Materials Research Proceedings 64 (2026) 171-178 https://doi.org/10.21741/9781644904091-21

A Hybrid πVAE-XGBoost Framework for High-Fidelity Simulation of Bifacial Solar Thermal Collectors

Imane FAKIR[1,a *], Mehdi NEJJAR[1,b], Mohammad Hachim ERAISSOUNI[1,c], Adam KESSAB[1,d] and Ahmed KHALLAAYOUN[1]

[1]Al Akhawayn University in Ifrane, Hassan II Avenue, Ifrane, 53000, Fes-Mekness, Morocco

[a]I.fakir@aui.ma, [b]M.nejjar@aui.ma, [c]M.eraissouni@aui.ma, [d]A.kessab@aui.ma

Keywords: Bifacial Solar Thermal Collectors, Hybrid Simulation, Deep Learning, πVAE, XGBoost, Renewable Thermal Energy, Nonlinear Modeling

Abstract. The following paper describes a new hybrid system which combines a deep learning-based pvae model with an XGBoost regressor to produce high-fidelity bifacial solar thermal collector high-fidelity simulation. Through modelling the proposed approach overcomes the above-mentioned challenges by capturing complex, nonlinear interactions between critical system variables, including temperature and irradiance and flow rates, as-sociated with the modeling of a pair-exposure collectors. The high-dimensional input space is effectively reduced to a strong latent representation using the pVAE and further predictions are refined by including the latent features along with the original measurements to the XGBoost component to make reliable predictions of heat transfer and outlet temperature. The validation based on a dataset of 147,396 sequential measurements, as well as comparisons of traditional MATLAB Simulink-CARNOT simulations, shows that the predictive performance is almost perfect, with the values of R2, Spearman, and Pearson correlation coefficients being over 99%. This mixed approach does not only increase the degree of simulation accuracy but also provides explainable information to optimize the design of solar thermal systems, which opens its way to the use in the further renewable energy systems.

Introduction

With the growing heating needs, renewable thermal energy technologies are becoming very critical towards sustainable energy systems. In this specific case, solar thermal energy has taken a center stage in its ability to offer heating solutions that are carbon-free, leading to massive developments in the design of the collectors [1].

Simulating and modeling the behavior of bifacial solar thermal system, nevertheless, remain to be difficult. The power production is determined by a combination of various factors (which are interdependent) and hence nonlinear and multivariable. Traditional physics-based simulation techniques need to consider these intricate interactions and may need the intensive computation [2]. It has been demonstrated that the conditions around the ground can dramatically influence the rear-side irradiance in the depth of the array, and thus, a finely detailed geometry and environmental modeling is required to give valid results. To overcome these challenges, scientists are becoming eager to resort to the use of data-driven methods (especially, deep learning) to increase the quality of the simulation and optimize the design. Recent studies have also shown that machine learning models could be used to enhance the forecast of key performance parameters in solar-thermal systems, including the non-linear behavior of the system that would be recreated with high fidelity [3].

Background

Alternatively referred to as bifacial solar thermal collectors and as the double-exposure flat-plate collectors, the latter have exterior surfaces on the front and the rear that enable them to capture solar energy on both sides [4]. This new design uses trans-clear glazing in place of conventional

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 license. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under license by Materials Research Forum LLC.

Materials Research Proceedings 64 (2026) 171-178 https://doi.org/10.21741/9781644904091-21

rear insulation and uses a reflector under the collector to reflect the sun onto its lower part as well as to a great degree increase solar collection over traditional single-faced systems. The experimental studies have shown bifacial collectors can provide about 41-66% more thermal performance than single equivalent monofacial units and simulation studies showed that a bifacial system with tunable reflector positions could capture more than 50 percent more solar energy in a year than the traditional counterparts [5]. Major features of the design are dual-sided absorber plates glazing both sides, and tilting mounting or reflector mechanisms to achieve maximum irradiation on both sides. In addition, these collectors have been successfully applied in agricultural practices like drying of crops.

Mathematical Model of Bifacial Collector

$$Q = \int [G_b(t)\,\mathrm{IAM}_{\theta}(t) + G_d(t)\,\mathrm{IAM}_{\mathrm{diff}}]\eta_{\mathrm{avg}}\ dt \quad (1)$$

Q in Equ.1 [6] is the total energy production (in kWh or J) by the combination of the total action of beam and diffuse irradiance over time. The beam (direct) solar irradiance, G_b, is in measurements of W/m^2 and modified by the incidence angle modifier $IAM(\theta)$ which is used to describe the performance of the collector with respect to the collector angle of incidence. In the meantime, diffuse solar irradiance G_b, also in W/m^2, is weighted by an independent modifier, IAM_{diff}, which can be uniform or empirical. The average thermal efficiency of the collector, η_{avg} (dimensionless) is the efficiency of the collector to harness the available solar energy and transform it into useful heat. Lastly, the dt represents the time interval (e.g. hourly, per minute) on which the energy contributions are energy-integrated to give the overall thermal output of the system.

Related Work

Bifacial solar thermal collectors are an advanced solar energy technology which attracts direct sunlight and reflected energy so as to enhance the thermal performance. In [7], the analysis in question is taken to a deeper level to depict increased energy yields, caused by improved absorber materials along with reflective surface conditions. These systems result in high efficiency because of maximized heat collecting capacity that reduce thermal losses that are mostly found in reflective settings.

Further, [6] explores the use of bifacial collector in the framework of computational mod-eling and economic analysis. Bifacial collectors are weather-adjustable, as well as reducing the long-term financial needs in solar thermal systems. The core advantage of the bifacial technology is based on the ability to recover more irradiance of reflected sunlight, based on the two research project findings. The results show that future studies should focus on refining system designs, creating strategies of produce reflectivity, and incorporation of bifacial collectors into hybrid energy systems to increase efficiency and sustainability.

Data Context

The data in this research was gathered and given with the consent of the Institute of New Energy Systems (InES) in Germany. It has 147,396 consecutive measurements of a bifacial solar thermal energy system, and these could be taken at intervals of 5 seconds between August 21 and August 29 of 2024.

Materials Research Proceedings 64 (2026) 171-178 https://doi.org/10.21741/9781644904091-21

Table 1. Definition and Units of Variables Used in the Study.

Variable	Description	Unit
T_in_1C	Inlet temperature	[°C]
T_out_1C	Outlet temperature	[°C]
T_mass_flow_1C	Mass flow temperature	[°C]
T_amb_outdoorC	Ambient temperature	[°C]
T_vent_1C	Vent duct temperature	[°C]
dT_1K	Temperature difference (ΔT)	[K]
Q_1kW	Thermal power	[kW]
Vol_flow_1lh	Volumetric flow rate	[l/h]
m_flow_1kgh	Mass flow rate	[kg/h]
rho_1kgl	Fluid density	[kg/l]
cp_1kJkgK	Specific heat capacity	[kJ/(kg·K)]
v_wind_speedms	Wind speed	[m/s]
G_trackerWm	Global irradiance on tracker	[W/m²]
G_diffuseWm	Diffuse irradiance	[W/m²]
IR_TrackerWm	Infrared radiation	[W/m²]
T_IR_Tracker	Tracker surface temperature	[°C]
zenith	Solar zenith angle	[°]
azimuth	Solar azimuth angle	[°]

The methodological Framework of Bifacial Collector Simulation

Classical Framework

A bifacial solar thermal collector simulation was first performed in MATLAB Simulink with the Carnot Toolbox. MATLAB Simulink data organization platform provides the opportunity to generate input files and process system reaction responses via data manipulation and visualization across various conditions. The Carnot Toolbox enables to simulate the energy collection of both sides as well as taking into account the ground albedo and irradiance distribution.

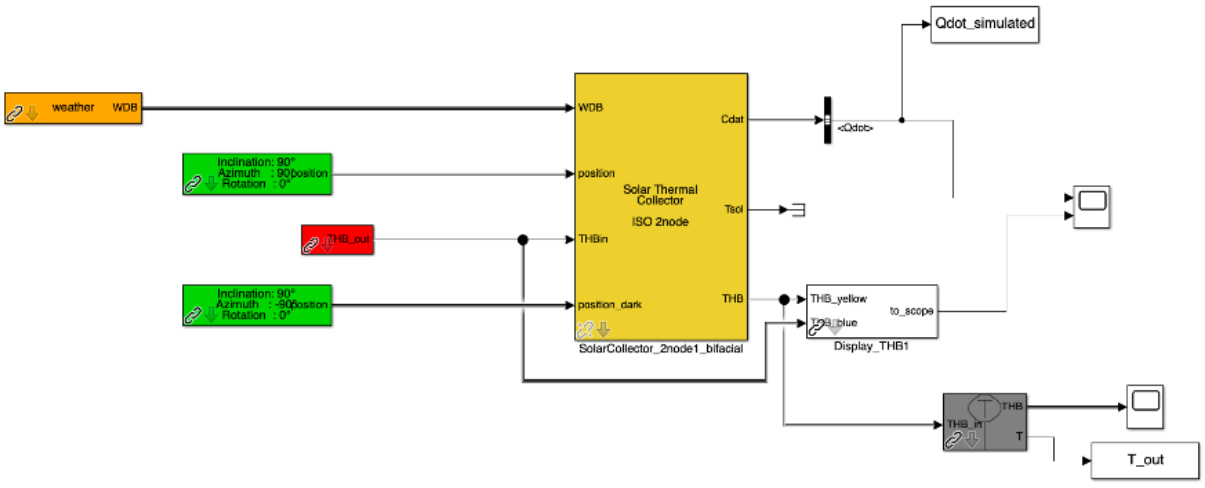


Fig. 1. External Structure of the Bifacial Model.

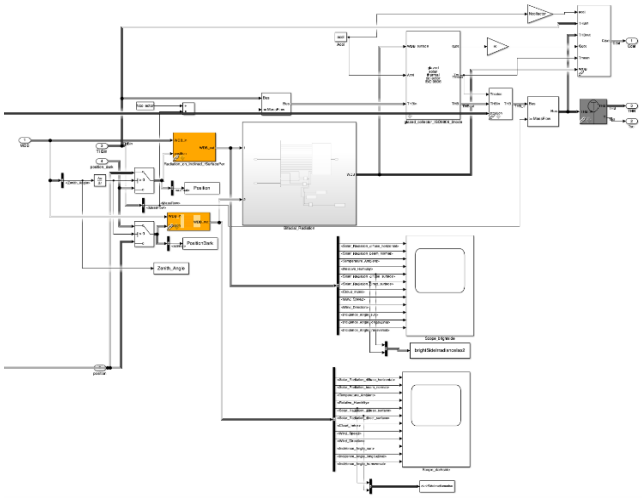


Fig. 2. Internal Structure of the Bifacial Collector Model.

The bifacial solar thermal collector (Fig. 1) is a device with an absorber plate, transparent cover and a switching system, which tilts +-90deg east-west to absorb maximum solar. It absorbs energy in direct, diffuse and reflected light both on the front and back surfaces. Addi-tional components regulate the flow of energy inside of the body and allow dual-surface absorption. The system has a Zenith Angle Block to position east-west and a WDBout and a WDBsurface connection to monitor rear-side irradiance, as presented in Fig. 2. It includes also two layers of energy absorption to replicate heat transfer, Scopebrightside and Scopedarkside blocks to monitor irradiance, MassFlow Bus to control water flow, as well as THByellow, THBblue, and Tmean node to track thermal performance.

Deep Learning Based Framework

Two complementary machine learning methods were used to completely model the work of the bifacial collector. To evaluate the input variables, first, a deep learning model (πVAE) was used to predict the heat transfer (Q 1kW) by taking advantage of the ability to model complex, nonlinear associations and quantify uncertainty in the latent representations of πVAE [8]. This deep learning workflow is the initial component of the framework as shown in Fig. 1. At the second stage, an XGBoost regression model is used to refine the prediction by combining the πVAE-calculated heat transfer and the initial input variables to predict the outlet temperature (T_{out}1C).

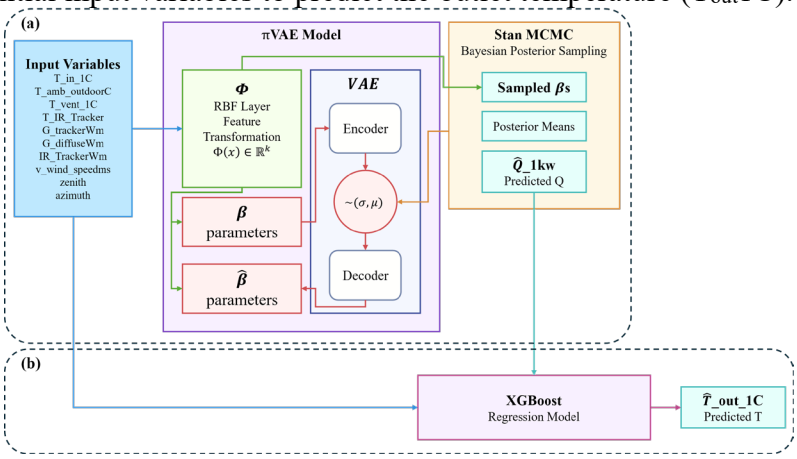


*Fig. 3. Flow diagram of the simulation of the bifacial collector workflow: (a) **π**VAE model involves predicting heat transfer Q1kW using the input variables. (b) XGBoost model makes use of the input variables and the predicted heat transfer to estimate T_{out} 1C.*

Materials Research Proceedings 64 (2026) 171-178 https://doi.org/10.21741/9781644904091-21

Dataset Processing

The data was split into training and testing groups. The training was done with data collected between August 21 and August 29, with the exception of August 28, which was completely held out to test. This design made sure that the test set is made up of a complete day of observations that offered an impartial evaluation of the models on unseen, real-world data hence facilitating a comprehensive evaluation of the models in terms of generalization.

***π**VAE*

The learning of the multifaceted correlation among the variables of the system inputs versus the heat exchange is performed in the first phase, where the πVAE model is trained. Fig. 3 mentions the input variables, which are temperature readings (inlet, ambient, ventilation), irradiance levels (global, diffuse, infrared on a tracker), wind speed, and solar zenith angle. These inputs are processed via a feature mapping function Φ(s) that forms a high dimensional representation that represents the underlying spatial correlations.

Secondly, the weights of each of the functions are then encoded using a variational autoencoder (VAE), denoted by β. Here, the encoder is used to compute the mean and the standard deviation of a Gaussian distribution at the β values, and, with the reparameterization trick, the latent variables Z are sampled. Decoder then recovers the β coefficients out of these latent samples. Most importantly, the result of the feature mapping Φ(s) (i.e. the basis functions) is multiplied with these reconstructed β values to create the resultant function output through inner product β^T Φ(s).

The latent space is then sampled using Hamiltonian Monte Carlo (HMC) to draw samples of the posterior distribution of Z. HMC is especially good at this task because it can easily explore the uncorrelated, low dimensional latent space. Lastly, the average of the posterior samples is also calculated and made the robust point estimate in generating the reconstructed function output. For each HMC sample $Z^{(i)}$, the decoder produces a corresponding output:

$$y_2^{(i)}(s) = d\left(Z^{(i)}\right)^T \Phi(s) \tag{2}$$

where:

$$\beta = d(Z) \tag{3}$$

When mean is taken of all the N samples, the predicted heat transfer is:

$$\widehat{Q}(s) = \frac{1}{N}\Sigma_{i=1}^{N} y_2^{(i)}(s) = \frac{1}{N}\Sigma_{i=1}^{N} d\left(Z^{(i)}\right)^T \Phi(s) \tag{4}$$

XGBoost Regressor

Having used the πVAE method on the whole dataset to estimate the heat transfer Q 1kW., we split the data as before. The predicted $\widehat{Q}$_1kW is then used together with the original input variables to predict the outlet temperature T_out_1C via an XGBoost regression model.

The XGBoost objective function minimizes the regularized loss [9]:

$$\mathcal{L} = \Sigma_{i=1}^{n} l\left(T_{\text{out_1C}}^{(i)}, \widehat{T_{\text{out_1C}}^{(i)}}\right) + \Sigma_{k=1}^{K} \Omega(f_k) \tag{5}$$

Where $\mathcal{L}$ is the loss function, f_k denotes each individual tree in the ensemble, and $\Omega(f_k)$ is the regularization term that penalizes model complexity.

Materials Research Proceedings 64 (2026) 171-178 https://doi.org/10.21741/9781644904091-21

Simulation in Practice

Classical Simulation vs. Real World Data

(a) (b)

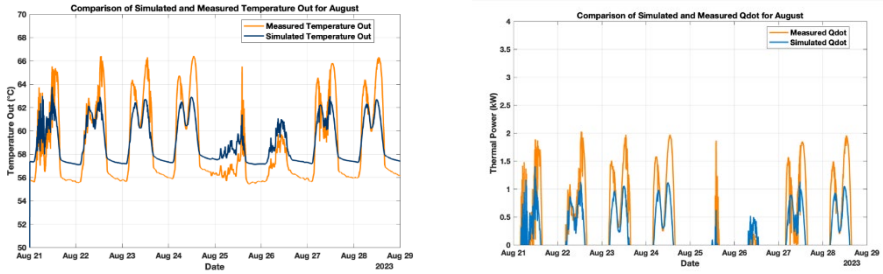


Fig. 4. Comparison of Simulated and Measured Values: (a) Temperature Out using CARNOT. (b) $\dot{\boldsymbol{Q}}$ using CARNOT.

The comparison between the simulated and experimental temperature values along with thermal power indicators proves the level of accuracy and reliability of the thermal model created. Through the two comparisons, the generated data and the experimental data are found to be similar and display the changes in solar irradiance as a repetitive trend in the data. The real temperature is sometimes in Fig. 4a. much higher than the simulated temperature up to 66-67°C compared to the simulated temperature of 63-64°C. The measured ($\dot{Q}$) of Fig. 4b can be described as having a higher peak values which can go up to 1.9-2.1 kW. The simulated ($\dot{Q}$) values are slightly below and oscillate around 1.0-1.2 kW.

Hybrid Framework Results

Training Results

(a) (b)

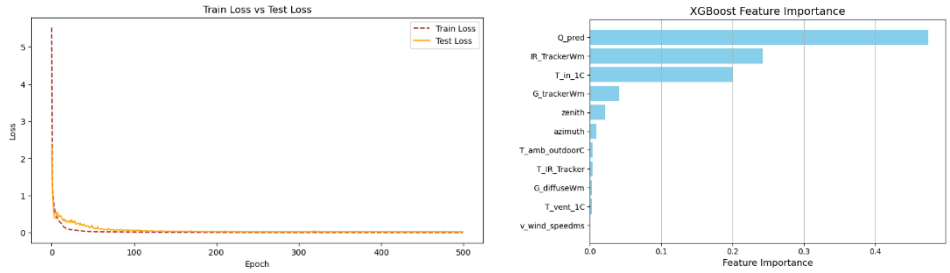


Fig. 5 πVAE-XGBoost Evaluation- Hybrid model: (a) loss- πVAE training as epochs, and (b) feature importance ranking of XGBoost.

Table 2. Models Performance Metrics: $\boldsymbol{R^2}$, Spearman correlation and Pearson correlation.

Model	R^2	Spearman	Pearson
πVAE	99.98%	99.86%	99.90%
XGBoost Regressor	100%	99.99%	100%
Hybrid Model	99.99%	99.92%	99.95%

The hybrid πVAE-Boost model has strong results on performance as seen in the evaluation metrics and the features analysis of Fig. 5. At the point of training, the πVAE component reaches almost perfect convergence (Fig. 5a) and its output Q turns out to be the most important aspect in the importance ranking of the XGBoost model (Fig. 5b). It concentrates on the effect of the pVAE to decrease high-dimensional input to low-dimensional, informative latent representation Q, which

Materials Research Proceedings 64 (2026) 171-178 https://doi.org/10.21741/9781644904091-21

decreases downstream predictive capacity. This synergy is measured in Table 2: The single πVAE brings impressive R^2: 99.98% Spearman: 99.86% Pearson: 99.90% and the XGBoost regressor uses Q and other characteristics (performance table is awaiting). The hybrid design, therefore, takes advantage of the representation learning features of the πVAE and the explainable weighted feature approach of XGBoost, which indicates that it is an effective architecture when faced with a task that requires both accuracy and interpretability.

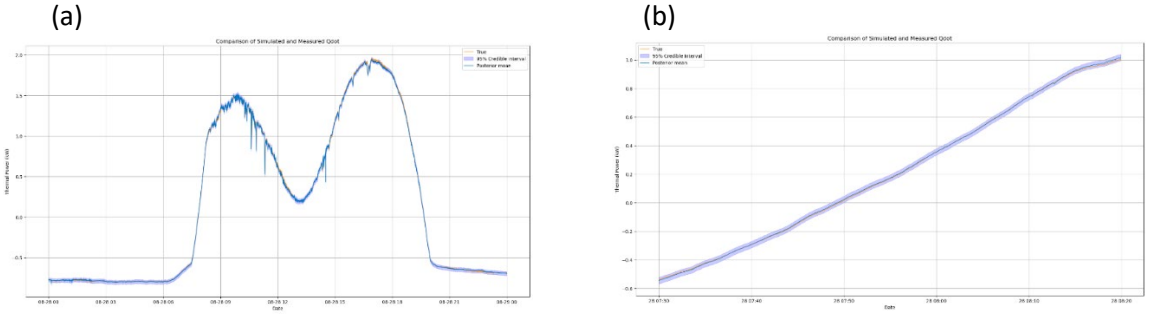


Fig. 6. Visual representation of predictions in the test set of the model: (a) Full simulation, 95% confidence interval, and (b) magnified section showing the quality of the fit.

Fig. 6 illustrates the high-fidelity predictions of the model with the entire simulation (**Fig. 6**a) data being closely aligned between simulated and measured values, and the magnified region (**Fig. 6**b) showing the accuracy of the fit within a 95% confidence interval.

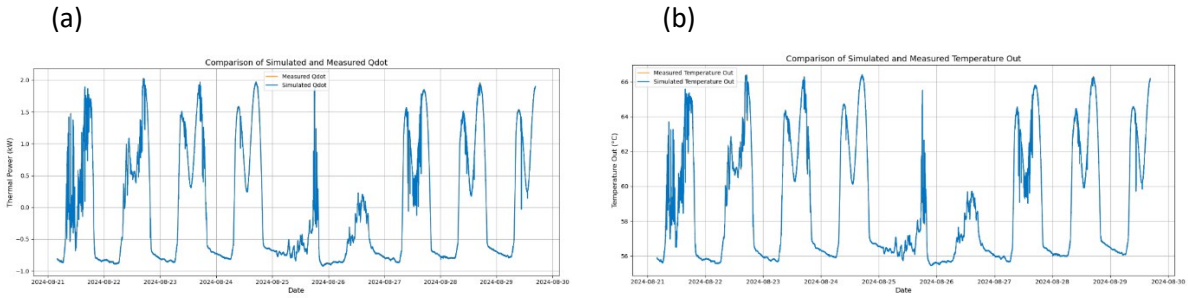


Fig. 7. The Bifacial Collector simulation results using Hybrid system. (a) indicates the simulated and the measured heat transfer. (b) represents the simulated and the measured outlet temperature.

Moreover, Fig. 7 demonstrates the simulation results of bifacial collector with the hybrid system, which represents the precision of this model in the heat transfer (Fig. 7a) and outlet temperature (Fig. 7b) parameters. The plots indicate a uniform recording of measured data, and the simulated curves follow the dynamic variations and the general tendencies of the experimented measurements.

Conclusion

After the presentation of this study, it is possible to argue that a hybrid model has been formed, which is a deep learning-based πVAE and an XGBoost regressor, simulating bifacial solar thermal collectors. Our model provides almost a perfect predictive accuracy by modeling the complex nonlinear interactions between the key variables and the simulated results have a significant correlation with the experimental data and the conventional MATLAB Simulink-CARNOT results. The πVAE is helpful in dimensionality reduction of data, whereas the XGBoost feature

improves the interpretability of features in the data. These findings indicate the possibility of the model to simplify the design of solar thermal systems, and the future studies aim at adaptive design and longer environmental integration.

References

[1] M. Benramdane, S. Abboudi, and M. Ghernaout, “Contribution to the simulation and parametric analysis of the operation of a solar concentration thermal installation,” International Journal of Heat and Technology, vol. 37, no. 2, pp. 446–456, Jun. 2019. https://doi.org/10.18280/ijht.370210.

[2] M. Ernst, G. E. J. Conechado, and C.-A. Asselineau, “Accelerating the simulation of annual bifacial illumination of real photovoltaic systems with ray tracing,” iScience, vol. 25, no. 1, p. 103698, Dec. 2021. https://doi.org/10.1016/j.isci.2021.103698.

[3] J. D. Osorio, M. De Florio, R. Hovsapian, C. Chryssostomidis, and G. E. Karniadakis, “Physics-informed machine learning for solar-thermal power systems,” *Energy Conversion and Management*, vol. 327, p. 119542, 2025. https://doi.org/10.1016/j.enconman.2025.119542.

[4] N. Nikolić and N. Lukić, “Theoretical and experimental investigation of the thermal performance of a double exposure flat-plate solar collector,” Solar Energy, vol. 119, pp. 100–113, Sep. 2015. https://doi.org/10.1016/j.solener.2015.06.038.

[5] N. Nikolic, N. Lukic, A. Nesovic, and D. Nikolic, “Optimal size and position of the planar back reflector moveable only in the direction normal to the bifacial solar collector plane,” Thermal Science, vol. 28, no. 6A, pp. 4483–4497, 2024. https://doi.org/10.2298/TSCI240123101N.

[6] T. Summ, M. Ehrenwirth, C. Trinkl, and W. Zörner, “Agri Solar Thermal Systems: A brief study on the energetic potential of bifacial solar thermal systems,” in Proc. EuroSun 2022 – ISES and IEA SHC Int. Conf. Solar Energy for Buildings and Industry, Kassel, Germany, 2022, pp. 1–8. https://doi.org/10.18086/eurosun.2022.12.16.

[7] T. Summ, B. Sturm, C. Scherer, C. Trinkl, and W. Zörner, “Analysis of thermomechanical stresses and strains for a novel design of insulating glass flat-plate collector,” in Proc. Solar World Congress 2023, New Delhi, India, 2023, pp. 1–11. https://doi.org/10.18086/swc.2023.04.14.

[8] S. Mishra, S. Flaxman, T. Berah, H. Zhu, M. Pakkanen, and S. Bhatt, “πVAE: A stochastic process prior for Bayesian deep learning with MCMC,” Statistics and Computing, vol. 32, no. 6, p. 96, Oct. 2022. https://doi.org/10.1007/s11222-022-10151-w.

[9] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” arXiv preprint, arXiv:1603.02754, accessed Mar. 25, 2025. [Online]. Available: https://arxiv.org/abs/1603.02754