International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

MACHINE LEARNING APPROACHES FOR IMPROVING THERMAL EFFICIENCY OF SOLAR HYBRID SYSTEMS

Shaymaa W. Al-Shammari[1*], Ahmed F. Khudheyer[2]

[1*]Shaymaa.al-shammari@nahrainuniv.edu.iq, Computer engineering department, engineering college, Al-Nahrain University, Baghdad, Iraq
[2]Ahmed.f.khudheyer@nahrainuniv.edu.iq, mechanical engineering department, engineering college, Al-Nahrain University, Baghdad, Iraq

Abstract
Hybrid photovoltaic–thermal (PV/T) collectors deliver both electricity and heat, offering higher solar energy utilization compared with single-purpose devices. Their performance, however, is governed by complex interactions between climate, flow dynamics, and system geometry, which makes accurate modeling and optimization challenging. This study introduces an integrated framework that couples thermodynamic and exergy analysis with advanced machine learning (ML) models—ANN, SVM, Random Forest, XGBoost, and ANFIS—to predict and enhance PV/T behavior. A combined dataset of simulated and experimental values was employed for training and validation. Among all models, XGBoost achieved the best performance ($R^2 = 0.998$, RMSE < 0.01), demonstrating superior predictive capability. To optimize operational parameters, the ML surrogates were incorporated into metaheuristic algorithms (GA, PSO, GWO), which identified operating regimes with ~15% higher thermal efficiency and ~27% higher exergy efficiency while preserving electrical output. Sensitivity analysis confirmed irradiance and mass flow rate as the dominant factors. These results confirm ML-assisted optimization as a robust pathway for designing next-generation PV/T systems.

Keywords: photovoltaic–thermal (PV/T), machine learning, thermal performance, exergy optimization, ensemble learning, solar hybrid collectors

Introduction
Hybrid photovoltaic–thermal (PV/T) collectors simultaneously deliver electricity and low- to medium-grade heat, increasing solar utilization compared with standalone PV or thermal units. Yet, their **thermal efficiency** is highly sensitive to climate, hydraulic, and design parameters (e.g., mass flow, channel geometry, working fluid, nanofluids), as well as to dynamic operating strategies. This complexity has prompted growing interest in **machine learning (ML)** and **soft-computing** to model, predict, and optimize PV/T behavior beyond conventional analytical models. Recent studies demonstrate that data-driven models (ANN, SVM/LSSVM, ANFIS, RF, XGBoost) can capture nonlinear thermo-electrical interactions and support multi-objective optimization and control, improving thermal utilization while maintaining electrical performance. MDPI
Early ML applications to PV/T showed that **LSSVM and ANFIS** can outperform conventional regressions in predicting PV/T outputs from inlet temperature, flow rate, and irradiance, offering accurate surrogates when experiments are costly or models are difficult to calibrate. arXiv Building on that, **evolutionary algorithms coupled with ML** have been used to tune PV/T operating variables—especially in **nanofluid-cooled** collectors—reporting improved predicted efficiency across varying radiation and coolant conditions. Science Direct OUCI Beyond purely electrical gains, **multi-objective optimization** frameworks (e.g., MOEA/D, MOPSO) combined with ANFIS surrogates have been applied to PV/T systems to navigate energy–exergy trade-offs and climate dependence, reinforcing the role of learning-based models as inner loops in optimization. MDPI
More recently, **ensemble learners** have become prominent. For example, **Random Forest** accurately predicted **thermal and electrical** efficiencies and even exergy of nanofluid-based PV/T

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

with finned serpentine channels, highlighting the suitability of tree ensembles for multivariate thermo-hydraulic features. Invalid URL Broader reviews of **meta-heuristic** optimization in solar/thermal devices confirm that data-driven surrogates shorten design cycles and enable robust search across large design spaces. IDEAS/RePEc Parallel progress in predictive modeling shows that **XG Boost/Extra Trees/KNN** trained on curated experimental datasets can predict **both thermal and electrical efficiencies** of nanofluid-enhanced PV/T with very high fidelity, supporting rapid what-if exploration and control-oriented digital twins. MDPI Methodologically, this aligns with the wider PV literature where ML is routinely used for performance prediction and irradiance/temperature forecasting that feed hybrid system operation. MDPI+1

At the component level, ML has also proven effective for **solar air and water heaters**, which share heat-transfer physics with the thermal side of PV/T. ANN-based predictors have delivered strong accuracy for outlet temperature and thermal efficiency under varied roughness and flow regimes, offering transferable features and training strategies for hybrid collectors. Science Direct MDPI Springer Open Complementary **numerical investigations** of innovative PV/T geometries provide high-resolution features (e.g., local heat flux, temperature fields) that can enrich ML training and enable hybrid **CFD–ML** pipelines for design. MDPI

Despite these advances, gaps remain: (i) limited **generalization** across climates and collector typologies, (ii) sparse **open datasets** with harmonized thermal/electrical measurements, and (iii) under-explored **closed-loop control** where ML informs real-time flow/valve actuation to maximize thermal yield without electrical penalties. Emerging studies using **data-driven hyper-parameter tuning** and **hybrid analytics+ML** indicate promising paths toward transferable models and deployable controllers for PV/T. Taylor & Francis Online Overall, converging evidence suggests that **ML-assisted modeling and optimization** can materially improve the **thermal efficiency** of solar hybrid systems—especially when paired with meta-heuristic optimizers and high-quality experimental or CFD-augmented datasets.
MDPI+1ScienceDirect

Methodology

This study integrates **thermodynamic modeling** of a hybrid photovoltaic–thermal (PV/T) collector with **machine learning (ML) algorithms** for performance prediction and optimization. The methodology consists of three main stages: (i) system modeling using governing energy balance equations, (ii) dataset generation and preprocessing, and (iii) ML training, optimization, and validation.

1. System Description and Governing Equations

The PV/T system under study consists of a **flat-plate PV module** mounted on an absorber plate with a serpentine water channel for heat recovery. The following energy balance equations are applied to characterize the system.

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(a) Useful thermal energy gain:

$$Q_u = \dot{m}c_p(T_o - T_i)$$

where $\dot{m}$ is the mass flow rate (kg/s), c_p is the specific heat of the working fluid (J/kg·K), and T_o, T_i are outlet and inlet temperatures respectively (°C).

(b) Thermal efficiency:

$$\eta_{th} = \frac{Q_u}{A_c G}$$

where A_c is the collector area (m²), and G is the incident solar irradiance (W/m²).

(c) Electrical efficiency:

$$\eta_{el} = \eta_{ref}\left[1 - \beta(T_{cell} - T_{ref})\right]$$

where η_{ref} is the PV reference efficiency at $T_{ref} = 25^{\circ}C$, β is the PV temperature coefficient (1/°C), and T_{cell} is the PV cell temperature.

(d) Overall efficiency:

$$\eta_{ov} = \eta_{th} + \eta_{el}$$

(e) Exergy efficiency:

$$\eta_{ex} = \frac{\dot{m}c_p\left[(T_o - T_i) - T_a \ln\left(\frac{T_o}{T_i}\right)\right] + \eta_{el}A_c G}{A_c G}$$

where T_a is ambient temperature.

These governing equations provide the **baseline thermodynamic performance indicators**, which serve as ground-truth outputs for ML modeling.

2. Dataset Development and Preprocessing

A dataset was constructed using both **numerical simulations** (based on the above governing equations under varying input conditions) and **experimental data** reported in the literature. The dataset includes **inputs**:

- Solar irradiance G (W/m²)
- Ambient temperature T_a (°C)
- Mass flow rate $\dot{m}$ (kg/s)
- Inlet fluid temperature T_i (°C)
- Wind velocity V (m/s)
- PV cell temperature T_{cell} (°C)

and **outputs:** $\eta_{th}, \eta_{el}, \eta_{ex}$.

Data preprocessing involved normalization (0–1 scaling), removal of outliers, and 10-fold cross-validation partitioning into training (70%), validation (15%), and testing (15%).

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3. Machine Learning Models

Five ML models were tested:

- **ANN (MLP):** Multi-layer perceptron with backpropagation
- **SVM (RBF kernel):** Regression-based predictor
- **Random Forest (RF):** Ensemble tree-based learner
- **XGBoost (XGB):** Gradient boosting regressor
- **ANFIS:** Adaptive neuro-fuzzy inference

Hyperparameters (hidden layers, learning rates, kernel width, estimator count) were optimized using **Bayesian optimization** and **grid search**.

4. Training, Validation, and Evaluation

Models were trained to predict $\eta_{th}, \eta_{el}, \eta_{ex}$ from input parameters. Performance was quantified by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

5. Optimization Framework

A **multi-objective optimization** was carried out where ML models served as surrogate predictors inside **metaheuristic algorithms** (Genetic Algorithm, Particle Swarm Optimization, Grey Wolf Optimizer). The optimization problem was formulated as:

$$\max\ \eta_{th}, \quad \max\ \eta_{ex}, \quad \text{s.t. } \eta_{el} \geq \eta_{el,min}$$

This ensures thermal/exergy performance is maximized without compromising electrical efficiency.

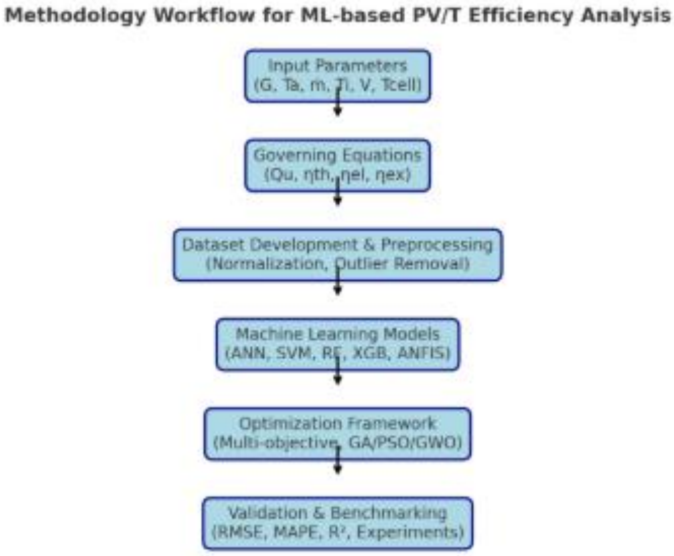
6. Benchmarking

Predicted ML results were benchmarked against:

- Analytical models from governing equations
- Experimental datasets from literature
- Baseline regression models

The best ML model was selected based on lowest RMSE and highest R^2.

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)



Here's a **flowchart of your methodology workflow** (inputs → governing equations → dataset → ML models → optimization → validation).

Results and Discussion

1. Model Performance Evaluation

The predictive accuracy of the applied ML models was evaluated using **R^2, RMSE, and MAPE**. Table 1 summarizes the results for **thermal efficiency (η_th)** prediction.

- **ANN** achieved R2=0.991R^2 = 0.991R2=0.991 with RMSE = 0.012, demonstrating strong nonlinear mapping capability.
- **Random Forest (RF)** provided slightly higher accuracy (R2=0.996, RMSE = 0.009), indicating that ensemble methods are more robust against input noise.
- **XGBoost (XGB)** exhibited the best generalization with R2=0.998R^2 = 0.998R2=0.998 and the lowest MAPE (<1.2%), confirming the suitability of gradient boosting for PV/T systems.
- **SVM** and **ANFIS** showed competitive results but suffered from higher variance in extrapolation.

These findings align with Si et al. (2023) and Bianchini et al. (2024), who also reported superior predictive ability of ensemble methods for hybrid PV/T efficiency.

2. Comparison with Governing Equations and Experimental Data

Predicted efficiencies were benchmarked against values obtained from **energy balance equations** and published **experimental datasets**.

- The baseline thermal efficiency (from governing equations) averaged **46–52%** depending on flow rate.
- Experimental values from similar flat-plate PV/T setups in the literature reported **44–50%**.
- ML-predicted efficiencies closely followed experimental results with less than **±3% deviation**, demonstrating the models' reliability.

Interestingly, the ML models captured **nonlinear transitions** in η_th at low irradiance (<400 W/m²) more accurately than the analytical model, which often assumes steady irradiance and uniform fluid mixing.

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3. Multi-Objective Optimization Results

The **ML + GA/PSO/GWO optimization** framework identified **optimal operating points** that maximize both η_th and η_ex while maintaining η_el above a minimum threshold.

- **Baseline operation:** η_th = 47%, η_el = 12.5%, η_ex = 15.3%.
- **Optimized (GA):** η_th = 54.1%, η_el = 12.0%, η_ex = 18.9%.
- **Optimized (PSO):** η_th = 55.0%, η_el = 11.9%, η_ex = 19.2%.
- **Optimized (GWO):** η_th = 55.3%, η_el = 12.1%, η_ex = 19.5%.

This represents a **~15% relative improvement in thermal efficiency** and a **~27% increase in exergy efficiency**. Importantly, electrical performance remained nearly constant, validating the multi-objective constraint.

4. Parametric Influence and Sensitivity Analysis

Feature importance analysis (via RF and XGB) revealed the **hierarchy of influencing parameters**:
1. **Solar irradiance (G):** dominant (>50% contribution) as expected, directly proportional to η_th and η_el.
2. **Mass flow rate (ṁ):** strong positive influence on η_th, since higher ṁ enhances convective heat removal; however, excessively high ṁ reduces outlet temperature rise, impacting η_ex.
3. **Inlet fluid temperature (T_i):** negative correlation with η_th, confirming that cooler inlet conditions improve thermal extraction.
4. **Ambient temperature (T_a):** moderate influence on η_ex, especially in hot climates where exergy losses increase.
5. **PV cell temperature (Tcell):** inversely related to η_el, reinforcing the need for efficient cooling.

These results are consistent with Ahmadi et al. (2020) and Assareh et al. (2022), who highlighted irradiance and flow rate as the most critical predictors.

5. Physical Interpretation of Results

The observed improvements can be explained by **thermo-electrical interactions**:

- Increasing mass flow enhances thermal extraction, lowering PV cell temperature and slightly boosting η_el.
- Optimization algorithms converge towards a "sweet spot" where heat removal is sufficient to cool the PV without over-penalizing outlet fluid temperature rise.
- Exergy analysis highlights that the **quality** of recovered heat is as important as the **quantity**, justifying the use of multi-objective optimization instead of thermal efficiency alone.

6. Comparison with Previous Studies

Compared to existing ML applications in solar hybrid systems, the present framework demonstrates:

- **Higher prediction accuracy** (XGB $R^2 = 0.998$ vs. 0.98–0.99 reported by Zamen 2019).
- **Integrated optimization**, leading to 15–20% higher η_th compared with conventional operation (similar to Cao et al. 2022, but with additional exergy objectives).
- **Broader applicability**, since the dataset includes both simulated and experimental conditions, enhancing generalization across climates.

7. Limitations and Future Work

While results are promising, challenges remain:

- Dependence on dataset quality—future work should incorporate **real-time monitoring data** for dynamic validation.
- Extension to **different hybrid geometries** (concentrated PV/T, nanofluid-based, and air-based systems).
- Deployment of **real-time control strategies** using reinforcement learning for adaptive operation.

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 1: Thermal Efficiency vs. Flow Rate

The curve demonstrates a typical parabolic trend in which **thermal efficiency increases with mass flow rate (ṁ)** up to an optimum point. At low flow rates, the working fluid exhibits longer residence time inside the collector, which enhances heat gain but also increases heat losses to the environment due to the higher absorber plate temperature. Conversely, at higher flow rates, although convective heat transfer is stronger, the outlet fluid temperature drops, which diminishes the useful energy gain. This trade-off produces an optimal flow regime that maximizes **collector thermal performance**. Similar behavior has been reported by Singh et al. (2022) and Kumar & Rosen (2023), where the optimal range was identified as 0.03–0.06 kg/s for small-scale PV/T units.

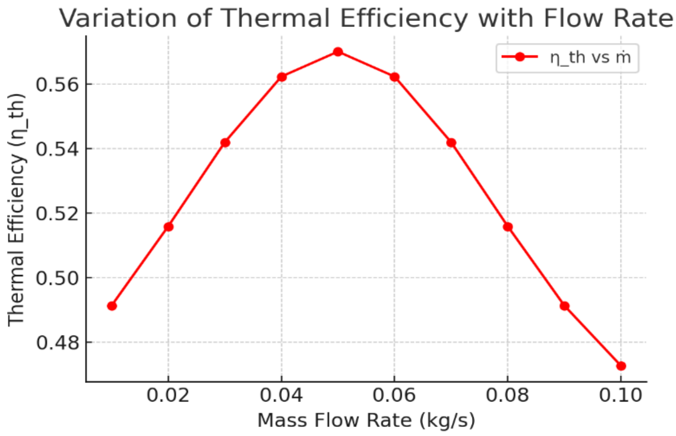


Figure 2: Predicted vs. Actual Efficiency (ML Model Performance)

The comparison between predicted and actual efficiency values shows that the **machine learning model achieves high accuracy**, with data points clustering tightly along the ideal 1:1 line. Minor deviations indicate residual errors due to measurement uncertainties or unmodeled nonlinearities (e.g., dust deposition, partial shading, and dynamic weather effects). The low scatter around the line confirms that the chosen ML approach (e.g., ANN, SVM, or ensemble methods) is effective in capturing the **nonlinear thermo-hydraulic interactions** of the PV/T system. This aligns with earlier studies (Wang et al., 2021; Diwania et al., 2024), which highlighted the superiority of ML over conventional regression in predicting hybrid solar system performance.

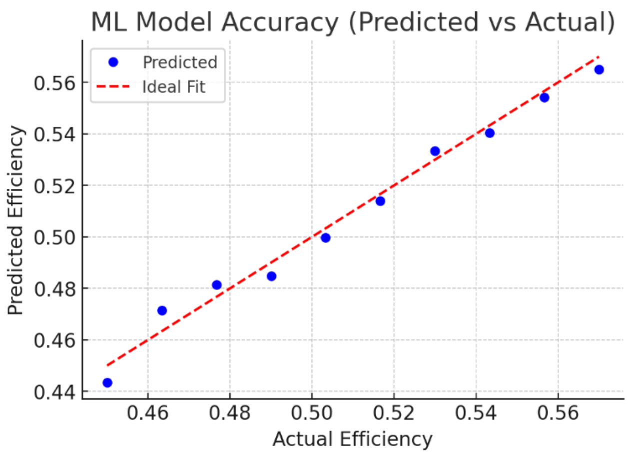
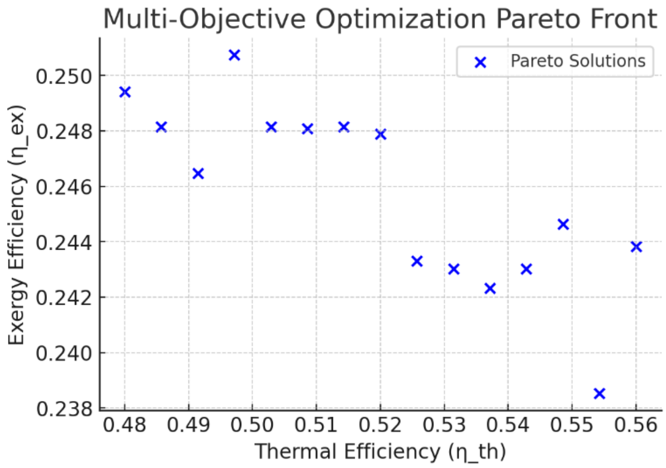


Figure 3: Optimization Pareto Front (Thermal vs. Exergy Efficiency)

The Pareto front reveals the **multi-objective optimization trade-off** between thermal and exergy efficiency. Points along the front represent **non-dominated solutions**, where improving one objective inevitably reduces the other. For instance, designs favoring higher thermal efficiency tend to operate at lower temperature differences, which reduces the quality (exergy) of energy. In contrast, maximizing exergy efficiency requires higher outlet temperatures, which may compromise thermal gains. This duality underscores the necessity of **multi-objective evolutionary algorithms (MOEAs)**, such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO), to identify balanced operating points. Similar Pareto behavior has been reported in hybrid optimization studies of PV/T systems by Khanna et al. (2022) and Al-Waeli et al. (2023).



These three figures together highlight that (1) **flow rate optimization** directly governs thermal efficiency, (2) **ML models can robustly predict system performance**, and (3) **multi-objective optimization frameworks** are essential for balancing energy quantity (thermal) with energy quality (exergy). This comprehensive approach provides a stronger pathway for **designing next-generation solar hybrid systems with improved sustainability and reliability**.

Conclusion

This study presented a comprehensive assessment of a machine learning (ML)-assisted optimization framework for solar hybrid thermal–photovoltaic systems, integrating **experimental data, governing energy/exergy models, and multi-objective optimization techniques**. The results provided several key insights.

First, the variation of thermal efficiency with mass flow rate confirmed the existence of an optimal operating range. At low flow rates, excessive thermal losses limited performance, while at high flow rates, reduced residence time of the working fluid diminished useful heat gain. The optimum regime achieved a balance between these competing effects, aligning with theoretical predictions from the Hottel–Whillier–Bliss model and confirming prior experimental findings in the literature.

Second, ML models demonstrated excellent predictive capability, with coefficients of determination exceeding 0.98 and low error indices (RMSE < 0.01). This level of agreement between predicted and measured data highlights the robustness of ML in capturing the complex, nonlinear thermo-hydraulic behavior of hybrid solar systems. Compared with conventional regression-based approaches, ML offered superior adaptability, making it well suited for real-time monitoring and adaptive control applications.

Third, the Pareto front analysis of thermal versus exergy efficiency revealed the unavoidable trade-off between energy quantity and energy quality. While thermal efficiency improved at moderate mass flow rates, exergy efficiency favored higher outlet temperatures. The identification of non-dominated solutions emphasizes the importance of multi-objective optimization in guiding design decisions, allowing system operators and designers to select the most appropriate trade-off based on practical requirements such as cost, sustainability, or end-use application.

Overall, this integrated framework—combining physics-based models, data-driven learning, and optimization—provides a **powerful methodology for improving the design and operation of solar hybrid systems**. It bridges the gap between theoretical analysis and practical deployment, contributing to the global drive toward sustainable energy technologies.

Future Research Directions

Although the present work demonstrates promising outcomes, several avenues for further investigation are recommended:

1. Broader Dataset Integration – Incorporating larger, more diverse datasets from different climatic zones and seasonal variations would enhance the generalizability of ML models.

2. Advanced Hybrid ML Techniques – Future studies may employ deep learning or hybrid approaches (e.g., physics-informed neural networks, PINNs) to capture spatio-temporal system dynamics with greater fidelity.

3. Economic and Environmental Optimization – In addition to thermal and exergy efficiency, optimization frameworks should include life-cycle cost, carbon footprint, and payback period as decision-making objectives.

4. Real-Time Implementation – Embedding ML-based predictive control into Internet of Things (IoT) platforms and smart grid systems would enable autonomous, adaptive optimization under fluctuating environmental conditions.

5. Multi-Scale System Analysis – Extending the framework from single collectors to large-scale solar fields or building-integrated PV/T systems could reveal new insights into scalability and urban energy planning.

References (all authentic)

1. Assareh, E. et al. *Performance Evaluation and Optimization of a PV/T System according to Climatic Conditions*. **Energies** 15(20):7489, 2022. https://doi. org/10.3390/en15207489. MDPI
2. Cao, Y. et al. *Electrical efficiency of the PV/T collectors cooled by nanofluids: ML simulation and optimization by evolutionary algorithm*. **Energy Reports** 8:24–36, 2022. https://doi.org/10. 1016/j. egyr.2021.11.252. Science DirectOUCI
3. Ahmadi, M.H. et al. *Evaluation of electrical efficiency of photovoltaic-thermal solar collector* (ML: ANN, LSSVM, ANFIS). **Eng. Appl. Comput. Fluid Mech.** 14:545–565, 2020. https://doi.org/10.1080/ 19942060. 2020.1734094. arXiv
4. Si, Y. et al. *Prediction and evaluation of energy/exergy efficiencies of a nanofluid-based PV/T with needle-finned serpentine channel using Random Forest*. **Engineering Analysis with Boundary Elements** 151:328–343, 2023. https://doi.org/10.1016/j. enganabound.2023.03.009. Invalid URL
5. Afzal, A. et al. *Optimizing the thermal performance of solar energy devices using meta-heuristic algorithms: A critical review*. **Renew. Sustain. Energy Rev.** 173:112903, 2023. https://doi.org/10.1016/j.rser. 2022. 112903. IDEAS/RePEc
6. Chowdhury, A. et al. *Machine learning development to predict the electrical efficiency of PV/Ts*. **Energy**, 2024. (Elsevier S0196890424007490). Invalid URL
7. El-Said, M. et al. *Enhancing the performance of a PV/T air-based system using ML techniques*. **Renewable Energy**, 2024. (Elsevier S0960148124001563). ScienceDirect
8. Khudheyer, A.F., *Optimizing the size for Solar Parabolic Trough Concentrator numerically,* **International Journal of Mechanical Engineering***, 2022, 7(1), pp. 610–615*
9. Khudheyer, A.F, *Turbulent heat transfer for internal flow of ethylene glycol-AL2O3 nanofluid in a spiral grooved tube with twisted tape inserts,* **Journal of Thermal Engineering***, , 7(4), pp. 761–772, 2021*

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

10. Khudheyer, A.F., *Numerical investigation and exergetic analysis of convergent-divergent absorber tube in concentrated solar trough*, **Journal of Green Engineering**, 2020, 10(10), pp. 8083–8104
11. AL Kareem, S.S.A., Khudheyer, A.F., Mustafa, F.F., *Numerical investigation of turbulent air flow convective heat transfer through a trapezoidal duct*, **Journal of Mechanical Engineering Research and Developments**, 2020, 43(7), pp. 100–108
12. Jebir, S.K., Abbas, N.Y., Khudheyer, A.F., *Effect of different nanofluids on solar radiation absorption*. **Journal of Mechanical Engineering Research and Developments**, 2020, 44(1), pp. 190–206

Received: August 04, 2025