

ORIGINAL RESEARCH

Open access

Enhancing Energy Demand Forecasting: A Comparative Analysis of LSTM, SVR, and Hybrid Models

Gabriel Costa^{1*}, Lucas Ribeiro¹, Rafael Mendes²

Abstract

Precise prediction of load profiles—most notably the accurate identification of peak demand—remains difficult owing to the highly variable and unpredictable patterns of energy use. This work evaluates three forecasting approaches: Long Short-Term Memory (LSTM), Support Vector Regression (SVR), and a hybrid ensemble combining SVR, Gated Recurrent Units (GRU), and Linear Regression (LR) to generate 24-hour-ahead load forecasts. Individual load categories—household (HH), heat pump (HP), and electric vehicle (EV)—were treated together using full one-year consumption records. The research focuses on improving the reliability of load-profile predictions by comparing LSTM, SVR, and the proposed ensemble across HH, HP, and EV demand patterns. An innovative correction procedure is presented that refreshes the forecast every 8 hours, thereby improving forecast trustworthiness. Results underscore the value of deep learning techniques for better anticipating electricity demand, with particular strength in recognizing peak events, which in turn supports smoother and more cost-effective power system management. Both graphical assessments and quantitative validation metrics were analyzed, covering model behavior during off-peak hours, peak hours, and across the complete daily cycle. Across the tested approaches, LSTM delivered marginally stronger overall metrics in the majority of cases and demonstrated especially good performance at detecting peaks. That said, the hybrid model slightly outperformed LSTM when predicting EV-specific power demand, achieving an on-peak mean absolute percentage error (MAPE) of 21.45%, compared with 29.24% (SVR) and 22.02% (LSTM). Even so, graphical inspection clearly illustrated LSTM's superior peak-capturing ability. This advantage was also evident in the on-peak mean absolute percentage error (MAPE) and symmetric mean absolute percentage error (SMAPE) values, where LSTM improved by roughly 3–5% relative to both SVR and the hybrid model. In the final application, LSTM was used to forecast next-day load profiles based on real measurements collected from four separate grid areas and consistently demonstrated excellent peak-detection capability, with MAPE below 10% across most of those datasets.

Keywords Load profile forecast, LSTM, SVR, Blended model, MAE, MAPE

*Correspondence:

Gabriel Costa

gabriel.costa@ufrj.br

¹ Department of Smart Grid Engineering, Faculty of Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

² Department of Energy Optimization Systems, Faculty of Engineering, University of Campinas, Campinas, Brazil

Introduction

The growing integration of renewable energy sources (RES) alongside modern end-user technologies, including electric vehicles (EVs) and heat pumps (HPs), places considerable strain on electrical grids. Minimizing the cost of infrastructure upgrades requires effective management of both demand and generation. Accurate forecasting plays a central role in achieving this goal. Energy demand forecasting is generally defined as the process of applying suitable statistical techniques to historical consumption records to predict future usage patterns [1]. Such energy demand forecasts (EDFs)

are essential for long-term scheduling, guiding decisions on infrastructure expansion, optimizing the operation of existing power plants, and designing appropriate energy pricing structures [2].

Traditional load forecasting has primarily focused on aggregate loads at the system or substation level. However, the widespread adoption of smart meters now provides access to large volumes of granular electricity usage data, enabling detailed forecasting at the individual consumer level [3, 4]. Compared with aggregated building-level loads that combine multiple households (HHs), single-household consumption profiles display significantly greater variability and unpredictability.

Forecasting models for energy demand can be grouped into three primary categories: engineering-based, statistical, and artificial intelligence-based [5, 6]. Engineering approaches rely on physical and thermodynamic laws and require detailed inputs on building characteristics and environmental conditions; these methods are often complex and resource-intensive to develop. In contrast, statistical and AI-based models generate predictions using influential variables such as temperature, humidity, weather information, and past consumption records [7].

In recent years, researchers have increasingly shifted toward recurrent neural networks (RNNs)—particularly long short-term memory (LSTM) architectures—for load forecasting [8, 9], moving away from conventional techniques like autoregressive integrated moving average (ARIMA) models and support vector machines (SVMs) [10, 11]. The superior performance of RNN-based models, especially LSTMs, over traditional approaches stems from their ability to capture temporal dependencies, which has led to their widespread use across diverse forecasting tasks.

Key strengths of LSTM include its capacity to derive features directly from timestamps, compatibility with clustering techniques, support for multi-input/multi-output sliding-window strategies, robustness when processing static or streaming data, mitigation of the vanishing gradient issue, and overall superiority relative to other RNN variants [12]. LSTMs are also well-suited for time-series sequences of varying durations [11]. Numerous studies in the literature have compared LSTMs (and RNNs more broadly) with classical models, with LSTM frequently emerging as the top-performing RNN variant due to their balance of accuracy and computational complexity [13, 14]. The present work reviews and builds upon existing comparisons between RNN-based and alternative forecasting methods to determine the most effective strategy for power load prediction.

Table 1 provides an overview of selected recent studies on load profile forecasting across various countries. A clear trend emerges: the majority of contemporary research employs machine learning and deep learning techniques. Model architecture is the most critical factor, since dataset characteristics (such as length and variability) strongly influence the choice of structure, making the identification of an optimal design highly important. In recent years, several efforts have focused on developing hybrid models that can adapt to diverse consumption behaviors. Careful model selection before deployment is therefore essential.

Table 1. Comparative research studies.

Ref.	Model(s)	Dataset	Evaluation metrics	Evaluation result	Other outcomes
Fayaz and Kim [15]	Deep extreme learning machine (DELM), Adaptive neuro-fuzzy inference system (ANFIS), and Artificial neural networks (ANNs)	4 residential buildings	RMSE, MAE, and MAPE	DELM (RMSE = 2.24, MAPE = 5.7, MAE = 2); ANFIS (RMSE = 2.46, MAPE = 6.38, MAE = 2.26); ANN (RMSE = 2.6, MAPE = 6.7, MAE = 2.39).	Variations in the number of hidden layers, hidden neurons, and activation function combinations enabled DELM to achieve superior optimization performance compared to the other approaches.

Wang et al. [16]	LSTM, ANN, and Support Vector Regressor (SVR)	Five building groups across two campuses	MAPE and RMSE	For dormitory buildings: LSTM (MAPE = 2.94%, RMSE = 0.27); ANN (MAPE = 9.09%, RMSE = 0.79). For SVR, MAPE = 8.48%, RMSE = 1.1.	LSTM delivered more accurate and superior predictions not only for dormitory buildings but also for other types, such as research and office buildings, outperforming the other tested models.
Hossain and Mahmood [17]	Extreme learning machine (ELM), Generalized regression neural network (GRNN), and LSTM	Electrical Reliability Council of Texas (ERCOT)	MAE, MAPE, and RMSE	LSTM (MAE = 55.42, RMSE = 63.81, MAPE = 4.79%); GRNN (MAE = 61.62, RMSE = 68.45, MAPE = 5.33%); ELM (MAE = 73.82, RMSE = 80.56, MAPE = 6.86%).	ELM and GRNN served as benchmarks to confirm and validate the superior performance of LSTM as an effective and leading model.
Jin et al. [18]	Decision tree, Random forest, SVR, and LSTM	Five households in the UK	MAE, MAPE%, and RMSE	At 30 min frequency: LSTM (MAE = 0.0842, RMSE = 0.1926, MAPE = 18.49%); Decision tree (MAE = 0.1923, RMSE = 0.2952, MAPE = 62.42%); Random forest (MAE = 0.1364, RMSE = 0.2216, MAPE = 40.87%); SVR (MAE = 0.834, RMSE = 0.2062, MAPE = 16.4%).	Although SVR showed stronger results at the 30 min frequency, LSTM outperformed the others at different frequencies. Additionally, the authors' proposed hybrid model—combining singular spectrum analysis with parallel LSTM—outperformed all the listed models.
Dubey et al. [19]	SARIMA, ARIMA, and LSTM	5567 London households	MAE	LSTM recorded the lowest MAE value among all the models tested.	The MAE for every model decreased progressively across training epochs, reaching its minimum value in the final epoch. Prediction errors were lowest during spring and highest in winter, likely because winter energy consumption patterns are less predictable and more difficult to model accurately.

To date, most forecasting investigations have focused on either real or simulated datasets, with primary attention on the overall load profile. However, relatively few studies have systematically modeled distinct load types (HH, EV, HP) both separately and in aggregated form as combined demand. Furthermore, as highlighted in **Table 1** and related publications, there remains a gap in research on developing and evaluating blended/ensemble models for load profile prediction, particularly in comparing their performance across individual load categories versus total demand. In response, this study introduces a novel blended model tailored to the heterogeneous nature of certain load profiles, which can significantly shape the resulting aggregate pattern.

Accordingly, the main objective of this research is to investigate the effectiveness of the LSTM model for forecasting power load profiles, considering both individual and combined demands. The paper also conducts a comparative evaluation of LSTM against SVR and a proposed blended model (SVR + linear regressor (LR) + GRU), aiming to identify the most accurate and efficient method for forecasting both synthetic and real-world measured data. The blended model is presented here for the first time, specifically designed in light of the datasets' characteristics, making it a meaningful benchmark against LSTM.

As an additional contribution, this work incorporates a forecast-correction mechanism that is applied every 8 hours to enhance model robustness, offering potential benefits for future optimization and peak-control applications. Although receiver operating characteristic (ROC)-style curves are uncommon in load forecasting literature, this study adopts such a metric by framing the evaluation as a classification task across on-peak, off-peak, and overall patterns. This approach enables assessment of model performance at various thresholds and provides a clearer measure of peak-capturing effectiveness.

Materials and Methods

Model processing

Drawing from the outcomes of model evaluation and selection, this section provides a detailed explanation of the long short-term memory (LSTM) model. The innovative blended model is also introduced due to its unique architecture. Additionally, a description of the support vector regression (SVR) model is included for completeness and comparison.

During model training, trainable parameters (weights) are assigned to the connections entering and exiting the LSTM gates, including several recurrent links. These parameters are iteratively refined over multiple training iterations, referred to as epochs in recurrent neural networks (RNNs), to improve prediction accuracy progressively. The presence of a dedicated forget gate, combined with the additive nature of the cell state gradients, allows LSTM to adjust its weights in a way that effectively overcomes the vanishing gradient issue [20, 21]. By employing specialized gates, LSTMs regulate the flow of information into and out of the memory cell while controlling the extent of information retention or forgetting (Figure 1), thereby enabling accurate modeling of sequential data [22].

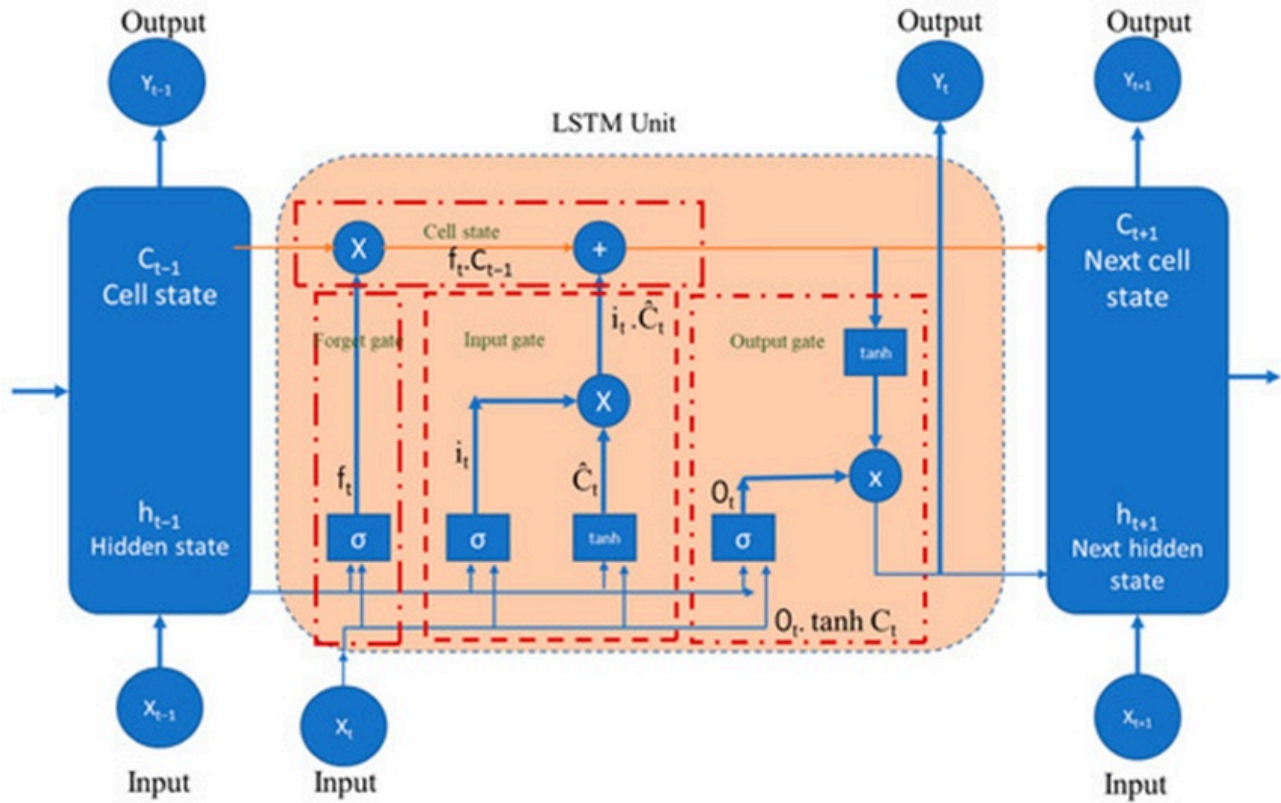


Figure 1. Structure of LSTM model [23].

As illustrated in **Figure 1**, the initial gate (forget gate) determines which prior information should be discarded (**Eq. 1**). The following gate (input gate) decides which elements from the current input should be incorporated into the cell state (**Eq. 2**). Ultimately, the output gate (**Eq. 5**) governs the information passed to the next time step. Within the input gate mechanism, the input gate controls the update to the cell state. At the same time, a tanh activation produces a candidate cell state value (**Eq. 3**) that is combined with the previous cell state to form the updated current cell state (**Eq. 4**). The final cell output at time t , denoted h_t , is computed through the output gate (**Eq. 6**) [24-26].

$$f_t = \sigma(W_f * x_t + U_f * h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma(W_i * x_t + U_i * h_{t-1} + b_i) \quad (2)$$

$$\hat{C}_t = \sigma(W_c * x_t + U_c * h_{t-1} + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (4)$$

$$O_t = \sigma(W_o * x_t + U_o * h_{t-1} + b_o) \quad (5)$$

$$h_t = O_t * \tanh(C_t) \quad (6)$$

$$\sigma(x) = \left(\frac{1}{1 + e^{-x}} \right) \quad (7)$$

$$\tanh(x) = \left(\frac{e^x - e^{-x}}{e^x + e^{-x}} \right) \quad (8)$$

f_t , i_t , $\hat{C}_t$, C_t , O_t , and h_t : forget gate at time t , input gate, candidate cell state, current cell state, output gate at time t , and hidden state (output) at time t , respectively.

σ : sigmoid activation function.

h_{t-1} : hidden state from the previous time step; x_t : current input vector.

W_f , U_f , W_i , U_i , W_C , U_C , W_o , and U_o : weight matrices.

b_f , b_i , b_C , and b_o : bias vectors.

The set of input features can be specified before training. However, if the relative importance (weights) of these inputs is not predefined, the LSTM architecture automatically learns and refines the contribution of each input feature during training through backpropagation through time (BPTT). For this forecasting task, the input vector (x_t) consisted of temperature measurements, sine and cosine transformations of the day of the week and the hour of the day, and the load profile from the preceding day.

The input layer consists of 100 neurons, followed by a hidden layer with the same number of neurons (100) employing a rectified linear unit (ReLU) activation function. The output layer is a dense layer containing a single neuron.

The optimal number of neurons for the measured data was determined after experimenting with various quantities (10, 20, 30, 50, 100, and 150) and evaluating the average error to select the most effective configuration. Upon analysis of the average error—particularly for the final day of the testing dataset—the configuration with 100 neurons achieved the best performance (error below 10%), while effectively balancing the risks of over- and under-fitting [19].

The model is trained for classification-based prediction using the adaptive moment estimation (Adam) optimizer—a widely recognized and robust method—with the mini-batch size adjusted according to the specific dataset. This optimizer consistently outperforms other stochastic gradient methods and demonstrates high efficiency with sparse or noisy data [19].

Model performance is assessed using multiple error metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) (Eqs. 9–12), to identify the superior model among those tested. According to Lewis's widely adopted benchmark for energy consumption forecasting, MAPE values below 10% indicate high accuracy, while 10–20% and 20–50% are considered good and reasonable accuracy levels, respectively [27].

It is important to note that error metric results should always be interpreted alongside visual comparisons of actual and predicted values to draw reliable conclusions. This step is especially critical when the dataset contains potentially high

error levels, including low or near-zero values. Analyzing patterns across multiple data points in actual-versus-forecasted plots provides essential insights for comprehensive evaluation [28, 29].

Furthermore, to address a key limitation of MAPE (division by zero when actual values approach zero), the Symmetric Mean Absolute Percentage Error (SMAPE) is also employed (Eq. 13). This metric offers a more stable and reliable accuracy measure in such cases.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i^* - y_i| \quad (9)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i^* - y_i)^2 \quad (10)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i^* - y_i}{y_i^*} \right| * 100\% \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^* - y_i)^2} \quad (12)$$

$$SMAPE = \frac{1}{N} \sum_{i=1}^N \left(\frac{|y_i^* - y_i|}{(|y_i^*| + |y_i|)/2} \right) * 100\% \quad (13)$$

N: number of observations.

y_i : predicted value; y_i^* : actual value.

To prevent excessive training epochs, an early-stopping mechanism is used to determine the optimal number of iterations. The dataset is partitioned into training, validation, and test sets. The model is implemented using Keras with a TensorFlow backend. Training is terminated if the validation set MSE fails to improve for a predefined number of epochs. In this work, early stopping monitors the validation loss with a patience of 3 epochs, halting training if no improvement occurs over 3 consecutive epochs. Additionally, the “Restore Best Weights” feature is enabled, allowing the model to revert to the weights corresponding to the epoch with the lowest validation loss.

The number of neurons was another key hyperparameter investigated to determine the optimal balance that avoids overfitting while preserving the capacity to model complex patterns. Multiple runs for each neuron count revealed that the lowest average error was consistently achieved with 100 neurons (Figure 2). Increasing the count to 150 resulted in a higher average error, likely due to vanishing or exploding gradients (causing unstable training from large weight updates) and overfitting (where the model memorizes the training data rather than generalizing effectively) [30].

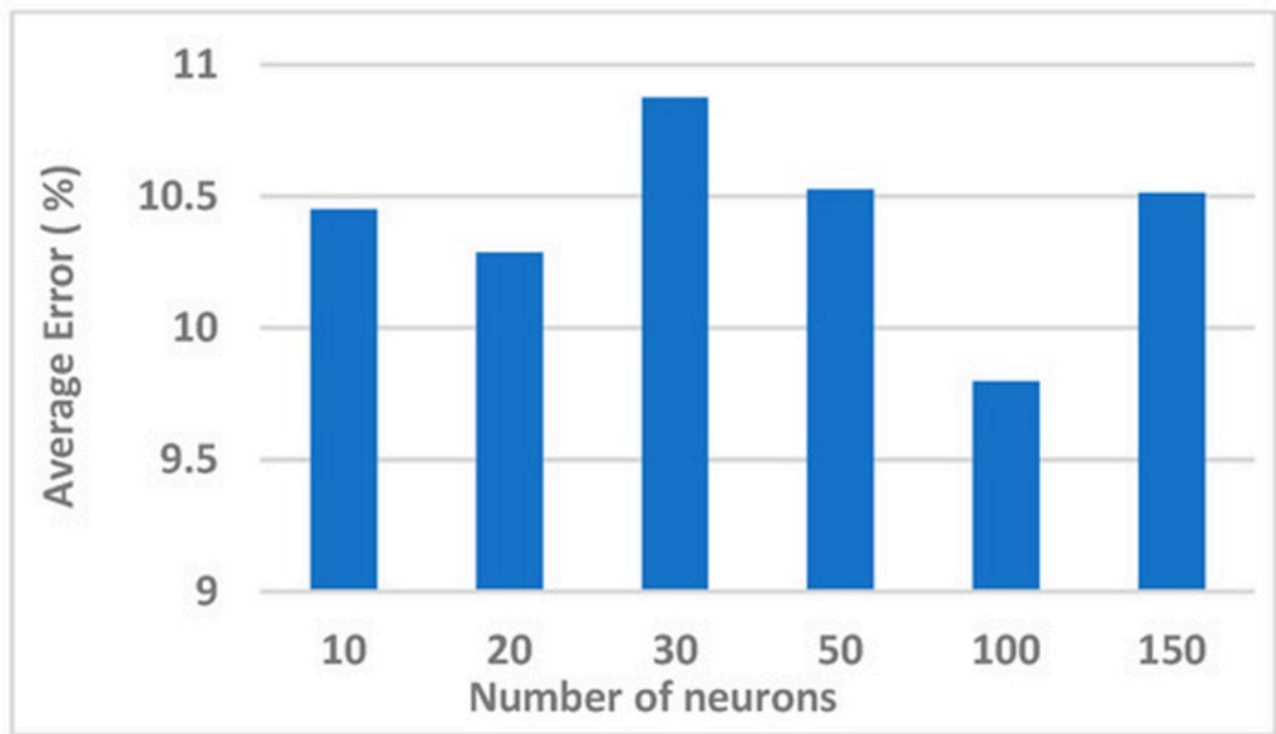
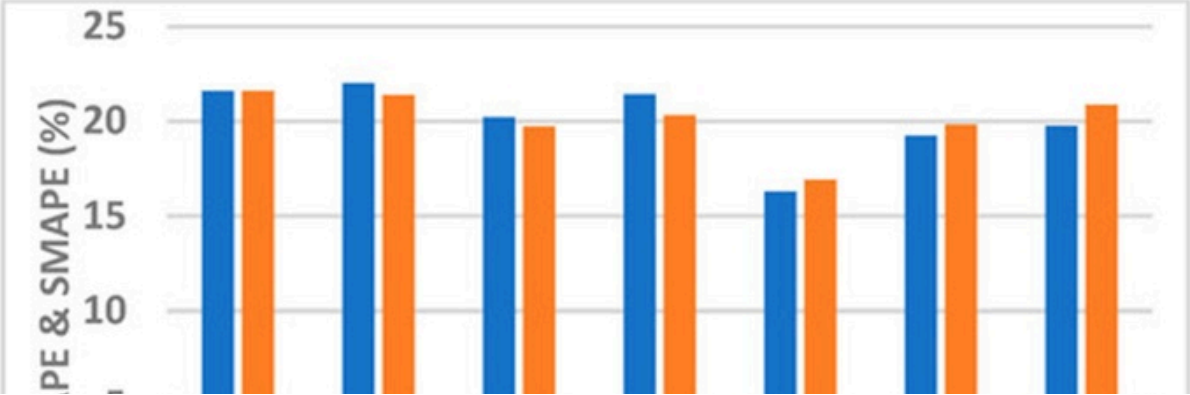
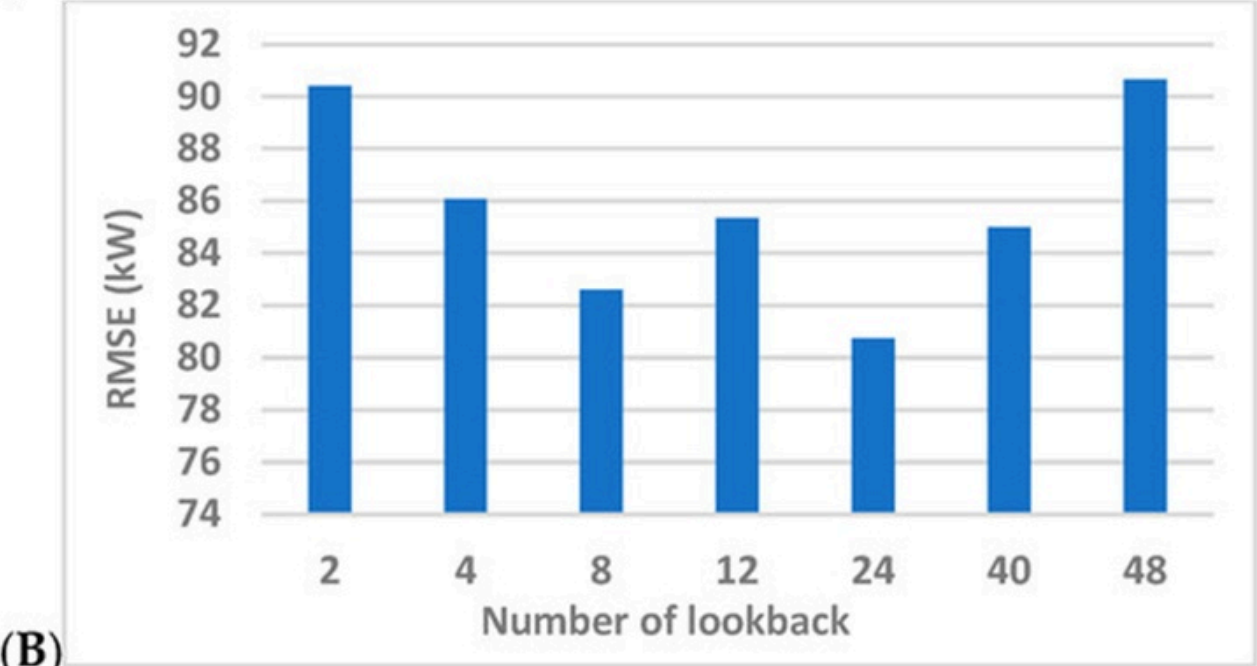
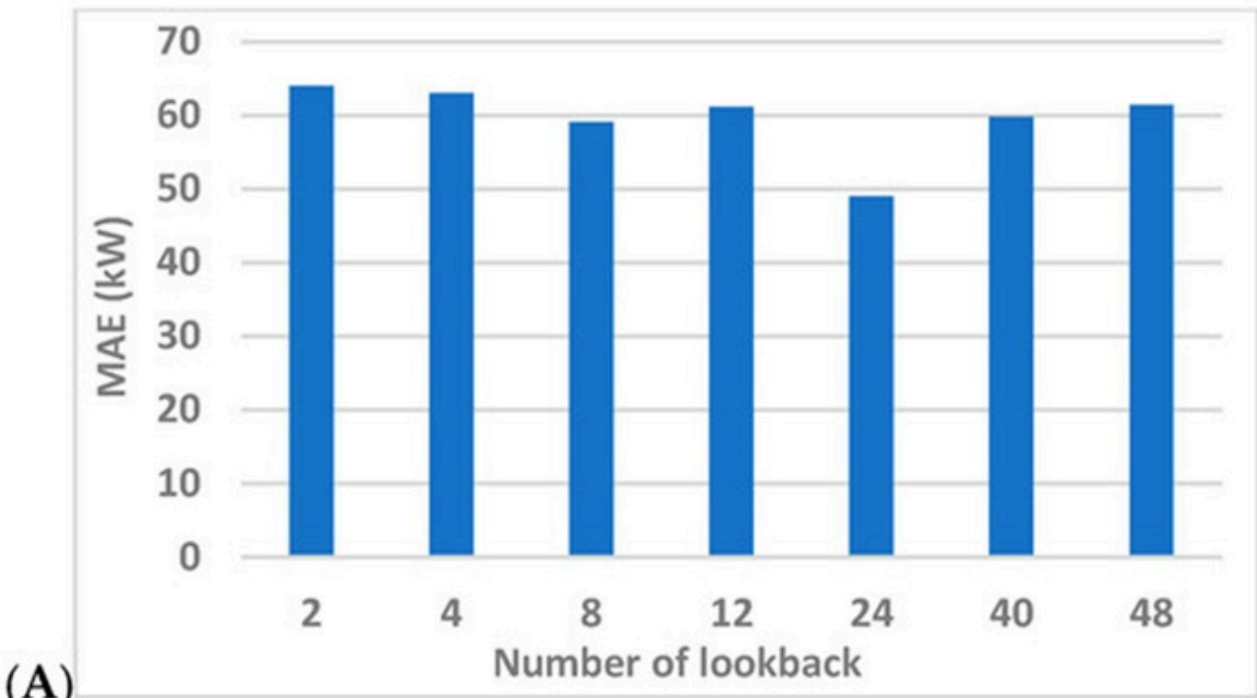
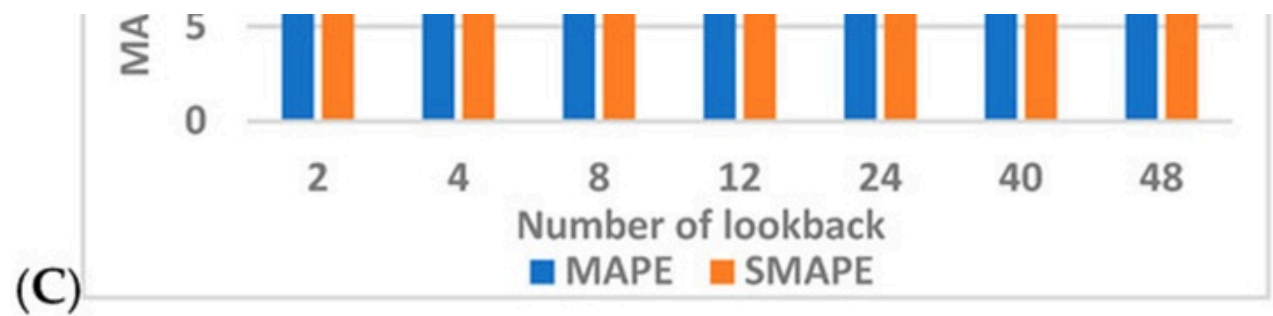


Figure 2. Influence of the number of used neurons on the average error.

In LSTM and sequential model architectures, the lookback period is a critical hyperparameter. Numerous studies on day-ahead forecasting at hourly resolution have demonstrated that a 24-hour lookback yields the lowest error [31, 32]. This occurs because the model can most effectively learn the recurring daily load pattern when provided with the preceding 24 hours of data.

To validate the superiority of 24 lookbacks, several alternative values were tested and their corresponding errors examined (Figure 3). The results clearly show that all error types reached their minimum with 24 lookbacks, after which errors increased as the lookback period grew longer. This pattern supports the hypothesis that an appropriate lookback length, combined with the input features, enables the model to detect cyclic patterns better and accurately identify trend changes (peaks and troughs) [31, 32].





perform better with unbroken sequential inputs, and masking can disrupt temporal relationships. Additionally, negative load values were set to zero. Dummy variables were also introduced to improve model performance; specifically, an on-peak time indicator was created as a binary feature to mark hours of the year with elevated load probability.

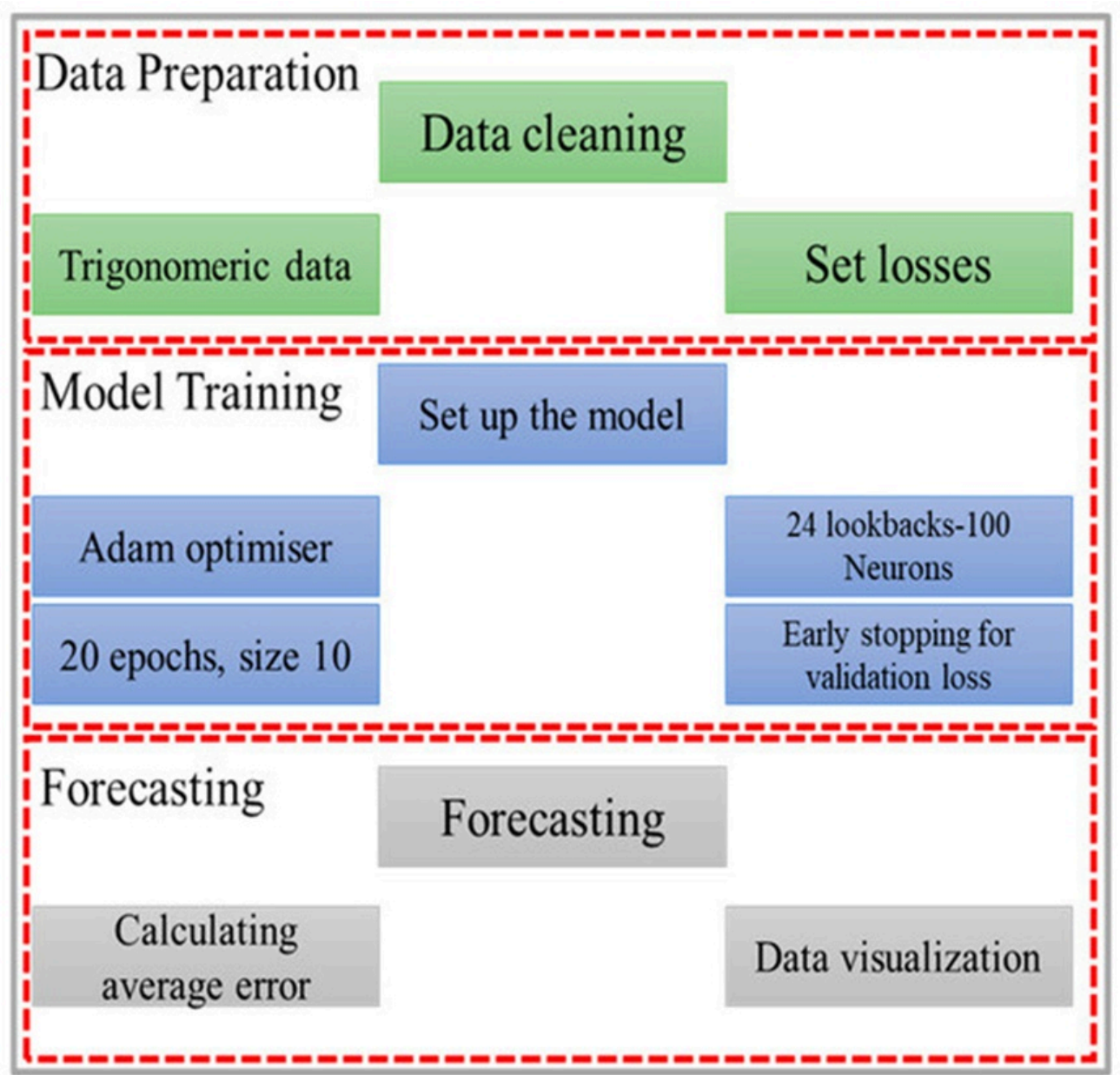


Figure 4. Flowchart for implementing forecasting for one-day-ahead loads.

During the feature engineering and selection stage, emphasis was placed on temporal features by using dummy variables to identify time-related patterns. In the initial dataset, peak demand periods were defined, and corresponding binary dummy variables were generated by assigning “1” to peak-hour instances and “0” otherwise. This approach enabled the model to account for diurnal variations in load behavior. For the EV dataset, dummy variables indicated whether charging occurred at various times; for example, charging during the morning hours was encoded as “0” to highlight periods of inactivity, thereby helping the model differentiate behavioral patterns that affect performance.

Additional processing steps included incorporating trigonometric features to model cyclic and seasonal variations, as well as applying set losses to enable early stopping and improve overall training efficiency. During model training, core architectural decisions were made, including selecting the Adam optimizer, determining neuron counts, lookback periods, and epoch limits. Each hyperparameter underwent dedicated experimentation to identify optimal settings—for instance, although an initial lookback of 24 hours was considered, tests extended to alternatives such as the preceding 8 hours or a full week; ultimately, 24 hours produced the lowest error across all three models evaluated.

The concluding phase of the forecasting workflow involved model assessment by computing average error metrics and visualizing the results.

As a comparative benchmark, SVR, an extension of the SVM framework, was employed. It leverages hinge loss to reduce forecasting discrepancies and seeks to define either strict or tolerant margins that enclose as many data points as possible, thereby improving robustness (Figure 5). Unlike artificial neural network (ANN) approaches, SVR is less prone to local minima traps and overfitting, often delivering superior results in energy consumption prediction tasks. Accuracy in SVR forecasting was further refined by addressing false neighbors within the reconstructed phase space, achieved by integrating time-series reconstruction characteristics and optimizing the local predictor via false-neighbor elimination [28]. The SVR implementation relied on sequential minimal optimization (SMO) as its core solver. However, performance gains could be achieved through advanced tuning methods such as Bayesian optimization, grid search, or evolutionary algorithms [34]. A Radial Basis Function (RBF) kernel was adopted to enable SVR to capture intricate nonlinear dependencies.

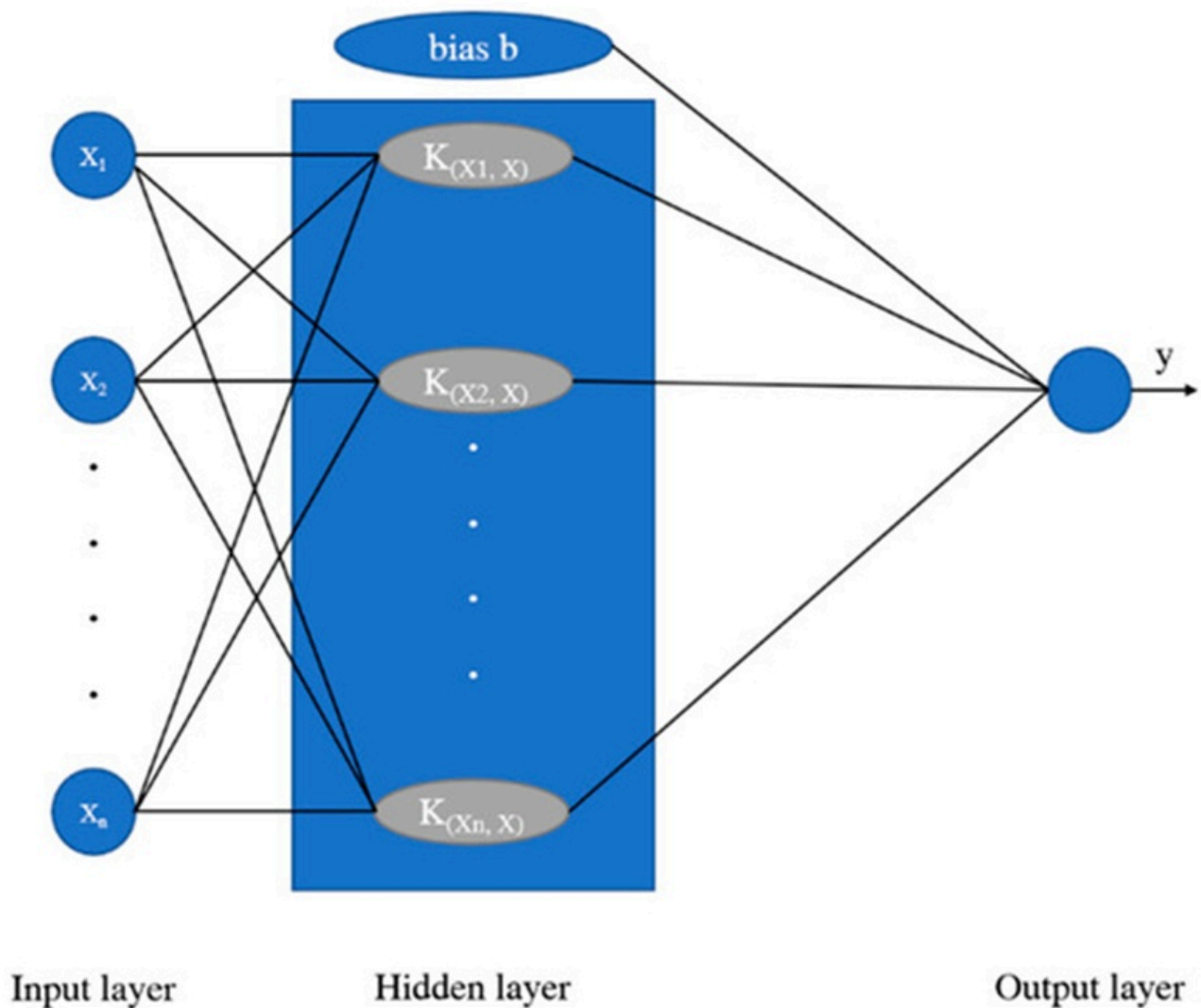


Figure 5. Structure of SVR model [35].

Researchers frequently apply ensemble or blending techniques to enhance model complexity and resilience [36, 37]. Multiple base learners can be combined to form a blended predictor. SVR, with its proficiency in modeling nonlinear associations, and LR, valued for its simplicity and interpretability, were consistently selected as high-potential components in such ensembles. Temporal dependencies represent another vital aspect in blended models for power load forecasting. One incorporated model was GRU, a streamlined RNN variant that is computationally lighter than LSTM. The inclusion of GRU and LR in the blended architecture stemmed from their complementary strengths: GRU adeptly captures sequential dependencies with lower resource demands than LSTM, while LR offers clear interpretability and effectively models linear trends in certain segments of the EV and HP datasets. SVR complemented these by addressing nonlinear patterns. This trio of models achieves a balanced trade-off among complexity, efficiency, and robustness, yielding improved forecasting across diverse load profile types. The hyperparameters for this blended model are detailed here for further specification: SVR employed an RBF kernel; GRU featured 50 neurons, utilized ReLU activation, was trained with the Adam optimizer over 30 epochs, and operated with a batch size of 32.

A standard GRU cell consists of two primary gates: the reset gate (R) and the update gate (Z) (Figure 6 and Eqs.14–17). Similar to an LSTM cell, the hidden state at time step t is computed by combining the previous hidden state from time $t -$

1 with the current input value at time t . The reset gate in GRU performs a function analogous to the LSTM forget gate [38].

Although many studies employing recurrent neural network (RNN) architectures prefer LSTM due to its ability to handle long-range dependencies, LSTM is generally more complex, contains a greater number of parameters, and requires longer training and inference times than GRU [39]. It has been demonstrated that LSTM effectively mitigates the vanishing gradient problem common in standard RNNs and successfully preserves information over extended sequences [24]. Thanks to its long-term memory mechanism, the LSTM model is widely used for forecasting future values from historical data and current inputs, and it is particularly effective at detecting anomalous patterns [40, 41]. **Table 2** summarizes the key characteristics of the models used in this study.

$$r_t = \sigma (W_r x_t + U_r h_{t-1} + b_r) \quad (14)$$

$$z_t = \sigma (W_z x_t + U_z h_{t-1} + b_z) \quad (15)$$

$$\bar{h}_t = \tanh (W_h x_t + U_h (r_t \odot h_{t-1}) + b_r) \quad (16)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \bar{h}_t \quad (17)$$

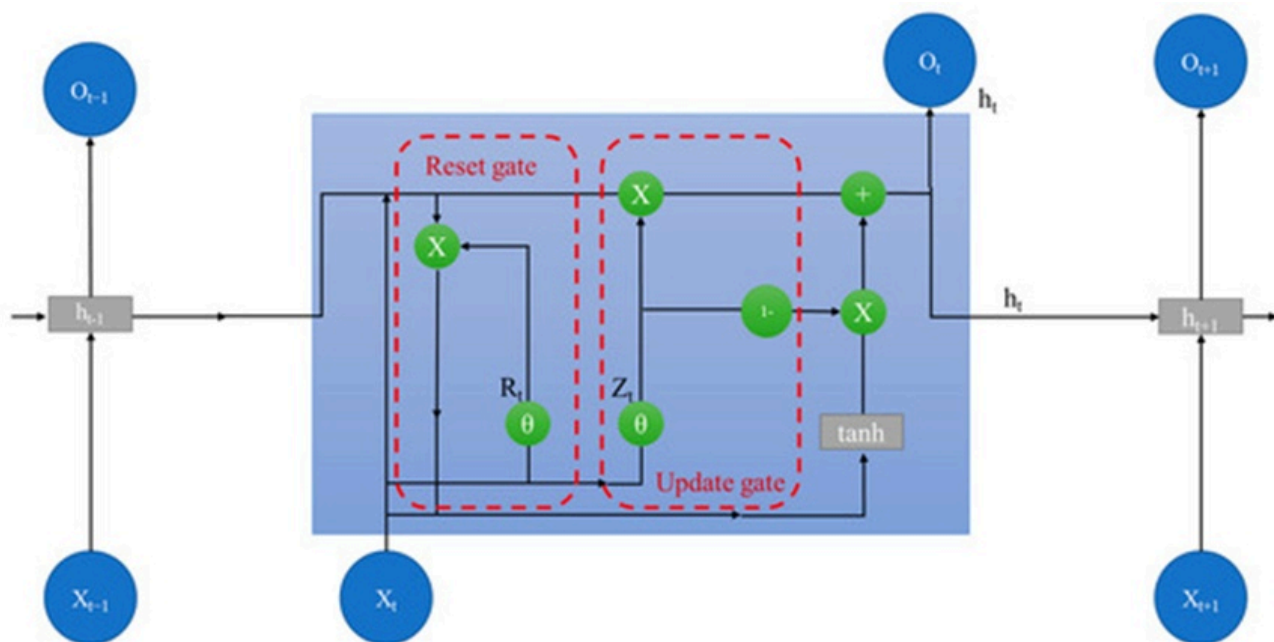


Figure 6. Structure of the GRU model [42].

Table 2. Important features of the models used.

Model architecture	Activation function	Optimization algorithm	Hyperparameters
LSTM	ReLU	Adam	1 input layer (100 neurons), 1 hidden layer (100 neurons), 1 output layer; Lookback: 24; batch size: 10; epochs: 20; early stopping (patience = 3)
SVR	RBF kernel (serves an analogous role to activation functions in neural network approaches)	Sequential Minimal Optimization	Kernel: RBF
Blended model	ReLU (applied to the GRU component)	Adam (applied to the GRU component)	SVR with RBF kernel + GRU (50 neurons, 1 hidden layer) + Linear Regression; GRU epochs: 30; GRU batch size: 32; early stopping (patience = 3)

X_t : input at time t .

h_{t-1} : previous hidden state.

W_r , W_z , W_h , U_r , U_z , and U_h : weight matrices for the current input and hidden state.

b_r , b_z , and b_h : bias vectors.

σ : sigmoid activation function.

$\odot$: element-wise (Hadamard) multiplication.

r_t : reset gate.

z_t : update gate.

h_t : hidden state at time t .

$\underline{h}_t$: candidate hidden state.

For the blended (ensemble) model, component selection was tailored to the specific characteristics and patterns present in each dataset. More complex load categories often exhibit multiple coexisting patterns, requiring a more sophisticated approach. For example, both EV and HP load profiles frequently include extended periods of zero or near-constant values. Previous research has noted that deep learning models—especially LSTM—can sometimes suffer from saturated forget gates, leading to excessive simplification of patterns and a loss of the ability to capture more intricate or dynamic behaviors.

To address this limitation, a blended modeling strategy was developed and evaluated to determine whether it could outperform standalone LSTM. In the proposed ensemble:

- Linear Regression (LR) captures basic linear trends,

- Support Vector Regression (SVR) handles nonlinear relationships through its kernel-based formulation,
- GRU models temporal dependencies and long-range dependencies effectively.

One of the most critical elements in any ensemble is the weighting scheme. In the method proposed here, a straightforward equal-weight averaging approach was adopted (Eq. 18), assigning equal contributions from each base model to the final blended prediction.

$$y_{Blended} = \frac{1}{3} (y_{SVR} + y_{LR} + y_{GRU}) \quad (18)$$

$y_{Blended}$: Blended prediction.

y_{SVR} , y_{LR} , and y_{GRU} : Predictions from the SVR, LR, and GRU models, respectively.

Data processing

The choice of time step significantly influences the predictive performance of forecasting models. In this study, a time step corresponding to one sensor reading per dataset spanning one full year was adopted. This granularity enables the model to detect short-term variations in load profiles effectively. Furthermore, such an hourly interval is widely accepted and commonly used in energy market operations and load forecasting.

To evaluate the performance of the proposed methodology, an empirical analysis was carried out using one synthetic dataset and four real-world datasets. All datasets consist of electricity consumption records collected over one year. The synthetic dataset simulates energy usage from 370 households, including 133 heat pumps (HPs) and 59 electric vehicles (EVs), reflecting a plausible future scenario for a low-voltage distribution network in central Europe.

- Because the monitored resource features exhibit substantial differences in scale, data standardization is essential to reduce disparities in feature magnitudes. The sklearn preprocessing library applies min–max standardization [16], which rescales every feature to the interval [0, 1]. This technique, also known as min–max scaling, performs a linear transformation of the raw data to map all values to the range [0, 1] (Eq. 19) [43].

$$y_i = \frac{x_i - \min_{0 \leq j \leq n} x_j}{\max_{0 \leq j \leq n} x_j - \min_{0 \leq j \leq n} x_j} \quad (19)$$

- To enhance model accuracy, categorical dummy variables were created, and seasonal effects were explicitly encoded. Dummy variables—particularly those representing categorical factors—are extensively employed to improve the precision of modeling and forecasting across diverse domains [44, 45]. One important feature is identifying peak consumption periods and the absolute maximum consumption hour to capture demand spikes better and reduce forecasting errors. These variables were derived by first computing the average consumption for each hour of the day, then defining two binary indicators: one marking the top 10 hours with the highest average consumption, and another flagging the single hour with the absolute highest average consumption. Notably, peak hours can be determined separately for each month or season. This per-period approach offers a clear advantage, as monthly or seasonal average consumption patterns vary considerably. Computing a single yearly average for peak identification can distort results by overlooking month-specific demand peaks, thereby degrading overall model performance.

• To properly handle the cyclical and discontinuous characteristics of temporal attributes—such as hour of the day, day of the week, month of the year, and season of the year—trigonometric transformations using sine and cosine functions were applied (Eqs. 20 and 21). These functions ensure a continuous, smooth representation of periodic values, preserving their periodicity. By integrating sine and cosine encodings, sequential models such as LSTM and other time-series architectures can more effectively learn and represent the repeating nature of temporal patterns, capturing meaningful dependencies across the cycle. Additionally, these trigonometric features increase the expressive capacity and complexity of the models [46].

$$x_{\sin} = \sin\left(\frac{2 * \pi * x}{\max(x)}\right) \quad (20)$$

$$x_{\cos} = \cos\left(\frac{2 * \pi * x}{\max(x)}\right) \quad (21)$$

Results and Discussion

This study evaluates the performance of the selected models in forecasting next-day load profiles by learning patterns from historical data. Model selection was guided by a combination of visual inspections and quantitative validation metrics to determine the most effective configuration. To boost predictive accuracy, various dummy variables—particularly those representing peak periods—were incorporated. Model performance was further enhanced by adopting a 24-hour lookback window and including the previous day's load values as input features. Additional explanatory data were deliberately incorporated to strengthen the overall model's capability. Given the primary objective of this research—minimizing the risk of bottlenecks through improved forecasting—the accurate detection of peak loads was assigned the highest priority. The robustness of the proposed approach was assessed using synthetic datasets and real measured data collected from two distribution grids and one shopping center.

Synthetic data

This section describes the load profile generator, the heat pump (HP) load generator, and the electric vehicle (EV) load generator [47] used to generate synthetic data for model evaluation.

EV load forecast

Among the considered categories, EV charging loads proved the most challenging to predict accurately, as validated by the results. This difficulty stems from highly variable user behavior regarding driving schedules and charging decisions, which introduces considerable randomness. Moreover, since EV load remains zero for the vast majority of hours, even minor absolute errors during charging periods translate into disproportionately large percentage errors.

As illustrated (Figure 7), day-ahead EV load forecasting on synthetic data achieved an acceptable level of accuracy, particularly when using the blended model. One contributing factor may be the unnecessary complexity of LSTM relative to the relatively structured nature of the data, contrasted with the simplicity and generalization strength of SVR. Examination of the patterns revealed that charging events tend to begin and end at roughly consistent times across days, although peak magnitudes vary. Consequently, the blended ensemble—combining linear regression, support vector regression, and GRU—proved more effective at capturing these variable peaks. In contrast, overly complex models applied to largely simple patterns risk overfitting, often requiring excessive layers to achieve precise fits [48].

Validation results showed a modest performance advantage for the blended model, though the edge was relatively small, and its ability to detect peaks in aggregate load remained somewhat limited. MAPE and SMAPE values across different

observation ranges for EV load forecasting are presented in [Table 3](#). Notably, on-peak periods exhibited lower SMAPE (especially for the blended model) compared with total load or off-peak periods. This improvement likely arises because the constituent models in the ensemble display differing sensitivities to noise, enabling the blended approach to more robustly extract underlying patterns.

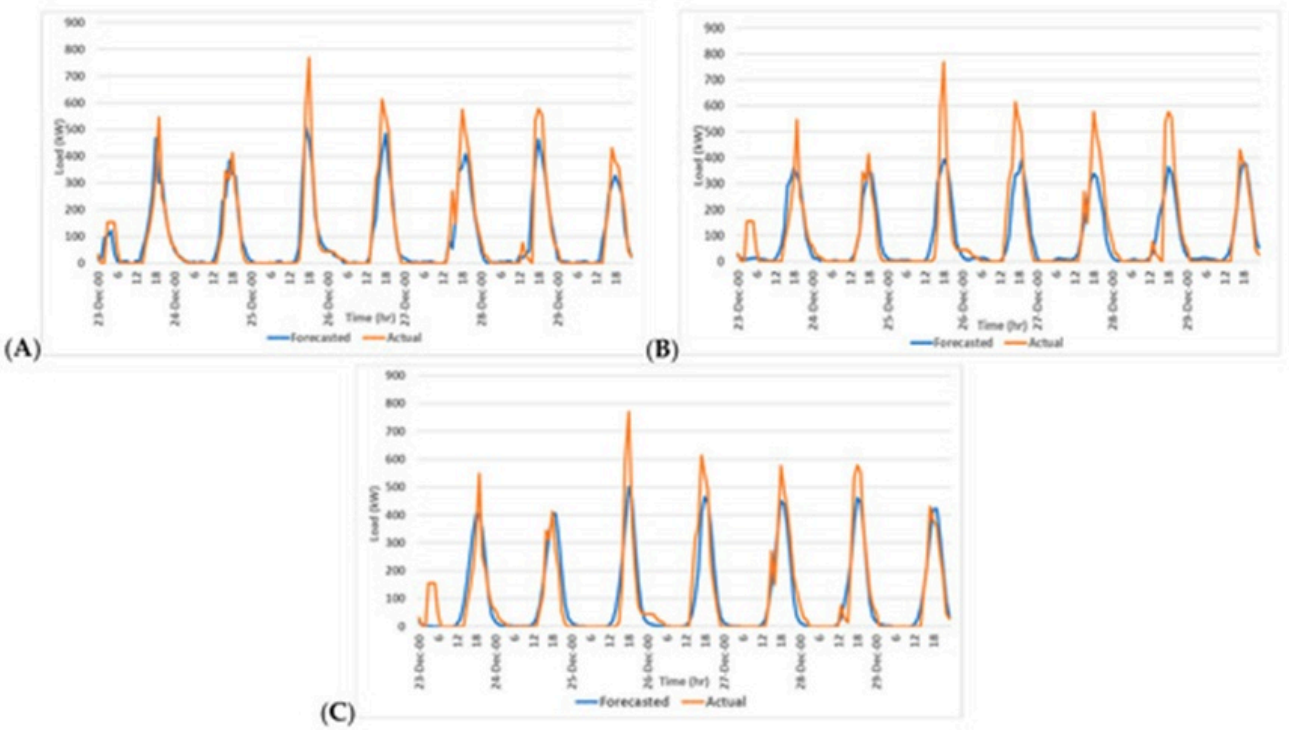


Figure 7. Actual vs. predicted EV load by the (A) blended, (B) SVR, and (C) LSTM models.

Table 3. MAPE and SMAPE for different observation ranges in EV power forecasting.

EV	Total		On-Peak		Off-Peak	
Model	MAPE	SMAPE	MAPE	SMAPE	MAPE	SMAPE
LSTM	inf	65.96%	22.02%	12.89%	inf	75.24%
SVR	inf	65.10%	29.24%	18.25%	inf	73.29%
Blended	inf	42.58%	21.45%	12.39%	inf	63.19%

Overall, error reductions were only marginal. Additionally, the frequent occurrence of zero values in actual EV load can occasionally lead to erratic behavior in error metrics. The key predictive challenge is to capture two dominant behaviors simultaneously: charging peaks and extended periods of zero consumption. Both the LSTM and blended models successfully reproduced these characteristics on the final test day. In comparison, SVR struggled to represent both patterns with comparable fidelity. As highlighted in [Figure 7A](#), the blended model demonstrated superior handling of fine-grained noise—for instance, accurately reproducing a minor peak observed on 23 December—a detail missed by both standalone SVR and LSTM.

HH load prediction

Following EVs, household (HH) consumption ranked second most difficult to predict, based on validation error metrics. This can be attributed to highly diverse, often unpredictable occupancy and appliance-use patterns across household

types.

Visual comparison of predictions revealed that the primary source of error in the LSTM model for total household consumption arises during periods of minimum load, resulting in elevated average errors—both across the full unscaled dataset and specifically for the final test day. In contrast, LSTM exhibited strong alignment with actual data during peak consumption periods. Interestingly, the opposite tendency was observed for EV and HH loads within the LSTM framework: minimum values were well captured, whereas significant deviations occurred during the peaks.

To reduce these inconsistencies, outliers were identified and addressed, and a dedicated dummy variable was introduced to improve forecasting accuracy. Care was taken to manage outliers in a way that avoids introducing information leakage into unseen test data. In particular, test-time inputs should only incorporate features that would realistically be available, such as season, month, day of the week, hour of the day, and statistically derived indicators of typical peak consumption (e.g., historical average maximum load). Relying on the actual daily maximum value for each test day as an input feature would artificially inflate apparent accuracy, since this information represents future knowledge unavailable during real-world deployment.

Although the overall MAPE of the LSTM model for households was approximately 22% (Table 4), this level remains acceptable for several reasons. However, it should first be emphasized that a 22% error is only tolerable if the model delivers low total consumption error or successfully captures peak loads; otherwise, the model should either be discarded or undergo architectural changes.

The primary reason is that one of the main goals of the forecasting task is to determine whether bottlenecks are likely to occur, and this analysis should primarily be performed at the grid level.

A second important reason is that household consumption patterns are inherently difficult to predict, so it is not unexpected that no model achieves an especially good fit for them.

To evaluate model performance across different power load ranges, MAPE and SMAPE were calculated for three categories: on-peak, off-peak, and total (Table 4). During off-peak periods, household power demand proved more predictable using SVR compared with the other models. This outcome is likely attributable to SVR's stronger performance when data patterns are relatively stable or exhibit linear behavior, which is frequently the case during off-peak hours.

In contrast, during on-peak periods, LSTM showed marginally better performance than both SVR and the blended model. This can be explained by the greater complexity and irregularity of patterns in the on-peak range, which favor a more sophisticated model such as LSTM.

Table 4. MAPE and SMAPE for different observation ranges in HH power forecasting.

HH	Total		On-Peak		Off-Peak	
Model	MAPE	SMAPE	MAPE	SMAPE	MAPE	SMAPE
LSTM	22.66%	11.36%	16.45%	8.79%	30.76%	14.70%
SVR	23.43%	10.63%	17.99%	9.96%	24.68%	11.55%
Blended	23.79%	11.28%	16.70%	8.98%	33.20%	14.34%

It is also important to note that as the number of households increases, the effect of individual prediction errors tends to be partially offset.

Finally, except during the summer season and to a lesser degree in spring, household consumption generally has only a minor influence on grid-level demand compared with other consumption categories, particularly heat pumps (HPs).

Another clear finding (**Figure 8**) is that the blended model—formed by combining an RNN with two regression models—may also be affected by data noise. It should be noted that **Figure 8** shows slightly better performance for the LSTM model, though the difference remains very small.

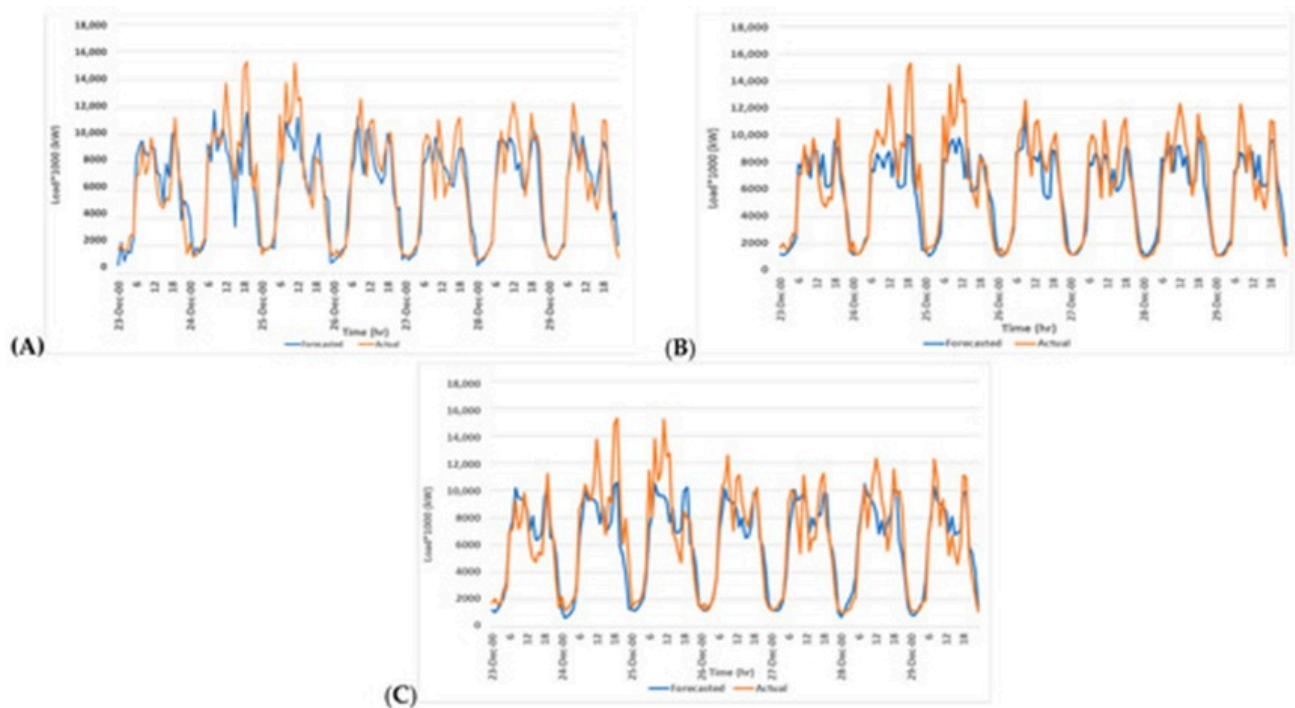


Figure 8. Actual vs. predicted HH load (load 1000) by the (A) blended, (B) SVR, and (C) LSTM models.

HP load prediction

Heat pump (HP) consumption exhibited the highest prediction accuracy (based on the validation results) among the three load categories. This can be attributed to its strong dependence on temperature and seasonal variations. Nevertheless, the underlying pattern is less clear-cut than in the other categories, since the peaks appear at unpredictable times each day and show little or no dependence on temperature, time of day, or other measured variables (**Figure 9**).

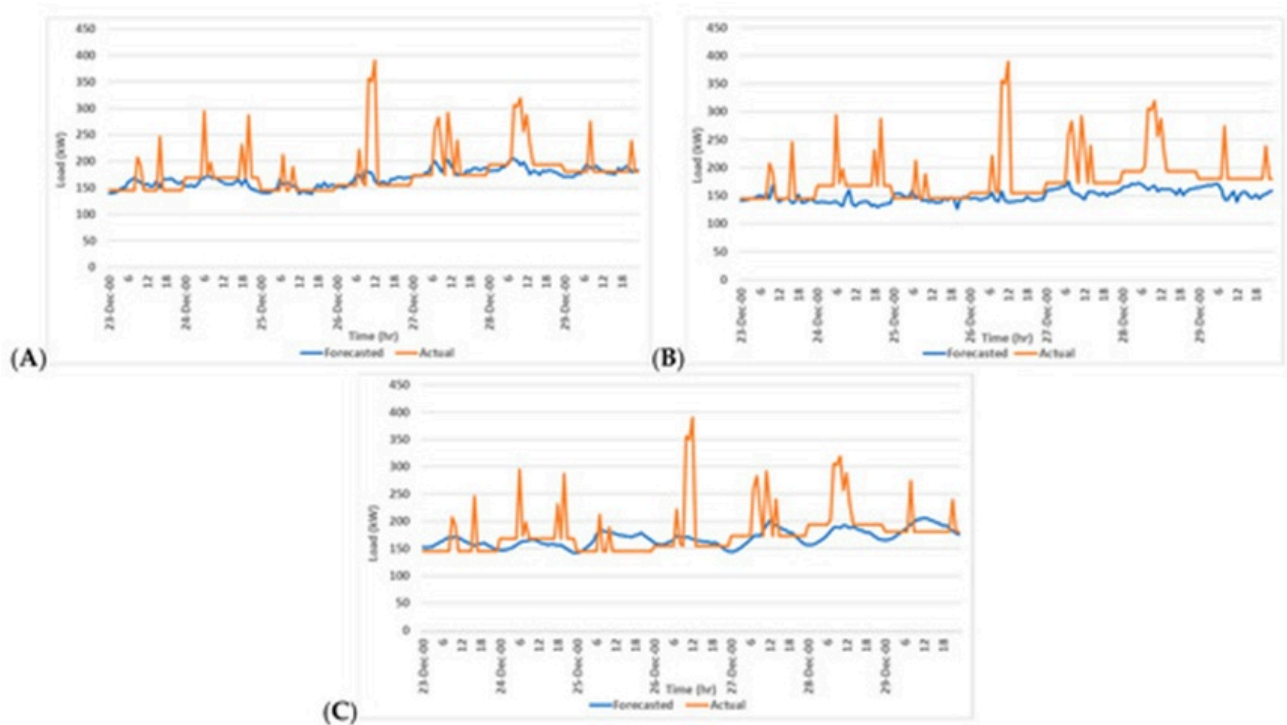


Figure 9. Actual vs. predicted HP load by the (A) blended, (B) SVR, and (C) LSTM models.

A distinct spike in electricity use is typically observed around 7 and 8 AM, followed by a pronounced drop from evening hours through the night. These two intervals are commonly referred to as transient periods. This behavior stems from the buildings' characteristics: the heat pump must raise the indoor air temperature to a comfortable level for occupants in the morning, while residual indoor heat helps sustain acceptable conditions throughout the rest of the day [49]. Furthermore, owing to the thermal inertia of building structures, heating and cooling demands exhibit delayed responses, which introduce unexpected fluctuations in energy consumption. Such variability differs across buildings, leading to peak occurrences at irregular, largely unpredictable times that often lack a strong correlation with the primary input features.

As discussed in the preceding section, model selection should jointly consider visual inspection and quantitative validation metrics. In HP load forecasting, nearly all reported errors originated during peak periods. Although the validation metrics indicated that the blended model outperformed both SVR and LSTM, visual evaluation concluded that none of the models fully met the intended performance objectives. To illustrate this more clearly, **Table 5** presents the results across different HP power load ranges. As observed, during off-peak periods, SVR and the ensemble (blended) model achieved superior performance. This advantage arises from SVR's lower model complexity and from the presence of both linear regression (LR) and SVR components within the blended model, enabling better capture of both linear and moderately nonlinear patterns. The absolute metrics (MAE and RMSE) and the percentage-based metrics (MAPE and SMAPE) showed only minor differences among the models. Still, the blended model consistently delivered the strongest overall performance, whereas SVR performed the worst.

Table 5. MAPE and SMAPE for different observation ranges in HP power forecasting.

HP	Total		On-Peak		Off-Peak	
Model	MAPE	SMAPE	MAPE	SMAPE	MAPE	SMAPE
LSTM	12.81%	6.89%	15.56%	8.95%	9.92%	4.73%
SVR	14.74%	8.61%	21.52%	12.95%	7.63%	4.06%

Blended	9.83%	5.42%	19.79%	7.44%	6.84%	3.36%
---------	-------	-------	--------	-------	-------	-------

Total load forecast

The results clearly indicate that transitioning from individual load categories to aggregated (cumulative) load at the transmission level progressively improves forecast predictability. The highest prediction accuracy, characterized by the lowest overall error, was observed when combining all energy consumption types.

A key consideration is that both the forecasting and testing datasets were gathered during the final days of December, encompassing Christmas Eve and Christmas Day. Some unexpected deviations are therefore reasonable and expected. This observation is further reinforced by the gap between training and testing MAE values: the testing MAE for HH, HP, EV, and total consumption stood at 0.863, 24.98, 32.12, and 47.19, respectively, compared with training MAE values of 0.66, 22.61, 29.34, and 39.38, respectively. Such discrepancies can generally be reduced by extending the training process to include multiple years of historical data.

One of the most critical aspects is the focus on peak periods. Accurate forecasting of peak loads is essential because the primary purpose of load profile prediction is to assess the likelihood of feeder-level bottlenecks during high-demand periods [50]. The forecasting errors during peak and off-peak intervals are detailed in Table 6. As shown, the blended model achieved lower MAPE and SMAPE values across the entire period compared with both LSTM and SVR. This outcome aligns with expectations, as the blended model integrates components with varying sensitivities, enabling it to handle minor data noise better. In contrast, LSTM exhibited greater robustness during on-peak periods, owing to its superior capacity to model complex, nonlinear patterns and long-term dependencies. This capability allows LSTM to track sudden fluctuations and irregular behaviors more effectively.

Visual inspection of the predictions (Figure 10) reveals that, in both the SVR model (Figure 10B) and the blended model (Figure 10A), the majority of errors occur during peak periods, with only minor errors during the lowest load intervals. Conversely, in the LSTM model (Figure 10C), a substantial portion of the error was associated with periods of minimal load. This is understandable: when actual values are very low, even small absolute deviations can produce disproportionately large percentage errors.

Another noteworthy point is that, as illustrated in Figure 10A–10C, the blended model demonstrated the strongest ability to capture noise among all three approaches. For example, on December 23 and 26, small secondary peaks appeared before the main peaks; neither LSTM nor SVR reproduced these features, whereas the blended model attempted to track them. This behavior is consistent with the general principle that model blending increases structural complexity [51], which in turn enables the model to reflect both major peaks and accompanying smaller fluctuations and noise—similar to the pattern previously observed in EV load forecasting.

In addition, this section examined two further performance dimensions: sensitivity and specificity. As previously emphasized, the model's most vital capability is correctly identifying peak load periods. Evaluating whether detected peaks genuinely reflect the model's predictive behavior—or whether they arise coincidentally—is therefore essential. A critical parameter in this analysis is the threshold, defined here as the grid capacity, which distinguishes peak from off-peak conditions. Because this evaluation used synthetic data, the threshold could be adjusted to assess how sensitivity and specificity vary across different threshold levels.

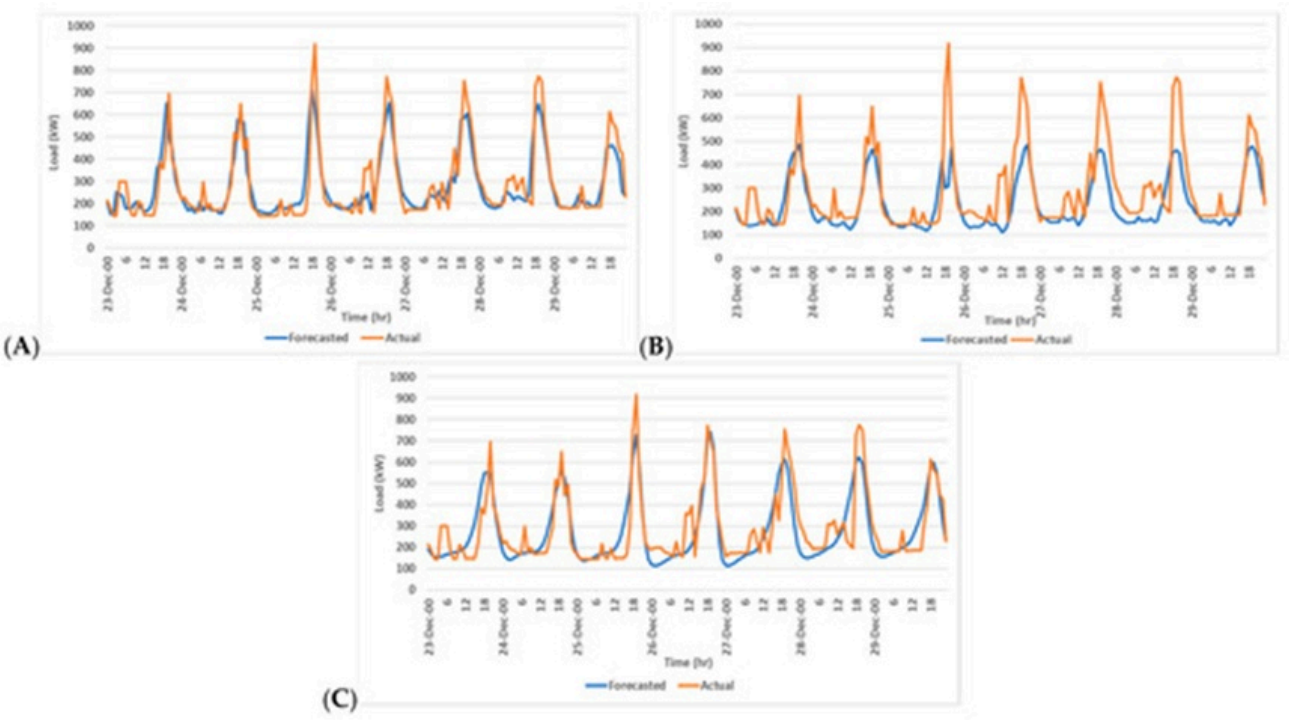


Figure 10. Actual vs. predicted total load by the (A) blended, (B) SVR, and (C) LSTM models.

Table 6. MAPE and SMAPE for different observation ranges in cumulative power forecasting.

Total	Total		On-Peak		Off-Peak	
Model	MAPE	SMAPE	MAPE	SMAPE	MAPE	SMAPE
LSTM	25.66%	12.28%	12.61%	6.79%	27.53%	13.07%
SVR	23.05%	13.25%	33.11%	20.70%	21.69%	12.19%
Blended	18.63%	8.85%	15.62%	8.66%	19.03%	8.88%

These findings were further corroborated by the comprehensive validation metrics presented in [Table 7](#). Overall, LSTM performed best across the evaluated characteristics, particularly in MAE and RMSE. As anticipated, SVR recorded the lowest accuracy, with an MAE value exceeding 60.

Table 7. Evaluation characteristics of the models for synthetic data.

Error	Model	HH	HP	EV	Total
MAE	LSTM	0.863	24.98	32.12	47.19
	SVR	1.4	25.72	39.43	63.30
	Blended	0.88	22.89	36.84	52.98
MSE	LSTM	1.42	1305	4673	4331
	SVR	1.5	1762	3528	6592

	Blended	1.23	1221	4812	6317
RSME	LSTM	1.19	36.13	68.35	65.81
	SVR	1.253	41.97	59.39	81.19
	Blended	1.1	34.95	69.36	79.47
MAPE	LSTM	22.66%	12.81%	inf%	25.66%
	SVR	24.43%	14.74%	inf%	24.05%
	Blended	23.79%	9.83%	inf%	18.63%
SMAPE	LSTM	22.66%	6.89%	65.96%	12.28%
	SVR	23.43%	8.61%	65.1%	13.25%
	Blended	23.79%	5.42%	42.58%	8.85%

LSTM demonstrates superior performance compared to other models for strongly nonlinear, highly time-dependent load patterns, particularly in accurately predicting peak loads. Nevertheless, simpler models such as SVR can sometimes deliver comparable or even superior performance during periods of greater stability, such as off-peak hours, owing to their reduced sensitivity to small-scale fluctuations and noise.

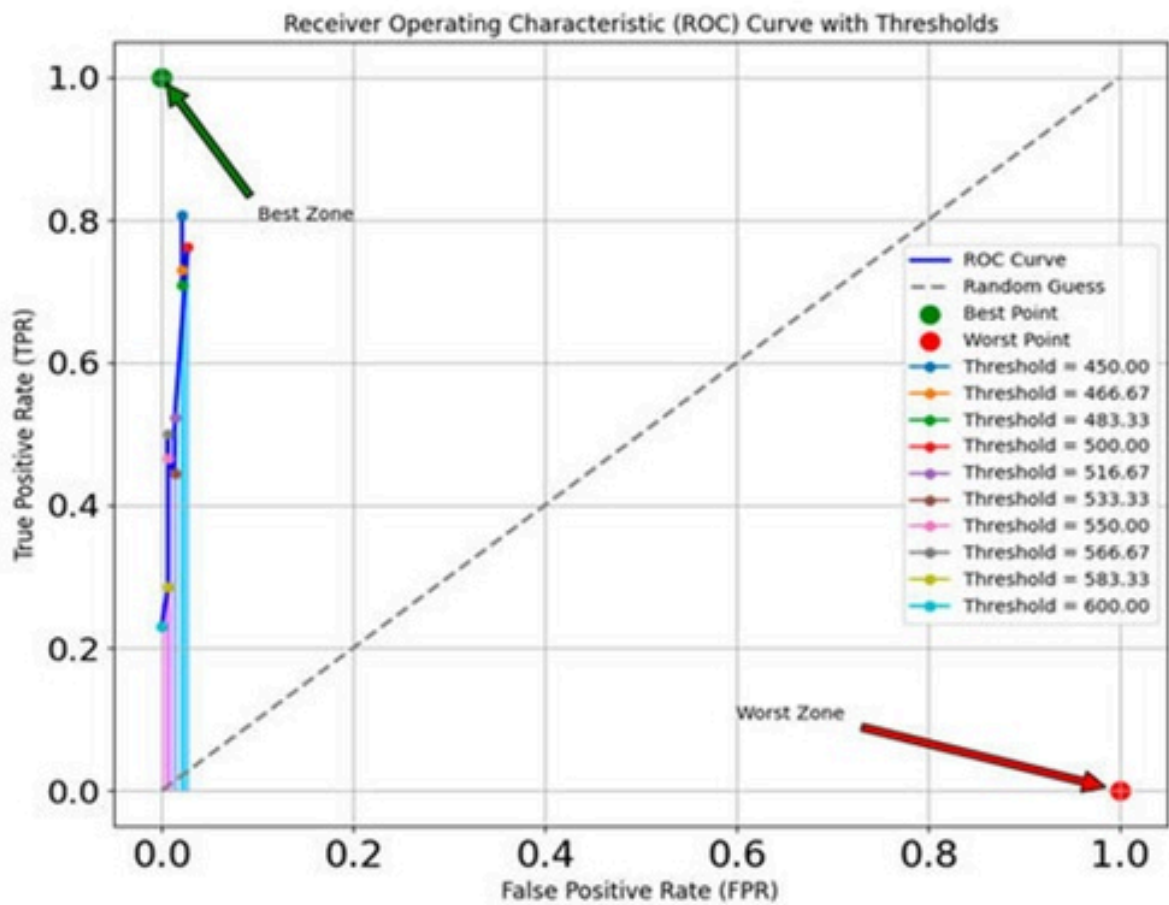
The primary motivation for evaluating model performance across three distinct load levels (on-peak, off-peak, and total) was to reveal how each model behaves in different operational regimes and, crucially, how well it handles outliers and extreme values. Across all load categories, SVR consistently outperformed LSTM—most notably in the household sector, where it achieved clearly better results than both LSTM and the blended model.

Another important observation is the strong agreement between quantitative validation metrics and qualitative visual assessments. For example, SVR showed notably stronger performance in the visual plots during off-peak intervals, which was fully consistent with the numerical validation results. Similarly, in household load forecasting with the blended model, the validation results indicated the weakest performance during off-peak periods; this was visually corroborated by the model's clear failure to accurately reproduce off-peak patterns during certain hours in the initial days ([Figure 8](#)).

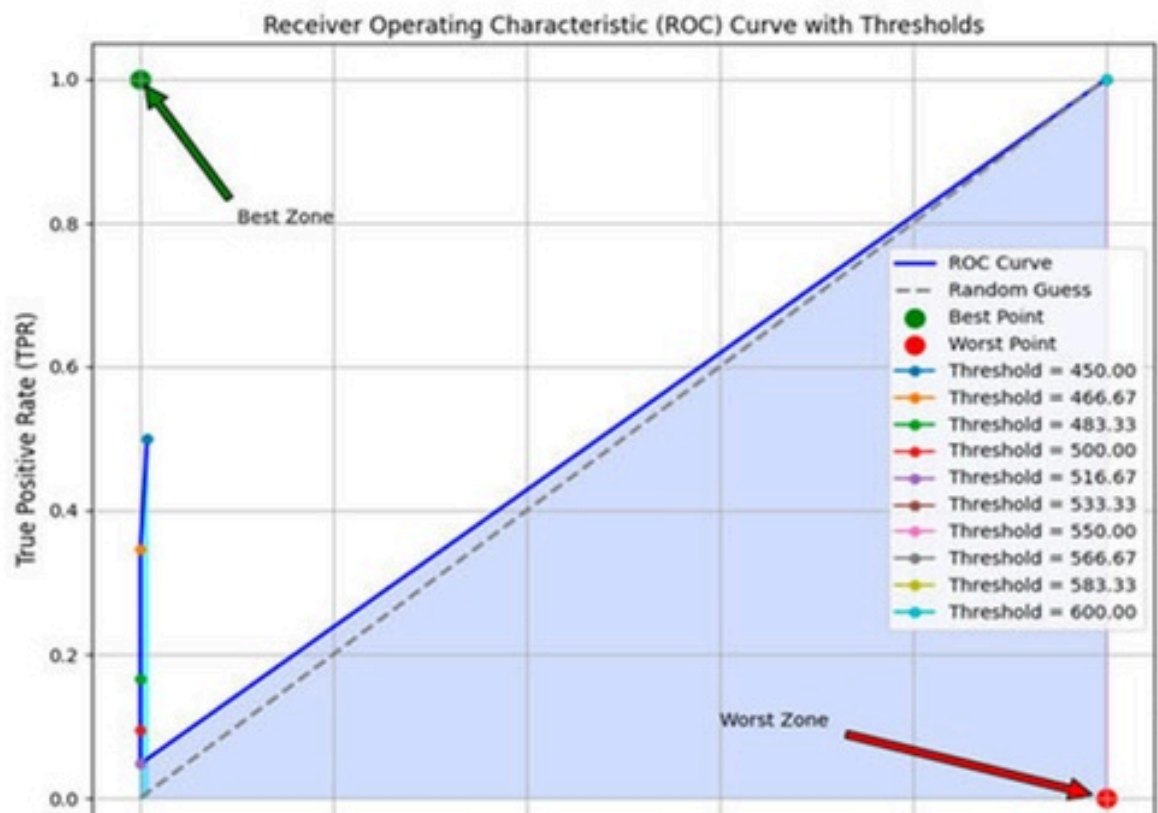
[Figure 11](#) presents the ROC-like curves and the specificity results for the LSTM model. The model's sensitivity fell within the ideal ("perfect") range, confirming that the identified peaks were not random and that the vast majority were true positives. Off-peak periods were also detected with very high accuracy, exhibiting almost no influence from random misclassifications.

$$\text{Sensitivity} = \frac{TP}{TP + FP} \quad (22)$$

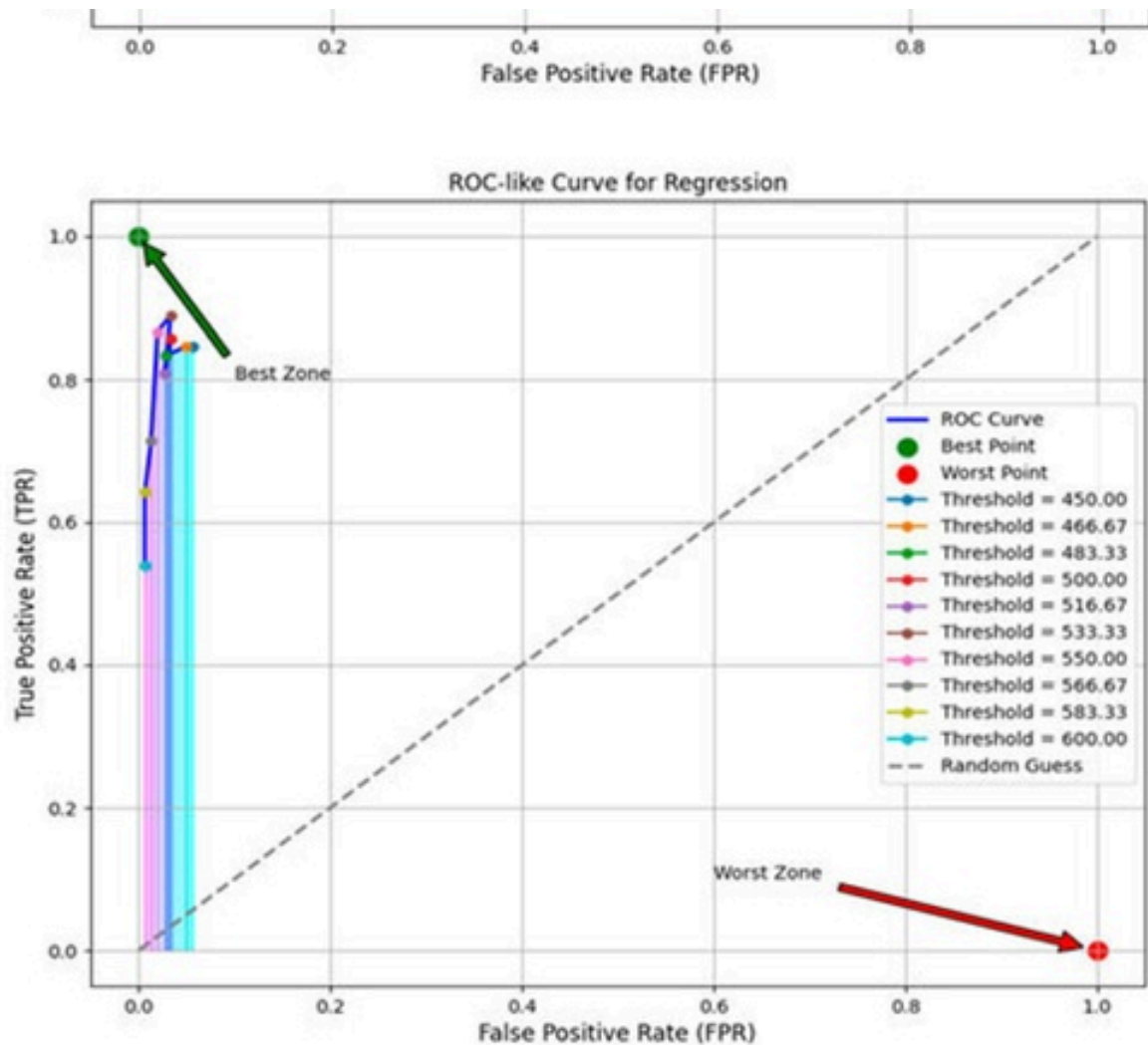
$$\text{Specificity} = \frac{TN}{TN + FN} \quad (23)$$



(A)



(B)



(C)

values, 450 yielded the best overall balance across all three models. Nevertheless, given the typical peak magnitudes observed throughout the year and the shape of the ROC-like curves, a recommended operational threshold lies between 533 and 550. At these higher levels, SVR performed poorly, and the blended model achieved only moderate results, whereas LSTM maintained a notably high true positive rate.

In conclusion, attaining greater robustness remains a key priority. In this context, developing a forecasting approach that allows for repeated, on-the-fly adjustments to enhance flexibility represents a meaningful advancement toward building a truly robust model.

The underlying assumption was that households and users would consume a predetermined quantity of energy over the next 24 hours. Accordingly, the aggregate (total) power demand was first forecasted, after which the power load profile was integrated over time to estimate the corresponding energy demand. This forecasted energy quantity is intended for consumption on the target day, yet it remains subject to modification by external incentives and dynamic price signals.

It was further assumed that electric vehicle (EV) owners would charge their vehicles sometime during the following day, although the precise timing could vary. In such scenarios, forecasting energy consumption (rather than instantaneous power load) provides a more informative indicator of overall energy demand. When a lower electricity price is offered, users are likely to shift their EV charging to that period, potentially creating an additional peak or displacing the original peak. At the conclusion of this study, a specific recommendation will be provided for strategies to manage and mitigate such peak loads.

As illustrated in **Figure 12**, the energy consumption trajectory was recalibrated every 8 hours to offset any surplus or deficit accumulated during the preceding 8-hour interval. The magnitude of the upward or downward adjustment applied in the subsequent hour depends linearly on the current cumulative energy level. For example, at hour 19—where energy consumption reached 553 kWh—the compensatory reduction required to balance the excess usage observed over the first 16 hours was approximately twice as large as the adjustment needed at hour 23, where only 281 kWh had been consumed in that hour. These adjustment relationships are formally defined in **Eqs. 24–27**. A linear adjustment mechanism was selected because no real post-incentive response data were available. Should actual behavioral data become available following several days of price offers, the adjustment logic could be refined to yield more accurate forecasts. It should be noted that the first and second adjustment steps relied on assumed (rather than observed) actual energy values.

$$E_f = \int_0^T P_{f(t)} dt \quad (24)$$

$$\Delta E(t) = \alpha (E_A(t) - E_f(t)) \quad (25)$$

$$\alpha(t) = \frac{E_f(t)}{\sum_{t=1}^T E_f(t)} \quad (26)$$

$$E_a(t) = \alpha(t) (E_t - E_{cf}(t)) \quad (27)$$

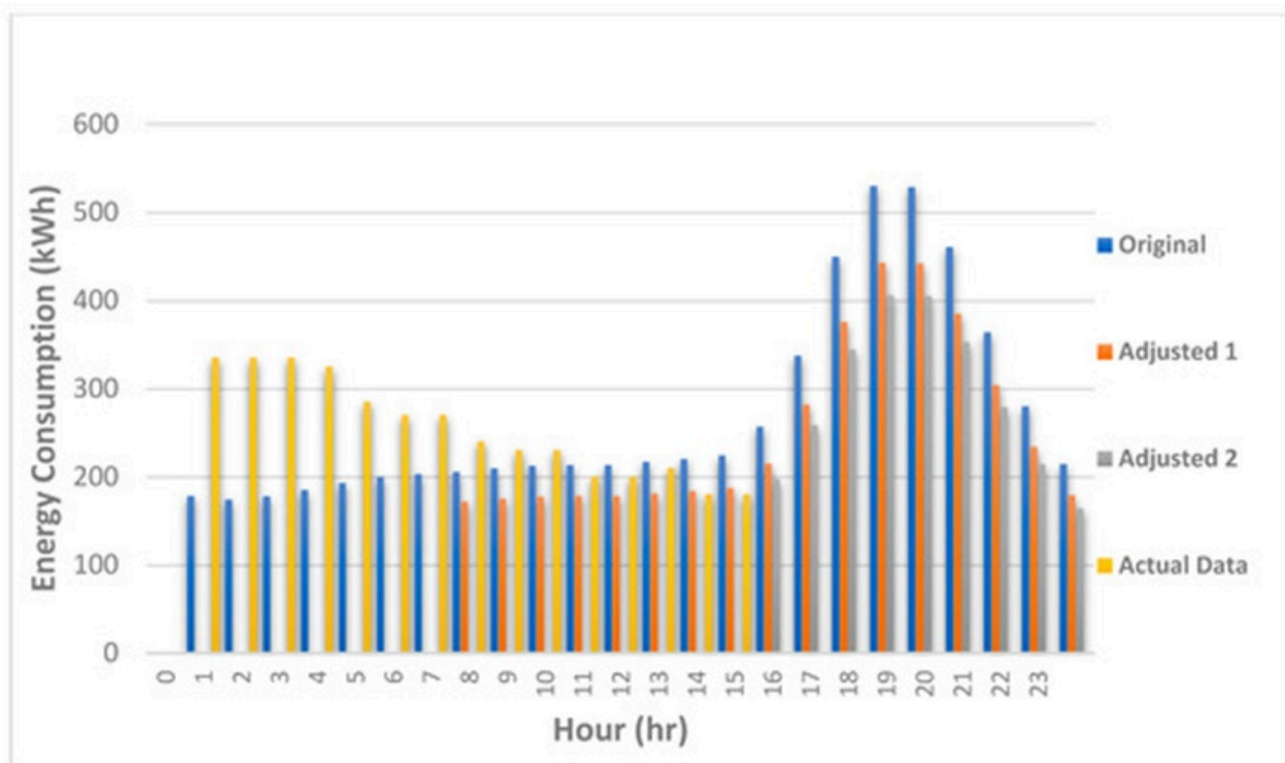


Figure 12. Original vs. adjusted energy consumption of the last testing data.

- E_f = forecasted energy

- $\sum_{t=1}^T E_f(t)$: This expression denotes the sum of forecasted (yet-to-be-revealed) energy portions. For the first adjustment, it corresponds to the sum over the last 16 hours; for the second adjustment, it covers the sum of forecasted energy over the last 8 hours.

- E_A = actual energy

- $\alpha(t)$ = adjustment factor

- E_t = total energy

- E_a = adjusted energy

- E_{cf} = cumulative sum of forecasted energy up to hour t

The adjustment factor is directly influenced by the observed (or assumed) actual energy consumption, resulting in a proportional effect on the magnitude and direction of the correction applied to the forecast.

Measured data

In this section, one-day-ahead load profile forecasting was conducted using measured data from three distinct grids and one shopping center located in different parts of Austria. All predictions in this part of the study were performed at the overall load level, corresponding to the substation level.

Unlike the synthetic dataset, which included a full year of historical data, the measured data available for these grids were limited to 2–3 months due to practical constraints. Consequently, the set of dummy variables employed here differed somewhat from those used in the synthetic data analysis. In particular, the influence of the month variable was not clearly observable, given the narrow range of seasonal and weather conditions covered. Nevertheless, both visual inspections and quantitative validation results indicated satisfactory accuracy—most notably with the LSTM model.

Following confirmation of LSTM's superior performance relative to the other models tested, it was selected to generate one-day-ahead load forecasts for measured data from two separate grids (A and B) as well as for a shopping center in Austria. As expected, LSTM delivered acceptable prediction accuracy for the one-day-ahead horizon (Table 8). In several cases, the model's performance on the real measured data even exceeded the accuracy levels previously achieved when trained exclusively on synthetic data.

Table 8. Evaluation of the characteristics of the models for the measured data.

Grid error	A		B	LCS
	Substation	Feeder		
MAE	5.25	0.717	1.77	11.24
MAE%	5.27%	11.99%	17.87%	5.40%
RMSE	6.88	0.9	2.604	16.67
RMSE%	6.05%	20.49%	25.14%	8.64%
MAPE	4.92%	12.56%	25.03%	6.57%

- At grid A, clear differences in performance emerged between the substation and feeder levels. The substation-level forecasts exhibited markedly higher accuracy, not only in terms of peak capture but also across the overall validation metrics. As previously discussed, this improvement arises primarily because aggregating consumption at lower to higher aggregation levels (i.e., moving from feeder to substation) inherently yields smoother, less volatile load profiles that are easier to predict.

- In contrast, grid B displayed the weakest validation performance among the measured cases. As shown in **Figure 13**, the model either successfully captured the peaks or produced a positive bias (i.e., the forecast systematically overestimated the actual load). While such overestimation does not necessarily disqualify the model from selection, it did reduce the overall accuracy scores in the validation phase.

A primary factor contributing to the reduced robustness of forecasts on measured data (compared with synthetic data) is the substantially shorter training period. The synthetic dataset spanned a full year, enabling better capture of seasonal patterns and input–output relationships, whereas the measured datasets covered no more than 2 months. With such limited historical context, the model faces greater difficulty in accurately determining the relative importance (weights) of input features, and seasonal effects become far less reliable.

- The shopping center exhibited the highest forecasting accuracy among all the measured cases. This strong performance was anticipated, as shopping center load profiles tend to be relatively stable and predictable, with day type among the most influential factors. More specifically, the fixed opening and closing hours largely dictate the consumption level, while the day type (weekday, Saturday, Sunday, holiday) determines whether that typical profile is realized [52]. It should be noted that shopping center operating hours and Sunday trading policies vary across countries: for instance, in the Czech Republic and Slovakia, centers are typically open on Sundays (often until 9 p.m.), whereas in Austria and Germany they generally remain closed on Sundays (and usually close earlier, around 7 p.m., on other days).

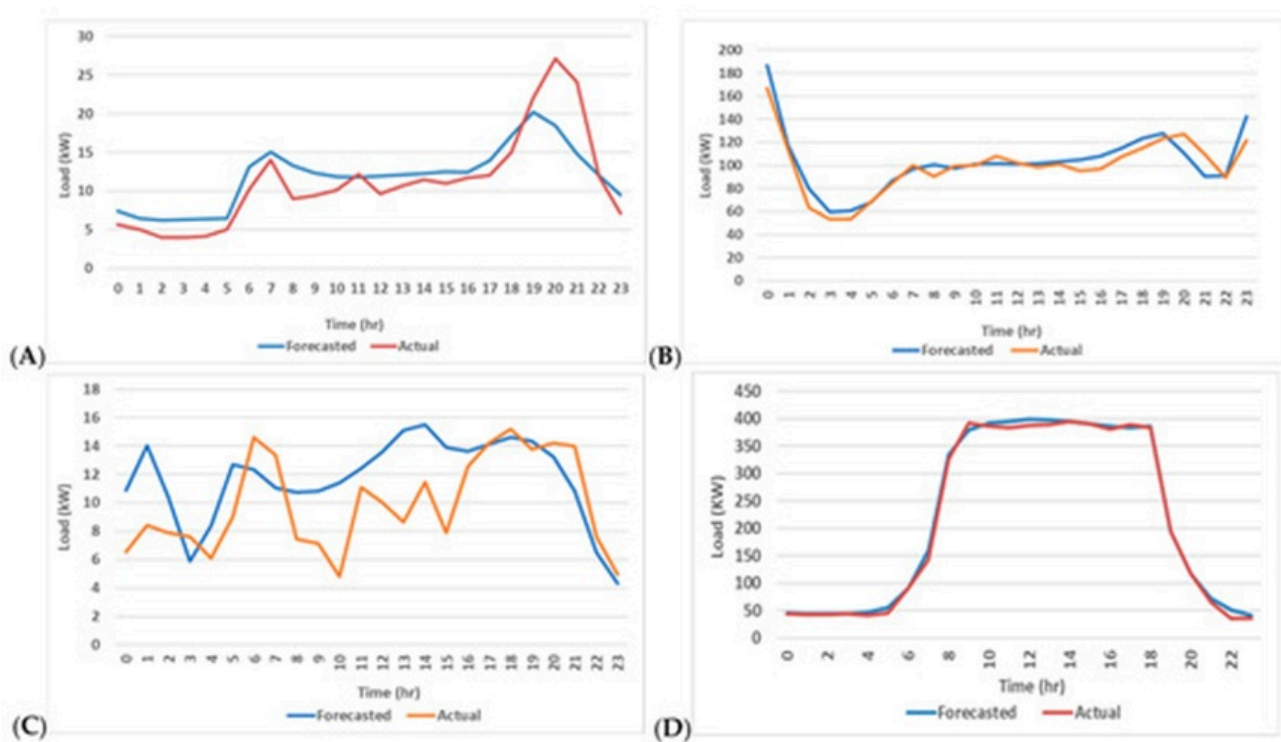


Figure 13. Actual vs. predicted total load of (A) grid A (feeder level), (B) grid A (substation level), (C) grid B, and (D) LCS shopping center.

Conclusion

This study examined the effectiveness of various forecasting models in predicting 24-hour-ahead load profiles across individual load categories and their aggregated (total) form. The central conclusion is that 24-hour-ahead load profiles can be forecasted with generally good—though model-dependent—accuracy. Furthermore, peak load periods were reliably identified, as confirmed by the ROC-like curve analysis.

The models evaluated here represent some of the most widely adopted approaches in load forecasting. The main findings of the research can be summarized as follows:

The impact of different input variables varies depending on multiple factors, notably the dataset type and its temporal length. Forecasting accuracy is typically higher for real-world measured data, where patterns tend to be more structured and consistent, particularly when the dataset encompasses a larger number of households, grids, or urban areas.

In this work, LSTM proved particularly capable of modeling intricate and nonlinear patterns thanks to its advanced recurrent architecture. In contrast, simpler patterns were better captured by simpler methods, such as SVR and GRU. Blended (ensemble) models, which integrate predictions from multiple architectures, occasionally excel at detecting minor peaks or background noise that individual models tend to overlook; however, this can come at the expense of accurately reproducing the dominant (main) peaks. It is important to emphasize that designing an effective blended model requires careful consideration of dataset-specific characteristics to control model complexity appropriately.

Similarly, throughout this research, visual inspection of forecast results proved essential and should be evaluated alongside quantitative validation metrics. This dual approach ensures a balanced assessment that aligns error minimization with the core research objectives—here, accurate peak detection. Consequently, visual evaluation often carried greater weight than numerical metrics alone. A ROC-like curve was additionally employed to assess peak-capture performance. Validation outcomes clearly indicated that LSTM most reliably identified peaks, followed by the blended model, which showed strong potential for future refinement. Overall evaluation metrics further confirmed LSTM's superiority in reproducing general consumption trends.

Regarding the blended model, the results were encouraging, though LSTM ultimately performed better. The proposed blending approach highlighted that different model types are naturally suited to different load regimes. For instance, a hybrid framework incorporating SVR, linear regression (LR), and deep learning components was able to more precisely capture small secondary peaks than LSTM alone, as each constituent model contributed complementary strengths. For continued development, it is strongly recommended that future studies implement dynamic weighting schemes tailored to the specific statistical and behavioral properties of the target dataset.

A significant insight from this research concerns the role of human behavior—an influential factor in energy markets driven by price signals and incentives. The study demonstrated that such behavioral effects become more detectable when forecasts are adjusted in near real-time. This capability could be further enhanced by performing hourly recalibrations, especially for highly variable loads such as electric vehicle (EV) charging. Realizing this improvement, however, would require access to real-time smart meter readings and grid response data.

Several avenues exist for enhancing LSTM performance in future applications. Among the most critical is optimizing the number of lookback time steps, which strongly affects the model's ability to track both rising and falling load trends. The inclusion of appropriately designed dummy (indicator) variables also remains valuable for most datasets, as they assist in pattern recognition. While the dummy variables in this study were constructed specifically to align with peak-related phenomena (consistent with the research focus), alternative formulations from different perspectives could be explored.

Looking forward, ensuring consistently high model performance will require careful attention to both architecture and data preprocessing. For LSTM and similar neural architectures, the number of neurons is a decisive hyperparameter. Monitoring training and validation losses is essential for detecting and preventing overfitting or underfitting. Hyperparameter tuning through methods such as Bayesian optimization, random search, or evolutionary algorithms offers substantial value, particularly in renewable-integrated forecasting tasks. To increase model interpretability, explainable AI techniques—including Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME)—can reveal feature importance and decision rationales. Feature selection strategies (e.g., correlation analysis, principal component analysis, recursive feature elimination) are likewise recommended to enhance predictive power in subsequent work. Implementing early stopping with varying patience values is highly advisable to mitigate overfitting, while addressing saturation issues in the forget gate—common in certain data regimes—should also be considered.

From a system-level perspective, developing optimization routines or formulations to suppress peaks that risk exceeding grid thresholds proactively offers considerable practical value. Given the sharp peak contributions often associated with EV charging, introducing a reservation/booking system for EV users—allowing them to schedule sessions with predefined power limits—could effectively smooth load profiles. Complementary demand-side measures, such as dynamic and threshold-based pricing, offer another promising pathway for peak management. In such a scheme, electricity prices would be tiered according to the current load level relative to generation capacity (the defined threshold). For example, prices could differ substantially when the load is below 50% of capacity compared with when it lies between 60% and 70%. This structure would incentivize EV owners to monitor prices frequently and shift charging to lower-cost periods. A critical design feature is that once a charging session commences, the price remains locked for its duration, providing users with price certainty and avoiding mid-session changes.

Abbreviations

ANN: Artificial Neural Network

ANFIS: Adaptive Neuro-Fuzzy Inference System

LSTM: Long Short-Term Memory

BPTT: Backpropagation Through Time

DELM: Deep Extreme Learning Machine

EV: Electric Vehicles

ERCOT: Electric Reliability Council of Texas

EDFs: Energy Demand Forecasts

ELM: Extreme Learning Machine

XAI: explainable AI

FN: False Negative

FP: False Positive

GRU: Gated Recurrent Units

GRNN: Generalized Regression Neural Network

HP: Heat Pump

HHs: Households

LIME: Local Interpretable Model-agnostic Explanations

LR: Linear Regression

MAE: Mean Absolute Error

MAPE: Mean Absolute Percentage Error

MLK: Masking Layer in Keras

MSE: Mean Squared Error

NaN: Not a Number

RBF: Radial Basis Function

ROC: Receiver Operating Characteristic

ReLU: Rectified Linear Unit

RNN: Recurrent Neural Networks

RES: Renewable Energy Sources

RMSE: Root Mean Squared Error

SMO: Sequential Minimal Optimization

SVM: Support Vector Machines

SVR: Support Vector Regression

SMAPE: Symmetric Mean Absolute Percentage Error

SHAP: Shapley Additive Explanations

TN: True Negative

TP: True Positive

Acknowledgements

None

Conflict of interest

None

Financial support

This research was funded by the Austrian Climate and Energy Funds (Klima- und Energiefonds) for the 'Friendly Charge Project' (No. 899917).

Ethics statement

None

Received: 11 Mar 2023

Revised: 22 Apr 2023

Accepted: 03 Jun 2023

Published online: 20 September 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Yeom S, An J, Hong T, Koo C, Jeong K, Lee J. Managing energy consumption and indoor environment quality using augmented reality based on the occupants' satisfaction and characteristics. *Energy Build.* 2024;311:114165.
- Bujalski M, Madejski P, Fuzowski K. Day-ahead heat load forecasting during the off-season in the district heating system using generalized additive model. *Energy Build.* 2023;278:112630.
- Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans Smart Grid.* 2019;10:841–51.
- Wang Y, Gao N, Hug G. Personalized federated learning for individual consumer load forecasting. *CSEE J Power Energy Syst.* 2023;9:326–30.
- Wang C, Li X, Li H. Role of input features in developing data-driven models for building thermal demand forecast. *Energy Build.* 2022;277:112593.
- Mendes N, Mendes J, Mohammadi J, Moura P. Federated learning framework for prediction of net energy demand in transactive energy communities. *Sustain Energy Grids Netw.* 2024;40:101522.
- Bui DK, Nguyen TN, Ngo TD, Nguyen-Xuan H. An artificial neural network expert system enhanced with the electromagnetism-based firefly algorithm for predicting the energy consumption in buildings. *Energy.* 2020;190:116370.
- Kanthila C, Boodi A, Marszal-Pomianowska A, Beddiar K, Amirat Y, Benbouzid M. Enhanced multi-horizon occupancy prediction in smart buildings using cascaded Bi-LSTM models with integrated features. *Energy Build.* 2024;318:114442.

Xie J, Zhong Y, Xiao T, Wang Z, Zhang J, Wang T, et al. A multi-information fusion model for short term load forecasting of an architectural complex considering spatio-temporal characteristics. *Energy Build.* 2022;277:112566.

Wang Y, Gan D, Sun M, Zhang N, Lu Z, Kang C. Probabilistic individual load forecasting using pinball loss guided LSTM. *Appl Energy.* 2019;235:10–20.

Rick R, Berton L. Energy forecasting model based on CNN-LSTM-AE for many time series with unequal lengths. *Eng Appl Artif Intell.* 2022;113:104998.

Hu W, Wang X, Tan K, Cai Y. Digital twin-enhanced predictive maintenance for indoor climate: a parallel LSTM-autoencoder failure prediction approach. *Energy Build.* 2023;301:113738.

Liu X, Lin Z, Feng Z. Short-term offshore wind speed forecast by seasonal ARIMA—a comparison against GRU and LSTM. *Energy.* 2021;227:120492.

Yang S, Yu X, Zhou Y. LSTM and GRU neural network performance comparison study: taking Yelp review dataset as an example. In: *Proceedings of the 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI 2020)*. Shanghai, China; 2020. p. 98–101.

Fayaz M, Kim D. A prediction methodology of energy consumption based on deep extreme learning machine and comparative analysis in residential buildings. *Electronics.* 2018;7:222.

Wang W, Hong T, Xu X, Chen J, Liu Z, Xu N. Forecasting district-scale energy dynamics through integrating building network and long short-term memory learning algorithm. *Appl Energy.* 2019;248:217–30.

Hossain MS, Mahmood H. Short-term load forecasting using an LSTM neural network. In: *Proceedings of the 2020 IEEE Power and Energy Conference at Illinois (PECI)*. Champaign, IL, USA; IEEE, Piscataway, NJ, USA; 2020.

Jin N, Yang F, Mo Y, Zeng Y, Zhou X, Yan K, et al. Highly accurate energy consumption forecasting model based on parallel LSTM neural networks. *Adv Eng Inform.* 2022;51:101442.

Dubey AK, Kumar A, García-Díaz V, Sharma AK, Kanhaiya K. Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustain Energy Technol Assess.* 2021;47:101474.

Du R, Chen H, Yu M, Li W, Niu D, Wang K, et al. 3DTCN-CBAM-LSTM short-term power multi-step prediction model for offshore wind power based on data space and multi-field cluster spatio-temporal correlation. *Appl Energy.* 2024;376:124169.

Smagulova K, James AP. A survey on LSTM memristive neural network architectures and applications. *Eur Phys J Spec Top.* 2019;228:2313–24.

Landi F, Baraldi L, Cornia M, Cucchiara R. Working memory connections for LSTM. *Neural Netw.* 2021;144:334–41.

Karijadi I, Chou SY. A hybrid RF-LSTM based on CEEMDAN for improving the accuracy of building energy consumption prediction. *Energy Build.* 2022;259:111908.

Fan D, Sun H, Yao J, Zhang K, Yan X, Sun Z. Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy.* 2021;220:119708.

Cai C, Tao Y, Zhu T, Deng Z. Short-term load forecasting based on deep learning bidirectional LSTM neural network. *Appl Sci.* 2021;11:8129.

Somu N, Raman G, Ramamritham MRK. A deep learning framework for building energy consumption forecast. *Renew Sustain Energy Rev.* 2021;137:110591.

Wei N, Li C, Peng X, Zeng F, Lu X. Conventional models and artificial intelligence-based models for energy consumption forecasting: a review. *J Pet Sci Eng*. 2019;181:106187.

Ju J, Liu FA. Multivariate time series data prediction based on ATT-LSTM network. *Appl Sci*. 2021;11:9373.

Hewamalage H, Ackermann K, Bergmeir C. Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Min Knowl Discov*. 2023;37:788–832.

Zucchet N, Orvieto A. Recurrent neural networks: vanishing and exploding gradients are not the end of the story. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2024;37:139402–43. <http://arxiv.org/abs/2405.21064>

Zou L, Munir MS, Kim K, Hong CS. Day-ahead energy sharing schedule for the P2P prosumer community using LSTM and swarm intelligence. In: *Proceedings of the 2020 International Conference on Information Networking (ICOIN)*. Barcelona, Spain;2020. p. 396–401.

Chamatidis I, Tzanes G, Istrati D, Lagaros N, Stamou A. Short-term forecasting of rainfall using sequentially deep LSTM networks: A case study on a semi-arid region. *Environ Sci Proc*. 2023;26:157.

Nguyen BN, Ogliari E, Pafumi E, Alberti D, Leva S, Duong MQ. Forecasting generating power of sun-tracking PV plant using long short-term memory neural network model: a case study in Ninh Thuan, Vietnam. In: *Proceedings of the 2024 IEEE 10th International Conference on Communications and Electronics (ICCE)*. Danang, Vietnam; 2024. p. 333–8.

Sultana N, Hossain SMZ, Abusaad M, Alanbar N, Senan Y, Razzak SA. Prediction of biodiesel production from microalgal oil using Bayesian optimization algorithm-based machine learning approaches. *Fuel*. 2022;309:122184.

Zhang Y, Wang Q, Chen X, Yan Y, Yang R, Liu Z, et al. The prediction of spark-ignition engine performance and emissions based on the SVR algorithm. *Processes*. 2022;10:312.

Ahmed N, Assadi M, Zhang Q. Investigating the impact of borehole field data's input parameters on the forecasting accuracy of multivariate hybrid deep learning models for heating and cooling. *Energy Build*. 2023;301:113706.

Huang X, Han Y, Yan J, Zhou X. Hybrid forecasting model of building cooling load based on EMD-LSTM-Markov algorithm. *Energy Build*. 2024;321:114670.

Zhang S, Gurusamy S, James-Chakraborty K, Basu B. Short-term office temperature forecasting through a data-driven approach integrated with bidirectional gated recurrent neural network. *Energy Build*. 2024;314:114231.

Shi J, Teh J. Load forecasting for regional integrated energy system based on complementary ensemble empirical mode decomposition and multi-model fusion. *Appl Energy*. 2024;353:122146.

ALMahadin G, Aoudni Y, Shabaz M, Agrawal AV, Yasmin G, Alomari ESR, et al. VANET network traffic anomaly detection using GRU-based deep learning model. *IEEE Trans Consum Electron*. 2023;70:4548–55.

Li B, Wu Y, Song J, Lu R, Li T, Zhao L. DeepFed: federated deep learning for intrusion detection in industrial cyber-physical systems. *IEEE Trans Ind Inform*. 2021;17:5615–24.

Yu H, Zhong F, Du Y, Xie X, Wang Y, Zhang X, et al. Short-term cooling and heating loads forecasting of building district energy system based on data-driven models. *Energy Build*. 2023;298:113513.

Yan M, Liang XM, Lu ZH, Wu J, Zhang W. HANSEL: Adaptive horizontal scaling of microservices using Bi-LSTM. *Appl Soft Comput*. 2021;105:107216.

Wang ZX, He LY, Zhao YF. Forecasting the seasonal natural gas consumption in the US using a gray model with dummy variables. *Appl Soft Comput*. 2021;113:108002.

Zhao D, Liu Y, Chen H. Are mini and full-size electric vehicle adopters satisfied? An application of the regression with dummy variables. *Travel Behav Soc*. 2024;35:100744.

Khaled A, Elsir AMT, Shen Y. GSTA: Gated spatial–temporal attention approach for travel time prediction. *Neural Comput Appl*. 2022;34:2307–22.

Thormann B, Kienberger T. Evaluation of grid capacities for integrating future e-mobility and heat pumps into low-voltage grids. *Energies*. 2020;13:5083.

Al Mamun A, Sohel M, Mohammad N, Sunny MSH, Dipta DR, Hossain E. A comprehensive review of the load forecasting techniques using single and hybrid predictive models. *IEEE Access*. 2020;8:134911–39.

Sun S, Chen H. Data-driven sensitivity analysis and electricity consumption prediction for water source heat pump system using limited information. *Build Simul*. 2021;14:1005–16.

Garrido-Valenzuela F, Cruz D, Dragicevic M, Schmidt A, Moya J, Tamblay S, et al. Identifying and visualizing operational bottlenecks and quick win opportunities for improving bus performance in public transport systems. *Transp Res Part A Policy Pract*. 2022;164:324–36.

André M, Perez R, Soubdhan T, Schlemmer J, Calif R, Monjoly S. Preliminary assessment of two spatio-temporal forecasting techniques for hourly satellite-derived irradiance in a complex meteorological context. *Sol Energy*. 2019;177:703–12.

Yuan Y, Chen Z, Wang Z, Sun Y, Chen Y. Attention mechanism-based transfer learning model for day-ahead energy demand forecasting of shopping mall buildings. *Energy*. 2023;270:126878.