

An Application of Deep Learning for Short-Term Electricity Load Forecasting

DRIADA MITRUSHI, VALBONA MUDA, IRMA BERDUF, URIM BUZRA, JOAN JANI

Department of Engineering Physics
Polytechnic University of Tirana,
"Sulejman Delvina Street",
1001 Tirana,
ALBANIA

Abstract: - Short-term electricity load forecasting is a challenging time-series problem due to nonlinear demand dynamics, strong temporal dependence, and recurring calendar effects. This paper evaluates deep learning models using open operational data for Albania (2024–2025). Recurrent architectures based on Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are used as baseline models, while a hybrid Transformer–LSTM is assessed to examine the contribution of attention mechanisms. Calendar effects are incorporated using trainable embeddings, and temperature is analyzed through a controlled ablation study based on ERA5-Land data. All models were trained and evaluated under a unified framework using standard regression metrics. The results show high accuracy ($\text{MAPE} < 2\%$, $R^2 > 0.99$) with LSTM without temperature providing the best average performance; however, an analysis per regime shows that the errors are concentrated in the ramping conditions where GRU and Transformer–LSTM are more robust, pointing to the importance of regime-aware evaluation.

Key-Words: - short-term load forecasting; one-hour-ahead forecasting; deep learning; LSTM; GRU; Transformer; regime-based evaluation; ramping analysis.

Received: March 13, 2026. Revised: May 12, 2026. Accepted: June 15, 2026. Published: September 21, 2026.

1 Introduction

Accurate short-term electricity load forecasting (STLF) is a fundamental requirement for secure and economic power system operation, supporting dispatch, balancing, reserve procurement, and real-time security actions at operational horizons ranging from minutes to days, [1]. In European power systems, one-hour-ahead forecasts are particularly critical for balancing processes and real-time decision-making, where transmission system operators must continuously manage system imbalances to maintain frequency stability, [2]. These operational requirements are implemented within harmonized European rules for system operation and cross-border coordination, where reliable short-horizon forecasts contribute directly to congestion management and system security, [3]. The STLF problem has become increasingly challenging due to evolving consumption patterns, higher penetration of variable renewable energy sources, electrification of end-use sectors, and changing consumer behavior. These developments amplify demand uncertainty and introduce stronger nonlinearities in load dynamics, increasing the operational cost of forecasting errors through inefficient reserve activation and higher balancing costs, [4]. Forecasting becomes more

challenging in small power systems, where the limited level of aggregation reduces the natural smoothing of demand and individual load variations have a stronger impact on short-term behaviour, [5]. In the Albanian power system, hydropower is the main generation. As most domestic electricity is supplied by hydroelectric plants, system operation is highly dependent on seasonal and hydrological conditions. During dry periods, the increased need for electricity imports makes the system more vulnerable to variations in regional market prices. In periods with low precipitation, the need for electricity imports increases, which makes the system more sensitive to fluctuations in regional market prices. Under these conditions, accurate short-term load forecasting becomes particularly important, as it directly affects cross-border scheduling, balancing operations, and the management of system imbalances, [6]. Albania's integration within the Western Balkans electricity framework, together with ongoing market coupling and regulatory alignment efforts, further highlights the relevance of advanced forecasting tools for operational efficiency and security of supply, [7], [8]. From a data perspective, the ENTSO-E Transparency Platform provides open and standardized operational datasets for European control areas, enabling

reproducible country-level STLF studies, [9], [10], [11]. Complementary meteorological information is available from the Copernicus Climate Data Store through the ERA5-Land reanalysis, which delivers hourly near-surface temperature and other land variables suitable for electricity demand forecasting applications, [12], [13], [14]. Deep learning has emerged as a dominant approach for STLF due to its ability to learn nonlinear temporal representations directly from data. Recurrent neural networks, particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, have shown strong performance in capturing short-lag autocorrelation and recurring demand patterns. More recently, attention mechanisms together with Transformer architectures have been proposed to enhance temporal representation learning. However, for very short forecasting horizons—such as one hour ahead—and especially in small power systems, the practical benefits of attention mechanisms remain an open question. Despite the growing literature on deep learning-based STLF, most studies emphasize average accuracy metrics without explicitly analyzing when forecasting errors occur and which operating regimes dominate operational risk. From a time-series perspective, one-step-ahead forecasts are expected to be largely driven by recent load history and calendar regularities, while architectural differences are more likely to manifest during dynamic transitions, such as rapid demand ramps, rather than during steady peak or valley conditions. This motivates the adoption of regime-based evaluation frameworks that link modeling assumptions to observed numerical behavior and operational relevance.

In this context, the present paper develops a deep learning-based framework for one-hour-ahead electricity load forecasting in Albania, comparing recurrent and attention-enhanced architectures under a unified experimental protocol. In addition to standard accuracy metrics, the analysis emphasizes regime-dependent error behavior under peak and ramping conditions. The incremental value of temperature is examined through a controlled ablation study to assess its relevance at short horizons in a hydro-dominated system.

The main contributions of this work are summarized as follows:

1. A unified comparison of recurrent deep learning models (LSTM and GRU) and a hybrid attention-recurrent architecture (Transformer-LSTM) under consistent experimental conditions.
2. Integration of calendar embeddings and a systematic temperature ablation study to

assess feature relevance in short-horizon forecasting.

3. Benchmarking against persistence and ENTSO-E day-ahead forecasts to provide operational context.
4. A regime-based error analysis reveals robustness differences that are masked by average accuracy metrics.

The remainder of the paper is organized as follows. Section 2 reviews related work on STLF and deep learning approaches. Section 3 describes the data sources, model architectures, and experimental setup. Section 4 reports the forecasting results together with the regime-based analysis. Finally, the main conclusions of the study are summarized in Section 5.

2 Literature Review

2.1 Classical Approaches to Short-Term Load Forecasting

Early studies on short-term electricity load forecasting (STLF) were mainly based on classical statistical methods derived from linear time-series analysis. Linear regression and decomposition techniques were among the first approaches used, mainly because of their simplicity and clear interpretation. Subsequently, autoregressive and moving average models were introduced to explicitly capture temporal dependence in load time series. Autoregressive integrated moving average (ARIMA) models became widely adopted for STLF because of their solid theoretical foundation and reasonable performance under stable operating conditions, [15], [16]. Exponential smoothing and state-space formulations were also applied to short-term demand forecasting, particularly for systems exhibiting strong seasonality and regular demand patterns, [17]. Despite their advantages, these statistical models rely on assumptions of linearity and weak non-stationarity, which are often violated in real electricity demand data. Electricity load exhibits nonlinear behavior, abrupt changes, and complex dependencies driven by human activity, weather conditions, and socio-economic factors, [18]. As a result, the predictive accuracy of purely statistical approaches is limited, especially for short forecasting horizons during volatile operating conditions, where rapid demand transitions dominate forecasting difficulty.

2.2 Machine Learning and Deep Learning Methods for Load Forecasting

To overcome the limitations of classical statistical models, machine learning techniques were

increasingly adopted for STLF. Artificial neural networks were among the first nonlinear models applied to electricity demand forecasting and demonstrated improved accuracy compared with linear methods. Support vector regression was later introduced to capture nonlinear relationships between load and explanatory variables while maintaining good generalization properties, [18], [19]. Tree-based ensemble methods, including random forests, gradient boosting algorithms, and their modern variants, further improved forecasting performance by combining multiple weak learners and modelling nonlinear interactions among inputs, [20], [21], [22]. While these machine learning approaches are effective in handling nonlinearities, they typically rely on manual feature engineering and treat temporal dependence implicitly rather than explicitly. For very short forecasting horizons, such as one hour ahead, this limitation can reduce their effectiveness when rapid temporal dynamics dominate. As a result, deep-learning methods have become one of the main approaches used in short-term load forecasting, mainly because they are able to learn complex temporal patterns directly from historical data without requiring explicit model formulation, [23], [24]. In particular, recurrent neural networks (RNNs) are well-suited for time-series prediction problems, since they process data sequentially and maintain information about past observations through their internal state, [25]. Among the different recurrent architectures, Long Short-Term Memory (LSTM) networks have been widely used in hourly load forecasting applications due to their ability to capture long-term dependencies in demand data. Their gated memory structure enables the capture of both short-term fluctuations and longer periodic patterns, such as daily and weekly cycles. Several studies have shown that LSTM-based forecasting approaches often provide better accuracy than traditional statistical techniques and standard machine-learning models, particularly in large interconnected power systems. The Gated Recurrent Unit (GRU) was later introduced as a more compact recurrent architecture, designed to reduce model complexity while retaining the ability to capture temporal dependencies in time-series data, [26], [27]. Reported results in the literature indicate that GRU models can reach similar, and sometimes higher, short-term forecasting accuracy, especially in systems where the load profile is highly correlated in time and the amount of training data is relatively limited, [26].

2.3 Attention Mechanisms and Transformer-Based Models

Attention mechanisms were introduced in neural network models to improve the handling of long input sequences by allowing the model to emphasize the most relevant parts of the data during the learning process, [28]. This idea was later extended in the Transformer architecture, which relies entirely on self-attention and removes the need for recurrent connections, making it possible to model long-range temporal relationships more directly, [29]. Transformer-based approaches have shown strong performance in several time-series forecasting applications, including electricity load prediction, [30], [31]. Nevertheless, fully attention-based models usually require large amounts of training data to reach their full potential, and they may be less effective in very short-term forecasting problems where local sequential patterns play a dominant role, [32], [33]. This limitation is particularly relevant for small power systems, where data availability is limited and short-lag dependencies are dominant. To overcome these limitations, several studies have proposed hybrid architectures that combine Transformer encoders with recurrent neural networks. In these approaches, the attention mechanism is used to capture long-range temporal dependencies, while the recurrent layers help preserve short-term sequential behaviour. Although such hybrid models have shown promising results in different forecasting applications, their effectiveness for one-hour-ahead load prediction in small, hydro-dominated power systems has not been extensively studied. The limited evidence available in the literature motivates further analysis.

2.4 Effect of Temperature and Weather Variables

Among the different external factors that affect electricity consumption, temperature plays a particularly important role because it determines the level of heating and cooling required by users. Numerous STLF studies incorporate temperature as an exogenous input and report improved forecasting accuracy, particularly during extreme weather conditions, [34], [35], [36]. The relationship between load and temperature is often nonlinear and region-dependent, with varying sensitivity across seasons and system characteristics, [37]. For very short forecasting horizons, however, recent studies suggest that the marginal value of weather variables may be reduced when strong load persistence and calendar effects dominate. In small and hydro-dominated systems, demand dynamics may be primarily driven by consumption regularities rather than by immediate

meteorological forcing, [38], [39], [40]. Only a limited number of studies explicitly perform temperature ablation analyses to quantify the incremental contribution of weather information under short-horizon conditions.

2.5 STLF in Small and Hydro-Dominated Power Systems and Evaluation Gaps

Most existing STLF studies focus on large interconnected power systems, where aggregation smooths demand variability and forecasting errors.

Small power systems require forecasting approaches that emphasize robustness, generalization, and reliable model evaluation, particularly under conditions of limited data availability and higher demand variability, [41], [42], [43]. In hydro-dominated systems, electricity demand forecasting interacts indirectly with hydrological constraints and reservoir management strategies, increasing the operational value of accurate short-horizon predictions. An additional limitation in the existing literature concerns evaluation practices. Many studies rely primarily on average error metrics, such as MAE or RMSE, which may obscure model weaknesses under operationally critical conditions, [44], [45], [46]. Recent work suggests that forecasting difficulty is often concentrated during ramping events and rapid transitions rather than during steady demand levels, [47], [48], [49], [50]. However, systematic regime-based error analyses remain relatively scarce, particularly for small power systems and very short forecasting horizons. Addressing this gap motivates the regime-aware evaluation framework adopted in the present study.

3 Methodology

This study investigates one-hour-ahead short-term electricity load forecasting (STLF) for Albania, a small and hydro-dominated power system characterized by limited demand aggregation, strong daily and weekly consumption patterns, and high operational sensitivity to short-horizon forecasting errors. These characteristics motivate the use of forecasting models capable of capturing short-lag temporal dependence, nonlinear demand dynamics, and recurring calendar effects, while maintaining robustness during rapid demand transitions.

Hourly electricity demand data were obtained from the ENTSO-E Transparency Platform for the Albanian control area. Two time series are used: Actual total load (target variable), and

Day-ahead load forecast (used strictly as a benchmark baseline and not as a model input unless

stated otherwise). Meteorological forcing is represented by 2 m air temperature from ERA5-Land (hourly). All time stamps are aligned on a common hourly index. To ensure a consistent merge, load, and temperature series are processed on the same time basis (UTC) and then joined by time stamp. The raw hourly dataset is screened for:

(i) duplicate time stamps, (ii) missing hours, and (iii) incomplete intra-hour coverage where applicable (using the available `n_15min_points` indicator). Only hours passing these consistency checks are retained. After cleaning and alignment, a unified dataset containing load, calendar variables, and temperature is formed.

3.1 Feature Engineering and Input Representation

The forecasting task is formulated as a direct one-step-ahead prediction problem, expressed as:

$$\hat{y}_{\{t+1\}} = f(X_t) \quad (1),$$

where y_t is the actual load at hour t , and X_t is a fixed-length historical input window:

$$X_t = [x_{\{t-L+1\}}, x_{\{t-L+2\}}, \dots, x_t] \quad (2),$$

$$X_t \in \mathbb{R}^{L \times d} \quad (3),$$

where L is the lookback window length, and d is the number of features per time step. Features include historical load values, calendar information (hour-of-day and day-of-week encoded via trainable embeddings), and when applicable temperature. A temperature ablation study is performed by training each model both with and without temperature inputs, while keeping all other features unchanged.

3.2 Experimental Protocol, Data Splits, and Reproducibility

Forecast performance was assessed using common regression metrics. These include the mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2), defined as follows:

Mean Absolute Error:

$$MAE = \frac{1}{N} \sum_{i=1}^N (|L_i - \hat{L}_i|) \quad (4)$$

Root Mean Squared Error:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (NL_{\text{true}}(i) - NL_{\text{pred}}(i))^2} \quad (5)$$

Mean Absolute Percentage Error:

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{NL_{true}(i) - NL_{pred}(i)}{NL_{true}(i)} \right| \quad (6)$$

Coefficient of Determination:

$$R^2 = 1 - \frac{\sum_{i=1}^N (NL_{true}(i) - NL_{pred}(i))^2}{\sum_{i=1}^N (NL_{true}(i) - NL_{mean})^2} \quad (7)$$

Forecast accuracy is evaluated on the test set using standard regression metrics: mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2).

3.3 Regime-Based Error Decomposition

To identify operating conditions that dominate forecasting difficulty, errors are decomposed by demand regime. Hourly load ramps are defined as

$$\Delta y_t = y_t - y_{t-1} \quad (8)$$

Ramp magnitude is given by,

$$|\Delta y_t| = |y_t - y_{t-1}| \quad (9)$$

Let $q_{0.90} = Q_{0.90}(|y_t - y_{t-1}|)$ denote the 90th percentile of absolute ramp magnitudes. Ramp regimes are defined as ramp-up events when

$$\Delta y_t \geq q_{0.90} \quad (10)$$

and ramp-down events when;

$$\Delta y_t \leq -q_{0.90} \quad (11)$$

Peak and valley operating conditions are identified using load percentiles, with peak hours defined by

$$y_t \geq Q_{0.90}y(t) \quad (12)$$

and valley hours by ;

$$y_t \leq Q_{0.10}y(t) \quad (13)$$

For each regime, MAE is computed separately, enabling a regime-aware assessment of forecasting robustness beyond aggregate accuracy metrics.

3.4 Deep Learning Models

Three architectures are evaluated: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and a hybrid Transformer–LSTM model.

3.4.1 Deep Learning Models and Conceptual Schematics

The selection of these architectures is motivated by their complementary inductive biases for modelling short-lag temporal dependence, nonlinear dynamics, and robustness under rapidly changing operating conditions.

The schematics in Figure 1 and Figure 2 illustrate functional data flow (inputs $\rightarrow$ temporal representation $\rightarrow$ output layer) and are intended as conceptual diagrams rather than implementation-level block diagrams.

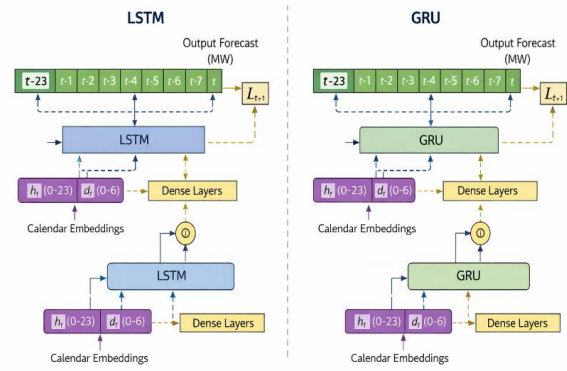


Figure 1: LSTM and GRU model architectures

Fig. 1: Conceptual architectures of the LSTM and GRU forecasting models

Source: created by the authors.

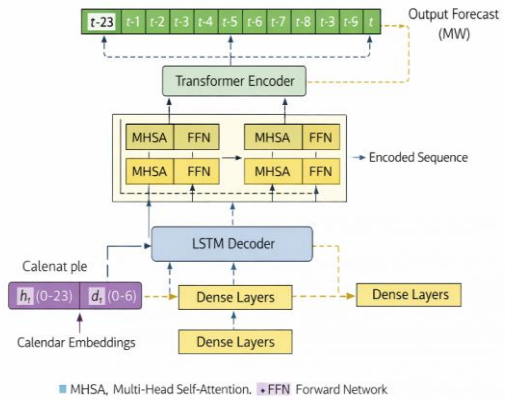


Figure 2: Transformer–LSTM hybrid model architecture

Fig. 2: Hybrid Transformer–LSTM model architecture

Source: created by the authors.

The Transformer encoder extracts context-aware temporal representations using self-attention, which are subsequently passed to an LSTM decoder to preserve short-lag sequential structure relevant for one-hour-ahead forecasting.

3.4.2 Recurrent Neural Network Models (LSTM and GRU)

The LSTM and GRU models are implemented as baseline recurrent architectures due to their proven effectiveness in short-term load forecasting and their ability to capture short-lag temporal dependence, which is theoretically dominant at the one-hour forecasting horizon. Both models process the historical input sequence X_t sequentially and generate a latent representation summarizing recent demand dynamics. The forecasting task is defined as a direct one-step-ahead mapping. The definition of the input

window X_t is given in equations (1), (2), (3). After processing the sequence, the last hidden state of the recurrent network is connected to a fully connected output layer that produces the forecast. The LSTM network uses memory cells and gating operations to control how information is stored and transferred between time steps, which helps preserve relevant past values and alleviates vanishing gradient issues. The GRU has a similar structure, but it relies on a simpler gating scheme with fewer parameters, making the model more compact while still able to represent temporal dependencies.

This reduced complexity can improve numerical stability and generalization, particularly in small power systems where demand variability is limited. Calendar information (hour-of-day and day-of-week) is incorporated via trainable embedding vectors and concatenated with the recurrent representation before the output layer. This allows the models to jointly learn short-term persistence and recurring daily and weekly demand patterns.

3.4.3 Hybrid Transformer–LSTM Model

To assess whether attention mechanisms improve robustness under dynamic operating conditions, a hybrid **Transformer–LSTM** architecture is implemented. In this model, a Transformer encoder is first applied to the input sequence X_t to learn context-aware representations, followed by an LSTM decoder that refines local temporal dynamics. The Transformer encoder is based on the scaled dot-product self-attention mechanism:

$$Attention(Q, K, V) = softmax\left(\frac{Q K^T}{(d_k)^{\frac{1}{2}}}\right) \quad (14)$$

and multi-head attention aggregation,

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W^O \quad (15)$$

where multiple attention heads enable the model to attend to different temporal aspects of the input sequence. Residual connections and layer normalization are applied to stabilize training,

$$head_i = Attention(Q W_i^Q, K W_i^K, V W_i^V) \quad (16)$$

followed by a position-wise feed-forward network,

$$Z_1 = LayerNorm(X + MultiHead(X, X, X)) \quad (17)$$

followed by a position-wise feed-forward network,

$$FFN(Z_1) = \max(0, Z_1 W_1 + b_1) W_2 + b_2 \quad (18)$$

and a second residual normalization step,

$$H = LayerNorm(Z_1 + FFN(Z_1)) \quad (19)$$

Equations, (16), (17), (18), (19) present the formulation of the Transformer encoder block employed in this study. The self-attention operation enables the model to account for temporal relationships over the entire input sequence, and the use of residual connections combined with layer normalization improves training stability. The encoder output H is subsequently passed to an LSTM decoder to preserve short-lag sequential dynamics relevant for one-hour-ahead forecasting. All projection matrices follow standard Transformer dimensionality and are learned during training, [30], [31].

3.4.4 Training Strategy and Numerical Stability

All models are trained using a direct one-step-ahead forecasting strategy, minimizing the mean squared error loss. This approach avoids recursive error propagation and ensures that observed forecasting errors reflect representation quality rather than multi-step accumulation effects. Training is performed using the Adam optimizer with early stopping based on validation loss, [51]. The models are implemented using the PyTorch deep learning framework, [52], while data preprocessing and evaluation utilities are handled using Scikit-learn, [53]. Fixed random seeds and consistent preprocessing are used to improve reproducibility. Model performance is evaluated exclusively on a held-out test set using the metrics defined in (4), (5), (6) and (7).

3.4.5 Theoretical Basis

The LSTM and GRU architectures follow their original formulations introduced in [25], [26]. The attention mechanism and Transformer encoder are based on the framework proposed in [29], and subsequent time-series forecasting adaptations [30], [31]. The hybrid Transformer–LSTM structure reflects the hypothesis that attention mechanisms may enhance robustness during rapid load transitions, while recurrent structures remain well-suited for capturing dominant short-lag dependence in one-hour-ahead load forecasting.

3.4.6 Methodological Workflow for One-Hour-Ahead Load Forecasting

Overall workflow of the proposed one-hour-ahead short-term load forecasting framework. Hourly electricity demand data from the ENTSO-E Transparency Platform and temperature data from ERA5-Land are pre-processed and combined through feature engineering, including lagged load variables and calendar embeddings. The resulting input representation is used to train deep learning models (LSTM, GRU, and Transformer–LSTM), and

forecasting performance is evaluated using standard accuracy metrics and regime-based error analysis. Figure 3 illustrates the overall workflow of the proposed short-term load forecasting framework, including data acquisition, pre-processing, feature engineering, model training, and evaluation.

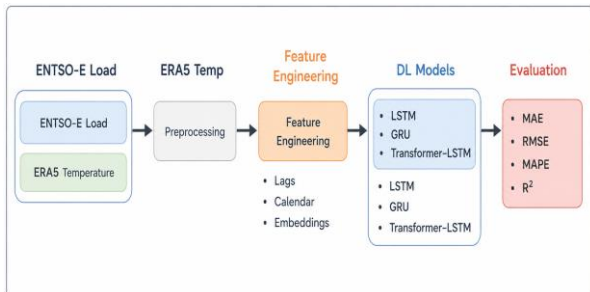


Fig. 3: Overall workflow of the proposed short-term load forecasting framework, including data acquisition, preprocessing, feature engineering, deep learning model training, and performance evaluation
Source: created by the authors.

The workflow concludes with the evaluation of model performance using standard regression metrics.

4 Results

4.1 Benchmark Forecasting Performance

Before analyzing the performance of deep-learning models, benchmark forecasting behavior is examined to contextualize the short-term forecasting challenge in the Albanian power system. In addition to the proposed models, two reference forecasts are considered.

The first is a persistence forecast obtained from the most recent load observation, while the second corresponds to the day-ahead load forecast reported by ENTSO-E.

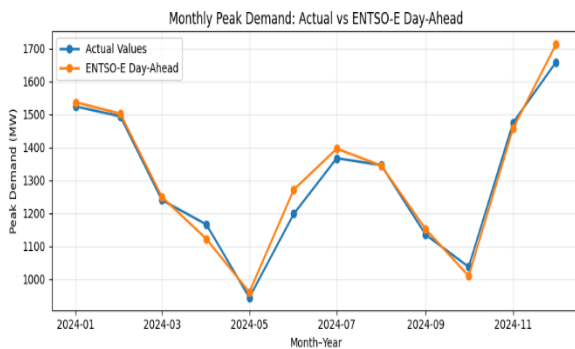


Fig. 4: Actual vs ENTSO-E Day Ahead Forecast
Source: created by the authors.

Figure 4 shows a comparison between the measured hourly electricity demand, the persistence forecast, and the ENTSO-E day-ahead prediction. Although the overall daily trend is represented reasonably well by both reference forecasts, differences appear near peak and low-load periods, indicating limited accuracy at short operational horizons. The mismatch becomes larger during highly variable periods, when sudden changes in demand cannot be captured by the baseline approaches. Table 1, summarizes the benchmark forecasting performance for the one-hour-ahead forecasting horizon.

Table 1. Main benchmark comparison (1h-ahead)

Model	MAE (MW)	RMSE (MW)	MAPE (%)	R ²
Persistence (t-1)	75.954	100.824	8.139	0.8760
ENTSO-E Day-Ahead	42.194	55.972	4.065	0.9618

Source: created by the authors.

The persistence forecast shows relatively large errors (MAE = 75.95 MW, RMSE = 100.82 MW), since it cannot account for short-term variations in demand. The ENTSO-E day-ahead forecast provides a clear improvement in accuracy (MAE = 42.19 MW, RMSE = 55.97 MW) and reproduces the general demand evolution more reliably, but it still has difficulty capturing extreme conditions and fast load changes. These results indicate that, although day-ahead forecasts are useful for overall system planning, they are not designed for the very short forecasting horizons required for real-time balancing and operational decision-making. For this reason, data-driven deep-learning models are investigated as a way to improve one-hour-ahead forecasting accuracy and to obtain more robust predictions under dynamic operating conditions.

4.2 Overall Forecasting Accuracy and Temperature Ablation

This subsection evaluates the overall one-hour-ahead forecasting accuracy of the deep learning models and examines the impact of temperature inclusion through a controlled ablation analysis. Table 2 reports the test-set results for the LSTM, GRU, and Transformer-LSTM models evaluated under two input configurations: with temperature and without temperature.

Table 2. Metrics Results

Model	Setting	MAE (MW)	RMSE (MW)	MAPE (%)	R ²
LSTM	With temp	18.49	24.17	1.83	0.9929
LSTM	Without temp	15.38	21.53	1.47	0.9944
GRU	With temp	16.89	22.03	1.73	0.9941
GRU	Without temp	17.48	24.20	1.69	0.9929
Transformer-LSTM	With temp	17.12	23.10	1.72	0.9935
Transformer-LSTM	Without temp	17.27	24.32	1.67	0.9928

Source: created by the authors.

All deep-learning models show very high forecasting accuracy, with mean absolute percentage error (MAPE) below 2% and coefficients of determination greater than 0.99, confirming a very close match between predicted and measured load values. The comparison also shows that adding additional input variables does not always improve short-term forecasting performance. This effect is particularly evident in small power systems, where load persistence and calendar-related patterns play a dominant role at short forecasting horizons. While temperature is an important driver of electricity demand at longer timescales, its contribution to one-hour-ahead forecasting accuracy is secondary and architecture-dependent in the present case.

4.3 Time-Series Forecasting Performance

To complement the aggregate accuracy metrics reported in Section IV.B, the time-series behavior of the evaluated deep-learning models is examined through visual inspection of their one-hour-ahead forecasts. This analysis provides insight into how well the models track the temporal evolution of electricity demand under both stable and rapidly changing conditions. Figure 5 shows the hourly electricity load over a representative multi-day interval for the Albanian power system.

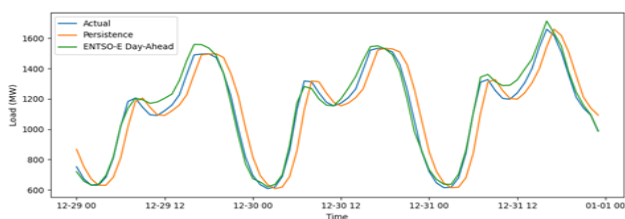


Fig. 5: Hourly electricity load over a representative multi-day period

Source: created by the authors.

A clear daily pattern can be observed, with recurring peaks during daytime hours and lower demand during the night, together with rapid transitions between different operating conditions. Such sudden variations make short-term load forecasting more difficult and require models that are able to capture short-term temporal dependencies. Figure 6 compares the daily peak electricity demand during the test period with the corresponding forecasts obtained using the LSTM, GRU, and Transformer-LSTM models.

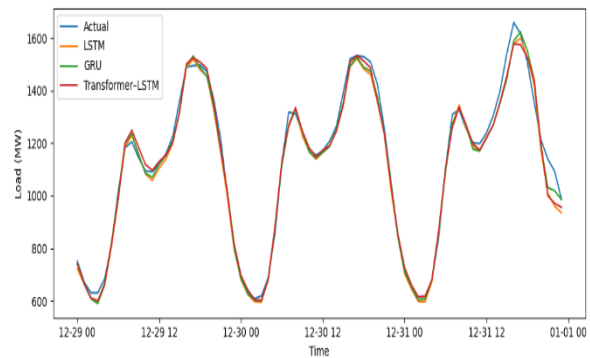


Fig. 6: Daily peak electricity demand during the test period, comparing actual observations with forecasts obtained using LSTM, GRU, and LSTM-Transformer models

Source: created by the authors.

All three models closely follow the observed evolution of peak demand, successfully capturing both the timing and magnitude of most daily maxima. Deviations are primarily observed during days with higher variability, indicating that forecasting errors are concentrated during dynamic operating conditions rather than during steady demand periods. Figure 7 shows the predicted and observed load values over the entire test period, together with a zoomed interval that makes it possible to examine both long-term tracking and short-term variations.

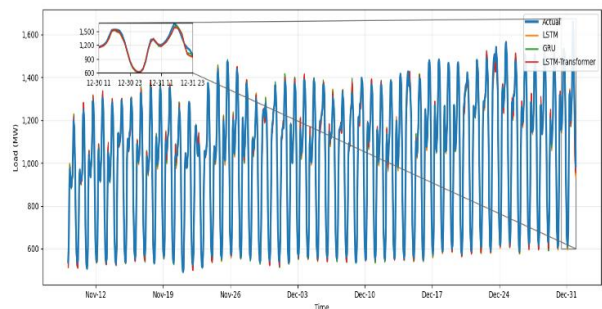


Fig. 7: Time-series forecasting behavior of the evaluated deep-learning models for the 1-hour-ahead horizon

Source: created by the authors.

Over the full horizon, all models follow the measured load profile closely, indicating stable performance and a good ability to reproduce the main temporal patterns of the demand signal. The zoomed-in view confirms that short-term variations and peak timing are reproduced with only minor discrepancies, particularly during rapid transitions. To further illustrate performance across multiple temporal scales, Figure 8 shows the forecasting behaviour of the GRU model for the full test period and for randomly selected subsets corresponding to 10% and 1% of the test samples.

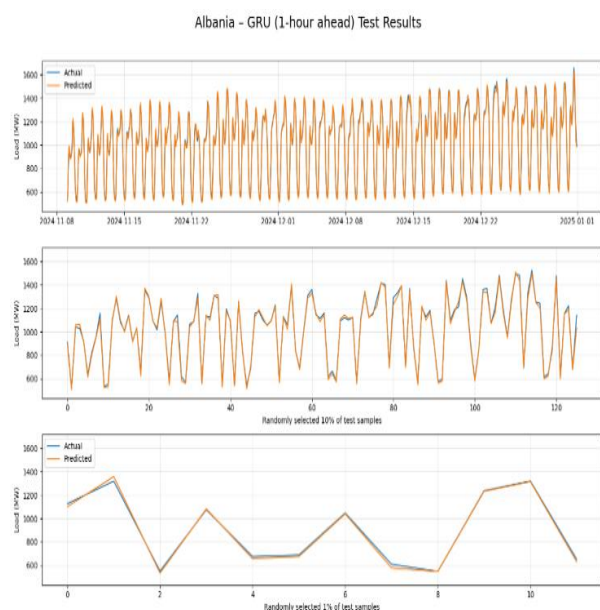


Fig. 8: Hourly electricity load forecasting results for Albania using the GRU model with a 1-hour-ahead horizon

Source: created by the authors.

The results show that the model preserves similar accuracy at different sampling resolutions, which suggests stable behaviour both over long periods and during short local variations. The top panel corresponds to the full test interval, while the middle and bottom panels show randomly selected subsets containing 10% and 1% of the samples. This representation makes it possible to evaluate model performance at different time scales.

In general, the time-series analysis shows that all tested deep-learning models follow the main diurnal pattern of electricity demand and produce reliable one-hour-ahead predictions under typical operating conditions.

The remaining discrepancies are mainly associated with periods of rapid load change, suggesting that forecasting difficulty is driven more by dynamic transitions than by absolute demand levels. These

observations motivate the subsequent analysis of peak-demand tracking and regime-dependent forecasting errors.

4.4 Peak-Demand Tracking

Accurate prediction of peak electricity demand is of particular operational importance, as peak errors directly affect reserve allocation, balancing actions, and system security. To assess model performance under these critical conditions, peak-demand tracking behavior is examined using weekly peak demand statistics over the test period. Figure 9 illustrates the weekly peak load over time and the corresponding predictions provided by the LSTM, GRU, and Transformer-LSTM models.

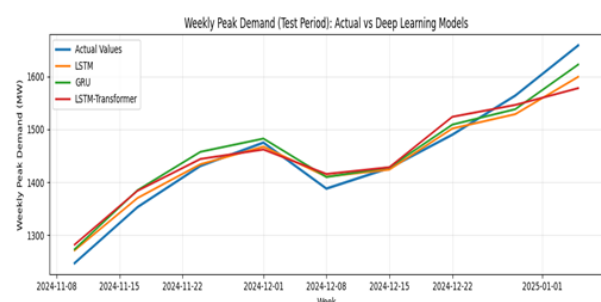


Fig. 9: Weekly peak load during the test period, comparing observed demand with predictions from the LSTM, GRU, and Transformer-LSTM models
Source: created by the authors.

All models successfully capture the overall trend and seasonal increase in peak demand toward the end of the year. However, differences between architectures become more evident at the weekly aggregation level, where deviations from observed peaks are more clearly distinguishable than in hourly time-series plots. Table 3 quantifies peak-tracking performance using the mean and maximum absolute weekly peak error.

Table 3. Weekly Peak Error Statistics (Test Period)

Model	Mean Peak Error (MW)	Max Peak Error (MW)
LSTM	20.46	59.45
GRU	21.84	36.25
LSTM-Transformer	28.17	80.96

Source: created by the authors.

The LSTM model shows the lowest mean absolute peak error (20.46 MW), which corresponds to the best average peak-tracking performance over the evaluated weeks. The GRU model provides a similar average

error (21.84 MW), but it behaves more reliably in extreme cases, with the smallest maximum absolute peak error (36.25 MW). In contrast, the Transformer-LSTM model shows a higher mean absolute peak error (28.17 MW) and the largest maximum deviation (80.96 MW), suggesting reduced reliability for extreme peak prediction under the present evaluation setup. These results suggest that average accuracy alone does not fully describe the operational performance of the forecasting models. Although the LSTM provides the best results on average, the GRU model shows better robustness during extreme peak conditions, which are often the most critical for system operation. The Transformer-LSTM model remains competitive in terms of overall accuracy, but it shows larger variations in peak prediction, indicating a trade-off between modeling flexibility and reliability in peak tracking. In general, the peak-demand analysis shows that the main differences between the forecasting architectures appear under extreme operating conditions rather than in the aggregated error metrics.

4.5 Regime-Based Error Decomposition

While aggregate accuracy metrics provide a useful summary of overall forecasting performance, they do not reveal when and under which operating conditions forecasting errors occur. To address this limitation, forecasting errors are decomposed by operating regime using the percentile-based ramp definitions. Hourly load ramps are classified into ramp-up, ramp-down, peak, and valley regimes based on load changes and load percentiles. This regime-aware analysis enables identification of operating conditions that dominate forecasting difficulty and highlights architecture-dependent robustness that is not visible in average performance metrics. Table 4 summarizes the MAE values obtained for each model under the different operating regimes.

Table 4. The Regime-Dependent Forecasting Errors

Operating regime	LSTM/MAE (MW)	GRU/MAE (MW)	Transformer-LSTM/MAE (MW)
Peak Hours	24.0	22.0	23.0
Valley Hours	12.4	14.0	11.3
Ramp-up	16.2	15.4	15.6
Ramp-down	12.1	8.6	8.9

Source: created by the authors.

The most challenging conditions are associated with ramping events, where rapid changes in demand occur over consecutive hours. During ramp-up periods, all models exhibit elevated errors, with MAE values in the range of approximately 15–16 MW, indicating comparable sensitivity to sudden demand increases. In contrast, more pronounced differences emerge during ramp-down events. The GRU and Transformer-LSTM models achieve substantially lower MAE (8.6 MW and 8.9 MW, respectively) compared with the LSTM model (12.1 MW), demonstrating improved robustness when demand decreases sharply. Figure 10 shows that errors are lowest during valley hours (approximately 11–14 MW across models) and increase during peak hours (approximately 22–24 MW).

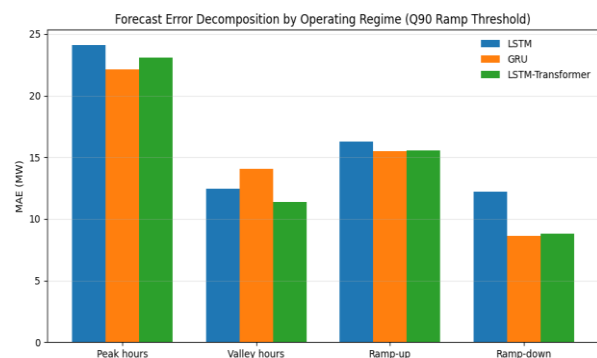


Fig. 10: MAE by operating regime (Q90 ramp threshold)

Source: created by the authors.

Overall, the regime-based decomposition reveals that forecasting difficulty is primarily driven by ramping dynamics, not by high or low demand levels per se. Moreover, the results demonstrate that gated and attention-enhanced architectures provide improved robustness during ramp-down events, a property that is not captured by aggregate accuracy metrics. This finding highlights the importance of regime-aware evaluation frameworks for short-term load forecasting in small power systems and provides actionable insight for operational model selection.

4.6 Representative Segment Comparisons

To verify forecasting stability across different operating intervals, model performance is illustrated using representative test segments. The segment-based comparisons confirm that all evaluated models consistently reproduce the dominant diurnal load pattern across multiple weeks, with deviations primarily occurring during peak and ramping periods. Differences between architectures are subtle at the hourly level and become more apparent only under dynamic conditions, consistent with the regime-based

analysis presented earlier. Figure 11 illustrates representative test segments comparing model forecasts with observed load values.

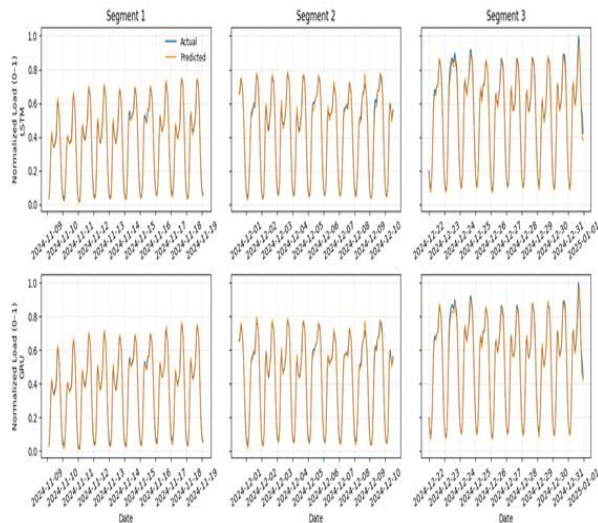


Fig. 11: Representative segment comparison of one-hour-ahead load forecasts obtained using the LSTM and GRU models

Source: created by the authors.

All models reproduce the dominant diurnal pattern across different operating intervals, with **small** deviations mainly occurring during peak and ramping periods. The comparison confirms stable behaviour across time and supports the robustness conclusions drawn from the aggregate and regime-based analyses.

5 Conclusion

The results demonstrate that deep-learning models provide accurate one-hour-ahead electricity load forecasts for the Albanian power system, while also revealing limitations of conventional evaluation practices based solely on aggregate accuracy metrics. Although all evaluated architectures achieve very low overall errors, meaningful performance differences emerge only when forecasting behavior is examined under specific operating conditions. The temperature ablation analysis shows that explicit temperature inputs do not consistently improve short-horizon forecasting accuracy and may even degrade performance for certain architectures, such as the LSTM. This confirms that, at the one-hour horizon, recent load history and calendar effects dominate the predictive information content in small power systems. The contribution of temperature is therefore secondary and architecture-dependent, highlighting the importance of context-aware feature selection rather than indiscriminate input expansion. A key insight of this study is that forecasting errors are

primarily driven by **dynamic load transitions**, rather than by high or low demand levels alone. Regime-based error decomposition reveals that ramping periods dominate forecasting difficulty, a behaviour that is not captured by standard accuracy metrics. During ramp-down events, the GRU and Transformer-LSTM models behave more robustly than the LSTM, which suggests that gated and attention-based architectures handle asymmetric temporal dynamics more effectively. These results also show that choosing a forecasting model based only on average accuracy may be misleading, since robustness during critical situations, such as peak demand or rapid load changes, is equally important for real operation. The proposed regime-aware evaluation framework provides a practical and transferable approach for assessing short-term forecasting models in small and hydro-dominated power systems.

The main findings from the results can be summarized as follows:

1. All deep-learning models achieve high one-hour-ahead forecasting accuracy for the Albanian power system.
2. Temperature inclusion does not consistently improve short-horizon accuracy, and its impact is model-dependent.
3. Aggregate accuracy metrics mask operational risk; peak-demand and worst-case behavior reveal meaningful differences between models.
4. Forecasting errors are dominated by ramping dynamics rather than by absolute demand levels.
5. GRU and Transformer-LSTM models exhibit improved robustness during ramp-down events, a behavior revealed only through regime-based evaluation.

Overall, this work demonstrates that reliable short-term load forecasting requires regime-aware evaluation frameworks and that model selection in small, hydro-dominated power systems must prioritize robustness under dynamic operating conditions rather than average accuracy alone.

Acknowledgement:

The authors thank the National Agency for Scientific Research and Innovation (AKKSHI) of Albania for financial support of this work under Contract No. Prot. 944, dated 20 June 2025, concluded between AKKSHI and the Polytechnic University of Tirana. The support provided by this project made possible the data collection, analysis, and preparation of the present study.

References:

- [1] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, 2020. Available: <https://doi.org/10.1109/OAJPE.2020.3029979>
- [2] European Commission, "Commission Regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing," *Official Journal of the European Union*, 2017. Available: <https://eur-lex.europa.eu/eli/reg/2017/2195/oj> (Accessed: September 3, 2025).
- [3] European Commission, "Commission Regulation (EU) 2017/1485 of 2 August 2017 establishing a guideline on electricity transmission system operation," *Official Journal of the European Union*, 2017. Available: <https://eur-lex.europa.eu/eli/reg/2017/1485/oj> (Accessed: September 3, 2025).
- [4] International Energy Agency, *World Energy Outlook 2023*. Paris, France: IEA, 2023. Available: <https://www.iea.org/reports/world-energy-outlook-2023> (Accessed: September 10, 2025).
- [5] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021, ISBN: 978-0987507112. Available: <https://otexts.com/fpp3/> (Accessed: September 10, 2025).
- [6] Energy Regulatory Authority (ERE), *Annual Report 2023*. Tirana, Albania: ERE, 2024. Available: https://ere.gov.al/media/files/2024/10/01/Annual_Report_2023.pdf (Accessed: October 1, 2025).
- [7] Energy Community Secretariat, *Annual Implementation Report 2023*. Vienna, Austria: Energy Community Secretariat, 2023. Available: <https://www.energy-community.org/publications/flagships/IRs.html> (Accessed: October 3, 2025).
- [8] European Commission, *Albania 2023 Report*, Commission Staff Working Document, Brussels, Belgium, 2023. Available: https://neighbourhood-enlargement.ec.europa.eu/albania-report-2023_en (Accessed: May 3, 2025).
- [9] ENTSO-E, *Transparency Platform*. Available: <https://transparency.entsoe.eu/> (Accessed: October 1, 2025).
- [10] ENTSO-E, *Transparency Platform Manual of Procedures (MoP)*, Version 3 Release 4, Feb. 2021. Available: <https://www.entsoe.eu/data/transparency-platform/mop/> (Accessed: October 12, 2025).
- [11] European Commission, "Commission Regulation (EU) No 543/2013 on submission and publication of data in electricity markets," *Official Journal of the European Union*, 2013. Available: <https://eur-lex.europa.eu/eli/reg/2013/543/oj> (Accessed: September 5, 2025).
- [12] Copernicus Climate Change Service (C3S), *ERA5-Land hourly data from 1950 to present*. Copernicus Climate Data Store (CDS), 2025. Available: <https://cds.climate.copernicus.eu/> (Accessed: October 10, 2025).
- [13] J. Muñoz-Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Bousssetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut, "ERA5-Land: A state-of-the-art global reanalysis dataset for land applications," *Earth System Science Data*, vol. 13, no. 9, pp. 4349–4383, 2021. Available: <https://doi.org/10.5194/essd-13-4349-2021>
- [14] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020. Available: <https://doi.org/10.1002/qj.3803>
- [15] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2015, ISBN: 978-1118675021. Available: <https://www.wiley.com/en-gb/shop/general-introductory-statistics/time-series-analysis->

- [forecasting-and-control-5th-edition-p-9781118675021](#) (Accessed: September 5, 2025).
- [16] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," *International Journal of Forecasting*, vol. 18, no. 3, pp. 439–454, Jul.–Sep. 2002, [https://doi.org/10.1016/S0169-2070\(01\)00110-8](https://doi.org/10.1016/S0169-2070(01)00110-8)
- [17] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, 2001. <https://doi.org/10.1109/59.910780>
- [18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995. <https://doi.org/10.1007/BF00994018>
- [19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. <https://doi.org/10.1023/A:1010933404324>
- [20] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001. <https://doi.org/10.1214/aos/1013203451>
- [21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
- [22] L. Schmid, A. Gerharz, A. Groll, and M. Pauly, "Tree-based ensembles for multi-output regression: Comparing multivariate approaches with separate univariate ones," *Computational Statistics & Data Analysis*, vol. 183, Art. no. 107628, 2023. <https://doi.org/10.1016/j.csda.2022.107628>
- [23] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. ISBN: 978-0262035613. Available: <https://mitpress.mit.edu/9780262035613/dee-p-learning/> (Accessed: June 30, 2025).
- [24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [25] K. Cho, B. van Merriënboer, C. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1724–1734. <https://doi.org/10.3115/v1/D14-1179>
- [26] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016. <https://doi.org/10.1016/j.ijforecast.2015.11.011>
- [27] J. Dong, Y. Jiang, P. Chen, J. Li, Z. Wang, and S. Han, "Short-term power load forecasting using bidirectional gated recurrent units-based adaptive stacked autoencoder," *International Journal of Electrical Power & Energy Systems*, vol. 172, Art. no. 110459, 2025. <https://doi.org/10.1016/j.ijepes.2025.110459>
- [28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>
- [29] B. Lim, S. Ö. Arik, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- [30] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, pp. 11106–11115, 2021. <https://doi.org/10.1609/aaai.v35i12.17325>
- [31] J. Zhao, F. Chu, L. Xie, Y. Che, Y. Wu, and A. F. Burke, "A survey of transformer networks for time series forecasting," *Computer Science Review*, vol. 60, Art. no. 100883, 2026. <https://doi.org/10.1016/j.cosrev.2025.100883>
- [32] T. Ciechulski and S. Osowski, "High precision LSTM model for short-time load forecasting in power systems," *Energies*, vol. 14, no. 11, Art. no. 2983, 2021. <https://doi.org/10.3390/en14112983>
- [33] Z. Zhao, C. Xia, L. Chi, X. Chang, W. Li, T. Yang, and A. Y. Zomaya, "Short-term load

- forecasting based on the transformer model,” Information, vol. 12, no. 12, Art. no 516, 2021. <https://doi.org/10.3390/info12120516>
- [34] N. B. Behmiri, C. Fezzi, and F. Ravazzolo, “Incorporating air temperature into mid-term electricity load forecasting models using time-series regressions and neural networks,” Energy, vol. 278, Art. no. 127831, 2023. <https://doi.org/10.1016/j.energy.2023.127831>
- [35] P. Bessec and J. Fouquau, “The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach,” Energy Economics, vol. 30, no. 5, pp. 2705–2721, 2008. <https://doi.org/10.1016/j.eneco.2008.02.003>
- [36] R. J. Hyndman and A. B. Koehler, “Another look at measures of forecast accuracy,” International Journal of Forecasting, vol. 22, no. 4, pp. 679–688, 2006. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- [37] C. Tofallis, “A better measure of relative prediction accuracy for model selection and model estimation,” Journal of the Operational Research Society, vol. 66, no. 8, pp. 1352–1362, 2015. <https://doi.org/10.1057/jors.2014.103>
- [38] F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy,” Journal of Business & Economic Statistics, vol. 20, no. 1, pp. 134–144, 2002. <https://doi.org/10.1198/073500102753410444>
- [39] C. Bergmeir, R. J. Hyndman, and B. Koo, “A note on the validity of cross-validation for evaluating autoregressive time series prediction,” Computational Statistics & Data Analysis, vol. 120, pp. 70–83, 2018. <https://doi.org/10.1016/j.csda.2017.11.003>
- [40] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, May 7–9, 2015. Available: <https://arxiv.org/abs/1412.6980> (Accessed: September 20, 2025).
- [41] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and machine learning forecasting methods: Concerns and ways forward,” PLOS ONE, vol. 13, no. 3, pp. e0194889, 2018. <https://doi.org/10.1371/journal.pone.0194889>
- [42] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” Journal of Machine Learning Research, vol. 13, pp. 281–305, 2012. Available: <http://jmlr.org/papers/v13/bergstra12a.html> (Accessed: September 3, 2025).
- [43] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, “DeepAR: Probabilistic forecasting with autoregressive recurrent networks,” International Journal of Forecasting, vol. 36, no. 3, pp. 1181–1191, 2020. <https://doi.org/10.1016/j.ijforecast.2019.07.001>
- [44] S. Smyl, “A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting,” International Journal of Forecasting, vol. 36, no. 1, pp. 75–85, Jan.–Mar. 2020. <https://doi.org/10.1016/j.ijforecast.2019.03.017>
- [45] H. Hewamalage, C. Bergmeir, and K. Bandara, “Recurrent neural networks for time series forecasting: Current status and future directions,” International Journal of Forecasting, vol. 37, no. 1, pp. 388–427, Jan.–Mar. 2021. <https://doi.org/10.1016/j.ijforecast.2020.06.008>
- [46] M. Liu, X. Sun, Q. Wang, and S. Deng, “Short-term load forecasting using EMD with feature selection and TCN-based deep learning model,” Energies, vol. 15, no. 19, p. 7170, 2022. <https://doi.org/10.3390/en15197170>
- [47] R. G. Allwyn, A. Al-Hinai, and V. Margaret, “A comprehensive review on energy management strategy of microgrids,” Energy Reports, vol. 9, pp. 5565–5591, 2023. <https://doi.org/10.1016/j.egyr.2023.04.360>
- [48] N. Khosravi, D. Çelik, H. Bevrani, and S. Echalih, “Microgrid stability: A comprehensive review of challenges, trends, and emerging solutions,” International Journal of Electrical Power & Energy Systems, vol. 172, Art. no. 110829, 2025. <https://doi.org/10.1016/j.ijepes.2025.110829>
- [49] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, “ARIMA models to predict next-day electricity prices,” IEEE Transactions on Power Systems, vol. 18, no. 3, pp. 1014–1020, 2003. <https://doi.org/10.1109/TPWRS.2002.804943>

- [50] H. Zhang, R. Zandehshahvar, M. Tanneau, and P. Van Hentenryck, "Weather-informed probabilistic forecasting and scenario generation in power systems," *Applied Energy*, vol. 384, Art. no. 125369, 2025. <https://doi.org/10.1016/j.apenergy.2025.125369>
- [51] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proceedings of the 7th International Conference on Learning Representations (ICLR)*, New Orleans, LA, USA, May 6–9, 2019. OpenReview.net. Available: <https://openreview.net/forum?id=Bkg6RiCqY7> (Accessed: September 20, 2025).
- [52] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, pp. 8024–8035. <https://doi.org/10.48550/arXiv.1912.01703>
- [53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. VanderPlas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. Available: <https://www.jmlr.org/papers/v12/pedregosa11a.html> (Accessed: September 20, 2025). The authors equally contributed to the present research, at all stages from the formulation of the problem to the final findings and solution.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed to the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

National Agency for Scientific Research and Innovation (AKKSHI)



Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US