MACHINE LEARNING AS A POWERFUL TOOL FOR PERFORMANCE PREDICTION AND OPTIMIZATION OF CONCENTRATED PHOTOVOLTAIC-THERMOELECTRIC SYSTEM

[1] **Aminu YUSUF** [iD] , [2, *] **Nevra BAYHAN** [iD] , [3] **Hasan TİRYAKİ** [iD] , [4] **Sedat BALLIKAYA** [iD]

[1,4] *Istanbul University-Cerrahpaşa, Engineering Faculty, Engineering Sciences Department, Istanbul, TÜRKİYE*
[2,3] *Istanbul University-Cerrahpaşa, Engineering Faculty, Electrical and Electronics Engineering Department, Istanbul, TÜRKİYE*
[1] aminu.yusuf@iuc.edu.tr, [2] nevra@iuc.edu.tr, [3] hasan.tiryaki@iuc.edu.tr, [4] ballikaya@iuc.edu.tr

Highlights

- PS, MOGA, and GOAL-MOGA optimizations were used to optimize a CPV-TE system.

- Machine learning algorithms were trained using datasets of MOGA and GOAL-MOGA.

- ANN-based ML algorithm has an average prediction error of 0.0692%.

- The ANN-based ML also has a correlation coefficient of $R^2 \approx 1$ on the test data.

- ANN-based optimization can accurately predict the performance of the CPV-TE system.

*Corresponding Author: Nevra BAYHAN, nevra@iuc.edu.tr

MACHINE LEARNING AS A POWERFUL TOOL FOR PERFORMANCE PREDICTION AND OPTIMIZATION OF CONCENTRATED PHOTOVOLTAIC-THERMOELECTRIC SYSTEM

[1] Aminu YUSUF [ID], [2,*] Nevra BAYHAN [ID], [3] Hasan TİRYAKİ [ID], [4] Sedat BALLIKAYA [ID]

[1,4] *Istanbul University-Cerrahpaşa, Engineering Faculty, Engineering Sciences Department, Istanbul, TÜRKİYE*
[2,3] *Istanbul University-Cerrahpaşa, Engineering Faculty, Electrical and Electronics Engineering Department, Istanbul, TÜRKİYE*
[1] aminu.yusuf@iuc.edu.tr, [2] nevra@iuc.edu.tr, [3] hasan.tiryaki@iuc.edu.tr, [4] ballikaya@iuc.edu.tr

ABSTRACT: Because there is a critical necessity to ensure the optimal operation of concentrated photovoltaic-thermoelectric (CPV-TE) systems, various optimization methods such as Paretosearch (PS), Multi-objective genetic algorithm (MOGA), and the hybrid Goal Attainment – Multi-objective genetic algorithm (GOAL-MOGA) are commonly employed. These approaches aim to enhance both the output power and energy efficiency of CPV-TE systems. By combining the Pareto fronts generated by MOGA and GOAL-MOGA, 19 distinct machine learning (ML) algorithms were trained. The findings demonstrate that the Artificial Neural Network (ANN) ML algorithm outperforms others, displaying an average prediction error of 0.0692% on the test dataset. In addition to its prediction capability, the ANN-based ML model can be viewed as an optimization model since it produces optimized outputs similar to those from MOGA and GOAL-MOGA. The ANN-based ML algorithm performs better when trained on a combined dataset from both MOGA and GOAL-MOGA compared to using either MOGA or GOAL-MOGA alone. To enhance the optimization capability of the ANN-based ML algorithm further, more Pareto fronts from other optimization techniques can be added.

Keywords: Concentrated Photovoltaic, Thermoelectric Module, Machine Learning, Optimization, Prediction

1. INTRODUCTION

Solar energy is a crucial green energy source with diverse applications, including domestic heating, crop drying, and electrical energy generation. Photovoltaic (PV) and thermoelectric (TE) devices are technologies used for converting solar energy into electrical energy. However, these technologies utilize different solar energy spectrums, resulting in each device's ability to convert only a portion of solar energy into electricity [1]. Combining these devices allows for full utilization of the solar spectrum, potentially enhancing energy output. Nonetheless, this integration doesn't always guarantee increased efficiency due to the PV's operational principles being nearly opposite to those of the thermoelectric generator (TEG). While PV performance declines with rising temperatures [2], TEG's performance improves under the same conditions [3]. Consequently, optimizing the combined concentrated photovoltaic-thermoelectric (CPV-TEG) system becomes essential to ascertain its optimal parameters.

For instance, the influence of TE leg geometry [4], load resistance [5], thermal resistance, Thomson coefficient [6], cooling fluid [7], and system configurations [8] on the CPV-TE system's performance has been studied. Many optimization studies concentrate on single-parameter optimization, which is time-consuming. This method often fails to identify the optimal operational point due to the system's dependency on numerous parameters. Conversely, multi-objective multi-parameter optimization is rapid and can identify the global optimum of the CPV-TE system [9, 10]. This approach yields non-dominative optimal solutions that strike the best balance between multiple system outputs. However, it is important to note that modelling complexity escalates with an increase in the number of parameters, and finding a global optimum is not guaranteed.

Concentrated PV systems are typically integrated with maximum power point tracking (MMPT) algorithms to harvest maximum output power throughout the day. Traditional methods like Perturb &

---

*Corresponding Author: Nevra BAYHAN, nevra@iuc.edu.tr

Observe (PO) and incremental conductance lack accuracy and response time, often causing oscillations around the maximum power point in a steady state [11]. To address these limitations, advanced control methods, including artificial intelligence such as fuzzy logic, neural networks, and genetic algorithms, have been introduced [12]. Experimental results demonstrate that using a fuzzy logic controller in a solar tracking system increases the efficiency of energy conversion of the PV panels [13]. In another study [14], an adaptive fuzzy logic MPPT controller outperforms conventional techniques in terms of efficiency, tracking ability, and harmonic reduction. It exhibits a faster response to the PV system, even without knowledge of the actual model, and ensures the optimal operating point remains stable without oscillation around the MPP.

Incorporating machine learning (ML) into multi-objective multi-parameter optimization could significantly advance research in mathematical programming. ML is a system employing complex algorithms and statistical models to discern patterns and trends in data [15]. Recent advancements have made ML an increasingly indispensable tool in various sectors. One notable application is predicting short-term and long-term output performance of stand-alone PVs [16-18], TEGs [19, 20], and CPV-TE systems [21]. Such predictions aid in preventing partial or total power blackouts by enabling strategic planning during energy system operations. Moreover, ML implementation has led to considerable reductions in computational time and modelling complexity [22-24].

Aligned with the aforementioned studies, this work aims to demonstrate ML's ability to predict and optimize a CPV-TE system. By amalgamating the Pareto front from various optimization techniques, a robust ML model can be developed. This study endeavours to bridge the existing research gap between ML, energy systems, and optimization techniques.

2. MATERIAL AND METHODS

2.1. Concentrated Photovoltaic-Thermoelectric System (CPV-TE)

The depicted system in Figure 1 comprises a linear solar concentrator, PV, TEG, and heat sink. While the PV can convert a portion of the concentrated solar energy into electrical power, the remaining energy takes the form of thermal energy used by the TEG. A heat sink is affixed to the TEG's cold side to establish a substantial temperature gradient. Additionally, thermal grease (8 $Wm^{-1}K^{-1}$) is assumed to be placed between the PV, TEG, and heat sink to mitigate thermal resistance among these components.
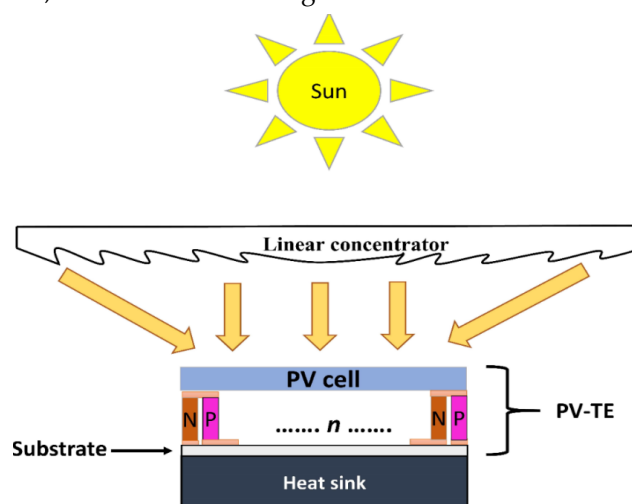

Figure 1. Proposed CPV-TE System

Should the upper ceramic plate of the TEG be eliminated, the temperatures of both the hot side of the TEG and the rear side of the PV become nearly identical. Consequently, the temperature of the hot side can be approximated as [25]:

$$T_h = T_a + Ge^{a+bv} \tag{1}$$

where G is the solar irradiance, T is the ambient temperature, and v is the wind speed measured at a height of 10 m. For a linear concentrator of 22, a and b are empirically determined to be -3.23 and -0.13, respectively. The temperature of the PV can be computed in relation to the temperature of the hot side as [25]:

$$T_{pv} = T_a + Ge^{a+bv} + \frac{G\,\Delta T}{G_s} \tag{2}$$

where G_s is the solar irradiance measured at a standard test condition, $\Delta T = T_{pv} - T_h$ is measured at a standard test condition of 1000 Wm^{-2}. For a solar concentration ratio of 22, the ΔT is empirically found to be 13.

From Equation 2, the electrical output of the PV can be given by:

$$P_{pv} = CGA_{pv}\tau_g\eta_r\left(1 - \beta\left(T_{pv} - 25\right)\right) \tag{3}$$

where A_{pv} is the surface area of the PV, β is the temperature coefficient of the PV, η_r is the reference efficiency of the PV, and τ_g is the transmissivity of glass.

A fraction of the thermal energy absorbed by the TEG gets transformed into electrical energy, whereas the remaining portion is dissipated through the cold side. Therefore, the electrical output of the TEG is calculated as follows:

$$P_{te} = NI[(S_hT_h - S_cT_c) - \tau(T_h - T_c) - IR] \tag{4}$$

where I is the electric current of one pair of thermoelement, $N = 127$ is the number of pairs of thermoelements, τ is the Thomson coefficient, R is the internal resistance of one pair of thermoelement, S_h and S_c are the Seebeck coefficients of the hot and cold sides of the TEG, respectively.

The internal resistance in Equation 4 is computed as:

$$R = \frac{L}{\sigma_p A_p} + \frac{L}{\sigma_n A_n} \tag{5}$$

where L is the length of thermoelements, A_p and A_n are the cross-sectional areas of the thermoelements, σ_p and σ_n are the electrical conductivities of the p- and n-type TE materials.

While the temperature of the cold side in Equation 4 depends on both the temperature of the hot side and the parameters of the heat sink as given by Equation 6:

$$T_c = \frac{T_h + hT_a A_{te} R_{kt}}{1 + h A_{te} R_{kt}} \tag{6}$$

where A_{te} is the surface area of the TEG, R_{kt} is the thermal resistance of the TEG, and h is the heat transfer coefficient. Similarly, the thermal resistance of the TEG is given as:

$$R_{kt} = \frac{L}{N\left(A_n k_n + A_p k_p\right)} \tag{7}$$

where k_n and k_p are the thermal conductivities of the n- and p-type TE materials. The TE materials are temperature dependent and are given in Table 1 while the parameters of the PV are given in Table 2.

Table 1. Temperature dependent TE materials

Variable	Bi₂Te₃ [26]
$S = Sp–(–Sn)$	$(44448 + 1861.2T_m – 1.981T_m^2) \times 10^{-9}$ V/K
$\sigma_p= \sigma_n$	$[(5112 + 163.4Tm + 0.6279T_m^2) \times 10 - 10\ (\Omega m)]^{-1-1}$
$k_p = k_n$	$(62605 – 277.7T_m + 0.4131T_m^2) \times 10^{-4}$ W/(m K)
$\tau=\tau_p-(-\tau_n)$	$(1861.2T_m – 3.962T_m^2) \times 10^{-9}$ V/K

Table 2. Parameters of the PV [27]

Parameter	Symbol	Value
Area of PV/TEG	A_{pv}	$16 \times 10^{-4}\ m^2$
Reference efficiency of PV	η_r	15.6%
Transmissivity of glass	τ_g	0.95
Efficiency temperature coefficient of PV	β	$45 \times 10^{-3}\ K^{-1}$

The electrical output and energy efficiency of the system are respectively given as:

$$P_{sys} = P_{pv} + P_{te} \tag{8}$$

$$\eta_{sys} = \frac{P_{pv} + P_{te}}{CGA} \tag{9}$$

The above equations are then used in the optimization process.

2.2. Optimization of the CPV-TE System

Due to the conflicting operational principles of photovoltaics and thermoelectric generators, the hybrid system consistently requires optimization to achieve peak performance. Furthermore, optimization efforts can contribute to material cost reduction and an extension of the hybrid system's lifespan. Regrettably, there exists no assurance that a specific algorithm will excel across all optimization problem types. To address this issue, researchers frequently compare multiple algorithms' performances on a particular optimization problem, selecting the algorithm that demonstrates the most favorable performance. In this context, three well-established algorithms—namely, the multi-objective genetic algorithm (MOGA) [28], Paretosearch (PS) [29, 30], and hybrid Goal Attainment – Genetic Algorithm (GOAL – MOGA) [31, 32] are employed to optimize the CPV-TE system using the MATLAB optimization toolbox. While MOGA and PS are individual optimization techniques, GOAL – MOGA represents a hybrid algorithm combining the strengths of both, thereby exhibiting greater potency than a single algorithm. The optimization variables considered in this study are given in Table 3.

Table 3. Optimization variables

Parameter	Symbol	Range
Solar irradiance	G	$200 – 1000$ Wm^{-2}
Wind speed	v	$2 – 11$ ms^{-1}
Ambient temperature	T_a	$15 – 40$ °C
Temperature between top and back surfaces of PV	ΔT	$3 – 20$ °C
Electric current	I	$1 – 100$ mA
Length of thermoelement	L	$0.5 – 4 \times 10^{-3}$ m
Cross-sectional area of thermoelement	$A_{n/p}$	$1 – 5 \times 10^{-6}\ m^2$
Heat transfer coefficient	h	$100 – 2000$ Wm^{-2}K^{-1}

The optimization process involves two primary objectives: maximizing output power and

maximizing energy efficiency of the system, as represented by Equations 8 and 9, respectively. The variables subject to optimization encompass both the system's parameters and prevailing weather conditions. Traditionally, system parameters are optimized under fixed weather conditions. However, since weather conditions fluctuate throughout the day and across seasons, optimal performance is achieved only when these conditions align with those during optimization. For instance, if the system is optimized based on summer weather, its performance will be suboptimal during winter, spring, and fall. To address this, our study incorporates weather conditions as dynamic optimization variables. This approach allows us to identify both the system parameters and weather conditions that yield maximum output. By revealing the Pareto optimal solution, our optimization results provide insights into the system's performance throughout the year. The Subset of the optimization outputs for the three algorithms is given in Appendix A.

2.3. Machine Learning Method

Each algorithm used for optimization in this study encompasses two objective functions, aiming to maximize the system's output power and energy efficiency. As these objectives cannot simultaneously reach their maximum values, the result is a collection of non-dominating solutions forming a Pareto front. These solutions represent optimal trade-offs between the two objective functions. To ensure a broad exploration of the search space and generate substantial data suitable for machine learning (ML), a large population size is selected for each algorithm. Moreover, running the three algorithms multiple times prevents them from getting stuck in local minima. Upon completion of the algorithms, both MOGA and the GOAL – MOGA hybrid produce 1400 datasets each, which are then merged to create a comprehensive dataset for ML. Consequently, the ML model acquires optimization capabilities akin to both the MOGA and GOAL-MOGA hybrid algorithms. This combined dataset of 2800 entries are employed to train various ML methods within the WEKA 3.8.5 program. The primary objective is to identify the ML method exhibiting superior predictive and optimization performance. In this scenario, only the optimized output power is predicted, considering that energy efficiency can be straightforwardly derived from the output power. The combined dataset undergoes classification into training (90% of the data) and testing (10% of the data) subsets. It is important to highlight that the ML algorithms are exclusively trained using the training data to predict outcomes on the test data accurately. To ensure an objective evaluation, the obtained training results from diverse ML methods are compared using performance criteria such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Correlation Coefficient (R^2), whose formulae are detailed in Table 4.

Table 4. Performance criteria

R^2	RMSE	**MAE**
$1 - \dfrac{\sum_{l=1}^{n}(y_i - \widehat{y_l})^2}{\sum_{l=1}^{n}(y_i - \widehat{y_{avg}})^2}$	$\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_l})^2}$	$\dfrac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_l}|$

where y_i, $\widehat{y_l}$, and $\widehat{y_{avg}}$ are the desired output, the predicted output, and the average of the desired output, respectively, and n represents each sample in the dataset [33].

2.3.1 Training phase

Table 5 showcases a random subset extracted from the complete training dataset, which comprises 2520 samples utilized to train 19 distinct ML algorithms. Each dataset within this subset comprises 8 input parameters and 1 output (representing output power), aligning with the optimization variables and objective function employed in the optimization algorithms.

Table 5. Sample training data

Parameter	Training Data 1	Training Data 2520
G (W/m^2)	200.0100902	999.9998231
v (m/s)	0.753583	0.500053
Ta (^{o}C)	29.99898876	29.99972218
I (mA)	99.96031788	1.054309357
L (mm)	3.999981986	0.583339642
Ap/n (mm^2)	1.000123794	4.999859674
h (W/(m^2 K))	656.4862416	573.603443
Delta T (^{o}C)	19.99098467	19.99995895
Output Power (W)	0.921398994	3.759967994

Table 6 presents the performance of the ML algorithms in respect to the prediction of the optimized output power on the training data. The success of ML algorithms is usually based on the R^2, RMSE, MAE, and computational time. For high prediction accuracy, it is desired that the R^2 approaches unity, while the RMSE and MAE should approach zero. Moreover, the computational time should be small.

Table 6. Training results

ML Algorithm	Duration (seconds)	R^2	RMSE	MAE
Artificial Neural Networks	**48.22**	**1.0000**	**0.0019**	**0.0012**
M5P	0.04	1.0000	0.0049	0.0027
M5Rules	0.55	1.0000	0.0055	0.0035
Random Committee	0.50	1.0000	0.0067	0.0052
Random Forest	0.93	0.9999	0.0095	0.0065
Bagging	0.43	0.9999	0.0087	0.0069
KStar	29.83	0.9998	0.0142	0.0036
Random Tree (RT)	0.01	0.9997	0.0171	0.0139
REPTree	0.01	0.9997	0.0187	0.0155
Linear Regression	0.09	0.9971	0.0557	0.0456
Support Vector Machines	10.88	0.9970	0.0602	0.0428
Simple Linear Regression	0.03	0.9969	0.0578	0.0477
Gaussian Process	54.42	0.9968	0.0594	0.4740
Regression by Discretization	5.60	0.9940	0.0804	0.6920
Decision Table	0.13	0.9937	0.0826	0.0707
Additive Regression	0.04	0.9796	0.1494	0.1220
Random Sub Space	0.38	0.9690	0.2168	0.1780
LWL	10.38	0.8840	0.3447	0.2935
Decision Stump	0.03	0.8571	0.3798	0.3233

In Table 6, it is worth observing that as the correlation coefficient of most of the ML algorithms approach unity, while concurrently, both RMSE and MAE tend towards zero. Remarkably, despite completing training in 48.22 seconds, the Artificial Neural Networks (ANN-Multilayer Perceptron) exhibit the most outstanding performance concerning the correlation coefficient, RMSE, and MAE. Figure 2 illustrates the prediction performance of the ANN-based ML concerning the actual training data, while Figure 3 depicts the correlation between the actual training data and the predicted data. Due to the high prediction accuracy of the ANN-based ML on the training data, the actual training and predicted output power remain indistinguishable. This performance is further solidified with Figure 3, where a very good agreement can be seen between the training and the predicted dataset because the correlation coefficient is $R^2 = 1$.
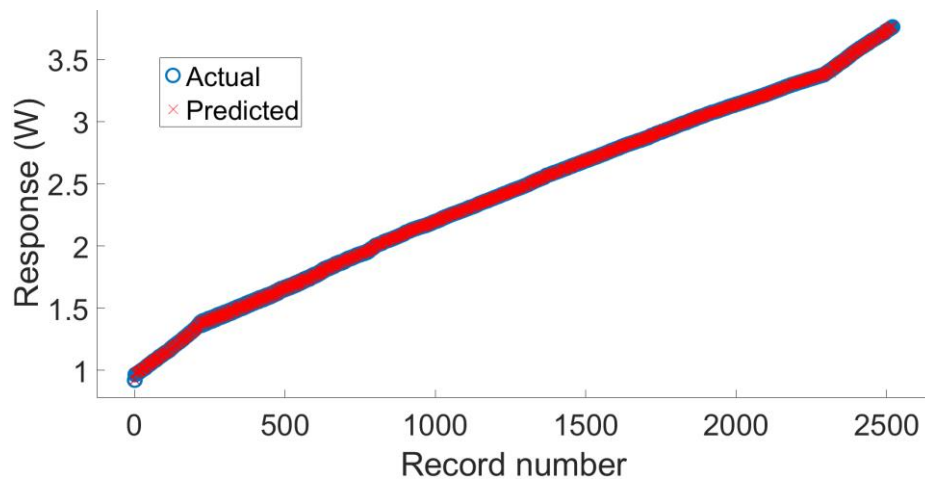
Figure 2. Actual and predicted training data

Figure 3. Correlation between the actual training data and the predicted data

2.3.1 Testing and prediction phase

All the ML methods used during training are also used during the testing phase. Table 7 presents a sample of the test data.

Table 7. Actual test data

Parameter	Test Data 1	Test Data 280
G (W/m²)	200.0003864	994.103292
v (m/s)	0.501314	0.500595
Ta (ᵒC)	29.99990794	29.99981498
I (mA)	1.804981179	1.09343441
L (mm)	0.507706014	0.64921773
Ap/n (mm²)	4.933844599	4.91306995
h (W/(m² K))	285.2222814	567.3205518
Delta T (ᵒC)	19.99887316	19.99919087
Output Power (W)	0.966334315	3.745774637

As previously stated, the test data (comprising 280 samples) remains entirely separate from the training and validation processes. Therefore, this test data is entirely new to the ML algorithms. During the testing phase, the ML methods predict the output power since the actual output of the test dataset is

not provided.

3. RESULTS AND DISCUSSION

Reiterating the two primary objectives of maximizing the output power and energy efficiency of the CPV-TE system, it is important to note that achieving the optimal value for both objectives simultaneously is unfeasible. Hence, a Pareto optimal solution needs to be identified. Each point situated on the Pareto front signifies a trade-off between the two objective functions. Figure 4 illustrates the Pareto front generated by the PS, MOGA, and GOAL-MOGA algorithms. Comparing the optimization performance of MOGA and PS reveals that MOGA tends to find higher energy efficiency, whereas PS excels in finding higher output power. This highlights the varying strengths of different optimization algorithms when applied to the same problem. The extreme values of the energy efficiency and output power for the MOGA were 13.8% and 3.35 W, respectively. Likewise, the extreme values of the energy efficiency and output power for the PS were 12.9% and 3.75 W, respectively. To enhance the optimization performance, MOGA was integrated with the Goal Attainment algorithm. The optimization process begins with Goal attainment, and upon its completion, MOGA takes over. This hybrid approach improves the output power in comparison with the output power searched by the MOGA alone, while maintaining higher energy efficiency than that of the PS. Thus, the hybrid algorithm has the highest energy efficiency of 13.15% and highest output power of 3.75 W. This indicates 11.9% power enhancement of the hybrid algorithm in comparison with MOGA alone and 1.9% energy efficiency enhancement in comparison with the energy efficiency of PS.
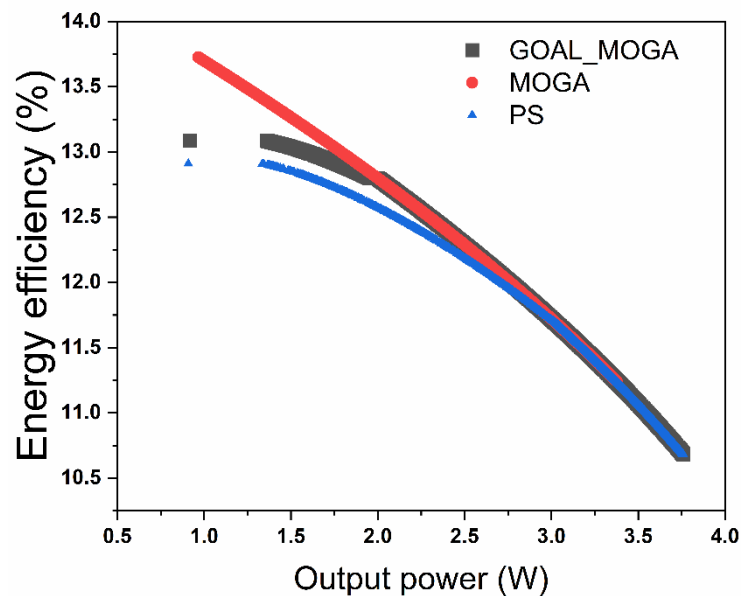
Figure 4. Pareto front: three algorithms

Figure 5 mirrors Figure 4, with the exclusion of the PS algorithm due to its inferior performance, hence not utilized in training the ML algorithms. The rationale behind combining the datasets from MOGA and GOAL-MOGA is to develop an ML model that surpasses any single optimization algorithm. This implies that the ML algorithm can predict the highest optimized energy efficiency similar to MOGA and the highest optimized output power similar to the GOAL-MOGA algorithm. This capability of the ML model can be improved by incorporating data from other optimization algorithms with varying performance levels.

Figure 5. Pareto front: two algorithms

Table 8 presents the prediction performance of the ML algorithms on the test dataset in descending order from the top to the bottom of the Table. For each algorithm, maximum, minimum, and average prediction errors were determined. The performance of the ML algorithms on the test data is similar to their performance on the training data as presented in Table 6. As can be seen in Table 8, the ANN-based ML has the lowest percentage of errors, indicating the highest prediction accuracy. On the other hand, the De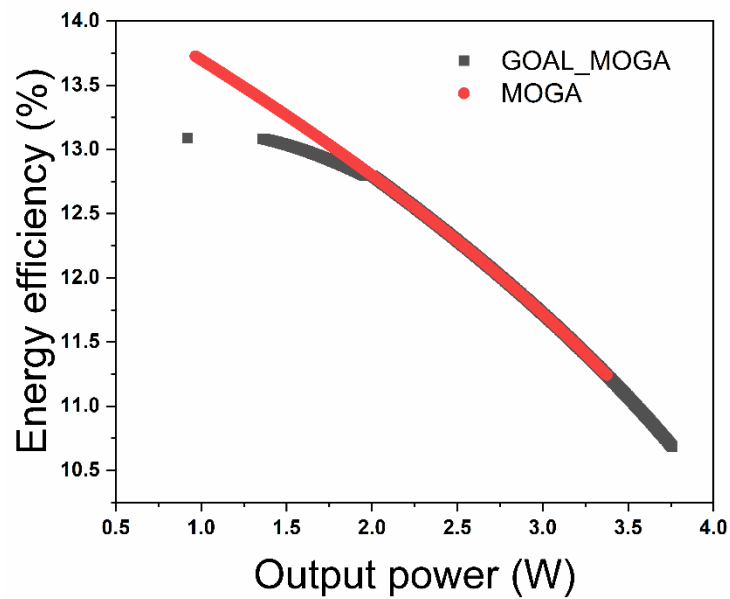cision Stump ML algorithm has the highest percentage of errors, indicating the least prediction accuracy. Due to its lowest prediction percentage error, the ANN is chosen for the prediction of the optimized output power.

Table 8. Prediction performance on the test dataset

ML Algorithm	Duration (seconds)	Max % Error	Min % Error	Average % Error
Artificial Neural Networks	**7.15**	**0.6898**	**0.0006**	**0.0692**
M5P	0.39	1.2589	0.0002	0.1873
M5Rules	0.22	1.2589	0.0011	0.1536
Random Committee	0.33	1.7246	0.0002	0.3580
Random Forest	0.17	2.6211	0.0048	0.7230
Bagging	0.23	12.3335	0.0040	3.2334
KStar	0.27	13.4517	0.0099	2.2435
Random Tree (RT)	0.22	13.7288	0.0157	2.3196
REPTree	24.11	15.7732	0.0103	2.3351
Linear Regression	10.74	16.9845	0.0210	2.2625
Support Vector Machines	0.25	17.7646	0.0103	3.4466
Simple Linear Regression	0.35	18.8024	0.0040	1.1230
Gaussian Process	0.24	19.1380	0.0058	5.6337
Regression by Discretization	0.17	22.2515	0.0024	1.1994
Decision Table	0.74	27.2764	0.0020	0.6261
Additive Regression	0.32	47.8816	0.0977	9.5229
Random Sub Space	4.42	50.5384	0.0039	5.9131
LWL	2.67	74.6807	0.0898	14.3199
Decision Stump	0.17	81.0967	0.0804	15.7659

Figure 6 illustrates the ANN-based ML model's prediction of the optimized output power on the test data. The high prediction accuracy results in the actual and predicted output power being nearly identical. The test data was randomly selected from the combined dataset as shown in Figure 5, leading to predicted optimized output power ranging from 1 W to 3.75 W.
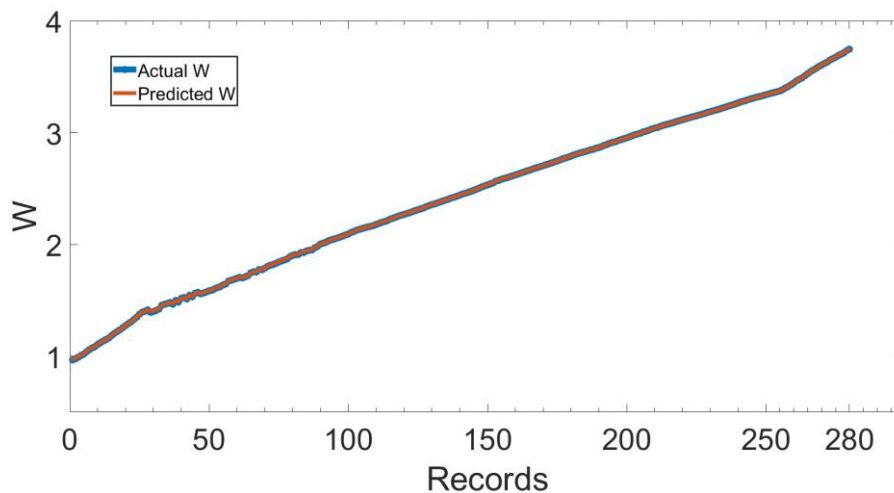
Figure 6. Actual and predicted test data

Figure 7 presents the correlation between the actual test data and the predicted optimized output power using the ANN-based ML model. As the correlation coefficient of the ANN-based ML stands at $R^2 = 1$, a strong agreement between the two responses is obtained. Consequently, a well-trained ANN-based ML algorithm can accurately predict the optimized output power of the CPV-TE system. This outcome is expected to match or potentially surpass that achieved by MOGA or GOAL-MOGA algorithms. Furthermore, the optimization process using the ANN-based ML will be notably swifter than the aforementioned metaheuristic optimizations. While enhancing the performance of a single optimization algorithm is typically challenging, augmenting the optimization capability of the ANN-based ML algorithm can be accomplished by integrating additional datasets from various optimization techniques. This study underscores the substantial value of ML in engineering applications.
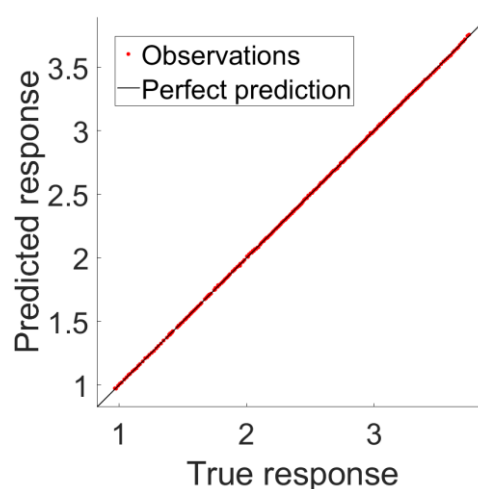
Figure 7. Correlation between the actual response and the prediction response

4. CONCLUSIONS

This study delves into the design and optimization of a concentrated photovoltaic-thermoelectric

hybrid system. Despite previous studies highlighting the potential for improved output performance by integrating photovoltaic and thermoelectric systems compared to standalone configurations, the conflicting operational characteristics of these subsystems necessitate the optimization of the hybrid system. Herein, PS, MOGA, and GOAL-MOGA metaheuristic optimizations are employed to optimize the output power and energy efficiency of the hybrid system. The optimization performance of PS was inferior to that of the MOGA and GOAL-MOGA algorithms, leading to its exclusion from further analysis.

Subsequently, the combined dataset from MOGA and GOAL-MOGA optimizations were utilized to train 19 distinct machine learning algorithms. Among these, an ANN-based ML algorithm trained in 48.22 seconds, exhibiting a correlation coefficient of $R^2 \approx 1$, an RMSE of 0.0019, and an MAE of 0.0012, emerges as the superior option. This ANN-based ML algorithm showcases an average prediction error of 0.0692% and a correlation coefficient $R^2 \approx 1$ on the test dataset, signifying its potential in predicting the optimized output power of the hybrid system. The ANN-based ML model performs better when trained on a combined dataset from both MOGA and GOAL-MOGA compared to using either MOGA or GOAL-MOGA alone. Furthermore, the optimization capability of the ANN-based ML algorithm can be further enhanced by incorporating additional datasets from other optimization techniques.

Declaration of Ethical Standards

Authors complied with all ethical guidelines including authorship, citation, data reporting, and publishing original research.

Credit Authorship Contribution Statement

AUTHOR1: Conceptualization, Methodology, Data collection, Validation, Writing- original draft.
AUTHOR2: Data curation, Visualization, Review, Investigation.
AUTHOR3: Data mining of training and testing phases, Software, Writing- original draft.
AUTHOR4: Supervision, Writing- review, and editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1]   A. Yusuf, S. Ballikaya, and H. Tiryaki, "Thermoelectric material transport properties-based performance analysis of a concentrated photovoltaic–thermoelectric system," *Journal of Electronic Materials*, vol. 51, no. 12, pp. 7198-7210, 2022.

[2]   K. Teffah and Y. Zhang, "Modeling and experimental research of hybrid PV-thermoelectric system for high concentrated solar energy conversion," *Solar Energy*, vol. 57, pp. 10-19, 2017.

[3]   A. Yusuf and S. Ballikaya, "Electrical, thermomechanical and cost analyses of a low-cost thermoelectric generator," *Energy*, Feb., vol. 241, p. 122934, 2022.

[4]   C. Maduabuchi, R. Lamba, H. Njoku, M. Eke, and C. Mgbemene, "Effects of leg geometry and multistaging of thermoelectric modules on the performance of a photovoltaic-thermoelectric system using different photovoltaic cells," *International Journal of Energy Research*, vol. 45, no. 12, pp. 17888-17902, 2021.

[5]   E. Yin, Q. Li, and Y. Xuan, "Experimental optimization of operating conditions for concentrating photovoltaic-thermoelectric hybrid system," *Journal of Power Sources*, Mar., vol. 422, pp. 25-32, 2019.

[6]   A. Yusuf and S. Ballikaya, "Thermal resistance analysis of trapezoidal concentrated photovoltaic–Thermoelectric systems," *Energy Conversion and Management*, Dec., vol. 250, p. 114908, 2021.

[7]   F. Rajaee, M. A. V. Rad, A. Kasaeian, O. Mahian, and W. M. Yan, "Experimental analysis of a photovoltaic/thermoelectric generator using cobalt oxide nanofluid and phase change material heat sink," *Energy Conversion and Management,* vol. 212, p. 112780, 2020.

[8]   A. Yusuf and S. Ballikaya, "Performance analysis of concentrated photovoltaic systems using thermoelectric module with phase change material," *Journal of Energy Storage*, vol.59, p. 106544, 2023.

[9]   E. Yin and Q. Li, "Device performance matching and optimization of photovoltaic-thermoelectric hybrid system," *Energy Conversion and Management: X*, vol. 12, p. 100115, 2021.

[10]   A. Yusuf, N. Bayhan, H. Tiryaki, B. Hamawandi, M.S. Toprak, and S. Ballikaya, "Multi-objective optimization of concentrated Photovoltaic-Thermoelectric hybrid system via non-dominated sorting genetic algorithm (NSGA II)," *Energy Conversion and Management*, vol. 236, p. 114065, 2021.

[11]   A. Menadi, S. Abdeddaim, A. Betka, and M. T. Benchouia, "Real Time Implementation of A Fuzzy Logic Based Mppt Controller for Grid Connected Photovoltaic System", *International journal of renewable energy research* Vol. 5, no. 1, 2015.

[12]   L. Suganthi, S. Iniyan, and A. A. Samuel, "Applications of fuzzy logic in renewable energy systems – A review", *Renewable and Sustainable Energy Reviews*, vol. 48, pp. 585–607, 2015,

[13]   H. Toylan, "Performance of Dual Axis Solar Tracking System Using Fuzzy Logic Control: A Case Study in Pinarhisar, Turkey", *European Journal of Engineering and Natural Sciences*, vol. 2, no. 1, 2017.

[14]   S.Choudhury, and P.K.Rout, "Adaptive Fuzzy Logic based MPPT Control for PV System Under Partial Shading Condition", *IJRER*, Vol. 5, no. 4, 2015.

[15]   C. Maduabuchi, "Thermo-mechanical optimization of thermoelectric generators using deep learning artificial intelligence algorithms fed with verified finite element simulation data," *Applied Energy*, vol. 315, p. 118943, 2022.

[16]   Z. He, M. Yang, L. Wang, E. Bao, and H. Zhang, "Concentrated photovoltaic thermoelectric hybrid system: an experimental and machine learning study," *Engineered Science*, vol. 15, pp. 47-56, 2021.

[17]   K. S. Garud, S. Jayaraj, and M.Y. Lee, "A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models," *International Journal of Energy Research*, vol. 45, no. 1, pp. 6-35, 2021.

[18]   J. H. Yousif and H. A. Kazem, "Prediction and evaluation of photovoltaic-thermal energy systems production using artificial neural network and experimental dataset," *Case Studies in Thermal Engineering*, vol. 27, p. 101297, 2021.

[19]   K. S. Garud, J. H. Seo, C. P. Cho, and M. Y. Lee, "Artificial neural network and adaptive neuro-fuzzy interface system modelling to predict thermal performances of thermoelectric generator for waste heat recovery," *Symmetry*, vol. 12, no. 2, p. 259, 2022.

[20]   I. S. Ameenuddin, K. Irshad, A. Algahtani, B. Azeem, V. Tirth, S. Algarni, and M. A. Abdelmohimen, "Machine learning-based modeling of thermoelectric materials and air-cooling system developed for a humid environment," *Materials Express*, vol. 11, no. 2, pp. 153-165, 2021.

[21]   H. Alghamdi, C. Maduabuchi, A. Yusuf, S. Al-Dahidi, A. Albaker, I. Alatawi, and M. Alkhedher, "Multiobjective Optimization and Machine Learning Algorithms for Forecasting the 3E Performance of a Concentrated Photovoltaic-Thermoelectric System," *International Journal of Energy Research*, vol. 2023, pp. 1-22, 2023.

[22]   R. A. Kishore, R. L. Mahajan, and S. Priya, "Combinatory finite element and artificial neural network model for predicting performance of thermoelectric generator," *Energies*, vol. 11, no. 9, p. 2216, 2018.

[23]   A. A. Angeline, L. G. Asirvatham, D. J. Hemanth, J. Jayakumar, and S. Wongwises, "Performance prediction of hybrid thermoelectric generator with high accuracy using artificial neural networks," *Sustainable Energy Technologies and Assessments*, vol. 33, pp. 53-60, 2019.

[24]   P. Wang, K. Wang, L. Xi, R. Gao, and B. Wang, "Fast and accurate performance prediction and optimization of thermoelectric generators with deep neural networks," *Advanced Materials Technologies*, vol. 6, no. 7, p. 2100011, 2021.

[25]   D. L. King, J. A. Kratochvil, and W. E. Boyson, "Photovoltaic array performance model, SANDIA Report," *Department of Energy (US)*, Dec., p. Report No.: SAND2004-3535, 2004.

[26]   R. Lamba and S. C. Kaushik, "Solar driven concentrated photovoltaic-thermoelectric hybrid system: Numerical analysis and optimization," *Energy Conversion and Management*, vol. 170, pp. 34-49, 2018.

[27]   P. Motiei, M. Yaghoubi, E. GoshtashbiRad, and A. Vadiee, "Two-dimensional unsteady state performance analysis of a hybrid photovoltaic-thermoelectric generator," *Renewable Energy*, vol. 119, pp. 551-565, 2018.

[28]   K. Deb, *Multi-objective optimisation using evolutionary algorithms: an introduction In: Multi-objective evolutionary optimisation for product design and manufacturing*. 1st ed., New York: John Wiley & Sons, 2011, pp. 3-34.

[29]   MATLAB & Simulink-MathWorks, "Paretosearch Algorithm," [Online]. Available: https://uk.mathworks.com/help/gads/paretosearch-algorithm.html. [Accessed September 14, 2023].

[30]   MATLAB & Simulink-MathWorks, "Find points in Pareto set," [Online]: https://uk.mathworks.com/help/gads/paretosearch.html. [Accessed September 22, 2023].

[31]   MATLAB & Simulink-MathWorks, "Effects of Multiobjective Genetic Algorithm Options," [Online]. Available: https://www.mathworks.com/help/gads/gamultiobj-options-effects.html. [Accessed September 22, 2023].

[32]   MATLAB & Simulink-MathWorks, "When to Use a Hybrid Function," [Online]. Available: https://www.mathworks.com/help/gads/when-to-use-hybrid-function.html. [Accessed September 22, 2023].

[33]   M. T. Akçay, A. Akgundogdu, and H. Tiryaki, "Prediction of travel time for railway traffic management by using the AdaBoost algorithm," *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 24, no. 1, pp. 300-312, 2022.

APPENDIX A

Table A1: Sample optimization output of the MOGA.

G (W/m²)	v (m/s)	Ta (⁰C)	I (mA)	L (mm)	$A_{n/p}$ (mm²)	h (WK⁻¹ m⁻²)	ΔT (ºC)	P (W)	Eff (%)
200.000386	0.5013	29.99991	1.80498	0.50771	4.933846	285.222281	19.998873	0.966334	13.726313
848.013699	0.5013	29.99924	1.41919	0.63165	4.852545	282.134138	19.997093	3.361266	11.260490
843.439852	0.5012	29.99876	1.79998	0.91075	4.862054	282.252476	19.997537	3.348482	11.278495
852.141534	0.5016	29.99814	1.92321	0.85974	4.826997	282.342032	19.995078	3.373186	11.245685
211.114279	0.5025	29.99791	1.62313	0.65108	4.810399	283.259538	19.988916	1.016926	13.684499
844.193420	0.5012	29.99883	1.70337	0.68983	4.603974	282.408527	19.998209	3.350511	11.275256
203.740383	0.5004	29.98741	1.64914	0.54866	4.814525	283.953076	19.981623	0.983456	13.713087
204.872787	0.5019	29.99917	1.48108	0.56830	4.804085	283.720155	19.976610	0.988563	13.708105
213.646280	0.5013	29.99945	1.52868	0.57289	4.791397	283.149118	19.995959	1.028365	13.674431
849.483425	0.5017	29.99122	1.90443	0.54373	4.923843	281.996238	19.995775	3.365675	11.255755
393.063516	0.5021	29.99864	1.52785	0.59747	4.764583	282.947521	19.996756	1.797536	12.991889
487.055165	0.5018	29.99905	1.55586	0.57390	4.776474	282.856822	19.993667	2.166057	12.634238
227.547113	0.5037	29.99904	1.78371	0.58734	4.792177	283.271018	19.996324	1.091067	13.621887
210.206866	0.5023	29.99421	1.68224	0.59594	4.791652	283.475314	19.992676	1.01282	13.688090
827.454683	0.5013	29.99917	1.68226	0.63185	4.840436	282.921384	19.997646	3.302599	11.338849
204.204056	0.5026	29.99921	1.50795	0.52660	4.850762	283.745195	19.997963	0.985498	13.710358
217.119467	0.5013	29.99911	1.90186	0.62686	4.682009	283.232852	19.996883	1.044096	13.661525
805.971713	0.5019	29.9983	1.62445	0.66246	4.815351	282.356393	19.994432	3.240161	11.420998
852.750068	0.5016	29.99813	2.50820	1.28033	4.867599	282.280735	19.997811	3.375201	11.244372
816.118559	0.5025	29.9978	1.60988	0.63171	4.770346	282.516046	19.998265	3.269847	11.382339
830.213961	0.5016	29.99899	1.75287	0.65546	4.808196	282.611144	19.997284	3.310606	11.328561
245.859458	0.5015	29.99902	1.39935	0.63212	4.779866	282.868678	19.995307	1.172816	13.551909
411.782129	0.5020	29.99871	1.42370	0.62363	4.799575	282.634199	19.996793	1.872808	12.920611
238.247811	0.5034	29.99879	1.39908	0.60337	4.7805135	282.874929	19.994354	1.138948	13.581016

Table A2: Sample optimization output of the Paretosearch.

G (W/m²)	v (m/s)	Ta (⁰C)	I (mA)	L (mm)	$A_{n/p}$ (mm²)	h (WK⁻¹ m⁻²)	ΔT (ºC)	P (W)	Eff (%)
866.9499512	0.5	30	1	0.5	5	100	20	3.413964	11.187221
978.9499512	0.5	30	1	0.5	5	100	20	3.708125	10.760964
751.8813477	0.5	30	1	0.5	5	100	20	3.07674	11.625157
738.3613281	0.5	30	1	0.5	5	100	20	3.034789	11.676612
942.378418	0.5	30	1	0.5	5	100	20	3.615767	10.90015
845.4499512	0.5	30	1	0.5	5	100	20	3.35365	11.269047
753.8613281	0.5	30	1	0.5	5	100	20	3.082843	11.617621
949.6210938	0.5	30	1	0.5	5	100	20	3.634342	10.872586
751.8881836	0.5	30	1	0.5	5	100	20	3.076761	11.625131
817.8613281	0.5	30	1	0.5	5	100	20	3.274442	11.374046
879.8881836	0.5	30	1	0.5	5	100	20	3.449662	11.13798
585	0.5	30	100	4	1	100	20	2.508172	12.180323
805.5820313	0.5	30	1	0.5	5	100	20	3.238532	11.420779
737.8598633	0.5	30	1	0.5	5	100	20	3.033223	11.678521
949.628418	0.5	30	1	0.5	5	100	20	3.634361	10.872558
849.8613281	0.5	30	1	0.5	5	100	20	3.366126	11.252258
773.5820313	0.5	30	1	0.5	5	100	20	3.143051	11.542567
946.9499512	0.5	30	1	0.5	5	100	20	3.627508	10.882752
903.878418	0.5	30	1	0.5	5	100	20	3.514668	11.046676
825.8813477	0.5	30	1	0.5	5	100	20	3.297678	11.343523
849.8813477	0.5	30	1	0.5	5	100	20	3.366183	11.252182
882.8110352	0.5	30	1	0.5	5	100	20	3.457665	11.126856
745.9135742	0.5	30	1	0.5	5	100	20	3.058283	11.647869
946.8305664	0.5	30	1	0.5	5	100	20	3.627202	10.883206
785.8813477	0.5	30	1	0.5	5	100	20	3.180074	11.495757
753.8813477	0.5	30	1	0.5	5	100	20	3.082904	11.617545

Table A3: Sample optimization output of the GOAL-MOGA.

G (W/m²)	v (m/s)	Ta (⁰C)	I (mA)	L (mm)	$A_{n/p}$ (mm²)	h (WK⁻¹ m⁻²)	ΔT (ºC)	P (W)	Eff (%)
360.871689	0.5077	29.99864	99.90044	3.99988	1.000282	625.866164	19.991230	1.647717	12.971411
297.968850	0.5198	29.99789	99.91607	3.99992	1.000063	641.151626	19.994881	1.372038	13.081355
310.217479	0.5128	29.99897	99.89149	3.99976	1.000062	626.451889	19.988589	1.426465	13.063276
966.859799	0.5004	29.99998	1.14896	0.64442	4.842421	568.170675	19.999821	3.678387	10.808146
470.015757	0.5018	29.99993	1.129458	0.64991	4.816721	566.007426	19.972198	2.101173	12.700086
429.168312	0.5090	29.99842	99.85233	3.99992	1.000315	626.862322	19.988118	1.935272	12.810663
297.349445	0.5245	29.99895	99.84141	3.99995	1.000276	630.252947	19.996356	1.369279	13.082236
344.413708	0.5135	29.99871	99.98328	3.99999	1.000148	626.810153	19.991762	1.576522	13.003999
298.667001	0.5086	29.99893	99.8912	3.99995	1.000392	629.731606	19.984575	1.375086	13.079766
295.832371	0.5147	29.99932	99.92162	3.99992	1.000181	632.171377	19.989962	1.362393	13.083198
308.529584	0.5062	29.99918	99.88103	3.99990	1.000246	626.863755	19.984508	1.418979	13.065817
324.074627	0.5132	29.99878	99.96889	3.99976	1.000109	626.635429	19.990582	1.487608	13.040700
331.722227	0.5054	29.99899	99.86777	3.99983	1.000279	626.240193	19.993239	1.521237	13.028053
393.990425	0.5046	29.99973	99.86743	3.99989	1.000472	627.116501	19.991165	1.788697	12.897586
420.964817	0.5092	29.99952	99.8517	3.99980	1.000442	627.869491	19.994003	1.901391	12.831667
995.647940	0.5013	29.99992	1.095738	0.65418	4.921296	569.789844	19.999608	3.749608	10.698858
952.070853	0.5009	29.99985	1.135866	0.63580	4.935311	570.002570	19.999601	3.64103	10.864565
445.617767	0.5006	29.99968	1.093316	0.65274	4.794118	567.037360	19.998832	2.006505	12.791899
420.780858	0.5096	29.999	99.88572	3.99996	1.000372	626.248573	19.990103	1.900599	12.831925
444.988961	0.5051	29.99969	1.127076	0.63524	4.759751	563.825234	19.999653	2.004148	12.794927
429.807035	0.5011	29.99895	99.83445	3.99986	1.000528	627.248351	19.991078	1.937828	12.808523
355.394880	0.5083	29.99952	99.89478	3.99986	1.000233	626.838490	19.993045	1.624126	12.982722
461.675030	0.5005	29.9997	1.082932	0.65799	4.790491	570.236830	19.991772	2.068908	12.730991
303.478657	0.5052	29.99853	99.88477	3.99992	1.000311	627.653728	19.987631	1.396517	13.073002