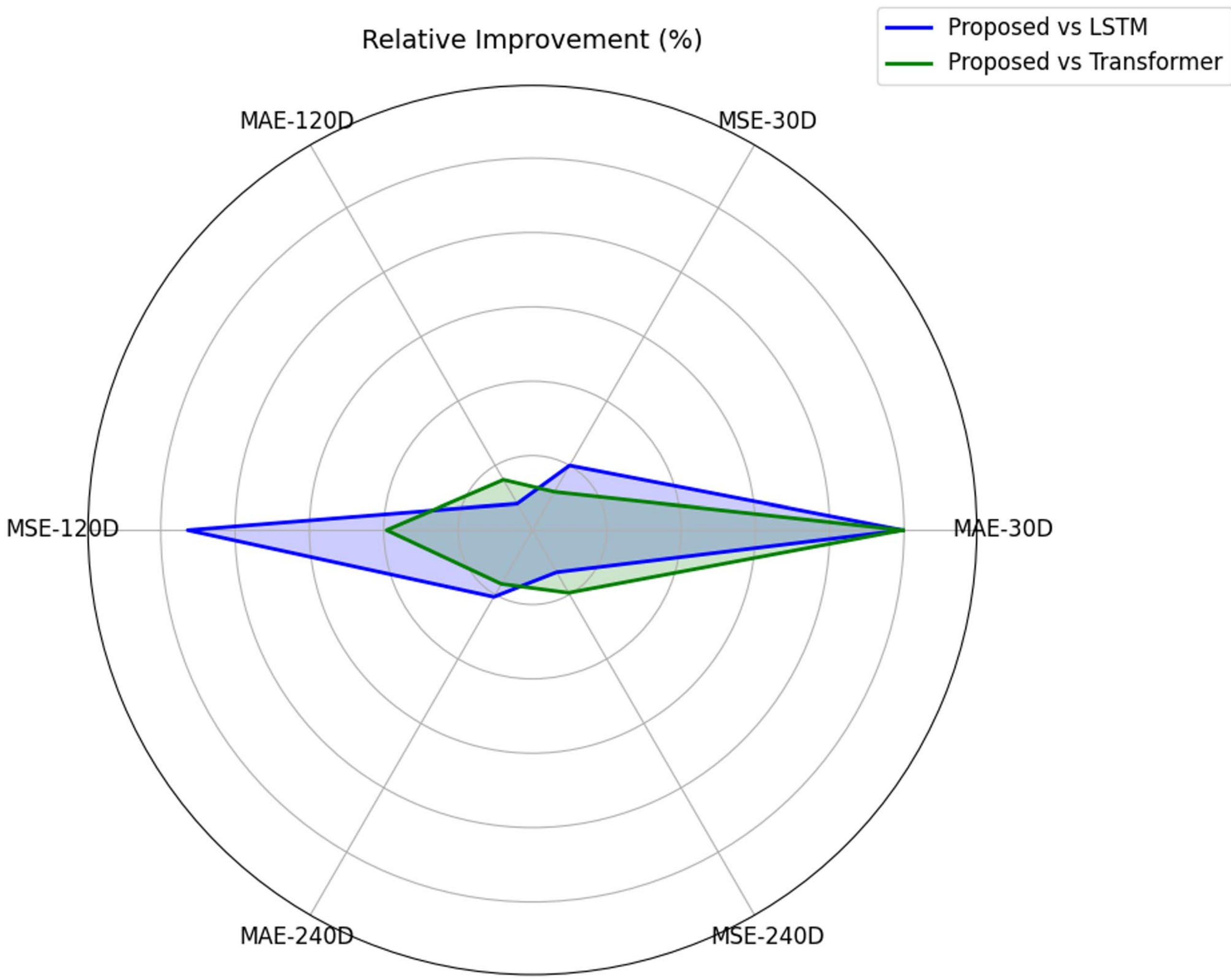

Fig. 30. Radar map of relative improvements for the proposed model compared with baselines using JPL data.

Data availability
The data will be made available upon request to the corresponding authors.

Received: 12 February 2025; Accepted: 15 September 2025
Published online: 21 October 2025

References
1. Breyer, C., Khalili, S., Rantanen, E. & Bogdanov, D. Solar photovoltaic capacity demand for a fully sustainable transport sector–how to fulfil the Paris agreement by 2050, in *IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)(A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)*, 2018: IEEE, pp. 1632–1637., 2018: IEEE, pp. 1632–1637. (2018).
2. Zhang, X., Xie, J., Rao, R. & Liang, Y. Policy incentives for the adoption of electric vehicles across countries, *Sustainability*, vol. 6, no. 11, pp. 8056–8078, (2014).
3. Zhang, X., Gao, F., Gong, X., Wang, Z. & Liu, Y. Comparison of climate change impact between power system of electric vehicles and internal combustion engine vehicles, in *Advances in Energy and Environmental Materials: Proceedings of Chinese Materials Conference 2017 18th*, : Springer, pp. 739–747. (2018).
4. Global, E. Global EV Outlook. https://www.iea.org/data-and-statistics/data-product/global-ev-outlook-2022. (accessed 2022).
5. Slednev, V., Jochem, P. & Fichtner, W. Impacts of electric vehicles on the European high and extra high voltage power grid. *J. Ind. Ecol.* **26** (3), 824–837 (2022).
6. Wu, X., Hu, X., Yin, X. & Moura, S. J. Stochastic optimal energy management of smart home with PEV energy storage. *IEEE Trans. Smart Grid.* **9** (3), 2065–2075 (2016).
7. Toquica, D., De Oliveira-De, P. M., Jesus & Cadena, A. I. Power market equilibrium considering an Ev storage aggregator exposed to marginal prices-a bilevel optimization approach. *J. Energy Storage.* **28**, 101267 (2020).
8. Feng, H. J., Xi, L. C., Jun, Y. Z., Ling, Y. X. & Jun, H. Review of electric vehicle charging demand forecasting based on multi-source data, in *IEEE Sustainable Power and Energy Conference (iSPEC)*, 2020: IEEE, pp. 139–146., 2020: IEEE, pp. 139–146. (2020).
9. Nalcaci, G., Özmen, A. & Weber, G. W. Long-term load forecasting: models based on MARS, ANN and LR methods. *Cent. Eur. J. Oper. Res.* **27**, 1033–1049 (2019).
10. Chen, Z., Xiong, R., Wang, K. & Jiao, B. Optimal energy management strategy of a plug-in hybrid electric vehicle based on a particle swarm optimization algorithm, *Energies*, vol. 8, no. 5, pp. 3661–3678, (2015).

11. Hannan, M. A., Azidin, F. & Mohamed, A. Hybrid electric vehicles and their challenges: A review. *Renew. Sustain. Energy Rev.* **29**, 135–150 (2014).
12. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning, *nature*, vol. 521, no. 7553, pp. 436–444, (2015).
13. Lim, B. & Zohren, S. Time-series forecasting with deep learning: a survey. *Philosophical Trans. Royal Soc. A*. **379** (2194), 20200209 (2021).
14. Zhu, J. et al. Electric vehicle charging load forecasting: A comparative study of deep learning approaches, *Energies*, vol. 12, no. 14, p. 2692, (2019).
15. Zhu, J., Yang, Z., Guo, Y., Zhang, J. & Yang, H. Short-term load forecasting for electric vehicle charging stations based on deep learning approaches, Applied sciences, **9**, 9, p. 1723, (2019).
16. Hu, T., Liu, K. & Ma, H. Probabilistic electric vehicle charging demand forecast based on deep learning and machine theory of mind, in *2021 IEEE Transportation Electrification Conference & Expo (ITEC)*, : IEEE, pp. 795–799. (2021).
17. Jahangir, H. et al. Plug-in electric vehicle behavior modeling in energy market: A novel deep learning-based approach with clustering technique. *IEEE Trans. Smart Grid*. **11** (6), 4738–4748 (2020).
18. Zhang, X. et al. Deep-learning-based probabilistic forecasting of electric vehicle charging load with a novel queuing model. *IEEE Trans. Cybernetics*. **51** (6), 3157–3170 (2020).
19. Arias, M. B. & Bae, S. Electric vehicle charging demand forecasting model based on big data technologies. *Appl. Energy*. **183**, 327–339 (2016).
20. Shen, X., Zhao, H., Xiang, Y., Lan, P. & Liu, J. Short-term electric vehicles charging load forecasting based on deep learning in low-quality data environments. *Electr. Power Syst. Res.* **212**, 108247 (2022).
21. Gao, Q. et al. Charging load forecasting of electric vehicle based on Monte Carlo and deep learning, in *IEEE Sustainable Power and Energy Conference (iSPEC)*, 2019: IEEE, pp. 1309–1314., 2019: IEEE, pp. 1309–1314. (2019).
22. Zhou, D. et al. Using bayesian deep learning for electric vehicle charging station load forecasting, *Energies*, vol. 15, no. 17, p. 6195, (2022).
23. Yi, Z., Liu, X. C., Wei, R., Chen, X. & Dai, J. Electric vehicle charging demand forecasting using deep learning model. *J. Intell. Transp. Syst.* **26** (6), 690–703 (2022).
24. Shanmuganathan, J., Victoire, A. A., Balraj, G. & Victoire, A. Deep learning LSTM recurrent neural network model for prediction of electric vehicle charging demand, *Sustainability*, vol. 14, no. 16, p. 10207, (2022).
25. Boulakhbar, M., Farag, M., Benabdelaziz, K., Kousksou, T. & Zazi, M. A deep learning approach for prediction of electrical vehicle charging stations power demand in regulated electricity markets: the case of Morocco. *Clean. Energy Syst.* **3**, 100039 (2022).
26. Vishnu, G. et al. Short-term forecasting of electric vehicle load using time series, machine learning, and deep learning techniques. *World Electr. Veh. J.* **14** (9), 266 (2023).
27. Hajirahimi, Z. & Khashei, M. Hybridization of hybrid structures for time series forecasting: A review. *Artif. Intell. Rev.* **56** (2), 1201–1261 (2023).
28. Zhu, J. et al. A novel LSTM based deep learning approach for multi-time scale electric vehicles charging load prediction, in 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), : IEEE, 3531–3536. (2019).
29. Liao, X., Kang, X., Li, M. & Cao, N. Short term load forecasting and early warning of charging station based on PSO-SVM, in *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, : IEEE, pp. 305–308. (2019).
30. Gang, W., Wu, L. & Xuan, G. A load forecasting method of electric vehicles charging station group based on gan-rf model, in *IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)*, 2021: IEEE, pp. 3076–3079., 2021: IEEE, pp. 3076–3079. (2021).
31. Ru, F., Yang, X., Zou, H., Zhang, L. & Xu, X. Electric Vehicle Charging Station Load Prediction Based on IWOA-LSSVM Combined Model with Improved Variational Mode Decomposition, in *Journal of Physics: Conference Series*, vol. 2320, no. 1: IOP Publishing, p. 012007. (2022).
32. Chung, Y. W., Khaki, B., Chu, C. & Gadh, R. Electric vehicle user behavior prediction using hybrid kernel density estimator, in *IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2018: IEEE, pp. 1–6., 2018: IEEE, pp. 1–6. (2018).
33. Lee, B., Lee, H. & Ahn, H. Improving load forecasting of electric vehicle charging stations through missing data imputation, *Energies*, vol. 13, no. 18, p. 4893, (2020).
34. Lu, K. et al. Load forecast method of electric vehicle charging station using SVR based on GA-PSO, in *IOP Conference Series: Earth and Environmental Science*, vol. 69, no. 1: IOP Publishing, p. 012196. (2017).
35. Wang, X. Electric vehicle charging load prediction with improved Kernel Extreme Learning Machine, in *4th International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2022)*, vol. 12257: SPIE, pp. 732–736. (2022).
36. Ullah, I., Liu, K., Yamamoto, T., Zahid, M. & Jamal, A. Electric vehicle energy consumption prediction using stacked generalization: an ensemble learning approach. *Int. J. Green Energy*. **18** (9), 896–909 (2021).
37. Jeon, Y. E., Kang, S. B. & Seo, J. I. Hybrid predictive modeling for charging demand prediction of electric vehicles, *Sustainability*, vol. 14, no. 9, p. 5426, (2022).
38. Li, Y., Huang, Y. & Zhang, M. Short-term load forecasting for electric vehicle charging station based on niche immunity lion algorithm and convolutional neural network, *Energies*, vol. 11, no. 5, p. 1253, (2018).
39. Ren, F., Tian, C., Zhang, G., Li, C. & Zhai, Y. A hybrid method for power demand prediction of electric vehicles based on SARIMA and deep learning with integration of periodic features, *Energy*, vol. 250, p. 123738, (2022).
40. Feng, J. et al. Load forecasting of electric vehicle charging station based on grey theory and neural network. *Energy Rep.* **7**, 487–492 (2021).
41. Eddine, M. D. & Shen, Y. A deep learning based approach for predicting the demand of electric vehicle charge. *J. Supercomputing*. **78** (12), 14072–14095 (2022).
42. Dabbaghjamanesh, M., Moeini, A. & Kavousi-Fard, A. Reinforcement learning-based load forecasting of electric vehicle charging station using Q-learning technique. *IEEE Trans. Industr. Inf.* **17** (6), 4229–4237 (2020).
43. Koohfar, S., Woldemariam, W. & Kumar, A. Prediction of electric vehicles charging demand: A transformer-based deep learning approach, *Sustainability*, vol. 15, no. 3, p. 2105, (2023).
44. Manzoor, T., Lall, B. & Panigrahi, B. Transformer Models for EV Charging Demand Forecasting: Comparing Attention Mechanisms, in *2024 Mediterranean Smart Cities Conference (MSCC)*, : IEEE, pp. 1–5. (2024).
45. Yılmaz, M., Çinar, E. & Yazıcı, A. A Transformer-Based model for state of charge Estimation of electrical vehicle batteries. *IEEE Access*, (2025).
46. Bouhamed, O., Dissem, M., Amayri, M. & Bouguila, N. Transformer-based deep probabilistic network for load forecasting. *Eng. Appl. Artif. Intell.* **152**, 110781 (2025).
47. Ke, F. & Wang, H. Divide-conquer transformer learning for predicting electric vehicle charging events using smart meter data, in 2024 IEEE Power & Energy Society General Meeting (PESGM), : IEEE, 1–5. (2024).
48. Thanh, P. Three-phase load prediction-based hybrid Convolution neural network combined bidirectional long short-term memory in solar power plant. *Int. Trans. Electr. Energy Syst.* https://doi.org/10.1155/2022/2870668 (2022).
49. Da, T. N., Cho, M. Y. & Thanh, P. N. Hourly load prediction based feature selection scheme and hybrid CNN-LSTM method for building's smart solar microgrid. *Expert Syst.* **41** (7), e13539. https://doi.org/10.1111/exsy.13539 (2024).
50. Nguyen, T. P. An incorporation of metaheuristic algorithm and two-stage deep learnings for fault classified framework for diesel generator maintenance. *Eng. Appl. Artif. Intell.* **151**, 110688. https://doi.org/10.1016/j.engappai.2025.110688 (2025).

Acknowledgements
A Special Thanks to the writing workshop by the International Education School of Chang'an University.

Author contributions
Conceptualization, A.H.; methodology, A.H., and A.A; software, V.E.; validation, A.H., A.A. and S.T.; formal analysis, A.A., and S.T; investigation, A.H., and V.E; resources, V.E.; data curation, A.A.; writing—original draft preparation, A.H.; writing—review and editing, A.A.; visualization, A.H., and S.T; supervision, A.H.; project administration, V.E.; funding acquisition, V.E. All authors have read and agreed to the published version of the manuscript.

Funding
This research received no external funding.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.H. or V.E.