

Hybrid LSTM-Transformer-ACO deep learning models for time series forecasting of solar power

Mehdi Sepahvand

Lorestan University

Ali Mir

Lorestan University

Vahid Mehrdad

`mehrdad.v@lu.ac.ir`

Lorestan University

Article

Keywords: Deep learning, Ant Colony Optimization, Hybrid models, Solar power, Transformer, LSTM

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10646057/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Hybrid LSTM-Transformer-ACO deep learning models for time series forecasting of solar power

Mehdi Sepahvand, Ali Mir, Vahid Mehrdad

Department of Electrical Engineering, University of Lorestan, Khorramabad, Iran

Corresponding author: Vahid Mehrdad

Email: mehrdad.v@lu.ac.ir

Abstract

Accurate solar power forecasting is essential for reliable renewable energy integration and intelligent grid management. This study proposes and evaluates novel hybrid deep learning frameworks that combine Long Short-Term Memory (LSTM) networks, Transformer (TF) architectures, and Ant Colony Optimization (ACO) for time-series forecasting of solar power generation. A comprehensive investigation of all feasible ACO-LSTM-TF hybrid configurations is conducted and compared with their standalone counterparts under multiple optimization algorithms. Experimental results demonstrate that the proposed LSTM-TF-ACO model achieves the best forecasting performance, attaining a mean absolute error (MAE) of 2.878% for solar power prediction when optimized using Nadam. For actual electrical load forecasting, the ACO-LSTM-TF model achieves an MAE of 0.5065%, outperforming competing benchmark and hybrid models. In addition to superior predictive accuracy, the proposed framework exhibits faster convergence and lower computational cost, making it well suited for real-time energy forecasting applications. The findings highlight the effectiveness of integrating metaheuristic optimization with deep learning architectures and emphasize the importance of model optimizer selection in improving forecasting performance. The proposed hybrid framework provides a robust and scalable solution for AI-driven renewable energy forecasting and smart grid applications.

Keywords Deep learning, Ant Colony Optimization, Hybrid models, Solar power, Transformer, LSTM

1. Introduction

Given the critical role of time series forecasting in deriving actionable insights and projections from historical data patterns, its application is particularly indispensable within renewable energy systems. This is especially true for solar power, which has emerged as a prominent clean and sustainable energy source owing to its ubiquity and ecological advantages. Consequently, the development of precise forecasting models is paramount for optimizing the utilization and seamless integration of solar energy into the electrical grid [1]. Accurately forecasting the actual power output of photovoltaic (PV) systems has been a relatively neglected area of research. The predominant body of work has concentrated predominantly on predicting solar irradiance. While essential, irradiance is merely a single variable within a multifaceted equation that determines the final electrical yield of a PV module. This output is contingent upon a complex interplay of numerous factors, such as the operational state and material composition of the solar cells, the system's design and configuration, the angle of incidence of sunlight, ambient weather conditions, and other pertinent parameters. A key exemplar is the operating temperature of the PV cells, a factor which significantly influences conversion efficiency and thus the net power generation. A comprehensive

forecasting model for solar power production, however, would inherently integrate the effects of all these contributing variables [2]. Accurate forecasting of solar energy output is critical for enhancing energy management, ensuring grid stability, and facilitating informed economic planning. These forecasts enable more effective balancing of the electrical grid and a reduced dependency on non-renewable sources by allowing for optimized scheduling of power generation, consumption, and storage. For instance, in a city reliant entirely on solar power, precise predictions are a fundamental prerequisite for the successful integration of this variable resource. An accurate forecast of a high-yield sunny day, for example, allows grid operators to proactively divert surplus energy to other regions or into storage facilities. Conversely, during periods of limited sunlight, such as overcast conditions, reliable forecasts facilitate preemptive measures to mitigate potential energy deficits. This capability is instrumental in maintaining the overall stability and promoting the sustainability of the power infrastructure. Consequently, the development of sophisticated solar forecasting models is essential for advancing renewable energy utilization and ensuring efficient grid management. Inaccurate predictions can result in the suboptimal operation of power systems, leading to financial inefficiencies and jeopardizing grid reliability. Thus, improved accuracy in solar power forecasting is indispensable for the effective and secure integration of renewable energy [3].

Accurate forecasting of solar power generation is particularly challenging due to the interdependent and dynamic nature of influencing factors. Solar power time series data exhibit nonlinear and nonstationary characteristics, primarily caused by weather conditions, shading, and fluctuating cloud cover. Conventional forecasting methods, including statistical models, often fail to capture the full complexity of these intricately linked relationships and variations. Consequently, generating reliable predictions amidst these inherent uncertainties is difficult. Effectively addressing these challenges necessitates advanced techniques specifically designed to model the complex patterns within solar power time series data [4]. Recent advances have demonstrated the considerable potential of deep learning in processing complex temporal data across a wide range of fields. This methodology has achieved significant breakthroughs in domains such as natural language processing, image recognition, and speech. Deep learning architectures, including Recurrent Neural Networks (RNNs) [5], Long Short-Term Memory (LSTM) [6], Convolutional Neural Networks (CNNs) [7], and Transformer (TF) models [8], are capable of handling nonlinear relationships, processing extensive datasets, and autonomously identifying salient features. Research has demonstrated that these techniques enhance predictive performance and effectively capture complex temporal dependencies in time series forecasting tasks [7]. Deep learning architectures have been extensively investigated for time series forecasting applications, particularly in the domain of solar energy generation. A wide array of deep learning models has been explored in the literature. These architectures demonstrate a strong capability for modeling temporal dynamics, processing sequential data, and enhancing predictive accuracy. However, each model presents specific limitations and architectural considerations, including computational efficiency during training, model interpretability, and robustness to noisy or incomplete datasets [9]. There is increasing focus on developing hybrid deep learning models for time series forecasting to overcome limitations and enhance the performance of single architectures. These hybrid approaches integrate two or three distinct network types to leverage their complementary strengths. The integration of diverse architectures, or the combination of deep learning with statistical methods, offers significant advantages including superior prediction accuracy, enhanced interpretability, and a more robust approach to limited data availability. This research direction holds considerable promise for advancing the field of solar power forecasting. Accurate forecasting of solar power generation is critically important within the domain of sustainable energy management. As the world transitions towards renewable energy sources, the ability to predict solar output becomes indispensable for enabling efficient grid integration, strategic energy planning, and optimal resource allocation. This study investigates the development and assessment of hybrid deep learning models for solar power time series forecasting. It specifically examines the mechanisms through which these hybrid models improve forecast accuracy, evaluates the performance of

architectures combining ACO, LSTM, and Transformer networks, and analyzes the impact of optimizer selection on model efficacy. Furthermore, the research aims to identify optimal configurations of ACO, LSTM, and Transformer components in conjunction with various optimizers. This study introduces novel hybrid deep learning architectures for forecasting solar power generation, systematically evaluating the performance of diverse integrations of ACO, LSTM, and TF models. The findings reveal the superior predictive capability of the LSTM-TF-ACO hybrid model and provide crucial insights into the influence of optimizer selection on forecast accuracy. Moreover, the proposed LSTM-TF-ACO framework combines the temporal memory capability of LSTM networks, the global dependency learning mechanism of Transformers, and the efficient search strategy of ACO, resulting in a robust architecture capable of capturing both short-term and long-term patterns in solar power generation data. The incorporation of ACO further enhances the optimization process and contributes to improved convergence behavior and model generalization. In addition to its superior forecasting accuracy, the proposed hybrid model provides a favorable balance between predictive performance and computational complexity, making it a promising candidate for practical deployment in renewable energy forecasting applications. A notable contribution of this work is the pioneering incorporation of Transformer networks within a hybrid framework for solar power prediction, thereby advancing the state-of-the-art in this field. Enhanced forecasting of solar power generation significantly informs both the operational strategies of renewable energy systems and the formulation of energy policy. By enabling more precise scheduling of electricity production and consumption, these forecasts are instrumental for grid operators in maintaining the critical balance between supply and demand. This capability becomes particularly vital as solar energy constitutes an increasingly substantial share of the overall power supply. Furthermore, the provision of dependable predictions exerts influence on policy formulation by furnishing evidence-based insights. These insights are instrumental in crafting renewable energy policies that account for the inherent intermittency of solar power generation. Consequently, precise forecasts empower policymakers to develop and implement regulatory frameworks and incentive structures that are congruent with the variable characteristics of solar energy, thereby facilitating its seamless integration into the power grid. The remainder of this paper is organized as follows: Section 2 provides a comprehensive review of the relevant literature, highlighting the research gaps addressed by this work. Section 3 details the proposed method, including its theoretical foundation and implementation framework. Section 4 presents and discusses the simulation and experimental results, demonstrating the effectiveness of the approach. Finally, Section 5 concludes the paper with a summary of key findings and suggestions for future research.

2. Literature review

The appropriate timeframe for solar power forecasts is dictated by their specific applications, spanning from minutes to days. Predicting abrupt changes in solar irradiance, termed ramp events, is a primary focus of very short-term and short-term forecasts. These rapid fluctuations can severely compromise the reliability and power quality of solar photovoltaic systems. Consequently, short-term forecasts are essential for estimating the most significant PV power ramp rates. In contrast, medium-term and long-term forecasts are instrumental in optimizing the business operations and enhancing the market integration of solar PV technology [10]. Research by Husein et al. [11] demonstrated that integrating day-ahead solar irradiance predictions into the operational strategy of commercial building microgrids results in significantly reduced annual energy expenditures. Consequently, the selection of a context-specific forecasting methodology is paramount for optimizing solar power applications. Recent studies in solar power forecasting have demonstrated a notable transition towards the adoption of deep learning methodologies to enhance predictive precision and operational efficiency. For daily solar power forecasting, Elsaraiti et al. [12] demonstrated the superiority of Long Short-Term Memory (LSTM) networks over Multi-Layer Perceptron (MLP) models. Their research, which involved 30-minute-ahead predictions using solar data from Halifax,

Canada throughout 2017, concluded that the LSTM algorithm excelled in all major accuracy and performance measures for time series analysis of solar power. In a similar approach, the study by Kim et al. [13] also investigates precise forecasting of photovoltaic power generation by evaluating seven different models. For the development of their time-series models, the input data were categorized according to seasonal variations and incorporated multiple parameters. The research utilized PV power generation data recorded from January 2017 to June 2021 in the cities of Ansan and Suwon. Among the evaluated models (including Holt-Winters, MLR, ARIMA, SARIMA, ARIMAX, and SARIMAX), the Long Short-Term Memory (LSTM) deep learning model demonstrated superior performance by achieving the lowest error rate in forecasting photovoltaic power generation that accounts for seasonal and weather influences. This outcome underscores a discernible shift within the research community, indicating a growing preference for deep learning methodologies in solar power time series forecasting. This trend aligns with the broader movement towards adopting sophisticated neural network architectures to enhance the precision and operational efficiency of solar power predictions. This research systematically examines three key deep learning techniques ACO, LSTM, TF, and all their possible hybrids, with Table 1 summarizing the principal merits and demerits of each model for time series tasks.

Table 1- Deep learning pros and cons for time series applications

Model	Refs.	Pros	Cons
LSTM	[14]	LSTM networks use memory cells to store and access information over long time intervals, enabling them to model long-range temporal dependencies and capture complex non-linear patterns in time series analysis.	Compared to simpler models, LSTMs demand significantly more computation especially for large-scale time series or deep architectures making both training and inference resource-intensive, requiring greater processing power and longer completion times.
TF	[15]	The Transformer model utilizes a self-attention mechanism, enabling it to recognize global interdependencies within input sequences. This capability renders it particularly effective for capturing long-range temporal relationships and intricate patterns in time series data.	Unlike sequential RNNs, Transformers process input in parallel, which makes capturing temporal dependencies explicitly more challenging. While self-attention captures global interconnections, it is less effective than LSTMs at learning the local, ordered dependencies inherent to sequential data.
ACO	[16]	ACO exhibits powerful emergent intelligence, efficiently solving complex combinatorial problems (e.g., routing, scheduling) through simple agent interactions and stigmergy, while offering robust adaptation to dynamic changes such as link failures.	ACO suffers from slow convergence, high sensitivity to parameter settings (e.g., evaporation rate, heuristic influence) leading to premature convergence on suboptimal solutions, and stochasticity that prevents guaranteeing deterministic or optimal results across runs.
Hybrid models	[17]	Hybrid models use the most effective parts of different architectures or methods to make better predictions than individual models. Combining different approaches allows hybrid models to find a broader range of patterns and relationships in time series data	The development of hybrid models is inherently more complex than that of singular models, necessitating greater investment in both design and implementation. Integrating multiple architectures or methodologies often presents significant challenges in training, optimization, and deployment.

In PV power generation systems, a hybrid deep learning framework integrating Long Short-Term Memory (LSTM), TF, and Ant Colony Optimization (ACO) is proposed. Unlike traditional methods, this approach employs LSTM to capture long-term temporal dependencies and mitigate vanishing gradient issues using its recurrent architecture and memory units, TF to model complex, multi-scale weather patterns and global dependencies via self-attention mechanisms, and ACO to optimally select input features, hyperparameters, or network architectures through swarm intelligence, thereby enhancing both prediction accuracy and computational efficiency. The solar power dataset used in this study was gathered from actual Limburg solar farms. According to simulation results, the proposed method achieves significantly lower prediction errors than the benchmark methods. Furthermore, while the study focuses on short-term prediction for PV power generation, the proposed method can potentially advance the use of solar energy in electrical power and energy systems.

3. Proposed Method

There are three primary approaches that have shown promise in capturing temporal dependencies and patterns in time series data, as well as optimizing predictive models. These methods are the LSTM, the Transformer model, and ACO. The majority of the research is focused on the deep learning methods LSTM and Transformer. More recently, hybrid models that bring these two deep learning architectures together with ACO based on their respective strengths are being researched in order to further raise the rate of accuracy for predicting. In the proposed method, ACO is employed to optimally select input features, hyperparameters, or network architectures for the LSTM-Transformer hybrid, thereby enhancing prediction performance while reducing computational complexity.

3.1. *Ant Colony Optimization*

Inspired by the collective foraging behavior of ants, ACO is a metaheuristic algorithm introduced by Dorigo in the early 1990s [18]. Despite the limited individual capabilities of ants, their collaborative behavior enables the emergence of a system capable of identifying optimal paths between their nest and food sources. Ants employ an initial random search strategy in the vicinity of their nest to locate nourishment. Upon discovering a viable food source, an ant assesses its value and transports a portion back to the colony. This return trip is marked by the deposition of a pheromone trail, the intensity of which is proportional to the resource's quality and abundance. This chemical deposit acts as a stigmergic signal (a form of indirect communication) that directs nestmates to the location, ultimately enabling the colony to optimize and identify the most efficient routes to sustenance. Fig. 1 illustrates this process. For continuous domains, the ACO strategy was adapted by maintaining a solution archive, which stores a population of candidate solutions to guide the algorithm via a Gaussian mixture probability model [19]. The evolutionary feedback loop between this model and the archive is the reason this approach is regarded as an Estimation of Distribution Algorithm (EDA) [20].

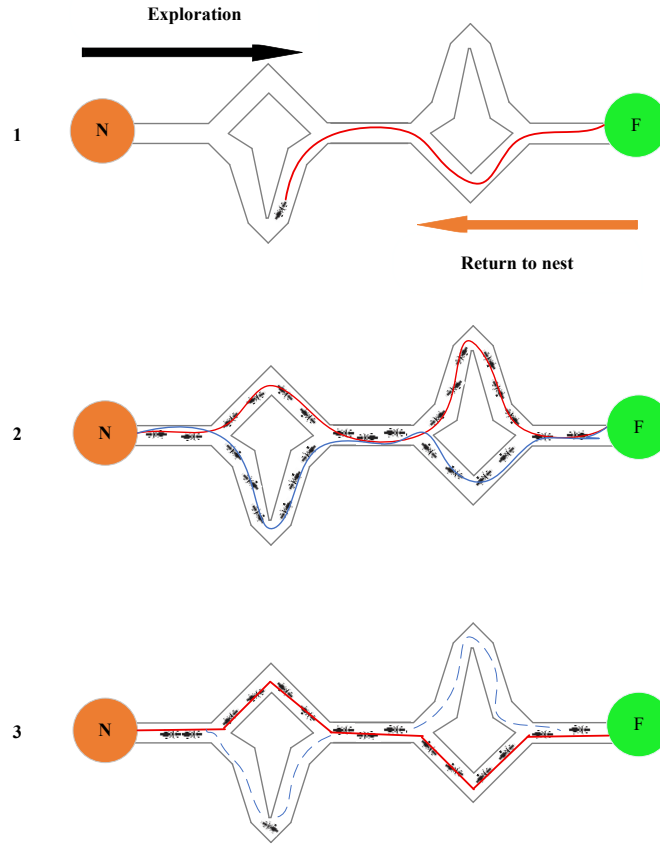


Figure 1. Selection of the shortest path by an ant colony. An ant finds a food source (F) then returns to the nest (N) by spreading the pheromone along its path.

The main steps of this algorithm are summarized as follows:

- a) **Initialization:** An initial colony comprising n ants is randomly dispersed across the search space, and their corresponding fitness values are computed.
- b) **Generation of the solution archive:** A fitness-based sorting procedure is applied to the initial solution set, where the top-performing and bottom-performing solutions are designated as x_1 and x_n , respectively.
- c) **Weight attribution:** A weight is assigned to each solution within the archive using Equation (1), which is based on a Gaussian distribution. The weights are then normalized according to Equation (2) to ensure that their sum equals one.

$$w_i \propto \frac{1}{\sqrt{2\pi\alpha n}} \cdot \exp\left[-\frac{1}{2}\left(\frac{i-1}{\alpha n}\right)^2\right] \quad (1)$$

and

$$\sum_{i=1}^n w_i = 1 \quad (2)$$

d) Generation of the probabilistic model: In the Gaussian mixture probabilistic model, the following equation is implemented:

$$G^k(x[k]) = \sum_{i=1}^n w_i N(x[k]; \mu_i[k], \sigma_i[k]) \quad (3)$$

$$N(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right] \quad (4)$$

The probabilistic framework is realized by computing the mean and standard deviation of the Gaussian mixture model, as defined by the following mathematical expressions, where k serves as the decision variable and $x[k]$ denotes the k^{th} element within vector x .

$$\mu_i[k] = x_i[k] \quad (5)$$

$$\sigma_i[k] = \frac{\eta}{n-1} \sum_{i'=1}^n [x_{i'}[k] - \mu_{i'}[k]]^2 \quad (6)$$

where the parameter $\eta > 0$ governs the trade-off between exploration and exploitation.

e) *Sampling*: New offspring are generated from the existing archive using a set of genetic operators, $g = (G_1, G_2, \dots, G_{nx})$. These newly created individuals are subsequently evaluated against a defined fitness function

f) *Selection*: During this stage, the solution archive is updated by integrating the generated offspring and the current best solutions. As a result, the elite solution preserved in the archive signifies the global best solution of the entire optimization procedure.

This entire process is repeated iteratively until a specific stopping criterion is satisfied.

3.2. Long Short-Term Memory

Long Short-Term Memory (LSTM), a specific architecture of recurrent neural networks, has demonstrated significant efficacy in discerning temporal dependencies within time series data. Consequently, it has been extensively adopted in research for forecasting applications, including the prediction of solar power generation. The LSTM model demonstrates a strong capability in analyzing the intricate dependencies of solar energy generation, which are influenced by meteorological factors, cloud density, and diurnal cycles [21]. According to reference [22], the core of the LSTM architecture is composed of three gate types: input, forget, and output gates. This gating mechanism empowers the model to control the passage of information through the memory cells, thereby making deliberate choices about which information to preserve or ignore, as shown in Figure 2. In LSTM Memory networks, the assimilation of new inputs by the memory cells is managed by the input gate. This gate's activity is controlled by a sigmoid function that integrates the present input signal with the preceding output state of the cell. The input is suppressed as the sigmoid output approaches zero, whereas it is maintained without attenuation when the output tends toward one. The forget gate in the LSTM architecture, activated by a sigmoid function, regulates the extent to which the previous cell state is retained for the current time step. Its decision is informed by both the present input and the preceding output of the cell. A sigmoid output value close to zero implies that the previous state will be almost entirely preserved. Conversely, as this output approaches one, the prior state is almost completely erased [22]. It is the function of the output gate to determine what information from the cell's present state will be exposed at the subsequent time step. This gating mechanism employs sigmoid and tanh functions to process a combination of the immediate inputs and the outputs from the previous step. The sigmoid function determines which segments of the memory cell's state are to be propagated forward, whereas the hyperbolic tangent regulates the amplitude of the LSTM's output [22]. The LSTM model utilizes its gating mechanisms to regulate the retention and discarding of information, enabling it to capture intricate temporal dependencies for solar power forecasting. This includes patterns across daily, weekly, and seasonal cycles, as well as complex nonlinear relationships between variables. The gate mechanism, a fundamental component of the LSTM architecture, has proven to be highly effective and is consistently achieving outstanding results in forecasting solar energy demand [23].

$$f_t = \sigma(W_{fh}h_{t-1} + W_{fx}s_t + b_f) \quad (7)$$

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}s_t + b_i) \quad (8)$$

$$c_t^\square = \tanh(W_{c^{\square}h}h_{t-1} + W_{c^{\square}x}s_t + b_{c^{\square}}) \quad (9)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c_t^\square \quad (10)$$

$$o_t = \sigma(W_{oh}h_{t-1} + W_{ox}s_t + b_o) \quad (11)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (12)$$

The mathematical formulation of the LSTM architecture can be depicted as follows:

The inputs, recurrent states, and outputs of each LSTM cell at time step t are denoted as x_t , h_t , and o_t respectively. Within the architecture, f_t signifies the forget gate, while c_t indicates the cell state. The weight matrices for the respective gates are represented by W_f , W_i , $W_{c\Box}$, and W_o . The symbol ' $\cdot$ ' denotes the Hadamard product (element-wise multiplication), and the bias vectors are given by b_f , b_i , $b_{c\Box}$, and b_o .

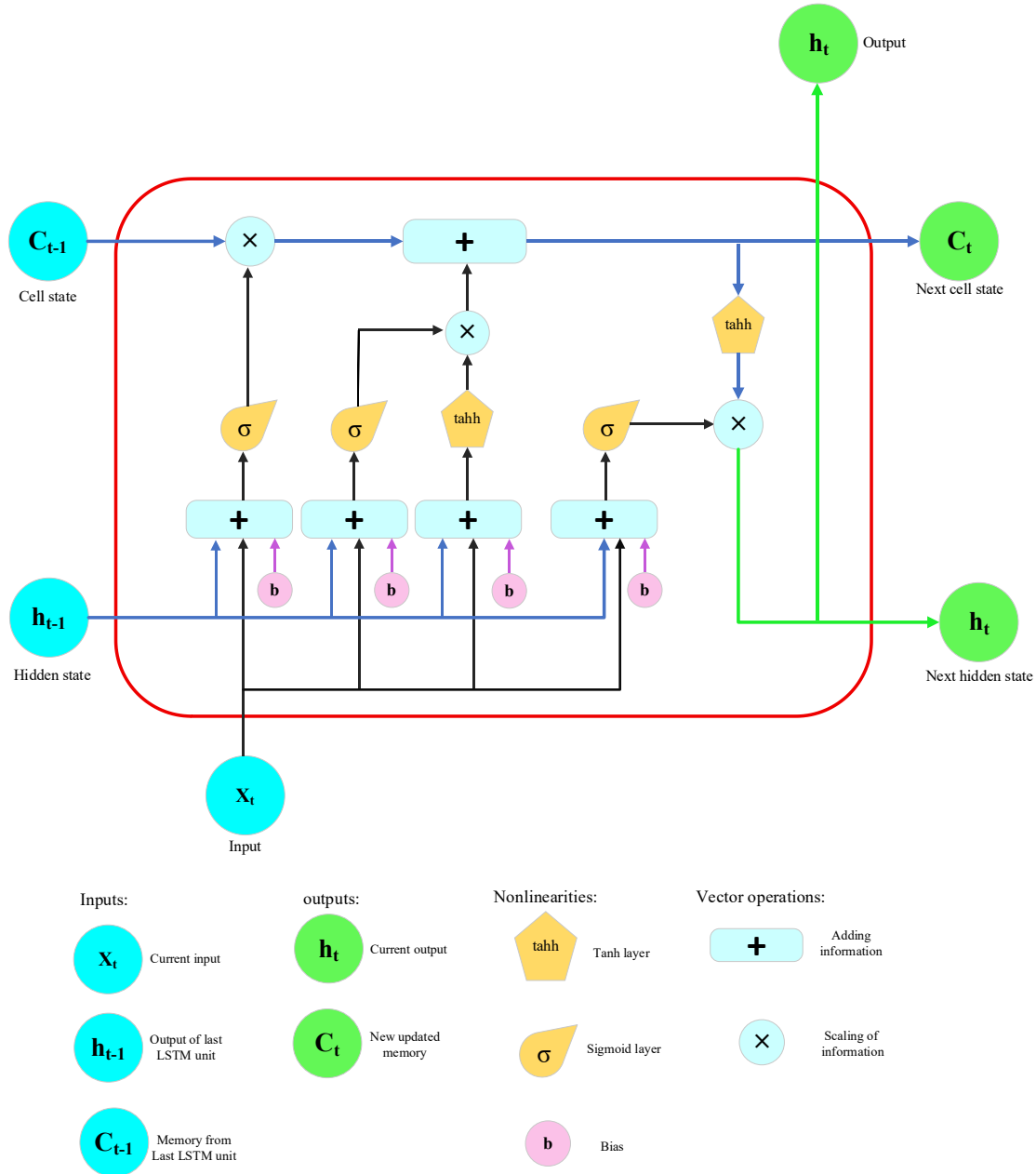


Fig. 2 LSTM architecture [24]

3.3. *Transformer model*

The TF architecture has fundamentally transformed the field of sequence modeling and is now extensively utilized in a wide range of applications, such as time series analysis [25]. The Transformer architecture effectively models complex dependencies within sequential data, thereby proving to be a powerful solution for sequential tasks. It leverages a self-attention mechanism to concurrently process various parts of a sequence, effectively overcoming challenges such as the vanishing gradient problem. This mechanism is fundamental to the TF, as it enables the integration of long-range dependencies and rich contextual cues. The model's ability to account for global contextual relationships stems from its attention to the full sequence and its use of associative weighting for auxiliary features. The incorporation of the attention mechanism empowers the TF model to concurrently process both short- and long-range dependencies, thereby enhancing the accuracy and consequent reliability of its predictions. In the TF architecture, both the self-attention mechanism and the position-wise feed-forward network perform nonlinear transformations on each position in the sequence independently. Consequently, the model can effectively represent complex temporal relationships, leading to superior learning outcomes and more accurate predictions. This capability further enables the model to comprehend sequential patterns. Owing to its parallelization of both training and inference, TensorFlow is exceptionally well-adapted for modeling large-scale sequences [25]. The ability of TF models to capture dependencies, configurations, and sequential patterns stems from their core components self-attention mechanisms and feedforward networks. Consequently, when properly configured, these models demonstrate exceptional proficiency in a wide range of sequence-based applications, such as interpreting input sequences and generating coherent output sequences. The encoding mechanism within the TF model processes an input sequence to generate its corresponding encoded representation. This transformation is accomplished through a series of feedforward neural networks and self-attention layers. In each layer, the self-attention mechanism initially analyzes the input to identify and model the intricate dependencies between the constituent data points. The self-attention mechanism enables every element in a sequence to dynamically weigh the importance of other elements, thereby integrating relevant contextual information. The encoder creates latent representations for the decoding process. In the Transformer model, the decoder's role is to interpret these representations to generate a sequential output. The decoder module itself is structured with a stack of layers, all equipped with both self-attention and feedforward neural network components. During the decoding phase, the model incorporates encoder-decoder attention, enabling it to dynamically access and focus on the most relevant parts of the encoded representations. This mechanism allows the decoder to leverage crucial contextual information from the input sequence effectively [25]. The architecture of the TF model is depicted in Figure 3.

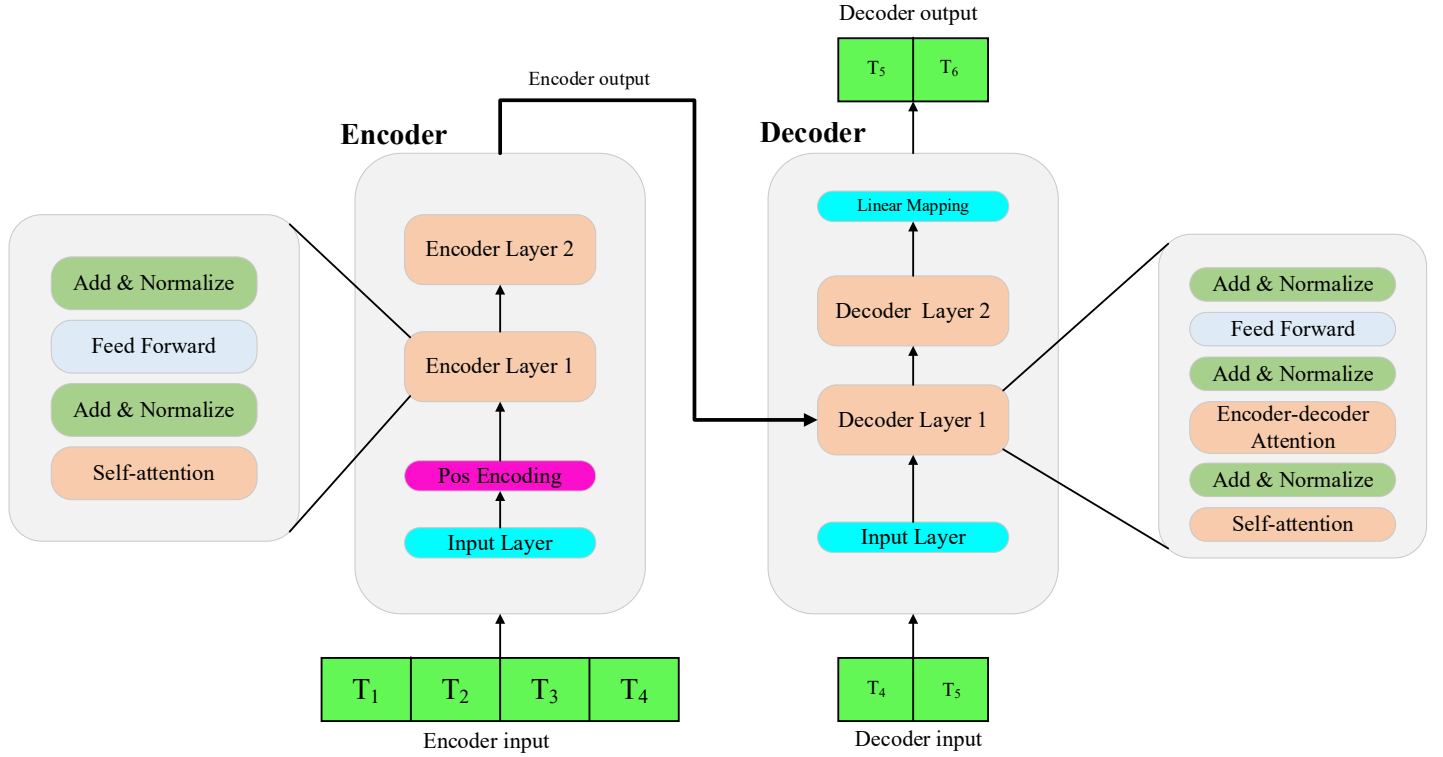


Fig. 3. Time series transformer forecasting based model architecture [26]

3.4. Hybrid models

The integration of deep learning methodologies in hybrid models leads to more accurate, interpretable, and robust forecasting for solar energy systems. Hybrid forecasting frameworks leverage advanced deep learning architectures, including Long Short-Term Memory (LSTM), ACO, and Transformers, to effectively model the complex, multi-faceted patterns inherent in solar power time series data. Each component addresses a specific aspect: LSTM networks excel at capturing long-term temporal dependencies, ACO algorithms are adept at feature selection and extracting spatial relationships, and Transformer models utilize self-attention mechanisms to represent global dependencies across the entire data sequence. Hybrid forecasting frameworks leverage the complementary strengths of constituent architectures in a synergistic manner to achieve superior predictive accuracy. This is accomplished through the strategic integration of these models. Furthermore, the performance of such hybrid models can be augmented by incorporating established forecasting methodologies, including autoregressive and regression models. The integration of these conventional approaches incorporates interpretability and domain expertise, potentially capturing intrinsic features of solar power systems that are often elusive to deep learning algorithms. Such hybrid frameworks are anticipated to not only improve forecasting accuracy but also to elucidate the key factors governing solar power generation. Integrating the interpretability of conventional models with the powerful feature extraction and nonlinear mapping capabilities of deep learning facilitates this objective. Enhanced time series forecasting through such hybrid approaches is crucial for the precise prediction of solar power generation and its seamless integration into the power grid. The availability of accurate forecasting enables more efficient resource management, contributes to grid

stabilization, and supports informed economic planning. Hybrid models facilitate the integration of renewable energy by resolving challenges in solar power forecasting, including the complex interrelationships among solar irradiance, meteorological conditions, and power output. Hybrid solar forecasting models facilitate the transition to renewable energy systems. This research seeks to enhance the precision and efficacy of forecasts by exploring a range of hybrid methodologies for time series analysis. The hybrid models under examination, such as LSTM-TF, TF-LSTM, LSTM-ACO, ACO-LSTM, ACO-TF, TF-ACO, ACO-LSTM-TF, ACO-TF-LSTM, LSTM-ACO-TF, LSTM-TF-ACO, TF-ACO-LSTM, and TF-LSTM-ACO, are each designed with specific architectures to discern and capitalize on distinct characteristics within the dataset. The LSTM-TF model leverages the strengths of both LSTM and Transformer networks, thereby facilitating the simultaneous extraction of local contextual features and global relational patterns from the dataset. In contrast, the TF-LSTM model prioritizes the Term-Frequency (TF) element to model interrelationships and sequential dependencies, while concurrently leveraging the Long Short-Term Memory (LSTM) network to identify complex, long-term temporal patterns. The LSTM-ACO hybrid architecture effectively amalgamates the proficiency of LSTM networks in modeling long-term temporal dependencies with the capability of the ACO algorithm to discern spatial features. The ACO-LSTM framework, however, employs a reversed order wherein the ACO algorithm takes the primary role in feature extraction, and the LSTM component then models the temporal dynamics. The integration of the feature extraction process of ACO and the self-attention mechanism of the TF within the ACO-TF model allows for a comprehensive consideration of global dependencies. The TF-ACO model synergistically combines the temporal dependency recognition capability of the TF's attention mechanism with the spatial data optimization prowess of the ACO. Through the synergistic combination of ACO, LSTM, and a Temporal Fusion module utilizing self-attention, the ACO-LSTM-TF algorithm facilitates a concentrated focus on critical temporal information. This integrated approach ensures the simultaneous modeling of both short-range and long-range dependencies within the data. The ACO-TF-LSTM framework leverages the transformer as its core element, while utilizing LSTM and ACO to capture local and global dependencies, respectively. The LSTM-ACO-TF model amalgamates the capabilities of LSTM, ACO, and TF, thereby harnessing their collective power to effectively address local, temporal, and global dependencies. Consequently, the resulting TF-ACO-LSTM structure is designed by integrating the core attributes of the LSTM, ACO, and TF models. The combined framework accounts for dynamically changing processes and captures interactions across local and global scales in the data. The synergy within these hybrid models harnesses the core advantages of each module, resulting in markedly improved accuracy and performance.

3.5. Evaluation metrics

Evaluation metrics provide quantitative criteria for assessing a predictive model's performance and efficacy in solving a specific problem. These measures, which can be derived through qualitative or quantitative methods, are essential for the comparative analysis of different models and algorithms. Moreover, evaluation metrics are instrumental in forming evidence-based conclusions regarding a model's suitability for a particular task and in assessing its overall efficacy [27]. Therefore, they also contribute to evaluating the adequacy of a solution's accuracy, reliability, and generalizability.

3.5.1. Mean Absolute Error

Mean Absolute Error (MAE) quantifies the discrepancy between observed and predicted data. As defined in Equation (13), its magnitude is contingent on the scale of the continuous variables being assessed, where a lower value denotes superior predictive accuracy.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \left| \hat{y}_i - y_i \right| \quad (13)$$

where the predicted output for epoch i is symbolized by $\hat{y}_i$, in contrast to the true value, denoted as y_i .

3.5.2 Mean square error (MSE)

This value is derived from the squared residuals, which are the differences between actual and expected outputs. Its quadratic nature ensures that larger prediction errors are penalized more severely.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N \left(\hat{y}_i - y_i \right)^2 \quad (14)$$

3.5.3 Root mean square error (RMSE)

This metric is employed to quantify the predictive accuracy of models. Its calculation necessitates deriving the square root of the average of squared errors between the actual observed data and the predicted values. Consequently, the RMSE is universally regarded as a fundamental criterion in performance assessment.

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

3.6 Working structure

Employing relevant time-series data marks the beginning of the process, capturing the evolution of solar power generation. This is followed by a preprocessing stage that augments data integrity for subsequent time-domain examination. The process of model selection encompasses a diverse array of architectures, including both standalone models, such as LSTM, ACO, and Transformer, and hybrid models that are specifically engineered for time series forecasting. Following this, the selected models are trained on the time-series power data, explicitly modeling the temporal dependencies. The incorporation of validation data improves the learning procedure and ensures robust model efficacy. The models' predictive performance for solar power generation is evaluated using an independent time series test set, with standard time series metrics MAE, MSE, and RMSE providing a complete assessment. The optimization commences with the selection of an optimizer tailored to the properties of the time series. The training and validation of the models were conducted utilizing a range of optimizers, namely Adam, Nadam, Adamax, RMSprop, and AMSGrad. Following the training phase, the models undergo rigorous evaluation on the testing time series dataset to assess their performance. The identification of the optimal model through a pivotal model comparison marks a critical juncture for time-domain-accurate solar power forecasting. The architecture of the solar power forecasting system, which methodically utilizes time series power data via interconnected processes and decisions, is presented in Figure 4.

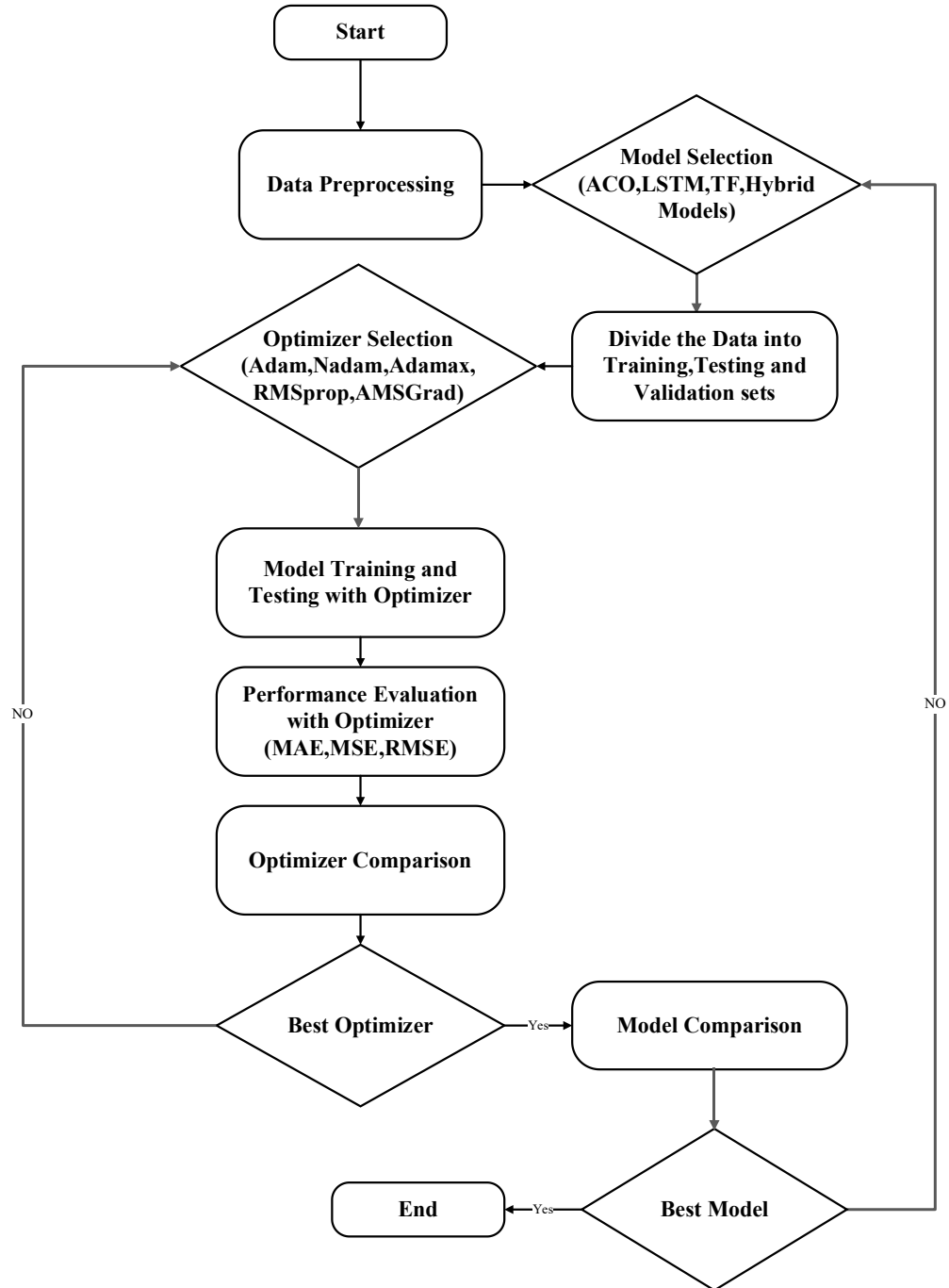


Fig. 4 Workflow and model selection flowchart

The proposed framework integrates ACO, LSTM networks, and a TF module in a unified hybrid architecture designed for high-accuracy time-series prediction. The flowchart illustrates the sequential and iterative mechanism through which the system adapts its hyperparameters and improves predictive performance. The process begins with the initialization of the ant population, where each ant is assigned a randomly generated hyperparameter configuration. This configuration typically includes LSTM-related parameters such as the number of layers, hidden units, dropout rates, learning rate, sequence length, and

batch size. In the next step, each hyperparameter configuration is evaluated by training an initial LSTM model and measuring its performance using accuracy metrics such as MAE, MSE, RMSE. These evaluation metrics determine the quality of each ant's solution in the hyperparameter search space. After the LSTM performance is computed, the framework introduces an additional enhancement: the extracted LSTM features are forwarded to a TF model for secondary learning. This TF module is trained on the hidden-state features generated by the LSTM, enabling the system to capture long-range temporal dependencies and global attention patterns that LSTM alone may not fully exploit. The performance of the combined LSTM-TF pipeline further influences the quality score assigned to each ant's hyperparameter configuration. Based on the obtained performance measures, the pheromone trails are updated. Hyperparameter sets that lead to better predictive outcomes deposit stronger pheromone values, increasing the probability that subsequent ants will explore similar configurations. Poor-performing configurations gradually lose pheromone concentration through evaporation, preventing premature convergence and encouraging exploration of new regions in the parameter space. In the following stage, ants utilize these updated pheromone trails to guide their selection of hyperparameters during the next iteration. This probabilistic decision-making mechanism forms the core optimization loop of the algorithm. The LSTM is re-evaluated with newly generated configurations and, once again, the TF module is trained on the corresponding LSTM feature outputs. The results continually influence pheromone adjustments. This iterative optimization cycle continues until the algorithm satisfies the convergence condition, indicating that the hyperparameters have stabilized around an optimal or near-optimal region. Once convergence is achieved, the final step involves training the LSTM model from scratch using the selected optimal hyperparameters. The optimized LSTM configuration ensures improved predictive accuracy, enhanced generalization ability, and reduced computational cost. Overall, in this hybrid framework, ACO searches for the best model settings, LSTM learns the important temporal patterns, Transformer captures deeper contextual relationships, and performance feedback continuously guides the system toward better configurations.

4. Results and discussion

In this section, the performance of the proposed hybrid deep learning framework integrating LSTM, Transformer, and ACO is thoroughly evaluated. The experimental setup, including data preprocessing, model configuration, and evaluation metrics, is first described. The results are analyzed under various weather conditions and seasonal scenarios to demonstrate the robustness and generalization capability of the proposed approach.

4.1 Data description

The research is underpinned by a four-year compilation of solar generation data for France, facilitating a fundamental understanding and detailed evaluation of the nation's solar power dynamics [28]. Analysis of the data reveals both annual and seasonal variations in solar power generation trends, as illustrated in Fig. 5. The inclusion of solar irradiance data in the dataset allows for a more precise characterization of power generation patterns. This study aims to develop models and strategies to achieve precise forecasting of future solar phenomena by utilizing patterns and trends extracted from historical power production data as primary inputs. This study primarily utilizes historical power generation data to forecast solar energy output; however, it does not explicitly account for external variables such as fluctuating weather conditions or advancements in solar technology. To gain deeper insight into the natural rhythms of solar power output, time series data were selected as the central analytical focus. The forecasting model incorporates critical meteorological variables, including temperature, solar irradiance, and humidity, by leveraging the established mathematical relationships between these external factors and solar power generation. The mathematical framework of the model accounts for these external variables, even though their direct

investigation falls outside the study's primary scope, which is to evaluate the methodology's efficacy for modeling and forecasting based solely on the available historical power data.

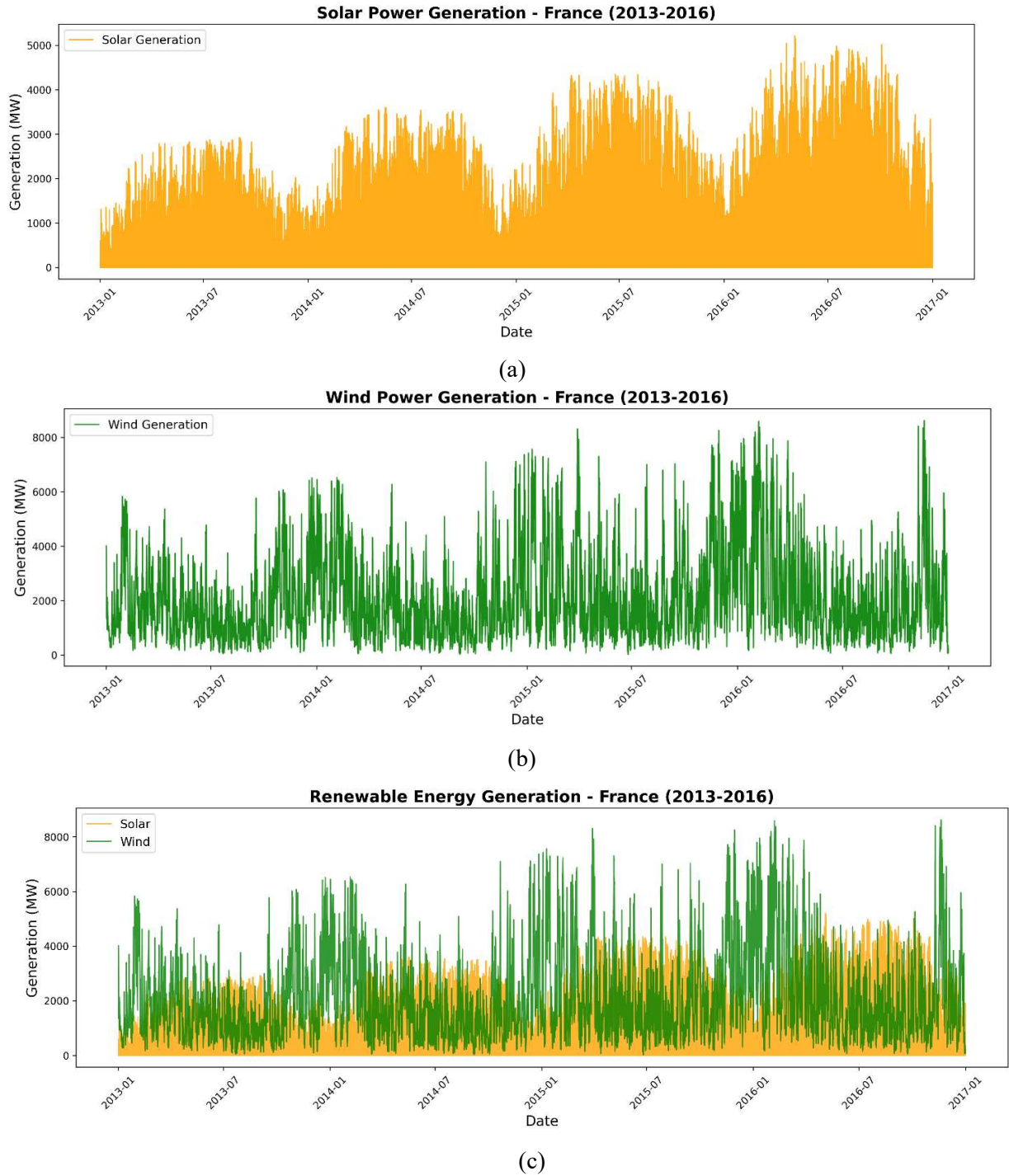
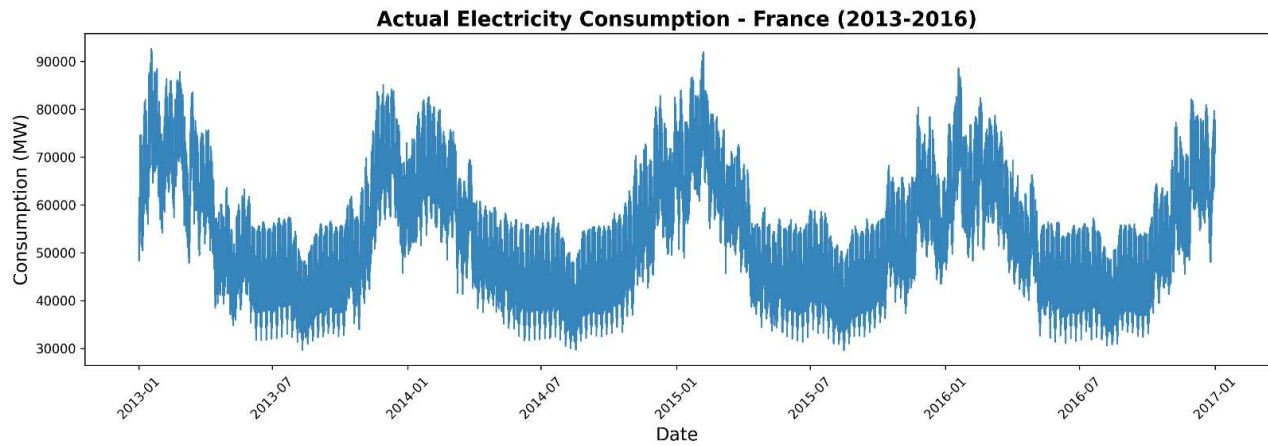


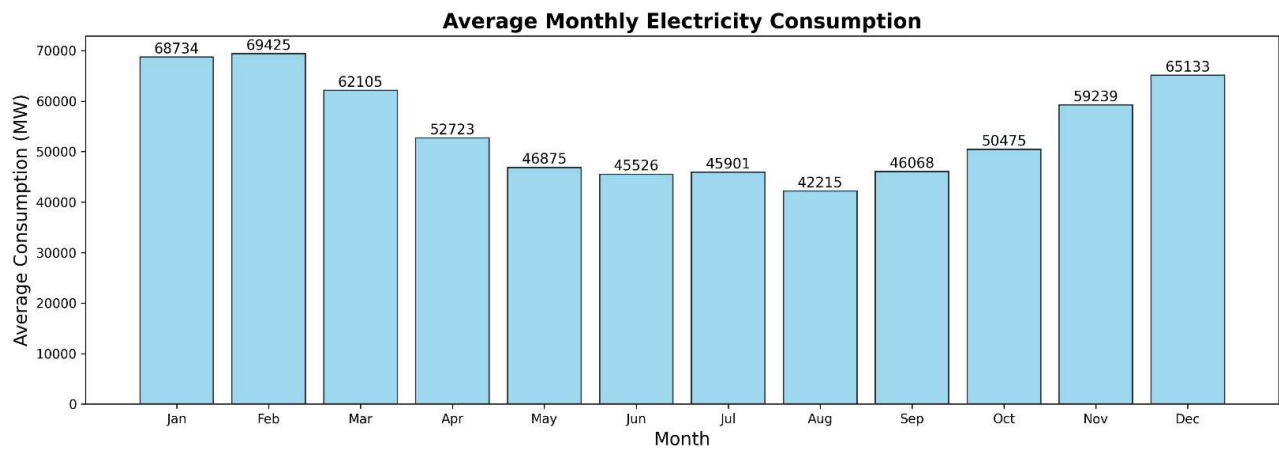
Fig. 5 Historical data for (a) Solar energy production, (d) Wind energy production and (c) Renewable energy generation

4.2 Numerical results and analysis

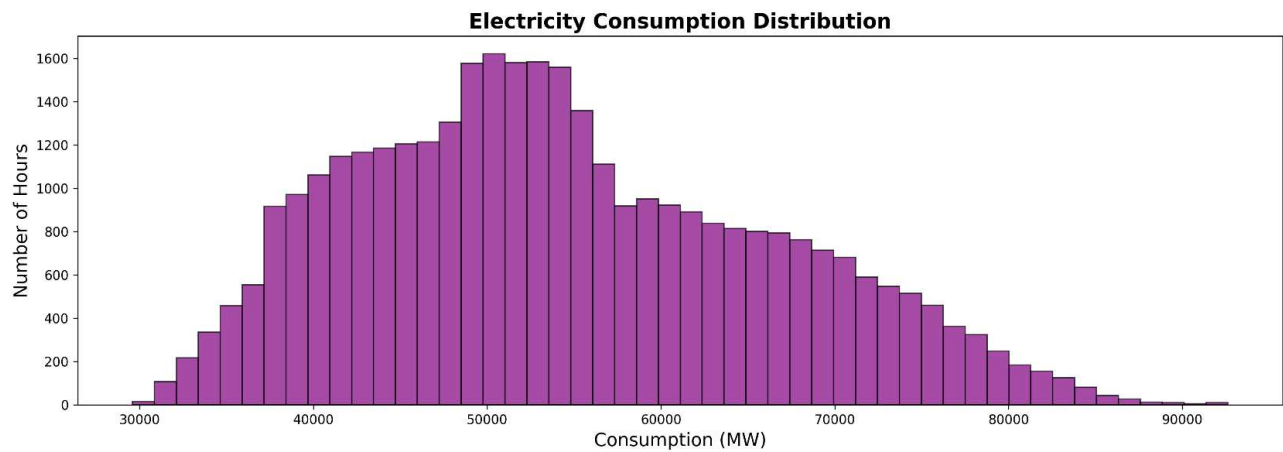
In this research, a four-year dataset containing hourly electric load measurements and corresponding meteorological variables was utilized Fig.6. The original time-indexed data were first merged using a nearest-timestamp matching strategy, followed by forward-backward filling to ensure temporal consistency. After preprocessing, the final dataset was passively divided into two groups: a training portion (80%) and a validation portion (20%), matching the structure of the original PyTorch implementation. The target variable represented the actual electrical load, while all other numerical variables including weather indicators and calendar-based features served as predictive inputs. All features and targets were normalized using independent Min/Max Scaler transformations to stabilize optimization and improve gradient behavior during training. A sliding-window strategy was applied to handle the sequential nature of the time-series data. Each sample was constructed by extracting a fixed-length historical window of 20 timesteps, which served as the model's input context. The forecasting horizon was set to one hour ahead, meaning that the model was tasked with predicting the immediate future load value based on the previous 20 hours of combined weather and energy-consumption history. This windowing method enables the model to leverage both short-term fluctuations and longer-term temporal dependencies. By creating overlapping windows, the training set becomes significantly more informative, allowing the deep model to better capture evolving seasonal, weekly, and daily patterns inherent in the data. To identify optimal architectural and training parameters, an ACO-based hyperparameter search was integrated into the pipeline. During this search phase, each ant corresponded to a candidate configuration in the hyperparameter space, sampling values such as the number of LSTM units, dropout rate, multi-head attention size, feed-forward dimensionality, Transformer depth, and learning rate. Each configuration underwent a short training cycle of four epochs, allowing rapid approximation of its effectiveness. The ants iteratively refined their solutions over multiple search rounds, progressively converging toward hyperparameters that minimized validation Mean Absolute Error. This biologically inspired exploration strategy improves robustness by avoiding premature convergence and enables the model to adaptively discover high-performance parameter sets without manual tuning. The final model adopted a hybrid architecture consisting of stacked LSTM layers followed by multiple Transformer encoder blocks. The LSTM layers capture long-range temporal dependencies and nonlinear patterns present in the time series, while the Transformer blocks enhance the model's ability to selectively attend to salient temporal features through multi-head self-attention mechanisms. After sequence encoding, an adaptive average pooling layer compresses timestep information, and a fully connected layer produces the final one-step-ahead prediction. This hybrid structure leverages both recurrent memory dynamics and attention-based feature selection, resulting in a more expressive and adaptive forecasting framework Fig.7. Following the ACO hyperparameter search, the best-performing configuration was selected for full-scale training. For the actual electrical load prediction task, the final model was trained for 200 epochs, whereas for solar power forecasting it was trained for 350 epochs. In both cases, training was performed using the Nadam optimizer, with Mean Squared Error (MSE) used as the loss function. The training process utilized shuffled mini-batches of size 64 and exploited GPU acceleration whenever available. Throughout training, both training and validation losses were monitored to ensure generalization stability. The extended training phase allowed the model to progressively refine its internal representation of the load-weather dynamics and minimize prediction error without overfitting.



(a)



(b)



(c)

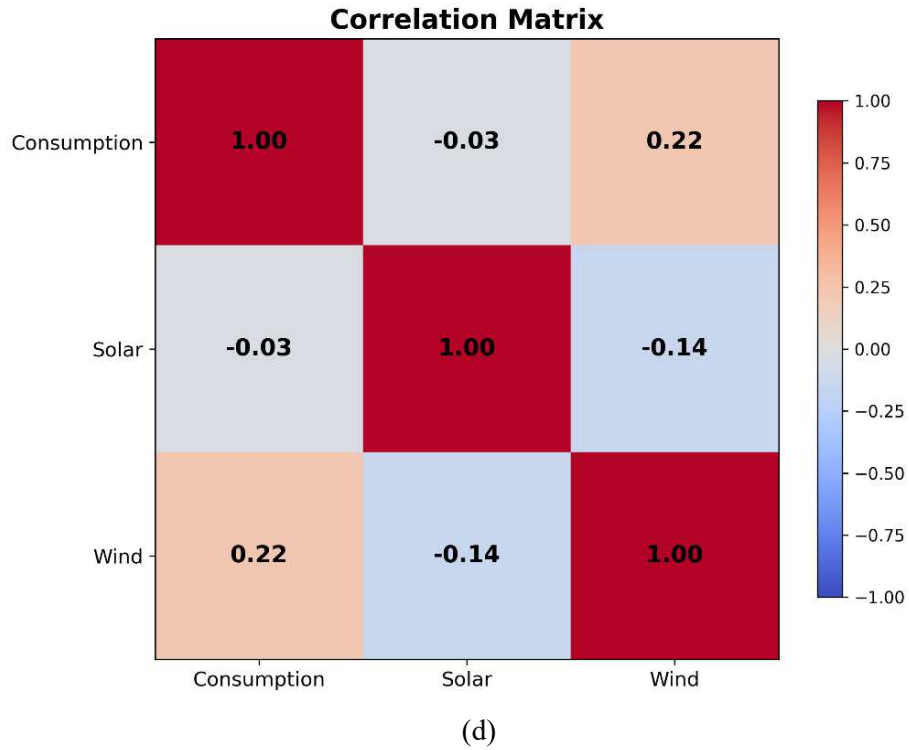


Fig. 6 Historical data for (a) Actual electricity consumption, (b) Average monthly electricity consumption (c) Electricity consumption and (d) Inputs and target dependencies and data relationship analysis using correlation matrix.

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	(None, 20, 23)	0	[]
lstm_1 (LSTM)	(None, 20, 128)	45,568	input_1[0][0]
dropout_1 (Dropout)	(None, 20, 128)	0	lstm_1[0][0]
lstm_2 (LSTM)	(None, 20, 128)	0	dropout_1[0][0]
dropout_2 (Dropout)	(None, 20, 128)	0	lstm_2[0][0]
transformer_1 (Transformer)	(None, 20, 64)	83,008	dropout_2[0][0]
global_avg_pool (GlobalAvgPool)	(None, 64)	0	transformer_1[0][0]
dense_output (Dense)	(None, 1)	65	global_avg_pool[0][0]

Fig.7. Summary of the LSTM-Transformer Model

Table 2 presents a comparative analysis of the performance of various time series actual electrical load forecasting models, including LSTM, ACO, and TF, as well as their hybrid combinations (e.g., LSTM-TF, LSTM-ACO). The evaluation of these models, which utilized different optimizers, is based on their MAE, MSE, and RMSE scores. The prediction model for time series actual electrical load forecasting was also evaluated using various optimizers, including Adam, Nadam, Adamax, and RMSprop. The results presented in the table unequivocally demonstrate the superior performance of the ACO-LSTM-TF model optimized with Nadam, which yields the minimum recorded error of 0.5065% in Mean Absolute Deviation (MAD). This outcome solidifies its efficacy as a premier forecasting tool for actual electrical load time series data. Furthermore, the selection of the optimizer is a critical factor that significantly influences model performance. In this case, the Nadam optimizer demonstrated superior performance for the ACO-LSTM-TF model in forecasting actual electrical load output. While Adam may not represent the optimal optimizer for the ACO-LSTM-TF model, its performance remains sufficiently noteworthy to warrant consideration. Despite Adam's satisfactory performance on the ACO-LSTM-TF model, the Nadam optimizer demonstrated superior prediction accuracy. These variations between Adam and Nadam can be attributed to the distinct convergence capabilities of each optimization algorithm in identifying optimal solutions tailored to the specific model architecture and dataset characteristics. The results presented in the table indicate that the TF-LSTM model utilizing the Nadam optimizer produced a MAE of 1.097%. This marks the highest MAE recorded in the study, signifying that the discrepancies between its predictions and the actual observed values were, on average, more substantial than those of all other tested configurations. The notably elevated MAE associated with the TF-LSTM and Nadam combination implies a limited capacity of this specific model-optimizer setup to effectively identify and learn the underlying patterns and trends within the actual electrical load time series data. Across all tested optimizers, the standalone LSTM model consistently outperformed the ACO and TF models in time series actual electrical load forecasting. A direct comparison of MAE, MSE, and RMSE metrics unequivocally demonstrated the superior predictive accuracy of LSTM over both ACO and TF. The ACO-LSTM-TF model exhibits a strong alignment with the actual electrical load data, whereas the TF-LSTM model appears to have certain constraints in precisely capturing its patterns. Figure 8 presents three subplots: (a) a frequency histogram of the ACO-LSTM-TF model's predicted values, (b) a scatter plot comparing predicted and actual values, and (c) a plot of training and validation losses. Together, these visualizations offer a detailed assessment of the ACO-LSTM-TF model, which is identified as the most accurate in terms of performance. Figure 8 a, displays the frequency histogram of the model's predictions. The distribution of these predicted values exhibits a strong alignment with the histogram of the actual values, indicating a well-fitted model. This close correspondence demonstrates the efficacy of the ACO-LSTM-TF model in accurately capturing the underlying patterns and trends inherent in the actual electrical load data. Figure 9 b, presents a scatter plot comparing predicted and observed values. The data points are tightly clustered along the diagonal, indicating a strong agreement between the model's predictions and the actual values, which demonstrates the model's accuracy and reliability. Figure 8 c, illustrates the training and validation loss, depicting the model's performance during its learning phase. The plotted loss values quantify the discrepancy between the model's predictions and the ground truth. The ACO-LSTM-TF model demonstrates a consistent downward trajectory in both training and validation loss, indicating its efficacy in minimizing the disparity between predicted outputs and actual values.

Table 2 Prediction models performance with respect to different optimizers

Optimizer Model	Adam			Nadam			Adamax			RMSprop			AMSGrad		
	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)
ACO	1.0369	0.0230	1.5162	0.9950	0.0214	1.4633	0.9849	0.0215	1.4677	1.3473	0.0356	1.8868	1.0745	0.0234	1.5301
LSTM	0.9957	0.0187	1.3688	0.9753	0.0176	1.3249	1.2747	0.0289	1.7000	1.6134	0.0376	1.9402	1.0574	0.0230	1.5162
TF	4.7888	0.3213	5.6685	3.9073	0.2251	4.7445	4.3593	0.2680	5.1766	4.2936	0.2438	4.9376	3.7120	0.2036	4.5119
ACO-LSTM	1.5817	0.0412	2.0300	1.2786	0.0297	1.7262	1.9304	0.0611	2.4734	5.5139	0.4454	6.6744	4.9407	0.3658	6.0481
LSTM-ACO	1.3792	0.0355	1.8858	3.8252	0.2241	4.6445	2.3914	0.1010	3.1786	3.1947	0.1695	4.1180	2.6523	0.1255	3.5431
ACO-TF	1.0672	0.0232	1.5232	1.1186	0.0232	1.5243	0.9897	0.0207	1.4390	1.3356	0.0334	1.8289	1.0444	0.0226	1.5020
TF-ACO	1.2567	0.0313	1.7690	1.2786	0.0320	1.7891	1.7179	0.0575	2.3982	1.7042	0.0517	2.2744	1.4522	0.0426	2.0649
LSTM-TF	1.4283	0.0356	1.8864	1.8949	0.0519	2.2791	1.4766	0.0394	1.9839	1.8389	0.0621	2.4924	1.8577	0.0627	2.5032
TF- LSTM	0.9788	0.0204	1.4281	1.0974	0.0249	1.5785	1.0704	0.0254	1.5942	1.5989	0.0454	2.1310	1.1497	0.0296	1.7212
ACO-LSTM -TF	0.7659	0.0130	1.1395	0.5065	0.0048	0.6929	1.0675	0.0212	1.4562	1.8346	0.0565	2.3771	0.9186	0.0175	1.3221
ACO-TF -LSTM	1.0157	0.0232	1.5221	0.9963	0.0225	1.5004	0.9521	0.0214	1.4620	1.4347	0.0410	2.0243	1.0291	0.0257	1.6021
LSTM-ACO -TF	0.7239	0.0121	1.0989	0.7394	0.0123	1.1079	0.8331	0.0153	1.2367	0.8527	0.0156	1.2506	1.0165	0.0212	1.4562
LSTM-TF - ACO	0.8782	0.0155	1.2437	0.7701	0.0137	1.1702	1.2891	0.0312	1.7668	1.3796	0.0359	1.8955	1.3812	0.0372	1.9281
TF- ACO - LSTM	0.7292	0.0124	1.1127	0.7896	0.0138	0.0138	0.7150	0.0119	1.0930	6.3076	0.6783	8.2360	0.7235	0.0120	1.0961
TF- LSTM - ACO	1.0522	0.0251	1.5853	1.0783	0.0266	1.6324	1.0882	0.0269	1.6390	1.5499	0.0486	2.2046	1.0987	0.0288	1.6976

Figure 9 comprises three panels: (a) a frequency histogram of predictions from the TF-LSTM model, (b) a scatter plot comparing predicted and actual values, and (c) a plot of the training and validation loss. Despite the TF-LSTM model yielding the least favorable results, the numerical data indicate a substantial deviation from the actual values. While the frequency histogram partially mirrors the distribution of the true values, the scatter plot reveals a significant dispersion of data points away from the diagonal line. This indicates that the model's predictions lack both accuracy and precision. Furthermore, the elevated training and validation loss values corroborate the discrepancy between the predicted and actual values.

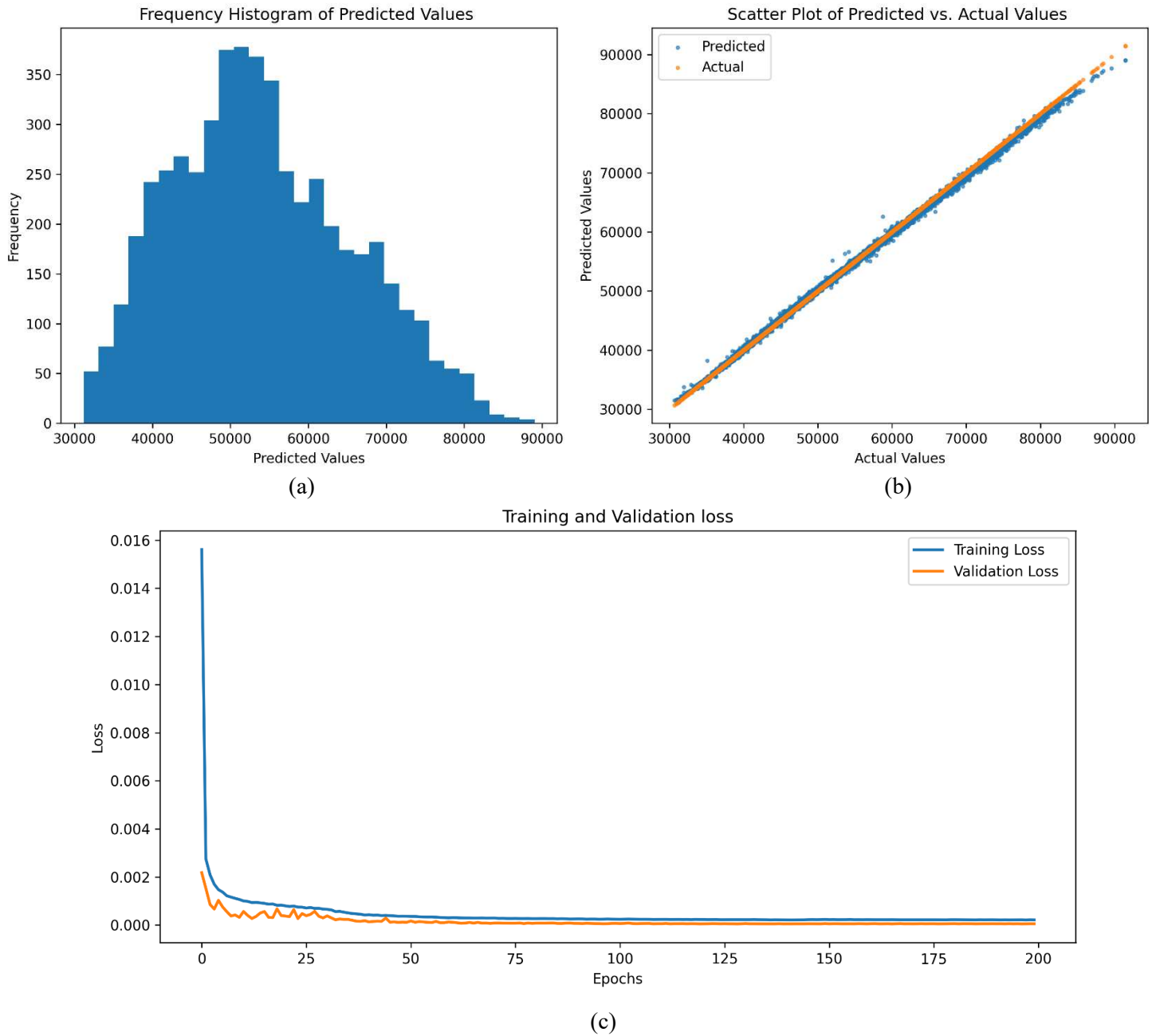


Fig. 8 ACO-LSTM-TF model performance, a) frequency histogram of predicted values. b) The scatter plot of predicted versus actual values. c) The training and validation loss plot.

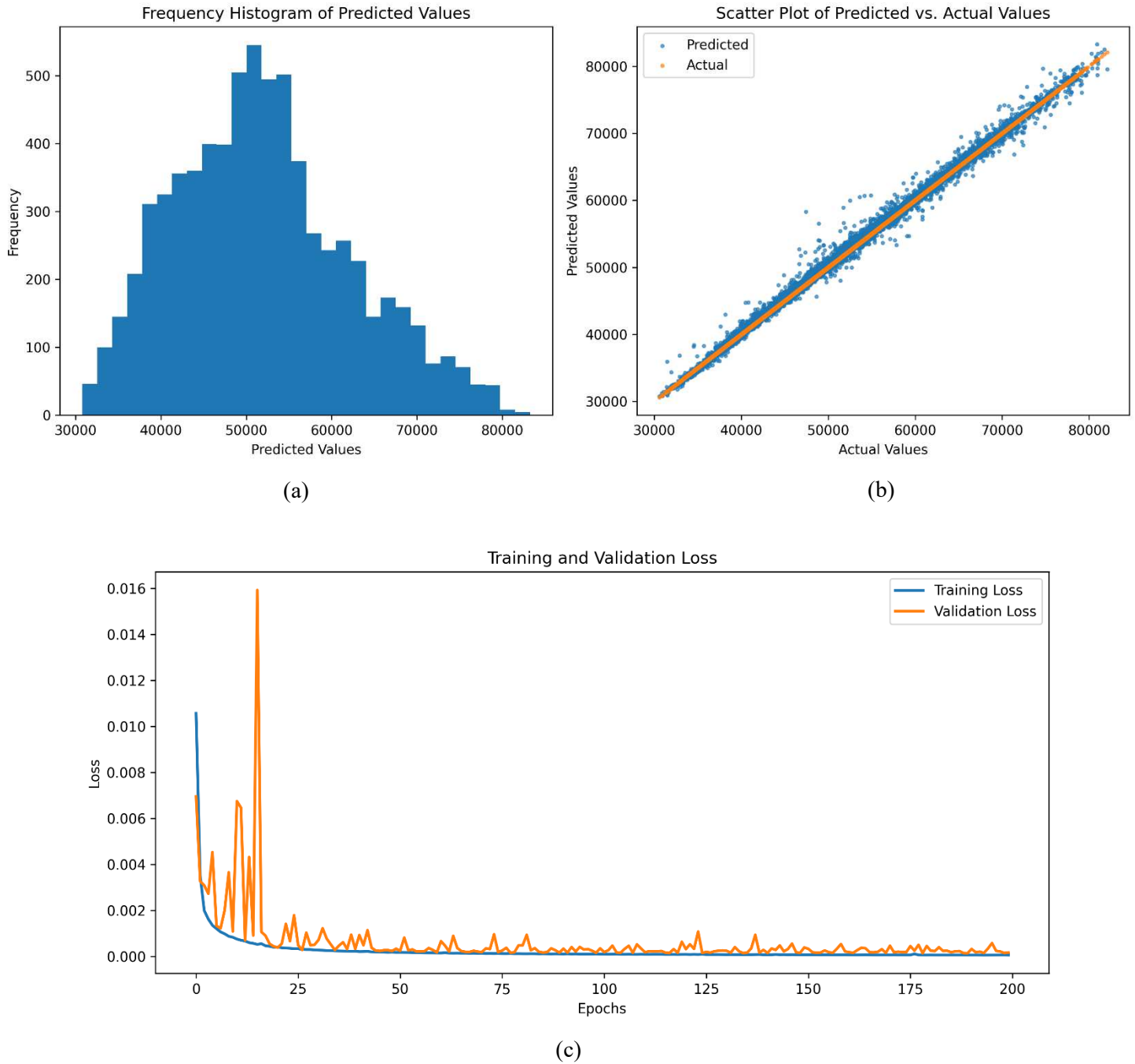


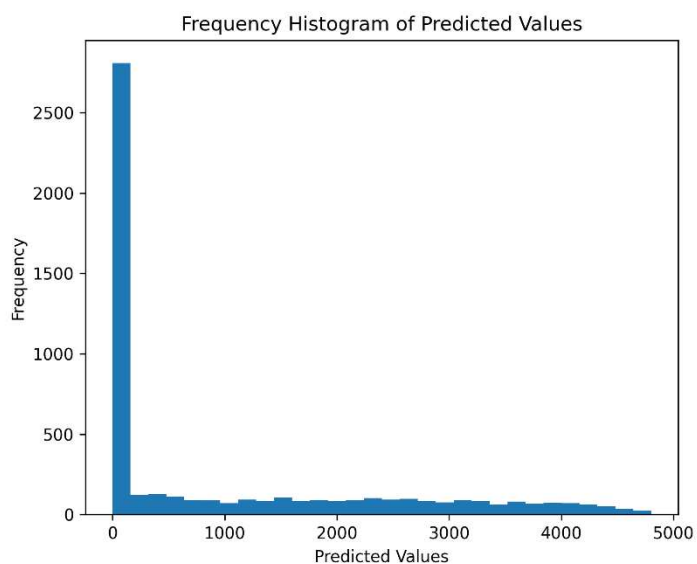
Fig. 9 TF-LSTM model performance, a) frequency histogram of predicted values. b) The scatter plot of predicted versus actual values. c) The training and validation loss plot

Table 3 presents a comparative analysis of the performance of diverse solar power forecasting models, evaluated with various optimizers, and provides a corresponding score for each configuration. The models considered in this study comprise LSTM, ACO, and TF, along with their hybrid combinations such as LSTM-TF, LSTM-ACO, TF-LSTM, ACO-LSTM, and ACO-TF, among others. To evaluate the predictive

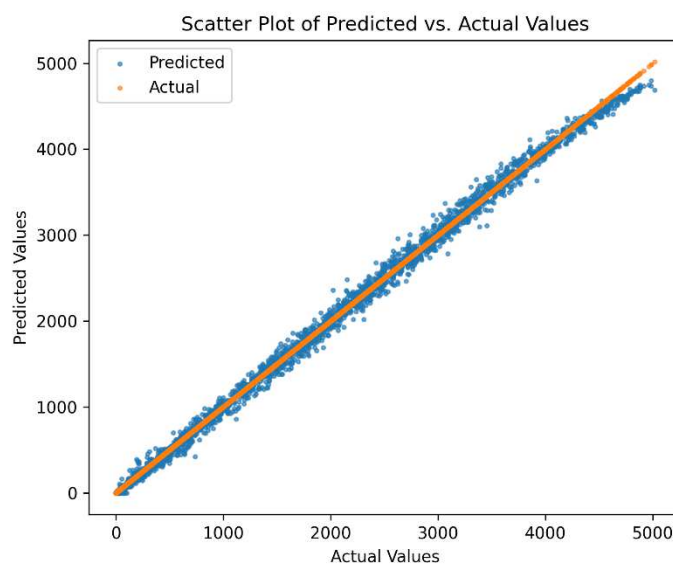
models for solar power time series forecasting, we employed MAE, MSE, and RMSE as performance metrics. The study also investigated the impact of various optimization algorithms, specifically Adam, Nadam, Adamax, and RMSprop, on the model's predictive accuracy. As presented in the table, the LSTM-TF-ACO model utilizing the Nadam optimizer significantly outperforms all other models. It records the lowest MAE of 2.878%, thereby establishing its effectiveness and robustness in forecasting solar power time series data. It should be noted that the choice of optimizer can influence model performance in various ways. In this case, Nadam was found to be the most effective optimizer for the LSTM-TF-ACO model applied to solar power forecasting. Adam may not be the most effective optimizer for the LSTM-TF-ACO architecture; however, its results are still worthy of mention. The Nadam optimizer demonstrates a clear advantage over Adam, providing even higher prediction accuracy for this specific model. Adam may not be the most effective optimizer for the LSTM-TF-ACO architecture; however, its results are still worthy of mention. The Nadam optimizer demonstrates a clear advantage over Adam, providing even higher prediction accuracy for this specific model. The results presented in the table indicate that the TF-ACO model utilizing the Nadam optimizer produced a MAE of 6.8054%. This constitutes the highest MAE observed in the study, indicating that its predictions deviated from the actual observed values, on average, more significantly than any other configuration tested. The suboptimal performance of the TF-ACO and Nadam combination, evidenced by its notably elevated MAE, indicates a deficiency in its ability to decipher and model the underlying trends inherent in the actual solar power forecasting dataset. The standalone LSTM model demonstrated superior performance compared to both the ACO and TF models across all optimizers evaluated for time-series forecasting of actual solar power. The superiority of the LSTM model in predictive accuracy was unequivocally established through a direct comparison of MAE, MSE, and RMSE with the ACO and TF models. The results indicate a strong correlation between the LSTM-TF-ACO model and the actual solar power time series, in contrast to the TF-ACO model, which exhibits limitations in precisely replicating its dynamic behavior. Figure 10 presents three subplots: (a) a frequency histogram of the LSTM-TF-ACO model's predicted values, (b) a scatter plot comparing predicted and actual values, and (c) a plot of training and validation losses. Figure 11 comprises three panels: (a) a frequency histogram of predictions from the TF-ACO model, (b) a scatter plot comparing predicted and actual values, and (c) a plot of the training and validation loss.

Table 3 Prediction models performance with respect to different optimizers

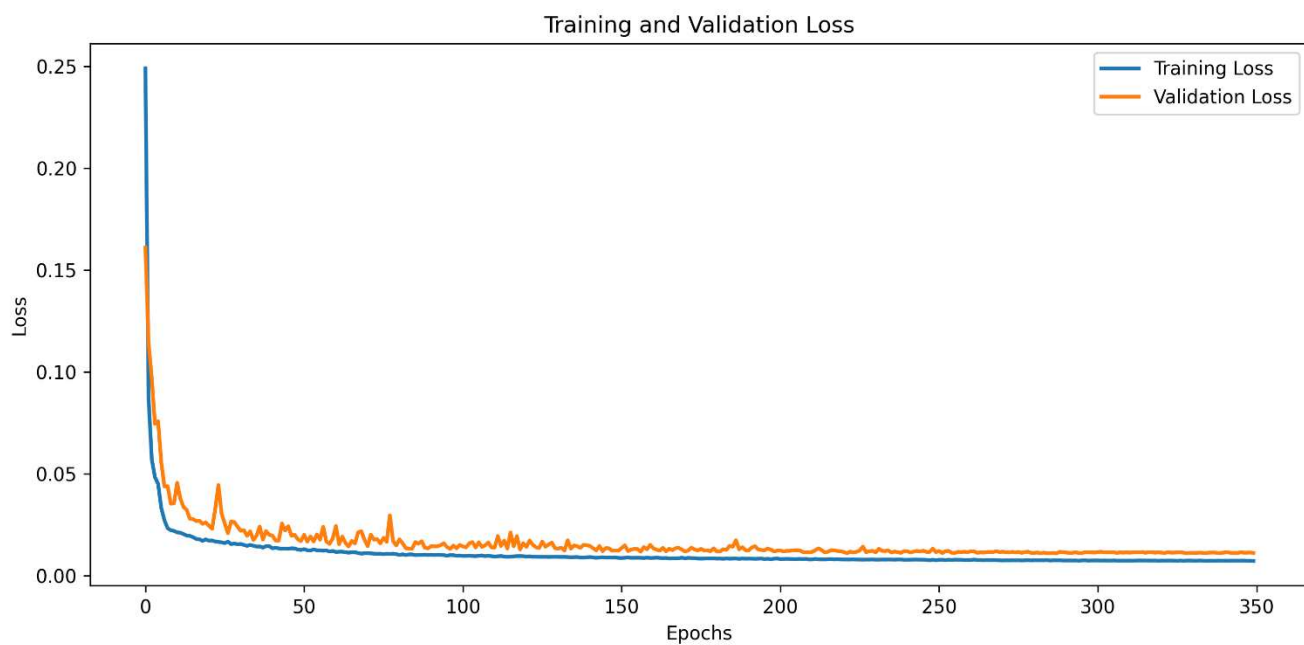
Optimizer	Adam			Nadam			Adamax			RMSprop			AMSGrad		
Model	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)
ACO-LSTM	5.8024	1.1678	10.806	5.5751	1.2989	11.396	6.3023	1.3753	11.727	5.8741	1.2820	11.322	6.5890	1.3830	11.760
LSTM-ACO	5.5477	1.1655	10.796	5.6547	1.2026	10.966	5.9499	1.2735	11.284	9.1132	2.4873	15.771	6.7078	1.5246	12.347
ACO-TF	9.6238	2.5783	16.057	12.531	4.6049	21.459	11.026	3.2205	17.945	10.122	2.9556	17.191	6.8427	1.5881	12.602
TF-ACO	6.8877	1.7651	13.285	6.8054	1.7659	13.288	8.6913	2.5183	15.869	7.7853	2.0778	14.414	7.4522	1.6426	14.064
ACO-LSTM-TF	5.7554	1.1602	10.771	5.4411	1.1485	10.717	7.8669	1.7642	13.282	30.359	39.091	44.600	7.0356	1.4849	12.185
ACO-TF-LSTM	6.4724	1.5310	12.373	6.3499	1.4265	11.943	33.6317	29.914	51.629	47.616	64.318	93.992	6.4338	1.4586	12.077
LSTM-ACO-TF	7.1072	1.5593	12.487	5.9103	1.2181	11.036	6.7906	1.6024	12.658	6.7244	1.4534	12.055	7.2637	1.5993	12.646
LSTM-TF-ACO	5.7015	1.1565	10.754	2.8780	0.2758	5.2518	5.8848	1.2302	11.091	5.5696	1.1901	10.909	6.1257	1.2675	11.258
TF-ACO-LSTM	6.3418	1.6090	12.684	6.5887	1.6044	12.6665	8.4340	1.9112	13.824	6.3076	0.6783	8.2360	7.5882	1.6692	12.919
TF-LSTM-ACO	27.748	35.989	59.991	23.342	25.430	50.429	24.851	28.373	53.267	24.605	27.693	52.624	25.783	29.683	54.482



(a)

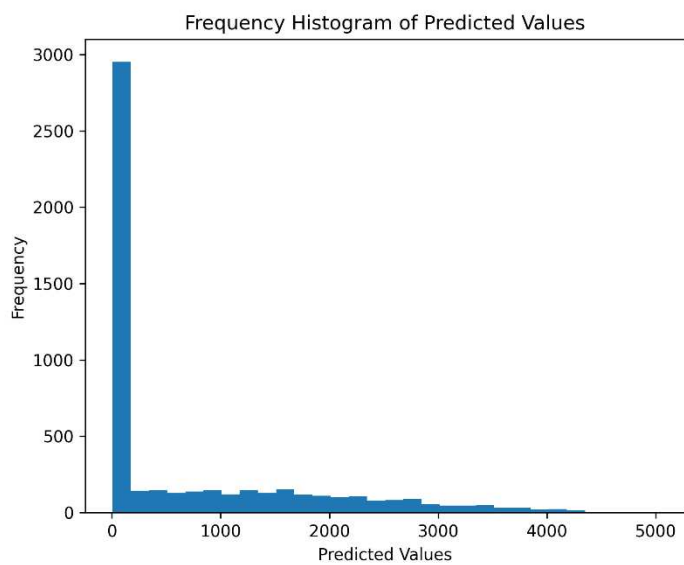


(b)

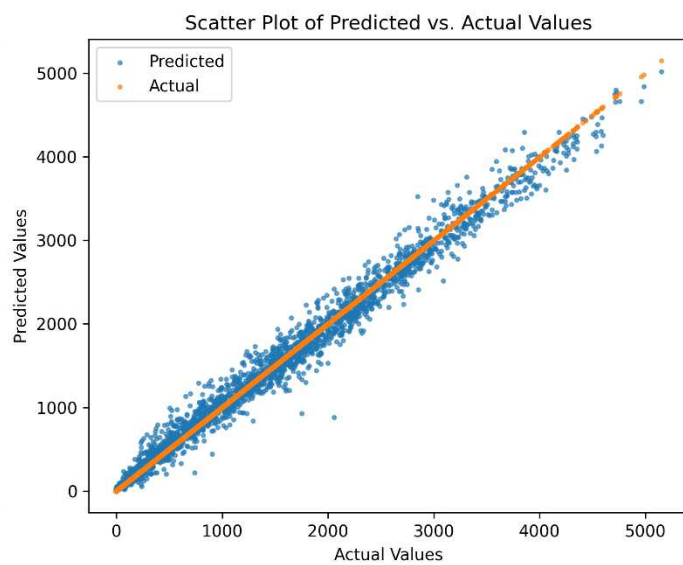


(c)

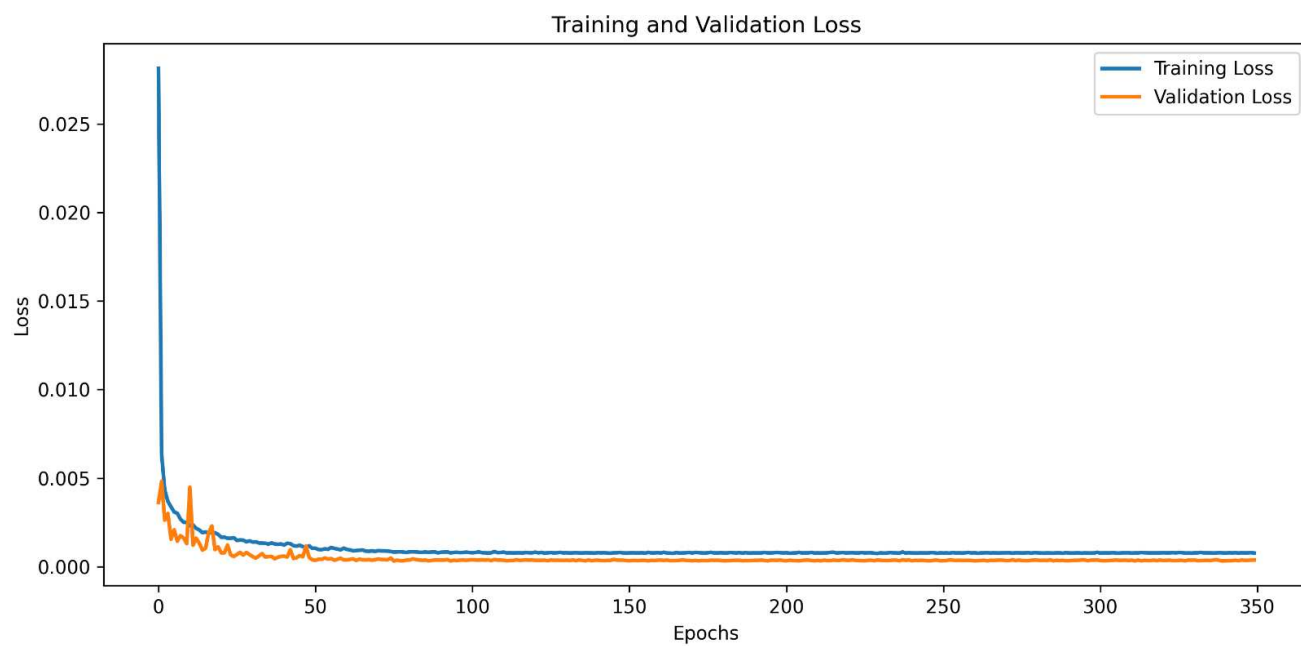
Fig. 10 LSTM-TF- ACO model performance, a) frequency histogram of predicted values. b) The scatter plot of predicted versus actual values. c) The training and validation loss plot



(a)



(b)



(c)

Fig. 11 TF- ACO model performance, a) frequency histogram of predicted values. b) The scatter plot of predicted versus actual values. c) The training and validation loss plot

Table 4 presents a comprehensive comparison of various deep learning and hybrid forecasting models for solar power generation prediction under five widely used optimization algorithms, namely Adam, Nadam, Adamax, RMSprop, and AMSGrad. Recent studies have emphasized the effectiveness of hybrid deep learning architectures for solar time series forecasting [29]. The performance assessment is primarily conducted based on the MAE, which is one of the most informative evaluation metrics in renewable energy forecasting because it directly quantifies the average prediction deviation from the actual generated power. A lower MAE value indicates higher forecasting accuracy and better model reliability for real-world photovoltaic applications. The experimental results reveal significant differences among the investigated architectures, highlighting the critical role of model design in determining forecasting performance. It can be observed that models relying exclusively on CNN and Transformer structures exhibit relatively poor prediction accuracy. The CNN-TF architecture produced MAE values ranging between 15.04% and 15.58%, whereas the TF-CNN model generated MAE values between 14.67% and 14.96%. These results suggest that the combination of convolutional feature extraction and attention mechanisms alone is insufficient for accurately modeling solar power generation dynamics.

Solar energy production is inherently a temporal process influenced by historical irradiance patterns, weather conditions, cloud movement, seasonal variability, and environmental uncertainties. Although CNN can effectively extract local patterns and Transformer can identify long-range dependencies through self-attention mechanisms, neither architecture possesses an explicit memory structure capable of preserving sequential information over time. Consequently, the absence of recurrent learning significantly limits the forecasting capability of these models. The importance of temporal modeling becomes evident when Long Short-Term Memory (LSTM) networks are incorporated into the forecasting framework. The CNN-LSTM model achieved MAE values close to 2.15%, representing a substantial improvement compared with CNN-TF and TF-CNN architectures. This remarkable reduction in forecasting error confirms that temporal information plays a dominant role in photovoltaic power prediction, a finding consistent with the results reported by Salman et al. [29], who demonstrated that integrating LSTM with CNN significantly enhances forecasting accuracy compared to attention-only models. The memory cells and gating mechanisms of LSTM networks enable the model to learn both short-term fluctuations and long-term temporal trends, which are essential for accurately forecasting solar energy output.

A further reduction in prediction error is observed in the LSTM-CNN architecture. This model achieved a minimum MAE value of 0.619%, making it one of the most accurate forecasting models presented in the study. Interestingly, the results indicate that the sequence of model components significantly affects forecasting performance. When LSTM is positioned before CNN, temporal information is extracted first and subsequently refined by convolutional operations. This architecture appears to preserve critical sequential characteristics while simultaneously enhancing feature representation quality. The MAE reduction from 2.149% in CNN-LSTM to 0.619% in LSTM-CNN corresponds to an improvement of approximately 71%, demonstrating the importance of architectural ordering in hybrid deep learning systems.

Among all investigated models, the CNN-LSTM-TF architecture achieved the best overall forecasting accuracy. Using the Nadam optimizer, this model obtained an MAE of only 0.551%, representing the lowest prediction error reported in Table 4. The superior performance of this architecture can be attributed to the complementary strengths of its three principal components. The CNN layer extracts local and spatial patterns from meteorological and historical power data, the LSTM layer captures temporal dependencies and sequential relationships, and the Transformer layer learns long-range interactions through self-attention. As a result, the model benefits simultaneously from local feature extraction, temporal memory, and global contextual awareness.

The performance improvement achieved by CNN-LSTM-TF over CNN-LSTM is particularly noteworthy. The MAE decreases from 2.149% to 0.551%, corresponding to an improvement of approximately 74.4%. Such a significant reduction demonstrates that Transformer-based attention mechanisms can substantially enhance forecasting accuracy when integrated after effective temporal feature extraction. Therefore, the results suggest that Transformer networks are most beneficial when employed as high-level contextual learning modules rather than standalone forecasting engines. Similarly, TF-LSTM-CNN and LSTM-TF-CNN also achieved excellent performance, with minimum MAE values of 0.589% and 0.608%, respectively. These findings further confirm that the integration of recurrent and attention-based learning mechanisms provides an effective framework for modeling the complex nonlinear behavior of photovoltaic systems. Although their forecasting accuracy is slightly inferior to CNN-LSTM-TF, the differences remain relatively small, indicating that several architectural configurations can successfully exploit the complementary characteristics of LSTM and Transformer networks.

Despite the superior MAE achieved by CNN-LSTM-TF, the proposed LSTM-TF-ACO model deserves particular attention because its contribution extends beyond pure forecasting accuracy. The best MAE obtained by the proposed model was 2.878% under the Nadam optimizer, while MAE values of 5.701%, 5.885%, 5.570%, and 6.125% were obtained under Adam, Adamax, RMSprop, and AMSGrad, respectively. Although these values are higher than those achieved by the best-performing deep learning architectures, the proposed framework introduces several important advantages that are highly relevant for practical renewable energy forecasting applications. The first major advantage of the LSTM-TF-ACO model is its enhanced optimization capability. Deep learning models typically contain numerous hyperparameters, including learning rates, hidden dimensions, attention heads, sequence lengths, and network depths. The performance of such models is highly sensitive to these parameters, and conventional gradient-based optimization methods often become trapped in local minima. By integrating ACO, the proposed framework introduces a population-based search mechanism capable of exploring a broader solution space. This characteristic increases the probability of identifying near-optimal parameter configurations and reduces the likelihood of premature convergence.

A second important advantage is improved robustness. Forecasting models deployed in real-world photovoltaic systems must operate under highly uncertain environmental conditions. Solar irradiance variability, cloud coverage, atmospheric disturbances, seasonal transitions, and sensor noise can significantly affect prediction performance. The stochastic search behavior of ACO provides an additional level of adaptability that enhances model resilience against such uncertainties. Consequently, the proposed architecture may exhibit superior stability when applied to datasets characterized by higher variability or greater noise levels. The third advantage relates to generalization capability. Models achieving extremely low training and testing errors may occasionally suffer from overfitting, particularly when dealing with limited datasets. The ACO-based optimization process promotes broader exploration of the parameter space and can prevent excessive dependence on local patterns within the training data. As a result, the LSTM-TF-ACO framework may offer improved generalization performance when transferred to unseen operating conditions, geographical locations, or climatic environments. Another notable advantage is the synergistic integration of temporal memory and global attention mechanisms. The LSTM component effectively captures sequential information and temporal dependencies, while the Transformer module identifies long-range interactions among input features through self-attention. Together, these components provide a comprehensive representation of photovoltaic generation behavior. The subsequent optimization by ACO further enhances parameter selection, creating a hybrid intelligence framework that combines deep learning and swarm intelligence principles. From an engineering perspective, the proposed model offers greater flexibility and scalability. Unlike purely deep architectures that depend entirely on gradient-based learning, the integration of ACO enables adaptation to a wider range of optimization objectives. Future extensions may incorporate multi-objective optimization strategies involving forecasting accuracy, computational

complexity, energy consumption, and model interpretability. Therefore, the proposed architecture establishes a foundation for next-generation intelligent forecasting systems within smart grid environments. The optimizer comparison presented in Table 4 provides additional insights. Nadam consistently produced the lowest MAE values across most architectures. For the proposed LSTM-TF-ACO model, the MAE decreased from 5.701% using Adam to 2.878% using Nadam, representing an improvement of nearly 50%. This behavior can be attributed to the combination of adaptive learning rates and Nesterov momentum incorporated within Nadam, which enables faster convergence and more efficient navigation of complex loss landscapes. These findings suggest that optimizer selection remains an important factor; however, its influence is generally smaller than the impact of architectural design. Overall, the experimental results demonstrate that architecture selection is the dominant factor affecting forecasting accuracy in solar power generation prediction. Hybrid models combining CNN, LSTM, and Transformer components achieved the lowest MAE values due to their ability to simultaneously exploit spatial, temporal, and global contextual information. Nevertheless, the proposed LSTM-TF-ACO framework offers unique advantages in terms of optimization intelligence, robustness, adaptability, and generalization capability. While its forecasting accuracy does not surpass the best-performing CNN-LSTM-TF architecture, its integration of swarm intelligence provides a valuable alternative for practical applications where optimization flexibility and resilience are equally important as prediction accuracy.

In conclusion, the results confirm that hybrid deep learning architectures represent a promising direction for advanced photovoltaic forecasting systems [29]. The CNN-LSTM-TF model achieved the highest prediction accuracy, whereas the proposed LSTM-TF-ACO framework introduced additional optimization and robustness benefits that may become increasingly important in real-world renewable energy environments. Consequently, the proposed model represents a meaningful contribution toward the development of intelligent, adaptive, and reliable forecasting systems for future smart grids and sustainable energy infrastructures.

Table 4 Comparative Analysis of Forecasting Accuracy

Optimizer Model	Adam			Nadam			Adamax			RMSprop			AMSGrad		
	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)	MAE (%)	MSE (%)	RMSE (%)
CNN-TF [29]	15.58	3.605	18.98	15.08	3.455	18.58	15.41	3.525	18.77	15.056	3.421	18.49	15.04	3.406	18.45
TF-CNN [29]	14.88	5.025	22.41	14.91	4.993	22.35	14.67	4.951	22.25	14.96	5.062	22.498	14.76	5.022	22.198
CNN-LSTM[29]	2.184	0.113	3.365	2.149	0.112	3.351	6.658	0.7393	8.598	2.689	0.164	4.0619	2.164	0.112	3.3501
LSTM-CNN[29]	0.683	0.014	1.2198	0.653	0.010	1.0375	0.927	0.0223	1.493	0.619	0.012	1.1187	0.735	0.014	1.245
LSTM-CNN- TF[29]	14.6	4.061	6.372	15.43	3.61	18.988	15.75	3.61	19.01	14.94	3.420	18.445	15.16	3.433	18.53
CNN-LSTM- TF[29]	0.562	0.010	1.039	0.551	0.01	1.001	0.826	0.017	1.328	0.584	0.010	1.0402	0.598	0.011	1.0644
CNN-TF- LSTM[29]	1.929	0.095	3.093	2.343	0.132	3.6393	2.293	0.1353	3.678	1.838	0.091	3.0244	2.295	0.126	3.559
LSTM-TF- CNN [29]	0.616	0.011	1.0878	0.725	0.016	1.2867	1.543	0.0718	2.681	0.683	0.013	1.1492	0.608	0.011	1.0507
TF-LSTM- CNN [29]	0.773	0.017	1.339	0.611	0.012	1.1023	0.714	0.0148	1.217	0.5886	0.011	1.0667	0.601	0.011	1.0701
TF-CNN- LSTM[29]	15.63	3.606	18.99	15.52	3.605	18.987	15.42	3.6058	18.98	15.54	3.605	18.987	15.59	3.605	18.988
LSTM-TF – ACO [This Work]	5.701	1.156	10.754	2.878	0.275	5.2518	5.8848	1.2302	11.091	5.5696	1.190	10.909	6.125	1.267	11.258

5. Conclusion

Accurate forecasting of solar power generation plays an essential role in the advancement of renewable energy and environmental sustainability. Given its abundance and high value, solar energy holds considerable promise for diversifying the global energy portfolio and reducing dependence on fossil fuels. Nevertheless, the inherent intermittency and unpredictability of solar power pose significant challenges to its seamless integration into electrical grids. Reliable solar power forecasting substantially facilitates the integration, planning, and operation of photovoltaic systems. It enhances grid stability, enables efficient energy management, simplifies participation in electricity markets, and supports evidence-based policymaking. Ultimately, such forecasting accelerates the global transition toward a sustainable, low-carbon future by optimizing the utilization of solar resources. Accurate prediction of solar power significantly enhances various aspects of energy management and strategic decision-making. Moreover, a comparative evaluation of deep learning approaches for time-series forecasting of solar energy systems underscores the superior capability of hybrid models in extracting the intricate, inherent data patterns. For solar power forecasting, the hybrid LSTM-TF-ACO model, when trained with the Nadam optimizer, outperforms all other models in terms of improvement, attaining an MAE of 2.878%. The data also show how the performance of the model varies depending on which optimizer is used, with Nadam and Adam consistently producing the best results. The obtained results underscore the critical role of choosing a suitable predictive model and optimization algorithm in solar power forecasting. In addition, these outcomes present meaningful implications for the optimal design and strategic planning of renewable energy infrastructures. Moreover, the derived conclusions offer practical guidance for academics, energy sector professionals, and decision-makers, enabling them to deploy such models effectively thereby enhancing both the operational efficiency and grid integration of solar energy systems. The integration of hybrid deep learning architectures in the present study enhances the precision of solar power forecasting. Nevertheless, it is essential to recognize and address the inherent limitations of the proposed approach. Given that the analysis relies on a specific time-series power dataset, the generalizability of these models to other contexts may be restricted. To overcome this shortcoming, subsequent investigations should prioritize the examination of diverse datasets in order to assess the robustness of the models under varying data conditions. Although the inquiry is helpful, it does not do an analysis of all of the hyperparameters, which may cause it to ignore the configurations that give the best results. A comprehensive investigation of hyperparameters is crucial for achieving a thorough understanding of the model's performance in subsequent research. These parameters play a key role in refining and broadening our methodologies for future research objectives. Future research in solar power forecasting is expected to focus on the adoption of advanced machine learning algorithms, the exploration of synergistic relationships among different predictive models, and the examination of temporal dynamics to achieve higher predictive accuracy. Moreover, assessing the influence of novel technologies such as enhanced weather prediction systems on solar forecasting holds significant potential for further methodological refinement and innovation in the approaches employed for estimating solar power generation.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

The data used in this study are publicly available from Figshare under the title “Historical Data in Simulation (Data in France)” at <https://doi.org/10.6084/m9.figshare.11341727>. These publicly available data were used for data preprocessing, model development, training, and evaluation presented in this study. The processed datasets and simulation results generated during the current study are available from the corresponding author on reasonable request.

Author contributions

M.S. conceived and designed the study, performed the formal analysis, conducted the investigation, visualized the data, and wrote the original draft. A.M. and V.M. contributed to the methodology, validation, and project administration, and reviewed and edited the manuscript. All authors reviewed and approved the final version of the manuscript.

References

1. Zheng J, Zhang H, Dai Y, Wang B, Zheng T, Liao Q, Liang Y, Zhang F, Song X. Time series prediction for output of multi-region solar power plants. *Applied Energy* 2020;257:114001. <https://doi.org/10.1016/j.apenergy.2019.114001>.
2. Sharadga H, Hajimirza S, Balog RS. Time series forecasting of solar power generation for large-scale photovoltaic plants. *Renewable Energy* 2020;150:797-807.
3. Wan C, Zhao J, Song Y, Xu Z, Lin J, Hu Z. Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems* 2015;1(4):38–46.
4. Ye H, Yang B, Han Y, Chen N. State-of-the-art solar energy forecasting approaches: Critical potentials and challenges. *Frontiers in Energy Research* 2022;10:868034.
5. Kumari P, Toshniwal D. Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production* 2021;318:128566.
6. Sharma J. A novel long-term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer: A case study of India. *Energy Science & Engineering* 2022;10(8):2909–2929.
7. Mulyadi A, Djamal EC. Sunshine duration prediction using 1D convolutional neural networks. In: *Proceedings of the 6th International Conference on Instrumentation, Control, and Automation (ICA)*; 2019. p. 77-81.
8. Al-Ali EM, Hajji Y, Said Y, Hleili M, Alanzi AM, Laatar AH, Atri M. Solar energy production forecasting based on a hybrid CNN–LSTM–Transformer model. *Mathematics* 2023;11(3):676. <https://doi.org/10.3390/math11030676>.

9. Benti NE, Chaka MD, Semie AG. Forecasting renewable energy generation with machine learning and deep learning: Current advances and future prospects. *Sustainability* 2023;15(9):7087.
10. Rajagukguk RA, Ramadhan RAA, Lee HJ. A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. *Energies* 2020;13(24):6623.
11. Husein M, Chung IY. Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. *Energies* 2019;12(10):1856. <https://doi.org/10.3390/en12101856>.
12. Elsaraiti M, Merabet A. Solar power forecasting using deep learning techniques. *IEEE Access* 2022;10:31692-31698. <https://doi.org/10.1109/ACCESS.2022.3160853>.
13. Kim EG, Akhtar MS, Yang OB. Designing solar power generation output forecasting methods using time series algorithms. *Electric Power Systems Research* 2023;216:109073. <https://doi.org/10.1016/j.epsr.2022.109073>.
14. Song X, Shi Y, Wang S, et al. Time-series well performance prediction based on long short-term memory neural network model. *Journal of Petroleum Science and Engineering* 2020;186:106682. <https://doi.org/10.1016/j.petrol.2019.106682>.
15. Zhao Y, Wang G, Tang C, Luo C, Zeng W, Zha ZJ. A battle of network structures: An empirical study of CNN, Transformer, and MLP. *arXiv* 2021. <https://doi.org/10.48550/arXiv.2108.13002>.
16. Priyadarshi R, Kumar RR. Evolution of swarm intelligence: A systematic review of particle swarm and ant colony optimization approaches in modern research. *Archives of Computational Methods in Engineering* 2025;32(6):3609-3650. <https://doi.org/10.1007/s11831-024-10164-5>.
17. Al-Abri ES, Samuels TA, Steidley C, Hossain MM. Modelling atmospheric ozone concentration using machine learning algorithms. *PLoS ONE* 2018;13(3):e0194889. <https://doi.org/10.1371/journal.pone.0194889>
18. Dorigo M. *Optimization, Learning and Natural Algorithms*. PhD dissertation. Milan, Italy: Politecnico di Milano; 1992.
19. Socha K, Dorigo M. Ant colony optimization for continuous domains. *European Journal of Operational Research* 2008;185(3):1155-1173. <https://doi.org/10.1016/j.ejor.2006.06.046>.
20. Xiao J, Huang Y, Lou HH. A population-based probability distribution estimation method for dynamic optimization. In: *Proceedings of the AIChE Annual Meeting*; San Francisco, CA, USA; 2006.
21. Liu CH, Gu JC, Yang MT. A simplified LSTM neural network for one-day-ahead solar power forecasting. *IEEE Access* 2021;9:17174-17195. <https://doi.org/10.1109/ACCESS.2021.3053797>.

22. Brahma B, Wadhvani R. Solar irradiance forecasting based on deep learning methodologies and multi-site data. *Symmetry* 2020;12(11):1883. <https://doi.org/10.3390/sym12111883>.
23. Kumar D, Mathur HD, Bhanot S, Verma A. Forecasting of solar and wind power using LSTM recurrent neural network for load frequency control in isolated microgrid. *International Journal of Modelling and Simulation* 2021;41(4):311-323. <https://doi.org/10.1080/02286203.2020.1773935>.
24. Le XH, Ho HT, Lee G, Jung S. Application of long short-term memory neural network for flood forecasting. *Water* 2019;11(7):1387. <https://doi.org/10.3390/w11071387>.
25. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Guyon I, von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Red Hook, NY: Curran Associates Inc.; 2017. p. 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>.
26. Wu N, Green B, Ben X, O'Banion S. Deep Transformer models for time series forecasting: The influenza prevalence case. *arXiv* 2020. <https://doi.org/10.48550/arXiv.2001.08317>.
27. Dalianis H. Evaluation metrics and evaluation. In: *Clinical Text Mining*. Cham: Springer; 2018. p. 45–53. https://doi.org/10.1007/978-3-319-78503-5_6
28. Huang W. Historical data in simulation (Data in France). *Figshare*. Dataset. 2019. <https://doi.org/10.6084/m9.figshare.9467144>.
29. Salman D, Direkoglu C, Kusaf M, Fahrioglu M. Hybrid deep learning models for time series forecasting of solar power. *Neural Computing and Applications* 2024;36(16):9095–9112. <https://doi.org/10.1007/s00521-024-09558-5>.