cigre Turkey

TEPES
TURKISH JOURNAL OF ELECTRICAL POWER AND ENERGY SYSTEMS

TEPES, Vol. 5, Issue. 3, 152-160, 2025 **DOI:** 10.5152/tepes.2025.25029

RESEARCH ARTICLE

Hybrid Deep Learning Framework for Short-Term Electricity Generation Forecasting in Türkiye Using Multi-Source Data

Hamdullah Karamollaoğlu

Ministry of Energy and Natural Resources, Electricity Generation Company, Ankara, Türkiye

Cite this article as: H. Karamollaoğlu, "Hybrid deep learning framework for short-term electricity generation forecasting in Türkiye using multi-source data," *Turk J Electr Power Energy Syst.*, 2025; 5(3), 152-160.

ABSTRACT

The growing complexity of electricity generation, driven by the diversification of energy sources and the integration of renewables, makes accurate short-term forecasting crucial for grid stability and energy security. This study proposes a deep learning-based hybrid forecasting model designed for Türkiye's dynamic energy landscape. Using hourly electricity production data from December 1, 2019, to March 1, 2025, sourced from the EPİAŞ Transparency Platform, the model analyzes generation patterns across 17 different sources, including both fossil fuels and renewables. The proposed architecture combines Long Short-Term Memory networks and Transformer models to effectively capture complex time-dependent relationships in electricity generation. To improve accuracy, preprocessing techniques such as time-based interpolation, normalization, and principal component analysis were applied. Experimental results demonstrate strong forecasting performance, achieving a mean absolute error of 589.50, a root mean squared error of 762.41, and a coefficient of determination (R^2) of 0.98017 for 1-hour ahead predictions, and an R^2 of 0.87813 for 1-day ahead predictions. These findings underline the model's potential to support operational planning, market regulation, and policy-making processes, particularly in emerging economies with dynamic and heterogeneous energy infrastructures.

Index Terms—Deep learning, electricity generation, short-term forecasting

I. INTRODUCTION

Electricity is essential for modern life and plays a vital role in economic development and social welfare [1]. Global electricity demand continues to rise due to industrialization, urbanization, and digitalization, requiring reliable and sustainable energy planning. Maintaining the balance between generation and consumption is strategically important for supply security and market stability [2, 3], and short-term generation forecasting is key for supply-demand equilibrium, market management, and operational planning [4, 5].

The increasing share of renewable sources, such as wind and solar, introduces variability and uncertainty, challenging system operators in load balancing, reserve management, and supply planning [6, 7]. These uncertainties pose both technical and economic risks, making accurate and reliable generation forecasts essential. In Türkiye, rising electricity demand is driven by population growth, industrialization, and infrastructure investments [7, 8]. The energy mix—comprising fossil fuels, hydro, solar, wind, and other alternatives—creates a technically and economically complex system requiring effective management [9–11]. Accurate modeling of temporal variability across sources is crucial for reliable total generation forecasts.

This study forecasts Türkiye's short-term total electricity generation using hourly data from 17 sources during 2019–2025, obtained from EPİAŞ [12]. Predictable fossil fuels provide baseload generation, while wind and solar introduce variability. Other sources like geothermal, biomass, and waste heat are relatively stable. The concurrent use of diverse sources creates a nonlinear generation profile [13–15], limiting traditional statistical models and motivating AI-based approaches.

A hybrid model combining Long Short-Term Memory (LSTM) and Transformer architectures is employed. LSTM captures short- and long-term dependencies, while the Transformer leverages attention mechanisms for time-independent information flow, enhancing prediction accuracy [16, 17].

The remainder of this paper is organized as follows: Section 2 reviews related literature, Section 3 details methodology. Section 4 presents

Received: July 16, 2025
Revision Requested: July 21, 2025
Last Revision Received: September 7, 2025
Accepted: September 9, 2025
Publication Date: October 20, 2025

Corresponding author: Hamdullah Karamollaoğlu **E-mail:** h.karamollaoglu@euas.gov.tr

experimental results and performance analysis, and Section 5 concludes with insights and future research directions.

Related Works

Accurate forecasting of electricity generation is of critical importance for ensuring effective energy management and maintaining the balance between supply and demand in modern power systems. In recent years, numerous studies have explored the integration of machine learning (ML) and deep learning (DL) models with various optimization algorithms to enhance prediction performance.

Li et al. [18] forecasted Türkiye's net electricity consumption by combining ML algorithms such as XGBoost and CatBoost with optimization techniques including Sparrow Search Algorithm (SSA), Phasor Particle Swarm Optimization (PPSO), and Hybrid Grey Wolf Optimization (GWO). Among these, XGBoost-SSA achieved the highest accuracy, emphasizing the impact of GDP, transmitted energy, and trade variables on predictive performance. For short-term load forecasting, Fan et al. [4] proposed a hybrid DL model, EWT-CNN-S-RNN + LSTM, which employed Empirical Wavelet Transform for feature extraction, CNN for spatial learning, and RNN with LSTM for temporal analysis, with hyperparameters tuned via Bayesian optimization. The model demonstrated high accuracy and strong generalization on datasets from Australia and Switzerland.

Aslam et al. [19] reviewed DL methods for forecasting power load and renewable generation in smart microgrids, evaluating models such as LSTM, Gated Recurrent Units (GRU), Inception Network, and Deep Belief Network (DBN). They highlighted that performance strongly depends on the quality and quantity of historical data and that uncertainty handling remains a notable research gap. To address limitations of conventional methods, Saxena et al. [20] proposed a hybrid KNN-SVM model, which outperformed standard techniques including LSTM in accuracy, precision, and specificity, underscoring the potential of solar energy forecasting for reliable power operations.

Wind energy forecasting challenges were reviewed by Sawant et al. [21], who emphasized its inherent variability and recommended hybrid approaches integrating multiple techniques for improved performance. Zhang et al. [22] introduced a framework combining Kolmogorov–Arnold Networks with TCN, BiLSTM, and Transformer architectures, using real-world UK data to enhance accuracy, robustness, and adaptability under economic and climatic variations. Wang et al. [23] developed CEEMDAN-SE-TR-BiGRU-Attention, decomposing complex wind signals with CEEMDAN and Sample Entropy, then processing them through Transformer-BiGRU-attention networks, achieving high accuracy across low- and high-frequency components under varying meteorological conditions.

For photovoltaic forecasting, Xiang et al. [24] proposed proposed a hybrid model combining Temporal Convolutional Networks (TCN) and GRU with an Efficient Channel Attention network (TCN-ECANet-GRU), which captures spatial and temporal features effectively, while Güldürek [25] improved short-term wind speed prediction using an Artificial Neural Network (ANN) combined with the Dragonfly Algorithm. Ibrahim et al. [26] introduced a CNN-LSTM autoencoder hybrid for short-term PV generation, showing strong performance metrics. Buratto et al. [27] addressed biomass generation variability with a Wavelet-CNN-LSTM model in Brazil, achieving a MAPE of 1.48%. Wu et al. [28] proposed the STCM model (CNN-LSTM) for ultra-short-term wind power forecasting, capturing both spatial and temporal dependencies, outperforming traditional models.

The existing literature demonstrates that hybrid ML/DL models can significantly enhance energy forecasting accuracy, particularly when integrated with optimization algorithms and signal decomposition techniques. Nevertheless, most studies are limited in scope, often focusing on a narrow set of energy sources and lacking real-world scalability. Addressing these limitations, this study proposes a comprehensive and scalable framework that forecasts electricity generation from 17 different energy sources using a novel LSTM-Transformer hybrid architecture. Leveraging real-world data and advanced preprocessing methods, the model is specifically designed to accommodate the high variability of renewable energy, offering strong practical relevance for energy planning in complex systems such as Türkiye's.

Main Points

- A hybrid Long Short-Term Memory-Transformer model is proposed for short-term electricity generation forecasting in Türkiye.
- The model uses hourly data from 17 fossil-based and renewable energy sources.
- Advanced preprocessing (interpolation, normalization, principal component analysis) enhances accuracy and efficiency.
- Real-world EPİAŞ data ensure practical and reproducible results.
- The model supports real-time energy planning and adapts to renewable energy variability.

II. METHODS

In this study, a comprehensive dataset obtained from the EPİAŞ Transparency Platform was utilized to model and forecast electricity generation in Türkiye. During the data preparation phase, preprocessing steps such as normalization and dimensionality reduction using PCA were applied. Subsequently, a hybrid DL model was developed, combining LSTM layers, which capture temporal patterns in time series data, with Transformer components that leverage attention mechanisms. The overall pseudo code of the proposed model is presented in Table I. As shown in Table I, the model architecture consists of four main stages: data preprocessing, LSTM layers, Transformer module, and output layers, all designed to enhance prediction accuracy. Detailed information regarding the pseudo code is provided in this section.

A. Dataset Description and Features

The dataset employed in this research was sourced from the EPİAŞ Transparency Platform [12] and comprises comprehensive electricity generation data pertaining to Türkiye. It spans the period from December 1, 2019, to March 1, 2025, encompassing approximately 5 years and 3 months of continuous hourly observations. Each instance within the dataset corresponds to the amount of electricity produced at a specific date and hour, providing a fine-grained

TABLE I.
PSEUDO CODE OF PROPOSED MODEL FOR ELECTRICITY GENERATION FORECASTING

Input:
$X \in R^{nx17}$: Hourly generation data (EPİAŞ)
$h \in \{1,24\}$: Prediction horizon (1 hour or 24 hours)

Output:
$\hat{Y} \in R^{nx1}$: Forecasted total generation
{MAE, RMSE, R^2} : Performance metrics

1. Data Preprocessing
$X' \leftarrow$ TimeSeriesInterpolation(X) // Handle missing values
$\hat{X} \leftarrow$ MinMaxNormalize(X') // Normalize to [0,1]
$U \leftarrow$ PCA($\hat{X}$, variance=0.99) // Dimensionality reduction ($U \in R^{nx13}$)

2. Model Architecture
$\Theta \leftarrow \{\theta_{LSTM}, \theta_{Transformer}\}$ // Trainable parameters

3. LSTM Component (Temporal Features)
For t = 1 to T:
$ht \leftarrow LSTM(u_t, h_{t-1}; \theta_{LSTM})$

4. Transformer Component (Long-Term Dependencies)
$H \leftarrow [h_1,\ldots, h_t]$
$Z \leftarrow$ MultiHeadAttention(Q=H, K=H, V=H; $\theta_{Transformer}$) // 4 heads, d_k=32

5. Prediction Head
$\hat{y}_t \leftarrow \sigma(W_{out} \cdot Pool(Z) + b_{out})$

6. Model Training
Objective: Minimize MSE($\hat{Y}$, Y)
Optimizer: Adam (lr=0.001, β_1=0.9, β_2=0.999)
Batch size: 64, TimeSeriesSplit(k=5)

7. Forecasting Procedure
for i = 1 to h:
$\hat{Y}_{t+i} \leftarrow$ Model($X_{t-w:t}$) // lookback window
If i < h: update X with $\hat{Y}_{t+i}$

end for
Return $\hat{Y}$, Evaluate($\hat{Y}$, $\hat{Y}_{test}$) // Return predictions and performance metrics

temporal resolution essential for time series modeling and analysis. The dataset consists of a total of 46 032 records, each capturing detailed production figures across a diverse range of energy sources, in addition to the overall generation amount.

The dataset represents electricity production from diverse generation technologies. "Natural Gas" corresponds to gas-fired plants, "Dam (Hydroelectric)" to large-scale dams, while "Lignite" and "Run-of-River (Hydroelectric)" reflect lignite coal and river-based hydro production. It also includes outputs from "Imported Coal," "Wind," "Solar," "Fuel Oil", "Geothermal", "Asphaltite Coal," "Hard Coal", "Biomass", "Naphtha", "LNG", and cross-border "International" imports. "Waste Heat" captures electricity from heat recovery processes. Each entry contains a "Datetime" timestamp, and the "Total (MWh)" column, aggregating hourly production from all sources, serves as the dependent variable, with individual sources as independent variables. The dataset's temporal coverage and source diversity provide a solid basis for analyzing Türkiye's electricity generation dynamics and developing advanced forecasting models.

For preprocessing, missing values (<0.01%) were interpolated to preserve temporal continuity. Outliers exceeding ±3 standard deviations within a 24-hour rolling window were cross-checked with EPİAŞ records; genuine operational anomalies were retained to reflect real-world variability. Exceptional variations due to regulatory changes or events such as the COVID-19 pandemic were preserved to maintain realistic operating conditions.

B. Normalization

In the normalization process, Min-Max normalization [29] was applied to both dependent and independent variables. For this purpose, all features were scaled to the range [0, 1] using the MinMaxScaler method. As shown in (1), this technique rescales each observation based on the minimum and maximum values of its corresponding feature.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Here, x represents the original value, x' denotes the normalized value, and x_{min} and x_{max} are the minimum and maximum values of the feature, respectively. Normalization mitigates scale disparities among variables and is essential for balanced and stable model training, thereby supporting accurate forecasting.

C. Principal Component Analysis

In this study, principal component analysis (PCA) was applied as a dimensionality reduction technique to transform the dataset into a more compact and informative structure. The PCA is a statistical method that projects high-dimensional data onto a lower-dimensional subspace using a new set of linearly uncorrelated variables known as principal components. This transformation aims to retain the maximum amount of original information (variance) while reducing the number of features [30]. The first step in implementing PCA involves computing the covariance matrix of the normalized dataset. As shown in (2), the covariance matrix R is derived by taking the average of the outer products of each data point's deviation from the mean vector.

$$R = \frac{1}{n}\sum_{i=1}^{n}(z_i - \mu)(z_i - \mu)^T \tag{2}$$

Here, z_i denotes the normalized feature vector of the ith sample, μ is the mean vector of all samples, and n indicates the total number of observations. To extract the principal components, an eigen decomposition is performed on the covariance matrix. As shown in (3), the eigenvectors and their corresponding eigenvalues are computed as follows.

$$Rv_i = \lambda_i v_i, i = 1,2,\ldots,m \tag{3}$$

Here, each eigenvector v_i represents a principal direction in the feature space, indicating a new axis along which the data varies the most. The associated eigenvalue λ_i quantifies the amount of variance explained along that direction. The parameter m denotes the original number of features in the dataset [31–33]. After computing the eigenvalues, they are sorted in descending order, and the top

principal components that collectively account for at least 99% of the total variance are selected. The cumulative variance explained by the selected principal components served as the basis for determining the number of features to retain. As a result of this dimensionality reduction process, the number of features was reduced from 16 to 13, satisfying the criterion of preserving at least 99% of the total variance. This transformation effectively compresses the data while maintaining the integrity of the original information, ensuring that critical temporal and source-specific patterns are preserved. Consequently, it improves computational efficiency and enhances the performance of learning algorithms by accelerating the training process and reducing the risk of overfitting, while interpretability is maintained since PCA components are linear combinations of the original features.

Upon the completion of the preprocessing phase, which included normalization and dimensionality reduction through principal component extraction, the dataset was suitably prepared for DL model development. The input data, consisting of the scaled and reduced feature matrix and the associated target values, satisfies the requirements for efficient and stable model training, offering a clean, consistent, and analytically tractable representation of the original dataset.

D. Long Short-Term Memory and Transformer Networks

Long Short-Term Memory networks are an advanced structure designed to overcome the core issues faced by traditional Recurrent Neural Networks (RNNs), such as vanishing and exploding gradients, which hinder learning long-term dependencies. Due to this robust architecture, LSTM models are highly effective in time-sensitive applications such as time series analysis, language modeling, speech recognition, and network-based intrusion detection systems. The LSTM cells consist of specialized units that transmit and regulate information over time steps. Each cell contains three key gates: the forget gate, the input gate, and the output gate. These gates are responsible for updating the cell state and hidden state [34].

The forget gate first determines which parts of the previous cell state should be retained or discarded. This process is carried out using a sigmoid activation function, as represented in (4).

$$\phi_t = \tilde{A}\left(W_{x\phi}\right).\left[z_{t-1}, u_t\right] + b_{x\phi} \tag{4}$$

Here, z_{t-1} is the previous time step's cell output, u_t is the input vector at the current time step, $W_{x\varphi}$ is the weight matrix, and $b_{x\varphi}$ is the bias term. The sigmoid function, σ, outputs values between 0 and 1, determining how much of the previous information should be forgotten. The second step involves the input gate, which controls how much new information will be added to the cell. This process consists of two parts. First, the gate (I_t) which determines what information will be updated is computed using the sigmoid function, as shown in (5).

$$It = \left(W_{xl}.\left[z_{t-1}, u_t\right] + b_{xl}\right) \tag{5}$$

Next, a candidate information vector (Ψ_t) is is created using the hyperbolic tangent function, as in (6).

$$\Psi_t = \tanh\left(W_{x\hat{E}}.\left[z_{t-1}, u_t\right] + b_{x\hat{E}}\right) \tag{6}$$

The updated cell state Ψ_t is obtained by combining the forget gate's impact on past information and the input gate's impact on new information. This process is represented in (7).

$$\Psi_t = \phi_t \quad \Psi_{t-1} + I_t \quad \Psi_t \tag{7}$$

Here, ⊙ denotes the Hadamard product (element-wise multiplication). The cell updates its state by considering both past and new information. Finally, the output gate ω_t determines which portion of the current cell state will be output. This gate is defined by a sigmoid function, as shown in (8).

$$\acute{E}_t = \tilde{A}\left(W_{x\acute{E}}.\left[z_{t-1}, u_t\right] + b_{x\acute{E}}\right) \tag{8}$$

Based on this value, the cell output (z_t) is calculated using (9).

$$z_t = \acute{E}_t \quad \tanh\left(\Psi_t\right) \tag{9}$$

This series of steps allows the LSTM cell to selectively remember and forget information, thus enabling effective learning of temporal dependencies [34-36].

The Transformer architecture, in contrast, was designed to address the limitations of traditional sequence-based models by overcoming the constraints of sequential data processing and improving the handling of long-range dependencies. Unlike RNNs, the Transformer does not rely on sequential processing and instead processes the entire input simultaneously, allowing for efficient parallelization [37, 38]. The core component of the Transformer is the self-attention mechanism, which enables the model to evaluate the importance of each input token relative to others, regardless of their position in the sequence. This mechanism enables the model to capture both local and global dependencies efficiently without the need for sequential data flow. The Transformer consists of two main components: the encoder and the decoder. The encoder processes the input sequence and generates a representation of the data, while the decoder uses this representation to produce the output sequence. Both the encoder and decoder utilize multiple layers of attention mechanisms and feed-forward neural networks, allowing the model to focus on various parts of the input sequence at each decoding step. A key feature of the Transformer is its use of scaled dot-product attention, which calculates similarity scores between query and key vectors and applies the softmax function to compute attention weights. These weights adjust the importance of corresponding value vectors, enabling the model to focus on the most relevant parts of the sequence. Additionally, multi-head attention is employed to apply attention mechanisms in parallel across multiple subspaces, which enhances the model's ability to capture diverse relationships in the data [39, 40].

In the proposed study, the Optuna method was used for hyperparameter selection [41]. Optuna is an open-source optimization framework that utilizes Bayesian optimization techniques to efficiently search for optimal hyperparameters. By leveraging this approach, the model's hyperparameters are selected based on their impact on the model's performance, ensuring more accurate and efficient results [41-43]. Approximately 10% of the earliest dataset

observations were temporarily set aside as a validation subset to prevent any future data leakage during hyperparameter tuning. The obtained hyperparameter values are as shown in Table II.

As seen in Table II, the model begins with sequential layers composed of LSTM cells, incorporating three LSTM blocks with 128, 64, and 32 units, respectively. These layers are designed to capture hierarchical temporal dependencies in the sequential data. The outputs of the LSTM blocks are first passed through a max pooling layer and then fed into the multi-head self-attention layer. This combination allows the model to integrate hierarchical temporal features captured by LSTM with long-range dependencies captured by the Transformer. This attention layer comprises 4 heads with a key dimension of 32, allowing the model to attend to multiple aspects of the sequence in parallel. The attention outputs are then passed through dense layers configured with 64 units and a dropout rate of 0.2. The ReLU activation function [44] is used in the dense layers to introduce nonlinearity, thereby enhancing the model's capacity to learn complex patterns. Finally, the learned feature representation is fed into the output layer, which employs a linear activation function for the prediction task. During training, the model is optimized using the Adam optimizer with a learning rate of 0.001 and a batch size of 64. The mean squared error (MSE) [45] is used as the loss function, aiming to minimize the difference between predicted and true values throughout the training process.

III. RESULTS

The performance of the proposed forecasting model was assessed using a real-world time series dataset on electricity generation. Given the sequential nature of time series data, it was critical to preserve the temporal order during the validation phase. For this reason, a time-aware cross-validation approach was adopted. In particular, the time-series split method from the scikit-learn library [46] was utilized. This method is specifically designed for time series problems and differs from standard k-fold cross-validation by avoiding data leakage from future to past.

For hyperparameter optimization using Optuna, a small subset corresponding to approximately 10% of the earliest observations in the dataset was temporarily set aside. This subset was used exclusively to select the optimal hyperparameters while ensuring no future data was exposed to the model. After tuning, the model was retrained using the remaining data for final evaluation.

The remainder of the dataset was then divided into five sequential folds for time-series cross-validation. In each fold, the training set was progressively expanded to include more historical data, while the test set always consisted of subsequent, unseen observations. This structure ensures that the model is trained on past data and evaluated on future data, mimicking a real-world forecasting scenario. The time-series split thereby maintains the chronological integrity of the data and provides a realistic estimate of the model's generalization performance in temporal tasks.

The development of the proposed model was performed using the Python programming language. For implementing the DL architecture, the TensorFlow 2.18 library was utilized, while data preprocessing was carried out with the aid of Pandas, NumPy, and scikit-learn libraries. Visualization tasks were executed using the Matplotlib library. The experiments were carried out on a desktop

TABLE II.
SELECTED HYPERPARAMETERS FOR THE PROPOSED MODEL

Layer/Module	Hyperparameter	Range/Values	Selected Value
LSTM layers	lstm_units	[32, 64, 128, 256]	128
	lstm_units	[32, 64, 128, 256]	64
	lstm_units	[32, 64, 128, 256]	32
	lstm_dropout	[0.1, 0.5]	0.2
Pooling	pooling_type	["max", "average"]	"max"
Multi-head attention	num_heads	[2, 8]	4
	key_dim	[16, 64]	32
Dense layers	dense_units	[32, 64, 128]	64
	dense_dropout	[0.1, 0.5]	0.2
	activation_function	["relu", "tanh", "gelu"]	"relu"
Training parameters	learning_rate	[1e-5, 1e-2]	0.001
	batch_size	[32, 128]	64
	optimizer	["adam", "rmsprop", "nadam"]	"adam"
Loss function	loss	["mse", "mae", "huber"]	"mse"

computer equipped with an Intel® Core™ i7-10700K CPU @ 3.80 GHz, 32 GB of RAM, and an NVIDIA GeForce RTX 3060 graphics processing unit. The training duration varied depending on the prediction horizon and dataset size; during the time-series split cross-validation experiments, the model predicting 1 hour ahead (horizon = 1) completed training in approximately 27 minutes, whereas the model predicting 1 day ahead (horizon = 24) required approximately 91 minutes.

The evaluation of the proposed model's performance is conducted using three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2) [47]. The MAE measures the average magnitude of errors in the predicted values, RMSE penalizes larger errors by squaring the differences between predicted and actual values, and R^2 indicates the proportion of variance in the actual data that is captured by the model, with values closer to 1 indicating a better fit.

A. One-Hour Ahead Electricity Generation Forecasting

The 1-hour ahead electricity generation forecasting results are summarized in Table III and illustrated in Fig. 1. Using five-fold time-series cross-validation, the model achieved an average MAE of 589.50 (1.67%), RMSE of 762.41 (2.15%), and R^2 of 0.98017, demonstrating high accuracy and robustness.

As shown in Fig. 1 and Table III, predicted values closely follow actual generation, effectively capturing short-term temporal patterns.

B. One-Day Ahead Electricity Generation Forecasting

The forecasting performance for the 1-day ahead horizon is presented in Table IV and visualized in Fig. 2. The model achieved an average MAE of 1328.24 (3.75%), RMSE of 1908.74 (5.39%), and R^2 of 0.87813.

As shown in Table IV, Fold 3 yielded the strongest results across all metrics, whereas Fold 5 recorded higher errors, likely reflecting local anomalies in the data. Compared to the 1-hour horizon, performance is lower as expected due to increased uncertainty in longer-term forecasts.

TABLE III.
PERFORMANCE METRICS OF THE MODEL FOR 1-HOUR AHEAD ELECTRICITY GENERATION FORECASTING

Fold No.	MAE	MAE (%)[a]	RMSE	RMSE (%)[a]	R^2
Fold 1	683.95	1.93	886.95	2.51	0.97351
Fold 2	492.68	1.39	635.46	1.80	0.98440
Fold 3	442.27	1.25	578.74	1.64	0.98676
Fold 4	476.40	1.35	617.77	1.75	0.98758
Fold 5	852.21	2.41	1093.14	3.09	0.96860
Average	589.50	1.67	762.41	2.15	0.98017

[a]Percentages relative to average generation (35381.73 MWh).

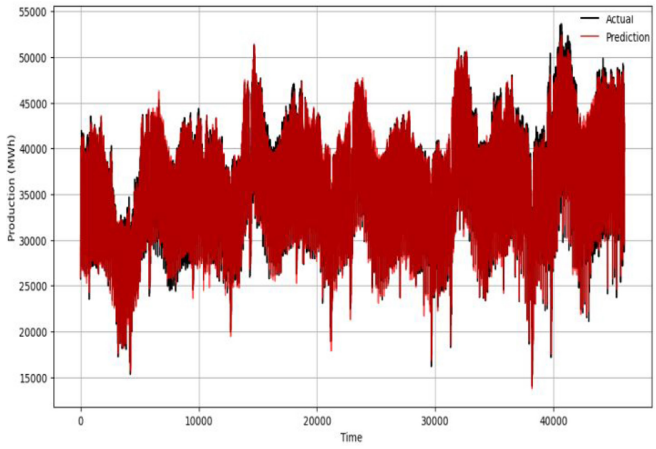

Fig. 1. Forecasting results of the 1-hour ahead electricity generation model.

As illustrated in Fig. 2, the model effectively tracks the overall trend of actual electricity generation, with only minor deviations around sharp peaks and troughs. The close alignment highlights the architecture's ability to capture seasonal and periodic patterns, supported by the attention and temporal pooling mechanisms that enhance long-range dependency modeling.

These metrics indicate that the model shows high performance for both 1-hour and 1-day ahead electricity generation forecasting tasks, with high accuracy for short-term predictions and satisfactory performance for longer-term forecasts. The 1-hour ahead forecast achieves an impressive R^2 of 0.98017, indicating excellent predictive accuracy, while the 1-day ahead forecast, with an R^2 of 0.87813, shows good performance, albeit with some increased error due to the complexity of predicting over longer time spans.

IV. DISCUSSION

In this study, we conducted a comprehensive evaluation of the proposed LSTM-Transformer hybrid model along with its individual

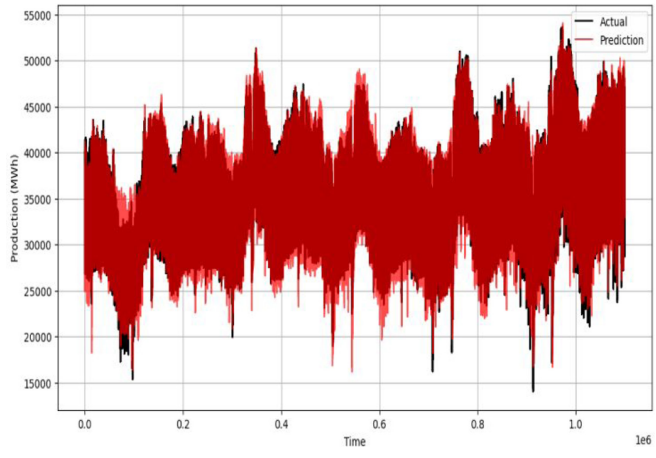

Fig. 2. Forecasting results of the 1-day ahead electricity generation model.

TABLE IV.
PERFORMANCE METRICS OF THE MODEL FOR 1-DAY AHEAD ELECTRICITY GENERATION FORECASTING

Fold No.	MAE	MAE (%)[b]	RMSE	RMSE (%)[a]	R^2
Fold 1	1480.95	4.18	2077.07	5.87	0.85545
Fold 2	1106.87	3.13	1637.75	4.63	0.89629
Fold 3	1112.20	3.14	1567.99	4.43	0.90281
Fold 4	1257.61	3.55	1820.28	5.14	0.89236
Fold 5	1683.56	4.76	2440.61	6.90	0.84376
Average	1328.24	3.75	1908.74	5.39	0.87813

[a]Percentages relative to average generation (35381.73 MWh).

components (standalone LSTM and Transformer models). The comparative results for 1-hour ahead forecasts are presented in Table V.

As shown in Table V, the proposed model achieves the lowest MAE and RMSE values and the highest R^2 scores. In contrast, the standalone LSTM and Transformer models exhibit higher error rates, particularly in daily predictions, indicating the superior predictive performance of the hybrid approach.

Additionally, this study provides a detailed examination of performance variations across different energy sources, as illustrated in Table VI. Only those with MAE (%) or RMSE (%) above 0.3, indicating significant impact on total generation error, are included. As shown in Table VI, short-term (1-hour ahead) forecasts exhibit notably low error rates for intermittent renewable sources such as wind and solar, indicating the model's effectiveness in capturing their short-term fluctuations. In contrast, fossil fuel-based sources like natural gas and imported coal display moderately higher error rates. Hydroelectric generation records the highest error among renewables, which is likely attributable to operational scheduling and water management practices that introduce variability independent of weather conditions, thereby increasing forecasting complexity.

The proposed model combines high forecasting accuracy with practical applications. In grid management, it supports real-time balancing of renewable fluctuations, optimizes hydropower releases, and plans reserve capacity. In market regulation, it enables day-ahead price forecasting, bidding strategy optimization, and early detection of abnormal price patterns. In policy-making, it aids capacity mechanism design, evaluation of renewable integration scenarios, and prioritization of infrastructure investments. The model's low inference time makes real-time deployment feasible, while the complexity of training favors offline execution.

TABLE V.
MODEL PERFORMANCE COMPARISON

Model	MAE	MAE (%)[a]	RMSE	RMSE (%)[a]	R^2
Proposed model	589.50	1.67	762.41	2.16	0.98017
Only LSTM	601.01	1.70	773.21	2.19	0.97935
Only transformer	666.60	1.88	856.33	2.42	0.97513

[a]Percentages relative to average generation (35381.73 MWh).

TABLE VI.
1-HOUR AHEAD FORECAST PERFORMANCE BY ENERGY SOURCE

Energy Source	MAE	MAE (%)[a]	RMSE	RMSE (%)[a]	R^2
Natural gas	372.55	1.05	493.01	1.39	0.983
Hydroelectric	421.77	1.19	552.98	1.56	0.959
Imported coal	170.79	0.48	253.24	0.72	0.980
Wind	130.79	0.37	169.77	0.48	0.993
Solar	46.58	0.13	73.05	0.31	0.931

[a]Percentages relative to average generation (35381.73 MWh).

While the case study focuses on Türkiye, the proposed framework is adaptable to other emerging economies with comparable energy system characteristics. Its transferability depends on factors such as the diversity of the energy mix, grid stability, and the availability of high-resolution operational data, which should be considered when applying the model in different national contexts.

V. CONCLUSION

The increasing complexity of modern energy systems, driven by the diversification of generation sources and the growing integration of renewables, has underscored the necessity for accurate short-term electricity generation forecasting. In response to this critical demand, the present study has demonstrated the efficacy of a DL-based forecasting framework that combines LSTM networks with a multi-head self-attention mechanism and robust regularization strategies. The proposed architecture effectively captures both short- and long-range temporal dependencies inherent in electricity generation time series, thereby enhancing predictive precision. Empirical evaluations conducted on one-hour ahead forecasting tasks revealed that the model consistently achieves high predictive performance, with an average MAE of 589.50 (1.67%), RMSE of 762.41 (2.15%), and R^2 of 0.98017 across five-fold cross-validation. These results indicate the model's robustness in capturing short-term fluctuations and daily-seasonal trends in electricity generation. Visual analyses further corroborate these findings by illustrating a strong alignment between predicted and actual generation values. For the one-day ahead forecasting horizon, the model attained an average MAE of 1328.24 (3.75%), RMSE of 1908.74 (5.39%), and R^2 of 0.87813. While a slight decline in accuracy was observed due to the inherent uncertainty in long-term predictions, the model maintained a satisfactory level of performance, demonstrating its potential applicability for medium-range planning tasks. Overall, the findings of this study affirm the model's capability to address the forecasting challenges posed by dynamic, multi-source energy

systems such as that of Türkiye. The model presents a promising tool for enhancing operational decision-making, ensuring supply-demand equilibrium, and supporting strategic planning in energy markets. Its integration into real-time monitoring and dispatch systems could significantly contribute to the resilience and stability of power systems, particularly in regions with high renewable energy penetration.

Future work could explore the incorporation of exogenous variables such as meteorological conditions, market dynamics, and demand-side behavior to further enhance forecasting accuracy. Additionally, adapting the model for real-time deployment and evaluating its performance under different seasonal or system stress scenarios would offer valuable insights for large-scale, data-driven energy management.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – H.K.; Design – H.K.; Supervision – H.K.; Resources – H.K.; Materials – H.K.; Data Collection and/or Processing – H.K.; Analysis and/or Interpretation – H.K.; Literature Search – H.K.; Writing – H.K.; Critical Review – H.K.

Declaration of Interests: The author have no conflicts of interest to declare.

Funding: The author declare that this study received no financial support.

REFERENCES

1. J. Zhang, "Energy access challenge and the role of fossil fuels in meeting electricity demand: Promoting renewable energy capacity for sustainable development," *Geoscience Frontiers*, vol. 15, no. 5, p. 101873, 2024. [CrossRef]
2. M. M. Islam et al., "Improving reliability and stability of the power systems: A comprehensive review on the role of energy storage systems to enhance flexibility," *IEEE Access*, 2024. [CrossRef]
3. S. Shahzad, and E. Jasińska, "Renewable revolution: A review of strategic flexibility in future power systems," *Sustainability*, vol. 16, no. 13, p. 5454, 2024. [CrossRef]
4. G.-F. Fan, Y.-Y. Han, J.-W. Li, L.-L. Peng, Y.-H. Yeh, and W.-C. Hong, "A hybrid model for deep learning short-term power load forecasting based on feature extraction statistics techniques," *Expert Syst. Appl.*, vol. 238, p. 122012, 2024. [CrossRef]
5. S. Luo, B. Wang, Q. Gao, Y. Wang, and X. Pang, "Stacking integration algorithm based on CNN-BiLSTM-Attention with XGBoost for short-term electricity load forecasting," *Energy Rep.*, vol. 12, pp. 2676–2689, 2024. [CrossRef]
6. A. I. Osman et al., "Cost, environmental impact, and resilience of renewable energy under a changing climate: A review," *Environ. Chem. Lett.*, vol. 21, no. 2, pp. 741–764, 2023. [CrossRef]
7. S. Mertens, "Design of wind and solar energy supply, to match energy demand," *Cleaner Eng. Technol.*, vol. 6, p. 100402, 2022. [CrossRef]
8. M. Bilgili, and E. Pinar, "Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye," *Energy*, vol. 284, p. 128575, 2023. [CrossRef]
9. D. Saygin, O. B. Tör, M. E. Cebeci, S. Teimourzadeh, and P. Godron, "Increasing Turkey's power system flexibility for grid integration of 50% renewable energy share," *Energy Strategy Rev.*, vol. 34, p. 100625, 2021. [CrossRef]
10. A. Telli, S. Erat, and B. Demir, "Comparison of energy transition of Turkey and Germany: Energy policy, strengths/weaknesses and targets," *Clean Technol. Environ. Policy*, vol. 23, no. 2, pp. 413–427, 2021. [CrossRef]
11. H. Tutar, and M. Atas, "A review on turkey's renewable energy potential and its usage problems," *Int. J. Energy Econ. Policy*, vol. 12, no. 4, pp. 1–9, 2022. [CrossRef]
12. *"Transparency Platform of the Energy Exchange Istanbul (EPİAŞ)."* [Online]. Available: seffaflik.epias.com.tr.
13. S. Inal, Ö. Yasar, and K. Aydıner, "Importance of domestic coal (lignite) reserves on Turkey's energy independency," *MT Bilimsel*, no. 19, pp. 11–32, 2021.
14. K. Kaygusuz, "Renewable and sustainable energy use in Turkey: A review," *Renew. Sustain. Energy Rev.*, vol. 6, no. 4, pp. 339–366, 2002. [CrossRef]
15. C. Kul, L. Zhang, and Y. A. Solangi, "Assessing the renewable energy investment risk factors for sustainable development in Turkey," *J. Cleaner Prod.*, vol. 276, p. 124164, 2020. [CrossRef]
16. A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, p. 132306, 2020. [CrossRef]
17. X. Wang, J. Dai, and X. Liu, "A spatial-temporal neural network based on ResNet-Transformer for predicting railroad broken rails," *Adv. Eng. Inform.*, vol. 65, p. 103126, 2025. [CrossRef]
18. X. Li, Z. Wang, C. Yang, and A. Bozkurt, "An advanced framework for net electricity consumption prediction: Incorporating novel machine learning models and optimization algorithms," *Energy*, vol. 296, p. 131259, 2024. [CrossRef]
19. S. Aslam, H. Herodotou, S. M. Mohsin, N. Javaid, N. Ashraf, and S. Aslam, "A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids," *Renew. Sustain. Energy Rev.*, vol. 144, p. 110992, 2021. [CrossRef]
20. N. Saxena et al., "Hybrid KNN-SVM machine learning approach for solar power forecasting," *Environ. Chall.*, vol. 14, p. 100838, 2024. [CrossRef]
21. M. Sawant et al., "A selective review on recent advancements in long, short and ultra-short-term wind power prediction," *Energies*, vol. 15, no. 21, p. 8107, 2022. [CrossRef]
22. Y. Zhang, L. Cui, and W. Yan, "Integrating Kolmogorov–Arnold networks with time series prediction framework in electricity demand forecasting," *Energies*, vol. 18, no. 6, p. 1365, 2025. [CrossRef]
23. S. Wang, J. Shi, W. Yang, and Q. Yin, "High and low frequency wind power prediction based on Transformer and BiGRU-Attention," *Energy*, vol. 288, p. 129753, 2024. [CrossRef]
24. X. Xiang, X. Li, Y. Zhang, and J. Hu, "A short-term forecasting method for photovoltaic power generation based on the TCN-ECANet-GRU hybrid model," *Sci. Rep.*, vol. 14, no. 1, p. 6744, 2024. [CrossRef]
25. M. Güldürek, "Short-term wind speed prediction using a hybrid artificial intelligence approach based on dragonfly algorithm: A case study of the Mediterranean climate," *TEPES*, vol. 4, No. 2, pp. 84–95, 2024. [CrossRef]
26. M. S. Ibrahim, S. M. Gharghory, and H. A. Kamal, "A hybrid model of CNN and LSTM autoencoder-based short-term PV power generation forecasting," *Electr. Eng.*, vol. 106, no. 4, pp. 4239–4255, 2024. [CrossRef]
27. W. G. Buratto, R. N. Muniz, A. Nied, C. F. de O. Barros, R. Cardoso, and G. V. Gonzalez, "Wavelet CNN-LSTM time series forecasting of electricity power generation considering biomass thermal systems," *IET Generation Trans. & Dist.*, vol. 18, no. 21, pp. 3437–3451, 2024. [CrossRef]
28. Q. Wu, F. Guan, C. Lv, and Y. Huang, "Ultra-short-term multi-step wind power forecasting based on CNN-LSTM," *IET Renew. Power Gener.*, vol. 15, no. 5, pp. 1019–1029, 2021. [CrossRef]
29. S. Patro, and K. K. Sahu, *"Normalization: A preprocessing stage," arXiv Preprint ArXiv:1503.06462*, 2015. [CrossRef]

30. T. Kurita, "Principal component analysis (PCA)," in *Computer Vision: A Reference Guide*, K. Ikeuchi, Ed. Cham: Springer International Publishing, 2021, pp. 1013–1016. [CrossRef]
31. S. Wang, and F. Xiao, "AHU sensor fault diagnosis using principal component analysis method," *Energy Build.*, vol. 36, no. 2, pp. 147–160, 2004. [CrossRef]
32. S. P. Mishra et al., "Multivariate statistical data analysis-principal component analysis (PCA)," *Int. J. Livest. Res.*, vol. 7, no. 5, pp. 60–78, 2017.
33. H. Abdi, and L. J. Williams, "Principal component analysis," *WIREs Computational Stats*, vol. 2, no. 4, pp. 433–459, 2010. [CrossRef]
34. G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, 2020. [CrossRef]
35. J. Cheng, L. Dong, and M. Lapata, *"Long short-term memory-networks for machine reading," arXiv Preprint ArXiv:1601.06733*, 2016. [CrossRef]
36. T. Fischer, and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, 2018. [CrossRef]
37. Q. Zhang, C. Qin, Y. Zhang, F. Bao, C. Zhang, and P. Liu, "Transformer-based attention network for stock movement prediction," *Expert Syst. Appl.*, vol. 202, p. 117239, 2022. [CrossRef]
38. T. P. Nguyen, H. Nam, and D. Kim, "Transformer-based attention network for in-vehicle intrusion detection," *IEEE Access*, vol. 11, pp. 55389–55403, 2023. [CrossRef]
39. S. Reza, M. C. Ferreira, J. J. M. Machado, and J. M. R. Tavares, "A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks," *Expert Syst. Appl.*, vol. 202, p. 117275, 2022. [CrossRef]
40. H. Chen, D. Jiang, and H. Sahli, "Transformer encoder with multi-modal multi-head attention for continuous affect recognition," *IEEE Trans. Multimedia*, vol. 23, pp. 4171–4183, 2021. [CrossRef]
41. J. Joy, and M. P. Selvan, "A comprehensive study on the performance of different Multi-class Classification Algorithms and Hyperparameter Tuning Techniques using Optuna," presented at the International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), IEEE, 2022, pp. 1–5. [CrossRef]
42. S. Hanifi, A. Cammarono, and H. Zare-Behtash, "Advanced hyperparameter optimization of deep learning models for wind power prediction," *Renew. Energy*, vol. 221, p. 119700, 2024. [CrossRef]
43. T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, *Optuna: A Next-Generation Hyperparameter Optimization Framework*," presented at the Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 2623–2631, 2019. [CrossRef]
44. C. Banerjee, T. Mukherjee, and E. Pasiliao Jr., "An empirical study on generalizations of the ReLU activation function." ACM Southeast Conference, 2019, pp. 164–167. [CrossRef]
45. T. O. Hodson, T. M. Over, and S. S. Foks, "Mean squared error, deconstructed," *J. Adv. Model. Earth Syst.*, vol. 13, no. 12, p, MS002681, e2021, 2021. [CrossRef]
46. S. Y. Y. Hamad, T. Ma, and C. Antoniou, "Analysis and prediction of bike-sharing traffic flow–Citi bike, New York."," presented at the 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), IEEE, 2021, pp. 1–8. [CrossRef]
47. D. Chicco, M. J. WARRENS, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, 2021. [CrossRef]