STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **x**, Month 202x, pp 0–16.
Published online in International Academic Press (www.IAPress.org)

# Metaheuristic-Optimized Deep Learning for Smart Solar Energy Management: A Comprehensive Review

Mohammed Aloqaily [1], Fadl Dahan [2,*]

[1] *Department of Management Information Systems, College of Business Administration - Hawtat Bani Tamim, Prince Sattam bin Abdulaziz University, Al-Kharj, 11942, Saudi Arabia*
[2] *Faculty of Education, Arts and Sciences (in Al-Turba), Taiz University, Taiz, 9674, Yemen*

**Abstract** The accurate forecasting of solar power generation is a cornerstone of smart energy management, directly affecting dispatch scheduling, battery storage operation, demand response, and the overall stability of grids with high photovoltaic (PV) penetration. Deep learning (DL) architectures, including long short-term memory (LSTM) networks, gated recurrent units (GRU), convolutional neural networks (CNN), and, more recently, attention-based transformers, have become the dominant tool for this task because of their capacity to model the nonlinear and non-stationary behaviour of solar irradiance and PV output. However, the predictive accuracy of these architectures is highly sensitive to their hyperparameters, and manual or grid-search tuning is computationally expensive and rarely converges to a global optimum. Metaheuristic optimization algorithms, inspired by evolutionary, swarm, physics-based, and human-based phenomena, have therefore been widely hybridized with DL models to automate and improve hyperparameter search, feature selection, and, in some cases, network weight optimization. This paper presents a structured and up-to-date review of metaheuristic-optimized deep learning algorithms for solar power and irradiance forecasting within smart energy management systems. We propose a taxonomy of the most commonly used metaheuristics, summarize the architectural characteristics of the deep learning backbones they are paired with, and synthesize quantitative results reported across recent studies, including Fire Hawk Optimization (FHO), the Improved Mountain Gazelle Optimizer (IMGO), the Evolutionary Mating Algorithm (EMA), the Reptile Search Algorithm (RSA), the Grey Wolf Optimizer (GWO), and the Coati Optimization Algorithm (COA). Comparative tables and figures illustrate the relative strengths, computational trade-offs, and reported accuracy gains of these hybrid frameworks, with several studies reporting coefficients of determination ($R^2$) above 0.98 after metaheuristic tuning. The review concludes by identifying open challenges, including computational overhead, limited generalization across climates and PV technologies, and the scarcity of standardized benchmarks, and it outlines promising research directions such as multi-objective and explainable metaheuristic optimization for real-time smart-grid deployment.

**Keywords** Metaheuristic algorithms, Deep learning, Solar power forecasting, Smart energy management, Hyperparameter optimization, Photovoltaic systems, Hybrid forecasting models.

**DOI:** 10.19139/soic-2310-5070-4304

## 1. Introduction

The rapid expansion of photovoltaic (PV) capacity worldwide has made solar energy one of the fastest-growing components of the electricity mix, driven by falling module costs, supportive policies, and the urgency of decarbonization. Yet the same attribute that makes solar power attractive from an environmental standpoint, namely its dependence on uncontrollable atmospheric conditions, also makes it one of the most challenging sources to integrate into a stable power grid. Cloud cover, aerosol content, humidity, and seasonal variation introduce sharp, sometimes unpredictable fluctuations in irradiance and, consequently, in generated power. Reliable short- and

*Correspondence to: Fadl Dahan (Email: fadl.dahan@taiz.edu.ye). Faculty of Education, Arts and Sciences (in Al-Turba), Taiz University, Taiz, 9674, Yemen, fadl.dahan@taiz.edu.ye.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

medium-term forecasting of solar power is therefore a prerequisite for smart energy management, since it underpins unit commitment, economic dispatch, battery energy storage system (BESS) scheduling, demand response, and voltage/frequency regulation in distribution networks [1]. Wen et al. [2] demonstrated this link directly, showing that deep learning-based solar power and load forecasts allow a community microgrid to schedule dispatch with markedly lower operating cost than schedules based on persistence or purely statistical forecasts.

For clarity, we define smart energy management in this review as the set of automated, data-driven functions that coordinate generation, storage, and consumption in a power system with significant PV penetration; specifically, unit commitment and economic dispatch, battery energy storage system (BESS) scheduling and control, demand response and load flexibility programs, and distribution-level voltage/frequency regulation and maximum power point tracking (MPPT). Accurate solar forecasting is treated throughout the review as an upstream enabler of these functions rather than as an end in itself, and Section 6 returns to each of these application areas in turn.

Machine learning and, more specifically, deep learning have progressively replaced purely physical and statistical forecasting approaches (e.g., numerical weather prediction, ARIMA, and its seasonal variants) because DL models can learn complex, nonlinear mappings between meteorological inputs and power output without requiring an explicit physical model of the PV system [3]. Recurrent architectures such as LSTM and GRU, convolutional architectures for spatial or multi-sensor feature extraction, and, increasingly, attention-based transformer models have all been applied to solar forecasting with considerable success [1]. A recurring finding across this literature, however, is that the performance of these deep networks is strongly dependent on hyperparameters such as the number of hidden units, learning rate, batch size, dropout rate, and window length, and that these hyperparameters are difficult to tune manually because the search space is large, non-convex, and computationally expensive to evaluate exhaustively [4].

Metaheuristic optimization algorithms offer a practical alternative to grid search, random search, or purely gradient-based tuning. Inspired by natural, biological, or social phenomena, metaheuristics such as genetic algorithms (GA), particle swarm optimization (PSO), the grey wolf optimizer (GWO), the whale optimization algorithm (WOA), and more recent nature-inspired variants (e.g., Fire Hawk Optimization) are population-based, derivative-free search procedures that balance exploration of the search space with exploitation of promising regions. When hybridized with deep learning models, these algorithms have been shown to substantially reduce forecasting error and improve the coefficient of determination ($R^2$) relative to non-optimized baselines, often at a moderate additional computational cost during the training (offline) phase.

This paper provides a comprehensive review of the intersection between metaheuristic optimization and deep learning for solar power and irradiance forecasting, framed explicitly within the broader context of smart energy management. Unlike earlier surveys that treat machine learning forecasting and optimization as separate topics, this review specifically synthesizes hybrid metaheuristic-deep learning frameworks, offers a taxonomy of the optimization algorithms involved, tabulates and visually compares the deep learning backbones used, and critically discusses the reported quantitative performance of representative hybrid models. The remainder of the paper is organized as follows. Section 2 reviews the principal deep learning architectures used for solar forecasting. Section 3 presents a taxonomy of metaheuristic optimization algorithms and their operating principles. Section 4 surveys metaheuristic-optimized deep learning frameworks reported in the recent literature, supported by a comparative summary table. Section 5 discusses comparative performance trends across studies. Section 6 examines applications within smart energy management. Section 7 discusses open challenges and future research directions, and Section 8 concludes the paper.

To position this contribution more precisely, it is useful to contrast it with two earlier, closely related reviews. Akhter et al. [43] reviewed machine learning and metaheuristic techniques for PV power forecasting, but predates the deep-learning-centric, nature-inspired algorithms (e.g., FHO, IMGO, COA) that dominate the recent literature synthesized here and does not organize its coverage around a deep-learning-specific taxonomy. Alkabbani et al. [1] reviewed machine learning and metaheuristic methods for renewable power forecasting more broadly, covering both wind and solar with a range of classical ML models (ANN, SVM, ELM); by contrast, the present review is restricted specifically to deep-learning backbones, is organized around an explicit taxonomy that maps metaheuristic search mechanisms to deep-learning architectures for solar/PV forecasting specifically, and synthesizes the most recent (2023-2026) wave of nature-inspired optimizers alongside a critical discussion of their limitations (Sections 4.4, 5, and 7) that is not the focus of either prior review. This distinction is what we mean when we describe this review as a taxonomy- and synthesis-oriented complement to, rather than a replacement for, these broader surveys.

### *1.1. Review Methodology*

The literature underlying this review was identified through systematic searches of Scopus, Web of Science, IEEE Xplore, and ScienceDirect, supplemented by Google Scholar to capture recent open-access outputs (e.g., Scientific Reports). Searches combined the terms "metaheuristic optimization", "deep learning", "solar power forecasting", "PV forecasting", "hyperparameter tuning", and the names of specific algorithms (e.g., Fire Hawk Optimization, Grey Wolf Optimizer, Coati Optimization Algorithm) in various Boolean combinations. The search was restricted mainly to studies published between 2019 and 2026, with a small number of foundational algorithm and architecture papers (e.g., LSTM, GA, PSO, GWO) included regardless of date because they define the methods used by later studies. Studies were included as core studies if they combined a metaheuristic optimization algorithm with a deep-learning model applied to solar irradiance or photovoltaic power forecasting and reported quantitative performance metrics. Studies addressing wind forecasting, household or behind-the-meter energy consumption, or non-deep-learning models were retained only as supporting or contextual literature, as clarified in Section 1.2, and are explicitly distinguished from the core studies throughout the review.

To make the screening process more transparent, the inclusion and exclusion criteria applied after the initial database search were as follows. Inclusion criteria: (i) peer-reviewed journal articles or conference papers published in English; (ii) the study applies at least one metaheuristic optimization algorithm to at least one deep-learning forecasting model; (iii) the forecasting target is solar irradiance or PV power (core studies), or a closely related renewable-forecasting or optimization task retained for context (supporting studies, per Section 1.2); (iv) the study reports at least one quantitative performance metric (e.g., $R^2$, RMSE, MAE, or MAPE). Exclusion criteria: (i) studies not available in full text; (ii) duplicate reports of the same experiment; (iii) purely theoretical or algorithmic papers that do not apply the method to a forecasting case study; (iv) studies unrelated to energy forecasting or optimization. Following a PRISMA-style logic, records were first identified through the database and keyword searches described above (identification), then screened by title and abstract against these criteria (screening), after which the remaining full texts were assessed for eligibility, and the final set retained is the one synthesized in Sections 2-7 of this review, split into core and supporting studies as defined in Section 1.2. Given the narrative, synthesis-oriented nature of this review, we report this process qualitatively rather than as a numerically tracked flow diagram; a fully quantified PRISMA flow diagram would be more appropriate for a future systematic-review or meta-analysis follow-up to this work.

### *1.2. Scope and Classification of Reviewed Studies*

To keep the scope of this review transparent, we distinguish two categories of cited work. Core studies are those that hybridize a metaheuristic optimization algorithm with a deep-learning backbone specifically for solar irradiance or PV power forecasting; these form the basis of Table 2 and the comparative analysis in Section 5. Supporting studies, including work on wind power forecasting, household or smart-home energy consumption, and non-deep-learning machine learning models, are cited only to provide methodological context (e.g., alternative optimizers, related application domains) and are identified as such at each point where they are discussed, so that they are not misread as direct evidence for solar-specific, metaheuristic-optimized deep learning performance.

## 2. Deep Learning Architectures for Solar Power Forecasting

Deep learning models used in solar forecasting can be broadly grouped into recurrent, convolutional, and attention-based families, in addition to hybrid combinations thereof. Table 1 summarizes the characteristics, typical inputs, and relative strengths and weaknesses of the architectures most frequently paired with metaheuristic optimizers in the reviewed literature.

### 2.1. Recurrent architectures: LSTM and GRU

The long short-term memory (LSTM) network [5] addresses the vanishing-gradient problem of conventional recurrent neural networks through gated memory cells that regulate the flow of information across time steps, making it well suited to the diurnal and seasonal periodicities present in solar irradiance data. Konstantinou et al. [7] showed that a stacked LSTM network alone can achieve a normalized RMSE below $10\%$ for PV power output forecasting, underscoring the architecture's baseline effectiveness even before any optimization is applied. The gated recurrent unit (GRU) [6] simplifies the LSTM gating mechanism into reset and update gates, offering comparable modeling capacity for temporal dependencies at a lower computational cost and faster convergence, which is advantageous when the network must be retrained repeatedly inside a metaheuristic search loop.

### 2.2. Convolutional and hybrid CNN-recurrent architectures

Convolutional neural networks (CNN) are effective at extracting local spatial or spectral features from multivariate meteorological inputs (e.g., irradiance, temperature, humidity, wind speed) or from decomposed sub-series obtained through signal-processing techniques such as variational mode decomposition. When combined with LSTM or GRU layers in a CNN-LSTM or CNN-BiLSTM configuration, the convolutional stage performs feature extraction while the recurrent stage models temporal dependencies, an arrangement that has repeatedly outperformed single-architecture baselines in PV power prediction tasks. Agga et al. [8] reported that combined CNN-LSTM and ConvLSTM models improved multi-day self-consumption PV power forecasts relative to standalone recurrent baselines, while Ladjal et al. [9] confirmed the consistent superiority of CNN-LSTM over purely statistical and single-architecture machine learning models for direct normal irradiance forecasting across sunny, cloudy, and mixed-sky conditions. Akhter et al. [10] further demonstrated that an RNN-LSTM hybrid reduces RMSE relative to a plain RNN across three distinct PV plant technologies, and Rai et al. [11] proposed a multi-directed differential-attention CNN-BiLSTM network that improved mean squared error by more than $70\%$ over comparative baselines. At the signal-processing level, Wu et al. [12]combined complete ensemble empirical mode decomposition with a CNN-GRU model (CEEMDAN-CNN-GRU) to explicitly address the nonstationarity and volatility of PV power series, decomposing the raw signal into more tractable sub-series prior to network training.

### 2.3. Attention and transformer-based architectures

Transformer architectures, originally proposed for natural language processing [13], dispense with recurrence in favour of self-attention mechanisms that can capture long-range dependencies more efficiently. Hybrid CNN-LSTM-transformer models [14] and transformer-infused recurrent networks [15] have recently been proposed for solar irradiance and PV power forecasting, reporting improved accuracy over purely recurrent baselines, particularly for longer forecasting horizons, at the cost of increased model complexity and training time, which in turn increases the computational burden of metaheuristic hyperparameter search.

Attention-based models remain comparatively underrepresented in the metaheuristic-optimized solar-forecasting literature relative to LSTM- and GRU-based architectures, and it is worth being explicit about why. The self-attention mechanism scales quadratically with sequence length, so transformer backbones have substantially more trainable parameters and a larger, more heterogeneous hyperparameter space (e.g., number of attention heads, embedding dimension, number of encoder/decoder layers, positional-encoding scheme) than recurrent models; searching this larger space with a population-based metaheuristic multiplies the already-considerable training cost discussed in Section 7, which likely explains why relatively few of the surveyed studies apply full metaheuristic tuning to a pure transformer rather than to a hybrid CNN-LSTM-transformer model. Where transformers have been used [14, 15], they tend to offer an advantage specifically for longer forecasting horizons, where their ability to weigh distant time steps directly can capture longer-range diurnal and multi-day dependencies that recurrent gating mechanisms compress into a fixed-size hidden state; the trade-off is reduced interpretability of the attention weights compared with the more transparent gating structure of LSTM/GRU cells. We view lightweight or sparse-attention variants, and surrogate-assisted metaheuristic tuning specifically adapted to transformer hyperparameter spaces, as a concrete direction for closing this gap, and return to this point in Section 7.

## 3. Metaheuristic Optimization Algorithms: Taxonomy and Mechanisms

Metaheuristic algorithms are population-based, derivative-free search procedures that iteratively improve a set of candidate solutions using stochastic operators guided by a fitness (objective) function, most commonly the mean squared error or root-mean-square error of the deep learning model on a validation set. Because they do not require gradient information about the hyperparameter space, they are well suited to the discontinuous, mixed-integer, and multimodal search landscapes typical of deep learning hyperparameter tuning. It is important to distinguish two related but technically different problems that metaheuristics are applied to in this literature, since they define different search spaces. Hyperparameter optimization searches over a comparatively low-dimensional space of architectural and training settings that are fixed before training (e.g., number of layers, hidden units, learning rate, window length); each candidate solution still requires a full backpropagation-based training run to be evaluated, so the objective function is expensive but the search dimensionality is modest. Weight (and, less commonly, architecture) optimization, as in EMA-DNN [31], instead searches directly over the network's trainable parameters themselves, replacing backpropagation with the metaheuristic; this search space is orders of magnitude higher-dimensional (one dimension per weight) but each candidate evaluation is comparatively cheap (a forward pass only), which is a very different cost-dimensionality trade-off from hyperparameter search. Most of the studies reviewed in Section 4 address the first problem; we flag the handful that address the second explicitly so that their reported costs and search efficiencies are not read as directly comparable to hyperparameter-tuning studies. Figure 1 illustrates the generic taxonomy of metaheuristics that have been hybridized with deep learning for solar forecasting, grouped into evolutionary-based, swarm intelligence-based, physics-based, and human/other-inspired categories.

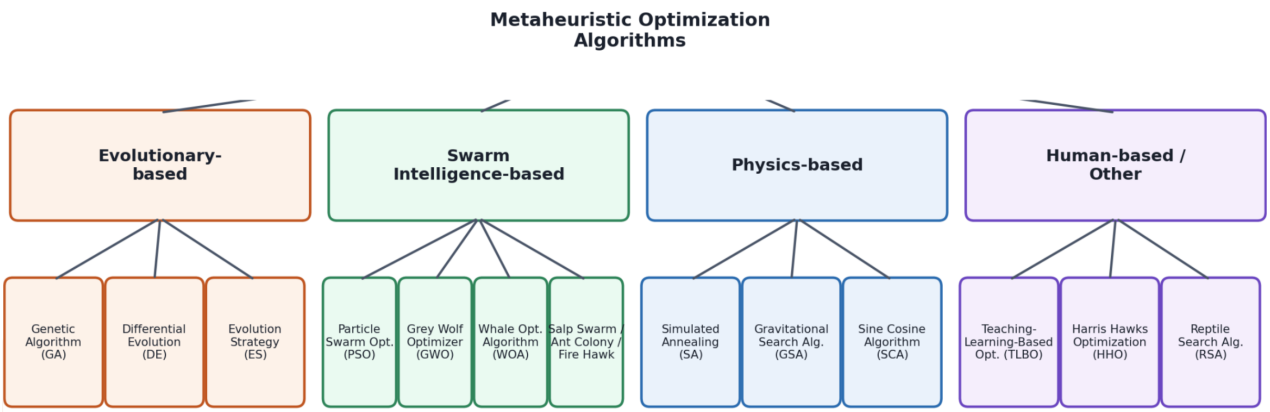


Figure 1. Taxonomy of metaheuristic optimization algorithms commonly hybridized with deep learning models for solar power forecasting. Boxes and labels indicate algorithm categories and representative members; as a categorical taxonomy diagram rather than a data plot, it does not carry numeric axes.

### *3.1. Evolutionary-based algorithms*

Evolutionary algorithms emulate the mechanics of natural selection. The Genetic Algorithm (GA) [16] encodes candidate hyperparameter sets as chromosomes and applies selection, crossover, and mutation operators across generations to evolve improved solutions. GA remains widely used because of its conceptual simplicity and its effectiveness for mixed discrete-continuous hyperparameter spaces, such as the joint optimization of layer counts, neuron numbers, and learning rates. Differential Evolution (DE) [17] is a related evolutionary approach that perturbs candidate solutions using scaled vector differences between population members, and is frequently used as a comparative benchmark against newer swarm-based optimizers in the solar-forecasting literature

### 3.2. Swarm intelligence-based algorithms

Swarm-based metaheuristics model the collective behaviour of decentralized, self-organized biological groups. Particle Swarm Optimization (PSO) [18] updates each candidate solution (particle) based on its own best-known position and the swarm's global best position, using velocity and inertia terms to balance exploration and exploitation. The Grey Wolf Optimizer (GWO) [19] mimics the leadership hierarchy and cooperative hunting behaviour of grey wolves, using alpha, beta, and delta wolves to guide the search toward the estimated position of the prey (the optimum). The Whale Optimization Algorithm (WOA) [20] simulates the bubble-net hunting strategy of humpback whales through a spiral-shaped movement toward the best solution found so far. More recent swarm-inspired algorithms, including the Fire Hawk Optimizer (FHO) [21], which mimics the fire-spreading hunting strategy of predatory birds, have been proposed to improve convergence speed and to reduce the risk of premature convergence to local optima that can affect classical swarm methods.

### 3.3. Physics-based and human/other-inspired algorithms

Physics-based metaheuristics, such as simulated annealing and the gravitational search algorithm, draw an analogy with thermodynamic or gravitational processes to control the acceptance of worse solutions during the search, which helps to escape local optima. The Equilibrium Optimizer (EO) [22] is a more recent physics-inspired algorithm based on control-volume mass balance dynamics that has been used to tune neural-network-based solar irradiance forecasters. Nature- and animal-inspired algorithms, including the Harris Hawks Optimization (HHO) [23] and the Coati Optimization Algorithm (COA) [24], are inspired by the hunting and escaping behaviours of hawks and coatis, respectively; these are not human-based methods, and the taxonomy in Figure 1 and Table 1 classifies them accordingly under the animal/nature-inspired category. Both algorithms combine multiple search behaviours (e.g., exploratory random walks and exploitative local search) within a single algorithm to improve robustness across different optimization landscapes. Table 1 compares representative algorithms from each category in terms of inspiration source, control parameters, and typical convergence behaviour reported in the solar-forecasting literature.

Table 1. Comparison of representative metaheuristic optimization algorithms used with deep learning for solar power forecasting.

| Algorithm | Category | Search Mechanism | Key Control Parameters | Reported Strength in Solar Forecasting |
|---|---|---|---|---|
| GA | Evolutionary | Selection, crossover, mutation across generations | Population size, crossover/mutation rate | Robust for mixed discrete-continuous hyperparameters; slower convergence. |
| DE | Evolutionary | Vector-difference perturbation of population members | Scaling factor, crossover rate | Simple, effective global search; common comparative baseline. |
| PSO | Swarm intelligence | Velocity update toward personal and global best | Inertia weight, cognitive/social coefficients | Fast convergence; prone to premature convergence in complex landscapes. |
| GWO | Swarm intelligence | Alpha/beta/delta hierarchy guides prey encirclement | Number of wolves, convergence factor $a$ | Strong global search and convergence reliability. |
| WOA | Swarm intelligence | Spiral bubble-net encircling of prey | Spiral shape constant, coefficient vectors | Good balance of exploration/exploitation; moderate convergence speed. |
| FHO | Swarm / bio-inspired | Fire-spreading hunting strategy of predatory birds | Number of fire hawks, territory size | High population diversity; strong global optimum discovery. |
| EO | Physics-based | Control-volume mass-balance dynamics | Generation-rate control parameter | Fast convergence; effective for continuous hyperparameter spaces. |
| HHO | Animal/nature-inspired | Surprise pounce with multiple attack strategies | Escape energy parameter | Effective escape from local optima; adaptable exploitation phase. |
| COA | Bio-inspired | Iguana-hunting and predator-escaping behaviour of coatis | Population size, escape probability | Competitive accuracy with fast convergence in joint PV/wind tasks. |

### 3.4. Additional algorithm variants reported in the solar-forecasting literature

Beyond the core algorithms summarized above, several additional metaheuristics have been proposed for solar and PV forecasting. Majumder et al. [25] combined a multi-kernel random vector functional-link network with an evaporation-based Water Cycle Algorithm, reporting a 13.82% improvement in forecasting accuracy after optimization. Aprillia et al. [26] paired a CNN with the Salp Swarm Algorithm, improving mean absolute percentage error by 48.46% over the non-optimized model, while Netsanet et al. [27] combined variational mode decomposition with Ant Colony Optimization and an artificial neural network, achieving $R^2$ above 97%. Lin et al. [28] used an improved Moth-Flame Optimization algorithm to tune support vector machine parameters for PV power generation prediction, and Peng et al. [29] proposed an enhanced Chaos Game Optimization algorithm combined with a locality-sensitive-hashing-based Informer model, reporting simultaneous improvements in $R^2$, mean squared error, and mean absolute error. Beyond forecasting itself, newly proposed metaheuristics continue to be applied to related PV planning problems, such as the Atan-Sinc Optimization Algorithm for optimal PV/D-STATCOM placement and sizing in distribution networks [45]. This diversity of algorithm proposals illustrates the breadth of the design space that researchers continue to explore in search of algorithms with better exploration-exploitation balance for this specific forecasting problem.

This proliferation of new, nature-inspired metaheuristics e.g., FHO, IMGO, COA, iHOW—which models human cognitive learning phases such as data absorption and adaptive knowledge development—, the Ninja Optimization Algorithm—which models a ninja's stealth-based scanning (exploration) and precision-refinement (exploitation) phases—) is worth examining critically rather than simply cataloguing. Some of this growth reflects genuine methodological contribution: newer algorithms such as FHO and COA introduce distinct search operators (e.g., fire-spreading exploration, dual exploration/exploitation phases tied to predator-escape behaviour) that demonstrably alter convergence behaviour relative to classical PSO or GWO in the studies reviewed here. At the same time, the field has been criticized elsewhere in the optimization literature for a pattern in which new metaphor-based algorithms reformulate existing search operators (e.g., a variant of adaptive step-size control or elitism) under a new biological narrative without a clearly demonstrated mechanistic advance, which the No-Free-Lunch theorem (Section 5) suggests should be evaluated empirically rather than assumed from novelty of framing alone. In practice, this means a newly proposed metaheuristic's value for solar forecasting is best judged by whether it is benchmarked against several established baselines (PSO, GWO, DE) on a shared dataset, as in Radzi et al. [30], rather than by algorithmic novelty alone; readers should treat single-algorithm, single-dataset proposals with the same caution urged in Section 5 for the broader accuracy comparisons in this review.

## 4. Metaheuristic-Optimized Deep Learning Frameworks for Solar Power Forecasting

Figure 2 depicts the generic pipeline followed by nearly all hybrid metaheuristic-deep learning studies reviewed here: meteorological and historical PV data are collected and preprocessed (normalization, missing-value imputation, and often signal decomposition), a deep learning backbone is defined, and a metaheuristic optimizer iteratively searches the hyperparameter space by repeatedly training and validating candidate configurations of that backbone until a stopping criterion (e.g., a maximum number of iterations or convergence of the objective function) is reached. The optimized model is then used to generate forecasts that feed downstream smart energy management functions.

### *4.1. Representative case studies*

Radzi et al. [30] proposed a serially connected GRU-LSTM architecture whose hyperparameters were tuned by the Fire Hawk Optimizer (FHO), benchmarking it against PSO-, GWO-, WOA-, and Owl Search Algorithm-optimized variants on two grid-connected PV arrays in Malaysia. The FHO-GRU-LSTM model achieved the best performance on both arrays, with $R^2$ of 0.9964 and 0.9966, and root-mean-square-error (RMSE) reductions of 12.67% and 23.29% relative to the non-optimized baseline, respectively, demonstrating that the choice of metaheuristic materially affects both accuracy and convergence stability even when the underlying deep learning architecture is held constant.

Sulaiman and Mustaffa [31] integrated the Evolutionary Mating Algorithm (EMA) with a feed-forward deep neural network (EMA-DNN) to optimize network weights and biases directly, rather than only hyperparameters, for forecasting AC power output from real PV plant measurements. The EMA-DNN model outperformed

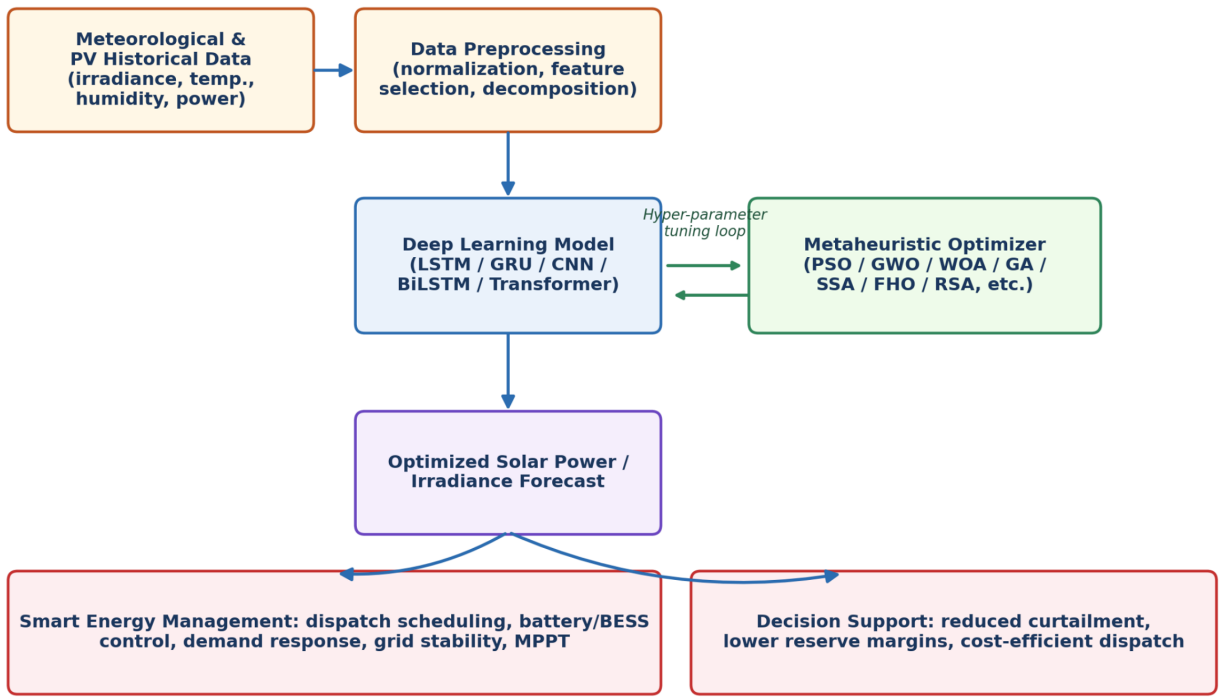


Figure 2. Generic framework of metaheuristic-optimized deep learning for solar power forecasting within smart energy management systems

comparative metaheuristic-DNN hybrids based on differential evolution, the barnacles mating optimizer, PSO, and the harmony search algorithm across multiple error metrics, illustrating that metaheuristics can be applied not only to hyperparameter search but also to the optimization of network parameters themselves as an alternative to gradient-based backpropagation.

Ridha et al. [32] combined the Improved Mountain Gazelle Optimizer (IMGO) with a multi-layer feedforward neural network and polynomial regression to predict PV system output current, reporting an RMSE of 0.0280 and $R^2$ of 0.9951, among the highest accuracies reported in the surveyed literature. Gaber et al. [33] proposed the iHOW optimization algorithm, which emulates human cognitive phases of data absorption and adaptive knowledge development, to tune a Dynamic Temporal Convolutional Network (DTCN) for green (solar and wind) energy forecasting; their study is notable for explicitly quantifying the stepwise contribution of feature selection and metaheuristic tuning, improving $R^2$ from 0.7265 (baseline DTCN) to 0.9005 after feature selection and to 0.9804 after iHOW-based hyperparameter optimization.

T. Singh et al. [34] applied the Grey Wolf Optimizer to tune a hybrid CNN-LSTM model for household-level energy consumption prediction and appliance scheduling in a smart-home energy management context, illustrating that GWO-based tuning generalizes beyond utility-scale PV forecasting to distributed and behind-the-meter applications. Abou Houran et al. [35] proposed the Coati Optimization Algorithm combined with CNN-LSTM (COA-CNN-LSTM) for joint PV and wind power forecasting in smart-grid applications, reporting improved accuracy over PSO- and GWO-optimized counterparts on the same dataset. Yassen et al. [36] introduced a Ninja Optimization Algorithm to tune a quantum-inspired temporal deep learning model for renewable energy forecasting, incorporating explainable AI (XAI) techniques to interpret feature contributions, reflecting a growing trend toward interpretability alongside raw accuracy.

### 4.2. Gradient-assisted and hybrid classical-metaheuristic optimizers

A related line of work combines metaheuristics with classical gradient-based optimizers rather than replacing them outright. Kothona et al. [37] introduced a hybrid Adam-UPSO algorithm that alternates gradient-descent updates with particle-swarm-guided search when tuning an LSTM forecaster, reporting a $15.21\%$ accuracy improvement over standard Adam optimization and a $22.60\%$ improvement over a persistence baseline. Tahir et al. [38] compared Bayesian optimization against random search for tuning ensemble regression trees, support vector machines, Gaussian process regression, and artificial neural networks, finding that Bayesian optimization consistently produced lower RMSE and MAE than random search across all four model families, which suggests that the choice of search strategy remains influential even for comparatively simple hyperparameter spaces. These findings collectively indicate that hybridizing metaheuristics with gradient information or Bayesian surrogate models is a promising middle ground between the computational cost of pure metaheuristic search and the local-optimum risk of purely gradient-based tuning.

It is also useful to situate population-based metaheuristics directly against simpler, non-population-based tuning strategies. Random search samples the hyperparameter space independently at each iteration and, unlike grid search, scales reasonably to higher-dimensional spaces, but it does not use information from previous evaluations to guide subsequent sampling, so it typically requires more objective-function evaluations than a metaheuristic to reach a comparable solution quality. Bayesian optimization, as applied by Tahir et al. [38], builds a probabilistic surrogate model (commonly a Gaussian process) of the objective function and uses an acquisition function to select the next candidate, which makes it markedly more sample-efficient than both random search and most metaheuristics when each training run is expensive; its main limitation is that the surrogate model itself becomes harder to fit and more computationally costly to update as the number of hyperparameters grows, which is precisely the high-dimensional, mixed discrete-continuous regime in which population-based metaheuristics such as GA, PSO, and GWO tend to be more robust. In practice, the choice is therefore a trade-off along two axes: sample efficiency (favouring Bayesian optimization when each DL training run is very expensive and the search space is low-to-moderate dimensional) versus robustness in large, irregular, mixed-type search spaces (favouring population-based metaheuristics). None of the reviewed studies directly benchmark a metaheuristic against both random search and Bayesian optimization on the same solar-forecasting dataset, which we flag as a concrete gap for future comparative work.

### 4.3. Context from non-metaheuristic hybrid baselines

It is useful to situate metaheuristic-optimized models against earlier, non-metaheuristic hybrid baselines. Bouzerdoum et al. [39] combined a SARIMA statistical model with a support vector machine (SARIMA-SVM) for short-term PV power forecasting, achieving $R^2$ of 0.9908 (we note this study is now over a decade old; we retain it not despite its age but because of it, as it usefully anchors the pre-metaheuristic-tuning era against which the more recent gains discussed above can be judged), a result that pre-dates the widespread adoption of metaheuristic tuning but is nonetheless competitive with several of the hybrid models discussed above, underscoring that architecture and feature engineering choices remain important even when a powerful optimizer is available. Beyond individual algorithm proposals, comparative studies provide useful broader context. Abdelsattar et al. [40] benchmarked multiple deep learning architectures for solar power prediction under a consistent experimental protocol, concluding that hybrid recurrent-convolutional architectures generally outperform single-family models, a finding that is reinforced by the metaheuristic-optimized hybrids surveyed above. Mbey [41] presented a comprehensive comparative review of deep learning methods for PV generation and electrical demand forecasting in smart power grids, underscoring that hybrid models combining feed-forward, recurrent, and metaheuristic optimization components (e.g., FFNN-LSTM with multi-objective PSO) consistently improve on single-architecture baselines for both generation and demand-side forecasting tasks relevant to smart energy management.

### 4.4. Critical Synthesis: Strengths, Limitations, and Practical Implications

Beyond summarizing individual results, it is useful to critically compare what the case studies in Sections 4.1-4.3 collectively imply. In terms of strengths, swarm-based algorithms (FHO, GWO, COA) consistently deliver strong accuracy with a manageable number of control parameters, which is practically attractive because it reduces the burden of meta-tuning the optimizer itself; this is a genuine advantage over evolutionary algorithms such as GA, whose crossover and mutation rates often require additional problem-specific tuning. Approaches that optimize network weights directly, such as EMA-DNN [31], avoid backpropagation's dependence on differentiable loss landscapes but scale poorly to the very deep or wide architectures now common in transformer-based forecasting, which limits their practical applicability to comparatively small feed-forward networks. Hybrid gradient-metaheuristic approaches such as Adam-UPSO [37] and the Bayesian/random-search comparison of Tahir et al. [38] represent, in our assessment, the more practically deployable middle ground for organizations with limited computational budgets, because they bound the number of expensive full training runs while still escaping some local optima that pure gradient descent would not. A recurring limitation across nearly all case studies, however, is that reported gains are demonstrated on a single dataset and PV technology; none of the surveyed papers repeat their central experiment across multiple independent sites or multiple random seeds, which, combined with the reproducibility concerns raised in Section 5, means the practical, deployment-ready evidence base for any single metaheuristic remains thinner than the volume of positive results might suggest. For practitioners, our reading of the evidence is that the choice of metaheuristic matters less than (i) whether the optimizer is benchmarked against at least one simple baseline (random search or default hyperparameters) on the practitioner's own data, and (ii) whether the computational budget allows for repeated runs to characterize variability, rather than reporting a single best run.

A further critical appraisal concerns the quality of the experimental designs in the primary studies themselves, which this review is not always able to verify from the published text alone. Based on what is reported, most studies use a chronological or random train/validation/test split from a single site, and it is often unclear whether the test set was held out from the metaheuristic search loop entirely or only from final model training, a distinction that matters for whether the reported test-set metric is a genuinely unbiased estimate of generalization. Baseline choices are also uneven: some studies compare only against other metaheuristic-optimized variants (e.g., FHO vs. PSO vs. GWO in [30]), which demonstrates relative but not absolute value, while comparisons against simple, non-learned baselines such as persistence or seasonal-naive forecasts are comparatively rare, appearing explicitly only in a minority of the studies discussed in Sections 4.2-4.3 [37, 39]. Relatedly, none of the surveyed studies report a formal statistical significance test (e.g., a paired test across cross-validation folds or multiple seeds) for the improvement of the proposed method over its baselines; reported gains of a few RMSE or MAPE percentage points are therefore best interpreted as point estimates from a single experimental run rather than as statistically established differences, and we recommend that future work in this area report variability across repeated runs together with an appropriate significance test. A notable exception outside our core solar-forecasting scope is a campus load-scheduling study using a Chimpanzee Optimization Algorithm, which reports paired significance tests ($p \leq 10^{-4}$) and an open-source reproducible codebase [46], illustrating the kind of reporting we recommend here.

## 5. Comparative Performance Analysis

Before comparing individual results, it is worth noting that the studies summarized in Table 2 do not report a single, harmonized accuracy metric: $R^2$, RMSE, MAPE, and MSE are each used by different subsets of studies, sometimes in combination and sometimes alone. This inconsistency is itself part of what makes cross-study comparison difficult, a point we return to in more depth in Section 7, and it means the trends below should be read as broad, metric-specific patterns rather than as a single ranked comparison.

Two consistent trends emerge from the quantitative results reported across the studies summarized in Table 2 and illustrated in Figure 3. First, metaheuristic hyperparameter tuning produces a measurable and often substantial

accuracy gain relative to a non-optimized (default or manually tuned) deep learning baseline. In the iHOW-DTCN study [33], for example, feature selection alone raised $R^2$ from 0.7265 to 0.9005, and the subsequent metaheuristic tuning stage raised it further to 0.9804, indicating that hyperparameter optimization and feature engineering contribute complementary, largely additive improvements rather than substituting for one another. Second, once a well-designed metaheuristic is applied, reported $R^2$ values across independent studies converge into a comparatively narrow high-accuracy band (approximately 0.98–0.997), suggesting diminishing marginal returns from further algorithmic refinement once basic preprocessing, architecture selection, and hyperparameter tuning have all been addressed.
At the same time, the ranking of algorithms is not universal across datasets. In the study by Radzi et al. [30], FHO consistently outperformed GWO, PSO, WOA, and the Owl Search Algorithm on both PV arrays studied, but the relative ranking of the remaining algorithms differed between the two arrays, and PSO exhibited signs of premature convergence in one array while remaining competitive in the other. This observation is consistent with the No-Free-Lunch theorem for optimization, which states that no single metaheuristic can be expected to outperform all others across every possible problem instance, and it reinforces the practical recommendation that multiple metaheuristics should be benchmarked on a given dataset rather than a single algorithm being assumed optimal a priori. This variability also highlights the value of integrated benchmarking efforts such as that of Guermoui et al. [42], who combined iterative filtering with an extreme learning machine and reported normalized RMSE below $10\%$ with $R^2$ above $98\%$ for intra-hour PV variance forecasting, providing an additional non-metaheuristic reference point against which optimizer-driven gains can be measured.

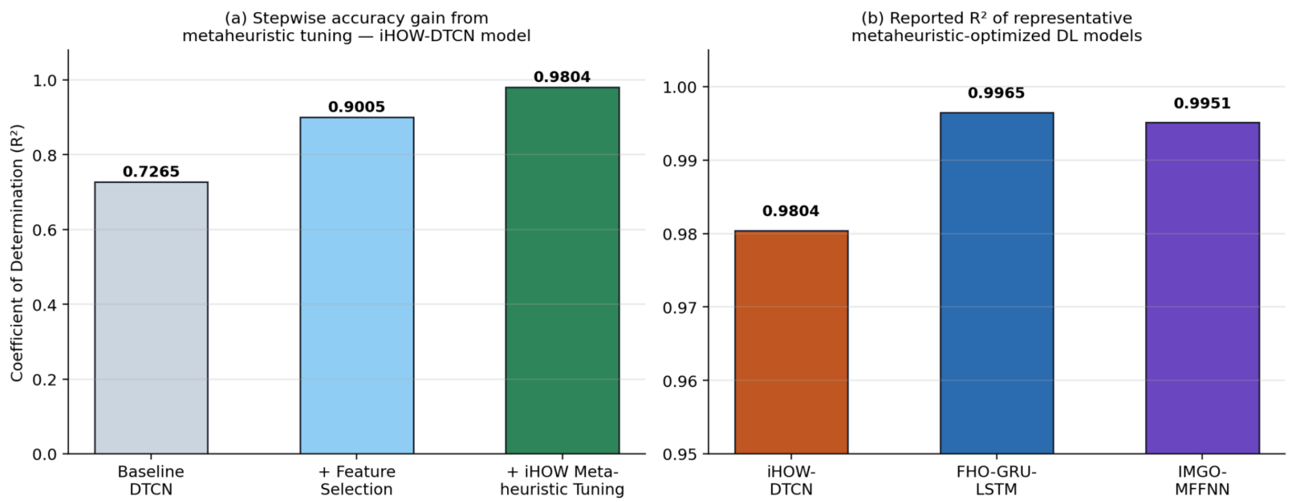


Figure 3. Reported accuracy of metaheuristic-optimized deep learning models in the solar forecasting literature. Values are drawn from separate, independent studies using different datasets, locations, and experimental protocols; they are presented for descriptive, side-by-side comparison only and should not be interpreted as directly equivalent or as a controlled benchmark. Both panels plot the coefficient of determination ($R^2$, a dimensionless ratio on a 0-1 scale), labeled on the y-axis, against the studies/stages identified on the x-axis.

It is important to qualify how these $R^2$, RMSE, and MAPE values should be read. The cited studies were conducted on different PV datasets, geographic locations, forecasting horizons, preprocessing pipelines, and experimental settings (e.g., train/test splits, sensor configurations), so the reported metrics are not obtained under a common, controlled benchmark. Table 2 and Figure 3 therefore present these values for descriptive, side-by-side comparison only; the narrow high-accuracy band observed across studies should be read as an indication of a general trend rather than as proof that any one algorithm or architecture is strictly superior to another. A direct, quantitative ranking of algorithms would require re-implementing the compared methods on a shared dataset under identical conditions, which is beyond the scope of this descriptive review.

Table 2. Summary of representative metaheuristic-optimized deep learning studies for solar and renewable power forecasting.

| Study | Metaheuristic | DL Backbone | Application | Dataset / Location | Forecasting Horizon | Optimization Target | Improvement over Baseline | Key Reported Result |
|---|---|---|---|---|---|---|---|---|
| Radzi et al. [30] | Fire Hawk Optimization (FHO) | GRU-LSTM (serial) | Grid-connected PV, hour-ahead forecast | Two grid-connected PV arrays, Malaysia | Hour-ahead | GRU-LSTM hyperparameters | RMSE reduced 12.7–23.3% | $R^2 = 0.9964$–$0.9966$; RMSE reduced 12.7–23.3%. |
| Sulaiman & Mustaffa [31] | Evolutionary Mating Algorithm (EMA) | Feed-forward DNN | PV AC power forecasting | Real PV plant AC power data | Not specified | DNN weights and biases | Lowest RMSE | Lowest RMSE compared with DE-, BMO-, PSO-, and HSA-DNN. |
| Ridha et al. [32] | Improved Mountain Gazelle Optimizer (IMGO) | Multi-layer FFNN + polynomial regression | PV output current prediction | PV system output current data | Not specified | FFNN + polynomial regression hyperparameters | Not reported | RMSE = 0.0280; $R^2 = 0.9951$. |
| Gaber et al. [33] | iHOW | Dynamic Temporal Convolutional Network | Green energy forecasting | Solar and wind datasets | Not specified | DTCN hyperparameters and feature selection | $R^2$ improved from 0.7265 to 0.9804 | Significant improvement in prediction accuracy. |
| Singh et al. [34] | Grey Wolf Optimizer (GWO) | CNN-LSTM | Smart-home energy forecasting | Smart-home dataset | Not specified | CNN-LSTM hyperparameters | Not reported | Improved accuracy over non-optimized CNN-LSTM. |
| Abou Houran et al. [35] | Coati Optimization Algorithm (COA) | CNN-LSTM | PV/Wind forecasting | Smart-grid PV/Wind dataset | Not specified | CNN-LSTM hyperparameters | Not reported | Outperformed PSO- and GWO-based models. |
| Yassen et al. [36] | Ninja Optimization Algorithm | Quantum-inspired temporal model | Renewable energy forecasting | Renewable energy dataset | Not specified | Model hyperparameters | Not reported | Improved forecasting accuracy with explainable features. |
| Kothona et al. [37] | Hybrid Adam-UPSO | LSTM | Day-ahead PV prediction | PV plant dataset | Day-ahead | LSTM hyperparameters | +15.21% vs. Adam | +22.60% compared with persistence baseline. |
| Majumder et al. [25] | Water Cycle Algorithm (EVWCA) | Multi-kernel RVFL | Short-term solar forecasting | Solar power dataset | Short-term | RVFL parameters | +13.82% | MAPE = 1.06%; 13.82% improvement after optimization. |
| Aprillia et al. [26] | Salp Swarm Algorithm (SSA) | CNN | Solar/PV power forecasting | Not specified in source study | Not specified in source study | CNN hyperparameters | MAPE reduced 48.46% vs. non-optimized CNN | MAPE improved 48.46% after SSA optimization |
| Netsanet et al. [27] | Ant Colony Optimization (ACO) | ANN (with VMD preprocessing) | Solar irradiance forecasting | Not specified in source study | Not specified in source study | ANN hyperparameters | Not quantified as % improvement in source study | $R^2$ above 0.97 with VMD–ACO–ANN |
| Lin et al. [28]* | Improved Moth-Flame Optimization (MFO) | SVM (non-DL)* | PV power generation prediction | Not specified in source study | Not specified in source study | SVM hyperparameters | Not quantified as % improvement in source study | Improved SVM prediction accuracy after MFO tuning |
| Peng et al. [29] | Enhanced Chaos Game Optimization (CGO) | Informer (LSH-based transformer) | Solar/PV power forecasting | Not specified in source study | Not specified in source study | Informer hyperparameters | Not quantified as % improvement in source study | Simultaneous improvement in $R^2$, MSE, and MAE |
| Tahir et al. [38]* | Bayesian Optimization (vs. random search) | ANN; also SVM/ensemble trees/GPR (non-DL)* | Solar PV power prediction | Not specified in source study | Not specified in source study | Hyperparameters of 4 ML model families | lower RMSE/MAE than random search (% not reported) | Outperformed random search across all 4 model families |

This descriptive comparison also motivates a more critical look at the limitations of metaheuristic optimization itself. Population-based search requires training and validating many candidate configurations, so the computational cost of the tuning stage can be substantial, particularly for deep, transformer-based backbones. Most metaheuristics are stochastic, so a single reported run can overstate or understate true performance; without repeated independent runs and reporting of variability (e.g., mean and standard deviation across seeds), a single high $R^2$ value may reflect a favourable random seed rather than a robust improvement. Extensive tuning against a fixed validation set also carries a risk of overfitting to that particular split, which can inflate reported accuracy relative to genuinely unseen data. Finally, few of the surveyed studies release code, data, or the exact hyperparameter search settings used, which limits reproducibility and makes independent verification of the reported gains difficult; we flag this as an open methodological issue for the field rather than a criticism of any individual study.

The suitability of $R^2$ and RMSE themselves for cross-study comparison also deserves comment. $R^2$ measures the proportion of variance in the target explained by the model relative to the variance of the target itself; because that variance depends on the site's climate, season, and sky-condition mix, two models can report similar $R^2$ values while differing substantially in absolute forecasting error, and a site with low irradiance variability can produce an inflated $R^2$ for a comparatively weak model. RMSE is scale-dependent, so an RMSE reported in kW or W/m² is not directly comparable across PV plants of different capacity or across irradiance versus power-output targets unless it is normalized (e.g., as a percentage of installed capacity or of the mean/peak target value); several of the studies discussed in Section 2.1 report a normalized RMSE for this reason. MAPE, used in a subset of the surveyed studies, is particularly sensitive to intervals with near-zero actual power (e.g., night-time or heavily overcast periods), which can inflate or destabilize the metric unless such periods are excluded or a modified formulation is used. For these reasons, Figure 3 and Table 2 should be read as a qualitative map of where reported performance clusters in the literature rather than as a precise, metric-equivalent ranking; a rigorous cross-study ranking would require re-computing a common, normalized metric from each study's raw predictions, which is not available for most of the surveyed papers.

## 6. Applications in Smart Energy Management

Accurate, metaheuristic-optimized solar forecasts feed directly into several smart energy management functions. In economic dispatch and unit commitment, improved forecasts reduce the reserve margins that grid operators must hold to compensate for forecasting error, lowering operating costs while maintaining reliability, as illustrated by the microgrid dispatch gains reported by Wen et al. [2]. In battery energy storage system (BESS) scheduling, higher-accuracy forecasts allow charge/discharge schedules to be optimized closer to the true PV output profile, improving round-trip efficiency and reducing curtailment [41]. In demand response programs, coordinated forecasting of both PV generation and electrical load, as demonstrated in multi-objective PSO-optimized FFNN-LSTM frameworks, enables utilities to better align flexible demand with periods of high solar availability, a pairing reinforced by evidence that metaheuristic (GA) tuning also improves the demand-side forecasts (e.g., SARIMAX-GA, GRU-GA) that such coordination depends on [44]. Finally, at the distributed and behind-the-meter level, metaheuristic-optimized forecasting models embedded in smart-home energy management systems [34] support automated appliance scheduling and maximum power point tracking (MPPT), extending the benefits of accurate forecasting beyond utility-scale operations to individual prosumers.

It is worth making the practical meaning of a high-accuracy metric concrete. An $R^2$ of, for example, 0.9964-0.9966 as reported for FHO-GRU-LSTM [30] indicates that the model explains all but a small fraction of the variance in PV output on the evaluated arrays; translated into operational terms via the RMSE reductions reported alongside it ($12.7 - 23.3\%$ relative to the non-optimized baseline), this corresponds to a proportionally smaller reserve margin that a grid operator must hold to cover forecast error, which, following the mechanism demonstrated by Wen et al. [2] for microgrid dispatch, plausibly translates into lower spinning-reserve and curtailment costs. We emphasize "plausibly" because none of the metaheuristic-optimized studies surveyed in Section **??** themselves quantify the downstream cost or reliability impact of their forecast-accuracy gains in monetary or operational-reserve terms; that translation is demonstrated only in dispatch-focused studies such as Wen et al. [2], which use deep-learning forecasts but not the metaheuristic-optimized models reviewed here. Connecting metaheuristic-optimized forecasting accuracy directly to a quantified operational or economic benefit therefore remains an open empirical gap rather than an established result, and we flag it as such rather than asserting the connection at a general level.

## 7. Challenges and Future Research Directions

Despite consistently positive results, several challenges limit the broader deployment of metaheuristic-optimized deep learning for solar forecasting. First, the computational overhead of population-based search, which requires training and validating dozens or hundreds of candidate model configurations, remains substantial, particularly for deeper architectures such as CNN-LSTM-transformer hybrids; this motivates research into surrogate-assisted or early-stopping-aware metaheuristics that reduce the number of full training runs required, an idea partly anticipated by the Bayesian-optimization comparisons of Tahir et al. [38]. In practical terms, this cost is incurred almost entirely offline, during the (re)training phase; once a metaheuristic-tuned model is deployed, inference for a new forecast is no more expensive than for any other trained deep network, so metaheuristic optimization is generally practical for real-world solar forecasting where retraining happens on a daily-to-monthly cadence rather than in real time. The overhead becomes a genuine deployment barrier mainly when frequent retraining is required (e.g., rapidly changing installations or fast-drifting sensor conditions) or when compute resources are constrained, as in small utilities or behind-the-meter systems. Several practical mitigations can reduce this overhead: (a) reducing population size and iteration count once a coarse optimum is found, combined with early-stopping of clearly unpromising candidate trainings; (b) surrogate-assisted search, where a cheap approximate model (as in Bayesian optimization) screens candidates before a small subset is fully trained; (c) warm-starting the metaheuristic from hyperparameters found on a related site or season rather than searching from scratch; and (d) parallelizing the population evaluation across available GPUs/CPUs, which is straightforward because candidate evaluations in most metaheuristics are independent of one another within a generation. Second, most studies evaluate a single geographic location or PV technology, and reported accuracy gains do not always transfer to other climates, sky conditions, or panel types, particularly in tropical or high-humidity regions where cloud variability is more pronounced [30]; systematic cross-regional benchmarking is still comparatively rare. Relatedly, as noted in Section 5, source code and trained model release remain the exception rather than the norm among the surveyed studies, and journals and conference venues in this space could meaningfully improve the field's reproducibility by encouraging or requiring it. Third, the field currently lacks standardized datasets, forecast horizons, and error-metric reporting conventions, which complicates like-for-like comparison across studies and contributes to the wide variation in reported performance gains. Fourth, while single-objective optimization of forecasting error dominates the literature, real smart-grid deployments often require multi-objective optimization that jointly considers accuracy, computational latency, model interpretability, and robustness to missing or noisy sensor data; multi-objective metaheuristics (e.g., multi-objective PSO) have been explored for combined generation-and-demand forecasting but remain underused for solar-specific applications. Finally, the growing use of explainable AI techniques alongside metaheuristic optimization [36] points toward a promising direction in which future frameworks not only optimize predictive accuracy but also provide actionable insight into which meteorological and temporal features drive forecast uncertainty, supporting more trustworthy integration into automated smart energy management decision loops.

### *7.1. Limitations of This Review*

In the interest of academic transparency, we note several limitations of this review itself. First, as stated in Section 1.1, this is a narrative, synthesis-oriented review rather than a formal systematic review or meta-analysis: we did not independently re-implement or statistically pool the results of the surveyed studies, and our screening process, while guided by explicit inclusion/exclusion criteria, was not conducted by multiple independent reviewers or numerically tracked in a PRISMA diagram. Second, the review is necessarily limited by what the primary studies report; where a source study did not specify a dataset location, size, or forecasting horizon, we have stated this explicitly (e.g., in Table 2) rather than estimating it, which leaves several comparative cells incomplete. Third, our search, described in Section 1.2, was conducted across a defined set of databases and search terms over a bounded time window (mainly 2019-2026); it is possible that relevant studies indexed elsewhere, published in venues outside our search terms, or appearing after our search was conducted were not captured. Finally, the critical assessments in Sections 4.4, 5, and 7 (e.g., on experimental design quality, statistical significance, and the practical significance of reported gains) are our own interpretive judgments based on what is reported in each primary study, not verified through access to the studies' underlying code or data, and should be read as a starting point for critical engagement with this literature rather than a definitive audit of it.

## 8. Conclusion

This review has examined the growing body of literature that hybridizes metaheuristic optimization algorithms with deep learning models for solar power and irradiance forecasting in the context of smart energy management. Across the studies surveyed, recurrent architectures such as LSTM and GRU, often combined with convolutional or attention-based components, remain the dominant deep learning backbone, while a diverse and rapidly expanding set of metaheuristics, ranging from classical algorithms such as GA and PSO to recent nature-inspired methods such as the Fire Hawk Optimizer, the Improved Mountain Gazelle Optimizer, and the Coati Optimization Algorithm, has been applied to tune their hyperparameters or, in some cases, their internal weights. The evidence consistently shows that metaheuristic optimization yields meaningful accuracy improvements over non-optimized baselines, with several hybrid models achieving $R^2$ values above 0.98, and that these improvements translate into tangible benefits for downstream smart energy management functions such as dispatch scheduling, battery management, and demand response. At the same time, no single metaheuristic dominates across all datasets and architectures, computational cost remains a practical barrier, and cross-regional generalization, standardized benchmarking, and multi-objective, explainable optimization remain open and worthwhile directions for future research.

## Acknowledgement

TThe authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/ 2026/01/ 44943).

## REFERENCES

1. H. Alkabbani, A. Ahmadian, Q. Zhu, and A. Elkamel, *Machine Learning and Metaheuristic Methods for Renewable Power Forecasting: A Recent Review*, Frontiers in Chemical Engineering, vol. 3, Art. 665415, 2021.
2. L. Wen, K. Zhou, S. Yang, and X. Lu, *Optimal Load Dispatch of Community Microgrid with Deep Learning Based Solar Power and Load Forecasting*, Energy, vol. 171, pp. 1053–1065, 2019.
3. J. Tian, R. Ooka, and D. Lee, *Multi-Scale Solar Radiation and Photovoltaic Power Forecasting with Machine Learning Algorithms in Urban Environment: A State-of-the-Art Review*, Journal of Cleaner Production, vol. 426, Art. 139040, 2023.
4. C. Stoean, R. Stoean, M. Atencia, R. Rodríguez-Labrada, and G. Joya, *Metaheuristic-Based Hyperparameter Tuning for Recurrent Deep Learning: Application to the Prediction of Solar Energy Generation*, Axioms, vol. 12, no. 3, Art. 266, 2023.
5. S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
6. K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, *Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation*, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734, 2014.
7. M. Konstantinou, S. Peratikou, and A. G. Charalambides, *Solar Photovoltaic Forecasting of Power Output Using LSTM Networks*, Atmosphere, vol. 12, Art. 124, 2021.
8. A. Agga, A. Abbou, M. Labbadi, and Y. El Houm, *Short-Term Self Consumption PV Plant Power Production Forecasts Based on Hybrid CNN-LSTM, ConvLSTM Models*, Renewable Energy, vol. 177, pp. 101–112, 2021.

9. B. Ladjal, M. Nadour, M. Bechouat, N. Hadroug, M. Sedraoui, A. Rabehi, M. Guermoui, and T. F. Agajie, *Hybrid Deep Learning CNN-LSTM Model for Forecasting Direct Normal Irradiance: A Study on Solar Potential in Ghardaia, Algeria*, Scientific Reports, vol. 15, Art. 94239, 2025.
10. M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. Mohamed Shah, *An Hour-Ahead PV Power Forecasting Method Based on an RNN-LSTM Model for Three Different PV Plants*, Energies, vol. 15, Art. 2243, 2022.
11. A. Rai, A. Shrivastava, and K. C. Jana, *Differential Attention Net: Multi-Directed Differential Attention Based Hybrid Deep Learning Model for Solar Power Forecasting*, Energy, vol. 263, Art. 125746, 2023.
12. S. Wu, H. Guo, X. Zhang, and F. Wang, *Short-Term Photovoltaic Power Prediction Based on CEEMDAN and Hybrid Neural Networks*, IEEE Journal of Photovoltaics, vol. 14, pp. 960–969, 2024.
13. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, *Attention Is All You Need*, Advances in Neural Information Processing Systems, vol. 30, 2017.
14. E. M. Al-Ali, Y. Hajji, Y. Said, M. Hleili, A. M. Alanzi, A. H. Laatar, and M. Atri, *Solar Energy Production Forecasting Based on a Hybrid CNN-LSTM-Transformer Model*, Mathematics, vol. 11, no. 3, Art. 676, 2023.
15. M. S. Naveed, M. F. Hanif, M. Metwaly *et al.*, *Leveraging Advanced AI Algorithms with Transformer-Infused Recurrent Neural Networks to Optimize Solar Irradiance Forecasting*, Frontiers in Energy Research, vol. 12, Art. 1485690, 2024.
16. J. H. Holland, *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI: University of Michigan Press, 1975.
17. R. Storn and K. Price, *Differential Evolution–A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces*, Journal of Global Optimization, vol. 11, no. 4, pp. 341–359, 1997.
18. J. Kennedy and R. Eberhart, *Particle Swarm Optimization*, Proceedings of the IEEE International Conference on Neural Networks (ICNN'95), vol. 4, pp. 1942–1948, 1995.
19. S. Mirjalili, S. M. Mirjalili, and A. Lewis, *Grey Wolf Optimizer*, Advances in Engineering Software, vol. 69, pp. 46–61, 2014.
20. S. Mirjalili and A. Lewis, *The Whale Optimization Algorithm*, Advances in Engineering Software, vol. 95, pp. 51–67, 2016.
21. M. Azizi, S. Talatahari, and A. H. Gandomi, *Fire Hawk Optimizer: A Novel Metaheuristic Algorithm*, Artificial Intelligence Review, vol. 56, pp. 287–363, 2023.
22. A. Faramarzi, M. Heidarinejad, B. Stephens, and S. Mirjalili, *Equilibrium Optimizer: A Novel Optimization Algorithm*, Knowledge-Based Systems, vol. 191, Art. 105190, 2020.
23. A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, *Harris Hawks Optimization: Algorithm and Applications*, Future Generation Computer Systems, vol. 97, pp. 849–872, 2019.
24. M. Dehghani, Z. Montazeri, E. Trojovská, and P. Trojovský, *Coati Optimization Algorithm: A New Bio-Inspired Metaheuristic Algorithm for Solving Optimization Problems*, Knowledge-Based Systems, vol. 259, Art. 110011, 2023.
25. I. Majumder, P. K. Dash, and R. Bisoi, *Short-Term Solar Power Prediction Using Multi-Kernel-Based Random Vector Functional Link with Water Cycle Algorithm-Based Parameter Optimization*, Neural Computing and Applications, vol. 32, pp. 8011–8029, 2020.
26. H. Aprillia, H.-T. Yang, and C.-M. Huang, *Short-Term Photovoltaic Power Forecasting Using a Convolutional Neural Network–Salp Swarm Algorithm*, Energies, vol. 13, Art. 1879, 2020.
27. S. Netsanet, D. Zheng, W. Zhang, and G. Teshager, *Short-Term PV Power Forecasting Using Variational Mode Decomposition Integrated with Ant Colony Optimization and Neural Network*, Energy Reports, vol. 8, pp. 2022–2035, 2022.
28. G.-Q. Lin *et al.*, *An Improved Moth-Flame Optimization Algorithm for Support Vector Machine Prediction of Photovoltaic Power Generation*, Journal of Cleaner Production, vol. 253, Art. 119966, 2020.
29. T. Peng *et al.*, *An Intelligent Hybrid Approach for Photovoltaic Power Forecasting Using Enhanced Chaos Game Optimization Algorithm and Locality Sensitive Hashing Based Informer Model*, Journal of Building Engineering, vol. 78, Art. 107635, 2023.
30. P. N. L. Mohamad Radzi, S. Mekhilef, N. Mohamed Shah, M. N. Akhter, M. Seyedmahmoudian, and A. Stojcevski, *Optimizing Solar Power Forecasting with Metaheuristic Algorithms and Deep Learning Models for Photovoltaic Grid Connected Systems*, Scientific Reports, vol. 15, Art. 40045, 2025.
31. M. H. Sulaiman and Z. Mustaffa, *Forecasting Solar Power Generation Using Evolutionary Mating Algorithm-Deep Neural Networks*, Energy and AI, vol. 16, Art. 100371, 2024.
32. H. M. Ridha *et al.*, *A Novel Prediction of the PV System Output Current Based on Integration of Optimized Hyperparameters of Multi-Layer Neural Networks and Polynomial Regression Models*, Next Energy, vol. 8, Art. 100256, 2025.
33. K. Sh. Gaber, M. M. Eid, and E.-S. M. El-kenawy, *Explainable AI for Intelligent Green Energy Forecasting: Deep Learning with iHOW Optimization Algorithm*, Scientific Reports, vol. 15, Art. 41158, 2025.
34. T. Singh *et al.*, *Grey Wolf Optimization Based CNN-LSTM Network for the Prediction of Energy Consumption in Smart Home Environment*, IEEE Access, 2023.
35. M. Abou Houran, S. M. S. Bukhari, M. H. Zafar, M. Mansoor, and W. Chen, *COA-CNN-LSTM: Coati Optimization Algorithm-Based Hybrid Deep Learning Model for PV/Wind Power Forecasting in Smart Grid Applications*, Applied Energy, vol. 349, Art. 121638, 2023.
36. M. A. Yassen, E.-S. M. El-kenawy, M. G. Abdel-Fattah, I. Ismael, and H. E. S. Mostafa, *Renewable Energy Forecasting Using Optimized Quantum Temporal Model Based on Ninja Optimization Algorithm*, Scientific Reports, vol. 15, Art. 14714, 2025.
37. D. Kothona, I. P. Panapakidis, and G. C. Christoforidis, *Day-Ahead Photovoltaic Power Prediction Based on a Hybrid Gradient Descent and Metaheuristic Optimizer*, Sustainable Energy Technologies and Assessments, vol. 57, Art. 103309, 2023.
38. M. F. Tahir, M. Z. Yousaf, A. Tzes, M. S. El Moursi, and T. H. M. El-Fouly, *Enhanced Solar Photovoltaic Power Prediction Using Diverse Machine Learning Algorithms with Hyperparameter Optimization*, Renewable and Sustainable Energy Reviews, vol. 200, Art. 114581, 2024.
39. M. Bouzerdoum, A. Mellit, and A. Massi Pavan, *A Hybrid Model (SARIMA–SVM) for Short-Term Power Forecasting of a Small-Scale Grid-Connected Photovoltaic Plant*, Solar Energy, vol. 98, pp. 226–235, 2013.
40. M. Abdelsattar, M. A. Azim, A. AbdelMoety, and A. Emad-Eldeen, *Comparative Analysis of Deep Learning Architectures in Solar Power Prediction*, Scientific Reports, vol. 15, no. 1, Art. 31729, 2025.
41. A. T. Mbey, *A Novel Deep Learning-Based Data Analysis Model for Solar Photovoltaic Power Generation and Electrical Consumption Forecasting in the Smart Power Grid*, Applied Computational Intelligence and Soft Computing, vol. 2024, Art.

9257508, 2024.
42. M. Guermoui, K. Bouchouicha, N. Bailek, and J. W. Boland, *Forecasting Intra-Hour Variance of Photovoltaic Power Using a New Integrated Model*, Energy Conversion and Management, vol. 245, Art. 114569, 2021.
43. M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. Mohamed Shah, *Review on Forecasting of Photovoltaic Power Generation Based on Machine Learning and Metaheuristic Techniques*, IET Renewable Power Generation, vol. 13, no. 7, pp. 1009–1023, 2019.
44. I. Lahjili, A. Lmakri, M. Hain, and H. Oukhouya, *Enhanced Electricity Demand Forecasting through Metaheuristic Optimization of Model Parameters*, Statistics, Optimization & Information Computing, vol. 15, no. 4, pp. 2527–2550, 2026.
45. O. D. Montoya Giraldo, L. F. Grisales Noreña, and R. Bolaños, *Optimal Placement and Sizing of PV Systems and D-STATCOMs in Medium-Voltage Distribution Networks Using the Atan-Sinc Optimization Algorithm*, Statistics, Optimization & Information Computing, vol. 14, no. 5, pp. 2365–2378, 2025.
46. H. S. Sadiq and M. A. M. Sadeeq, *Intelligent Energy Scheduling for Distributed Duhok Polytechnic University: A Chimpanzee Optimization Approach for Load Reduction*, Statistics, Optimization & Information Computing, vol. 15, no. 1, pp. 423–442, 2025.