Research Article

Multiobjective Optimization and Machine Learning Algorithms for Forecasting the 3E Performance of a Concentrated Photovoltaic-Thermoelectric System

Hisham Alghamdi,[1] **Chika Maduabuchi**,[2,3] **Aminu Yusuf**,[4,5] **Sameer Al-Dahidi**,[6] **Abdullah Albaker**,[7] **Ibrahim Alatawi**,[7] **Theyab R. Alsenani**,[8] **Ahmed S. Alsafran**,[9] **Mohammed AlAqil**,[9] **and Mohammad Alkhedher**[10]

[1]*Electrical Engineering Department, College of Engineering, Najran University, Najran 55461, Saudi Arabia*
[2]*Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA*
[3]*Artificial Intelligence Laboratory, University of Nigeria Nsukka, Nsukka, 410001 Enugu, Nigeria*
[4]*Department of Engineering Sciences, Istanbul University-Cerrahpaşa, Avcılar, Istanbul 34320, Turkey*
[5]*Department of Electrical and Electronics Engineering, Federal University-Dutsinma, Katsina, P.M.B. 5001, Nigeria*
[6]*Department of Mechanical and Maintenance Engineering, School of Applied Technical Sciences, German Jordanian University, Amman 11180, Jordan*
[7]*Department of Electrical Engineering, College of Engineering, University of Ha'il, Ha'il 81451, Saudi Arabia*
[8]*Department of Electrical Engineering, College of Engineering in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia*
[9]*Department of Electrical Engineering, College of Engineering, King Faisal University, Alahsa, 31982, Saudi Arabia*
[10]*Department of Mechanical and Industrial Engineering, Abu Dhabi University, Abu Dhabi 59911, UAE*

Correspondence should be addressed to Chika Maduabuchi; chika691@mit.edu, Aminu Yusuf; aminu.yusuf@iuc.edu.tr, Sameer Al-Dahidi; sameer.aldahidi@gju.edu.jo, and Theyab R. Alsenani; t.alsenani@psau.edu.sa

Received 6 December 2022; Revised 20 January 2023; Accepted 13 February 2023; Published 15 March 2023

Academic Editor: Viknesh Andiappan

Previous theoretical research efforts which were validated by experimental findings demonstrated the thermo-economic benefits of the hybrid concentrated photovoltaic-thermoelectric (CPV-TE) system over the stand-alone CPV. However, the operating conditions and TE material properties for maximum CPV-TE performance may differ from those required in a standalone thermoelectric module (TEM). For instance, a high-performing TEM requires TE materials with high Seebeck coefficients and electrical conductivities, and at the same time, low thermal conductivities (k). Although it is difficult to attain these ideal conditions without complex material engineering, the low k implies a high thermal resistance and temperature difference across the TEM which raises the PV backplate's temperature in a hybrid CPV-TE operation. The increased PV temperature may reduce the overall system's thermodynamic performance. To understand this phenomenon, a study is needed to guide researchers in choosing the best TE material for an optimal operation of a CPV-TE system. However, no prior research effort has been made to this effect. One method of finding the optimum TE material property is to parametrically vary one or more transport parameters until an optimum point is determined. However, this method is time-consuming and inefficient since a global optimum may not be found, especially when large incremental step sizes are used. This study provides a better way to solve this problem by using a multiobjective optimization genetic algorithm (MOGA) which is fast and reliable and ensures that the global optimum is obtained. After the optimization has been conducted, the best performing conditions for maximum CPV-TE energy, exergy, and environmental (3E) performance are selected using the technique for order performance by similarity to ideal solution (TOPSIS) decision algorithm. Finally, the optimization workflow is deployed for 7000 test cases generated from 10 features using the optimal machine learning (ML) algorithm. The result of the optimization chosen by the TOPSIS decision-making method generated an output power, exergy efficiency, and CO_2 saving of 44.6 W, 18.3%, and 0.17 g/day, respectively. Furthermore, among other ML algorithms, the Gaussian process regression was the most accurate in learning the CPV-TE performance dataset, although it required more computational effort than some algorithms like the linear regression model.

1. Introduction

The adoption of renewable energy systems is on the rise because of the increasing pollution from nonrenewable energy systems based on fossil fuel combustion [1]. Environmental pollution is more prevalent in African countries like Nigeria where the most polluted city in the world (Onitsha) was reported to have a poor air quality, 30 times more than the world health organization's recommended levels [2]. The reason behind this level of pollution is due to the primary dependence on fossil fuel-based engines for electricity generation since the electric distribution grid of the country is characterized by numerous failures with 2 nationwide blackouts in April 2022 alone [3, 4]. To alleviate these problems, several start-ups and government agencies have contributed by installing solar panels in different residential homes, offices, and tertiary institutions in the country [5]. Worthy of mention is the commencement of the biggest solar farm project in West Africa, the 200 MW Ashama solar farm in Delta State, Nigeria, that is expected to save 200 Mt in carbon dioxide emissions from fossil-powered engines [6]. These advances are very intriguing for Nigeria and Africa; however, the thermal operating conditions in Nigeria are quite harsh for solar panels, resulting in their untimely deterioration, performance losses, and ultimate premature failure [7, 8].

The high temperature incident on the solar panel reduces the obtainable efficiency from the crystalline silicon solar cell due to its thermal sensitivity [9, 10]. However, researchers have proposed several means of cooling the solar panel when exposed to harsh environmental conditions like concentrated solar using active, passive, phase change material, and thermoelectric cooling techniques [11, 12]. Among these proposed techniques, the thermoelectric cooling method stands out due to the relative effectiveness, commercial availability, and affordability of the thermoelectric module (TEM) in cooling the concentrated solar photovoltaic (CPV) system [13–15]. Some researchers have experimentally demonstrated how the CPV-TE hybrid system was capable of generating a higher power and efficiency than the stand-alone CPV system by utilizing the entire broad solar spectrum [16–19]. Khan et al. [16] recorded a 5.5% drop in the solar cell temperature and a subsequent 17% increase in the overall system's efficiency when TEMs were integrated with PVs. Mahmoudinezhad et al. [17] concluded that the power generation rate of the hybrid CPV-TE was 11.67% higher than that of a stand-alone CPV system. Li et al. [18] improved the CPV-TE's overall efficiency by 8%, while Yin et al. [19] concluded that operating the CPV in tandem with the TEM improved the hybrid system's performance by 8.7%, especially under concentrated solar applications.

Further research avenues were proposed in the current literature regarding the possibilities of attaining higher performances from the CPV-TE when the properties of the thermoelectric and photovoltaic materials are optimized. Thus far, several parameters like the thermoelectric structure/geometry [20, 21], electrical parameters [22], and solar cell material properties [23] have been thoroughly investigated. The attention on the TEM geometry has been quite intense because of the cooling effect of this device on the hybrid system's performance. For instance, Shittu et al. [24] showed how the thermoelectric element shape can significantly alter the thermal transfer from the solar panel backplate temperature to the heat sink. In addition to the thermoelectric geometry, the thermoelectric material property significantly affects the performance of the hybrid system. Nevertheless, the interplay of the optimum thermoelectric material still remains uncovered.

A major component in the CPV-TE is the energy conversion material, like crystalline silicon in the solar PV or bismuth-telluride in the TEM which directly converts the incident solar photons or thermal energy, respectively, to electricity. Since the focus is on the cooling of the PV using TEM technology, the impacts of the thermoelectric material properties on the cooling of the CPV-TE are discussed. This means that it is crucial to determine the suitable property combinations needed to derive the optimum material. There seems to be a bottleneck in the discovery of more efficient thermoelectric materials [25, 26] due to the reliance of the thermoelectric figure of merit, ZT, on the Seebeck coefficient (S), electrical (σ), and thermal (k) conductivities according to the expression, $ZT = TS^2\sigma/k$. From this expression, there must be a reduction in the lattice thermal conductivity and an increase in the electrical conductivity for there to be an increase in the figure of merit. However, it is difficult to increase the electrical conductivity while decreasing the thermal conductivity since both properties are almost directly proportional in most materials. Hence, previous research efforts have focused on introducing impurities in the conventional bismuth-telluride compound to achieve this effect [27–29].

Despite these advances, the efficiencies obtained from thermoelectric modules are still in the range of 5-8% [30], indicating the need for more research thrusts in this domain. Most importantly, the thermoelectric material property complexity comes into play when the thermoelectric module is integrated in a hybrid photovoltaic-thermoelectric (PV-TE) system. This is because the thermal conductivity of the thermoelectric material is of design essence since the aim of integrating the TEM in the PV is to eradicate waste heat and increase the PV-TE's overall performance. This implies that the TEM with low thermal conductance will imply a higher thermal resistance and, consequently, a high temperature difference between the TEM's hot and cold junctions which will result in the overheating of the solar cell and the performance reduction of the PV module. To solve this problem, an investigation is needed to disclose the thermoelectric materials with optimal combinations that are needed for maximum CPV-TE operation. One method of achieving this is by parametrically varying one or more of the thermoelectric transport properties (S, σ, and k) until an optimum value is found. This method is time consuming and might not be able to obtain the global optimum point if a large incremental step size is used. Therefore, a better approach to this problem is to use the multiobjective genetic algorithm in addition to different machine learning models to be able to ensure the fast and reliable attainment of the global optimum point.

The effectiveness of the multiobjective genetic algorithm (MOGA) in optimizing the performance of several energy conversion systems like stand-alone TEMs [31–33] and PVs [34–37] has already been thoroughly investigated. Nevertheless, scarcely has the MOGA been applied in optimizing the performance of CPV-TEs. Akbar et al. [38] experimentally investigated the performance of a PV-TE comprising a solar PV module, a TE module, and a cooling system for waste heat removal from the assembly. The study involved using different mixtures as cooling fluids and assessing their impacts on the overall performance of the hybrid system. The experimentally obtained results were used as an initial population to develop a MOGA optimization of the hybrid system to obtain the optimal Pareto set for maximizing the system's performance. The GA optimization results were compared to those of the experimental results, and the results indicated that, irrespective of the cooling media used, the combined system's power and efficiency was always maximized. Yusuf et al. [22] utilized the nondominated sorting genetic algorithm (NSGA II) to optimize the performance of a CPV-TE operated with nanostructured thermoelectric materials for higher performance considering seven crucial operating parameters. Four distinct models of the CPV-TE were developed, and the best operating system was chosen using the TOPSIS decision-making algorithm. Obtained results concluded that models 1 and 2 were the best for the CPV-TE system, yielding peak power and efficiencies of 426.5 W, 11.45%, and 461.12 W, 10.77%, respectively. Similarly, the TOPSIS decision-making algorithm proved that model 4 was the best in optimizing the thermoelectric module's performance. In the same vein, Yin and Li [14] focused on conducting the multiobjective performance optimization and device matching of a CPV-TE using the genetic algorithm. Their results indicated that the least thermoelectric figure of merit was determined when the hybrid system was operated under low concentration ratios and good cooling conditions. Nevertheless, no previous scholarly effort has been made in the available literature to uncover the best combinations of thermoelectric materials to maximize the performance of CPV-TEs.

After the optimum performance of the CPV-TE has been obtained using optimization and decision-making algorithms like the MOGA and TOPSIS, there is need for the optimization process to be deployed using tools like machine learning (ML). ML is a subfield of artificial intelligence that seeks to learn patterns and trends in data, just like the brain, and make predictions based on the observed trends in the data [39]. The field of ML has gained increased interest in recent years, especially with the advent of neural networks that seek to transition the world from artificial narrow intelligence to artificial general intelligence [40]. The merits of ML over the conventional modelling tools used to analyze the performance of energy conversion devices like stand-alone PVs [41–43] and TEMs [44–46] have been comprehensively discussed in literature. Issues like modelling complexity, computational time consumption, massive energy requirements, and even deployment ease have been reportedly solved using ML [47–49]. Furthermore, the application of ML has further been extended in the energy system community to forecast weather parameters of cities and even countries, with the possibility of replacing meteorological stations due to the high accuracies and reliabilities reported in the ML models [50–52]. These studies disclose that the field of ML is fast growing and will keep finding crucial relevance in the coming years. Despite these advantages provided by ML, there is a deficit of literature showing how ML can be applied in forecasting the performance of solar CPV-TEs. This will be the focus of this research paper.

It is known that the performance of a PV reduces as the surface temperature of the PV increases above the reference temperature, typically 25°C. On the other hand, the performance of the TEM increases with an increase in the temperature gradient. Usually, parametric studies are conducted to determine the optimal point of operation of combined CPV-TE systems. The parametric study is time consuming, and there is no guarantee that global optimal points may be found. This brought about the need to incorporate an evolutionary algorithm to optimize the CPV-TE system. In this regard, many parameters will be optimized concurrently, and optimization goals will be achieved in a short period of time. The literature survey indicates that scarce efforts have been made to highlight the effectiveness of MOGA in optimizing the performance of CPV-TEs. Furthermore, the multiobjective optimization of the thermoelectric material combination needed to ensure efficient cooling of CPV-TEs has never been attempted before. Finally, although ML has been applied in predicting the performance of stand-alone PVs and TEMs, it is yet to find its application in predicting the performance of CPV-TEs. Therefore, the aim of this work is to conduct a multiobjective optimization and TOPSIS decision-making to determine the optimal operating conditions for maximum power production, exergy efficiency, and carbon dioxide savings per day. Thereafter, the data generated during the optimization scheme is fed to diverse machine learning models to uncover the best model for deployment and upscaling in industrial applications.

2. Methodology

2.1. Multiobjective Genetic Algorithm (MOGA) and TOPSIS Decision Method. Multiobjective evolutionary algorithms are attracting attention in scientific and nonscientific optimization problems due to their ability to generate multiple Pareto optimal fronts in a single run. One of the most popular evolutionary algorithms used in multiobjective optimization is an elitist nondominated sorting genetic algorithm (NSGA-II) [53]. Herein, the NSGA II is chosen because of the success it recorded in optimizing many engineering problems. In NSGA-II, an offspring population is generated from the initial parent population using evolutionary processes: mutation and recombination. The two populations are combined and then classified using a nondominated sorting method. In this way, the elitist population made up of both the parents and the offspring is retained and moved to the next generation. In a minimization problem, the front with the lowest fitness values is ranked in the first category and considered the best. This continues until all the solutions are assigned a front in descending order. A new

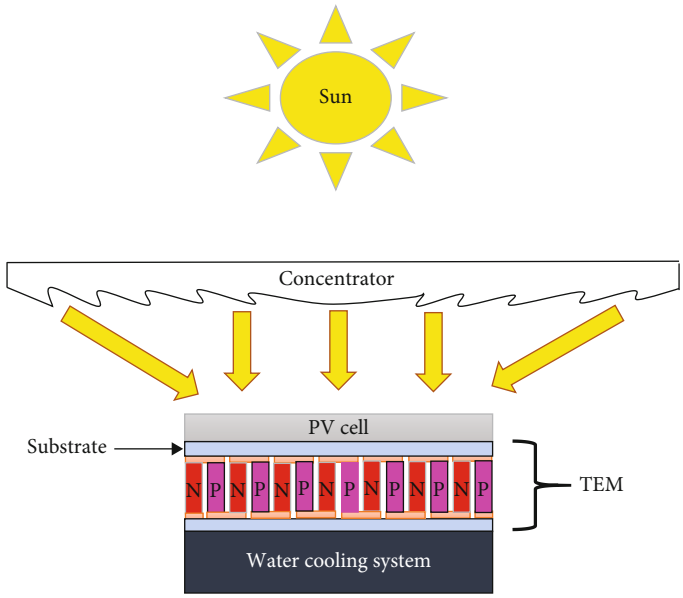


Figure 1: Schematic of the CPV-TE system.

population is formed by filling in the solutions of different nondominated fronts starting with the solutions of the best front and then moving in descending order. If the number of the slots to be filled is lower than the number of the solutions in the last front to be considered, a crowded tournament selection would be applied to fill in the remaining slots. A nondominated rank and crowded distance are the two criteria used in crowded sorting. Let us assume that solution A is compared with solution B; solution A wins the tournament either if it has a better rank than solution B or if they have the same rank, but the crowding distance of solution A is higher than that of solution B. The crowding distance of a solution measures the distance in the search space around that solution and any other solution in the population. Having filled all the slots of the next population, the above steps are repeated until any of the stopping criteria is met. On termination, the algorithm reveals several nondominated outputs known as Pareto optimal solutions. To ensure that the final solutions are the global optimum, the algorithm can be run several more times. Herein, to optimize the transport properties of the TE materials and to generate enough data for the machine learning, a population size of 10000, maximum number of generations of 9000, mutation probability of 0.125, crossover distribution index of 10, mutation distribution index of 20, and crossover probability of 0.9 are used, and the optimization was conducted using MATLAB® environment.

Once the Pareto optimal solutions are revealed, a decision-making method is required to select the point with the best compromise between the objective functions. A TOPSIS decision-making method [54] has been used successfully in both engineering and nonengineering fields. Since the Pareto optimal solutions lie between the ideal and nadir solutions, the TOPSIS solution is the point that is closest to the ideal solution. This point strongly depends on the relative importance of each objective function defined by the user; the TOPSIS solution will change by changing the relative importance of the objective functions.

2.2. Mathematical Formulations and Boundary Conditions. Figure 1 shows the schematic of the CPV-TE system; the sun energy passing through a concentrator is focused on the glass of the solar cell where part of the energy is converted into electrical energy. Through the thermalization process, the other part of the energy is converted into heat energy and, if not properly dissipated, will heat up the PV. This affects the performance of the PV as well as may reduce the lifespan of the PV. However, this effect can be mitigated by coupling a TEM with the PV. The TEM utilizes the excess thermal energy by converting it into electrical energy. Moreover, incorporating a cooling system can further improve the performance of the CPV-TE system.

One-dimensional energy balance (Equations (1)–(3)) can be used to model the CPV-TE system [55]. Equation (1) models the thermal and electrical processes from the top surface of the PV to the hot side of the TEM with consideration of convection and radiation losses, while Equation (2) gives the thermal and electrical processes from the hot side to the cold side of the TEM. Similarly, the heat transfer by conduction from the TE legs to the cold side is represented by Equation (3) [55]:

$$CGA_{pv}\tau_g\alpha_{pv} - \varepsilon A_{pv}\gamma\left(T_{pv}^4 - T_a^4\right) - h_{conv}A_{pv}\left(T_{pv} - T_a\right) - \frac{T_{pv} - T_h}{R_{pv-te}} - CGA_{pv}\tau_g\eta_r\left(1 - \beta\left(T_{pv} - 298\right)\right) = 0, \tag{1}$$

$$\frac{T_{pv} - T_h}{R_{pv-te}} - nR_L\left(\frac{S(T_h - T_c)\left(A_L\sigma_p\sigma_n\right)}{L\left(\sigma_p + \sigma_n\right) + R_L\left(A_{TE}\sigma_p\sigma_n\right)}\right)^2 - \frac{T_c - T_w}{R_{c-hs}} = 0, \tag{2}$$

$$\frac{NA_{\mathrm{L}}(T_{\mathrm{h}}—T_{\mathrm{c}})(k_p + k_n)}{L} - \frac{T_{\mathrm{c}}—T_{\mathrm{w}}}{R_{\mathrm{c-hs}}} = 0, \quad (3)$$

where C is the concentration ratio, G is the solar irradiance, A_{pv} is the surface area of the PV, τ_{g} is the transmissivity of glass, α_{pv} is the absorptibity of the PV, ε is the emissivity of PV cell, γ is the Stefan Boltzmann constant, S is the average Seebeck coefficient of the hot and cold sides, n is the number of TE couples, A_{TE} is the surface area of a TE leg, σ_p and σ_n are electrical conductivities of p- and n-type TE legs, k_p and k_n are thermal conductivities of p- and n-type TE legs, L is the length of the TE leg, $R_{\mathrm{pv-te}}$ is the thermal resistance between the top surface of the PV to the hot side of the TEM, $R_{\mathrm{c-hs}}$ is the thermal resistance between the cold side and the heat sink, R_{L} is the load resistance of a unicouple, h_{conv} is the convective heat transfer coefficient, T_{a} is the ambient temperature, T_{pv} is the surface temperature of the PV, T_{h} is the hot side temperature of the TEM, T_{c} is the cold side temperature of the TEM, β is the efficiency temperature coefficient of the PV, η_{r} is the reference efficiency of the PV, and T_{w} is obtained by taking the average of temperatures of water before and after passing through the heat sink, and in this study, it is assumed to be 300 K. If the temperatures of the PV (T_{pv}), hot (T_{h}), and cold (T_{c}) sides of the TEM are the only unknowns in Equations (1)–(3), the Newton-Raphson method can be used to compute the values of the temperatures and subsequently the output powers and energy efficiencies of the PV and the TEM. However, since there are other unknown variables, the Newton-Raphson method is coupled with a multiobjective genetic algorithm (NSGA-II) to determine the temperatures and, at the same time, the optimized three objective functions. The first objective function is the output power of the system given by Equation (4). This equation is obtained by adding up the power contributions of the PV and the TEM. Note that the output power of the TEM is given in terms of the transport properties of the TE materials. A thermosyphon system is assumed to be the cooling system, hence no power consumption by a water pump.

$$P_{\mathrm{sys}} = CGA_{\mathrm{pv}}\tau_{\mathrm{g}}\eta_{\mathrm{r}}\left(1 - \beta\left(T_{\mathrm{pv}} - 298\right)\right) + nR_{\mathrm{L}}\left(\frac{S(T_{\mathrm{h}} - T_{\mathrm{c}})(A_{\mathrm{L}}\sigma_p\sigma_n)}{L(\sigma_p + \sigma_n) + R_{\mathrm{L}}(A_{\mathrm{TE}}\sigma_p\sigma_n)}\right)^2. \quad (4)$$

The second objective function is the exergy efficiency of the system and is given as

$$\Psi_{\mathrm{sys}} = \frac{P_{\mathrm{sys}}}{CGA_{\mathrm{pv}}\left[1 - (4T_{\mathrm{a}}/3T_{\mathrm{s}}) + (1/3)(T_{\mathrm{a}}/T_{\mathrm{s}})^4\right]}. \quad (5)$$

Herein, an exergy efficiency is chosen over an energy efficiency because it measures the quality of energy that is in part destroyed or lost [56]. It can also be used to maximize the work done by the system by minimizing the exergy destruction in the system [57]. T_{s} is the Sun's temperature.

The amount of carbon saving per day (g/day) had it been the energy is generated by a gas thermal power plant is another parameter that is used to analyze the system and is given by:

$$\mathrm{CO}_{\mathrm{2saving/day}} = xE_{\mathrm{sys}}, \quad (6)$$

where E_{sys} is the daily energy generated by the system and is obtained by multiplying the P_{sys} by 8 h (assuming 8 h of suns per day) and x is the coefficient of CO_2 saving given as 0.475 kg/kWh [58].

The performance of a TEM operating in a CPV-TE system can be improved by either determining the optimum load resistance, geometrical optimization, minimizing radiation, and convection losses, minimizing contact losses, or optimizing the TE materials. In our previous work [22], the geometry of the TEM was optimized, and the optimum load resistance was determined. Since the transport properties of TE materials have a strong influence on the performance of the TEM, the goal of the optimization of this study is to determine the optimum values of the transport properties of n- and p- type TE materials and load resistance that will result in the highest output performance of the CPV-TE system. Thus, a maximum concentration ratio of up to 500 is considered in this study. Likewise, the internal resistance of a unicouple can be as low as $1 \times 10^{-3}\,\Omega$ and as high as $0.1\,\Omega$ depending on the geometry, fabrication method, and the properties of the TE materials. Moreover, Seebeck coefficients, electrical conductivities, and thermal conductivities of Bi_2Te_3-based TE materials can vary in the following ranges [59, 60]:

$$10 \le C \le 500, \quad (7)$$

$$\pm 100 \le S_{p/n} \le \pm 250\,\mu\mathrm{V/K}, \quad (8)$$

$$0.5 \le k_{p/n} \le 2\,\mathrm{W/(m \cdot K)}, \quad (9)$$

$$5 \times 10^4 \le \sigma_{p/n} \le 15 \times 10^4\,\mathrm{S/m}, \quad (10)$$

$$1 \times 10^{-3} \le R_L \le 0.1\,\Omega. \quad (11)$$

These variables can be directly seen in Equations (1)–(3). The assumptions made in this study are as follows:

(i) Steady state heat transfer from the top surface of the PV down to the heat sink via the TEM

(ii) Adiabatic side boundaries of the PV

(iii) Contact resistance within the TEM is about 10% of its internal resistance

(iv) Convection and radiation losses within the TEM are neglected

(v) To avoid damaging the PV, the input energy is limited by the concentration ratio, and thus, only a small temperature gradient can be created across the TEM. By that, the TE materials can be considered temperature independent

Table 1: Parameters of the CPV-TE system.

Parameter	Symbol	Value	Reference
Absorptivity of PV	α_{pv}	0.9	[55]
Area of PV	A_{pv}	16×10^{-4} m^2	[55]
Area of TE leg	A_L	2.25×10^{-6} m^2	Assumed
Atmospheric temperature	T_a	298 K	[55]
Efficiency temperature coefficient of PV	β_{pv}	3×10^{-3} K^{-1}	[61]
Efficiency at standard test conditions	η_r	30%	[61]
Length of TE leg	$\boldsymbol{L}$	1.8×10^{-3} m	Assumed
Number of thermocouples	$\boldsymbol{n}$	127	[61]
Solar radiation	G	1000 W/m^2	[62]
Stefan Boltzmann constant	σ	5.67×10^{-8} W m^{-2} K^{-4}	[55]
Wind speed	v	1 m s^{-1}	[55]
Thermal resistance between PV-TEM	R_{pv-te}	0.1 K/W	Calculated
Thermal resistance between TEM-heat sink	R_{c-hs}	1 K/W	Calculated
Transmissivity of glass	τ_g	0.95	[55]
Convective heat transfer coefficient	h_{conv}	9.5 W/(m^2 K)	Calculated

The parameters of the CPV-TE system used in this study are given in Table 1, where the convective heat transfer coefficient and the thermal resistances are calculated from Equations (12) and (13), respectively

$$h_{conv} = 5.8 + 3.7v, \tag{12}$$

$$\frac{R_{pv-te}}{R_{c-hs}} = \frac{t}{kA}, \tag{13}$$

where v is the wind speed, t is the thickness between the surface of the PV cell and the hot side of the TEM/cold side of the TEM and the top surface of the heat sink, A is the contact surface area, and k is the total thermal conductivity of the materials. Likewise, some parameters such as the surface area and length of the TE legs have been assumed. The assumed values are very similar to that of a commercial TEM.

The transport properties have an interdependency relationship with different forms of materials as shown in Figure 2 [63, 64]. Generally, an insulator material may have high Seebeck coefficient and very low electrical and thermal conductivities. On the contrary, a metal material may have high electrical and thermal conductivities but a very low Seebeck coefficient, while the transport properties of a semiconductor material (typically TE material) lie between that of an insulator material and a metal material. If a semiconductor is doped towards changing into metal, then the semiconductor may have large values of electrical and thermal conductivities while the Seebeck coefficient may be low. Ideally, the TE material should have large values of Seebeck coefficient and electrical conductivity, and a very low value of thermal conductivity. Due to the difficulty in obtaining a TE material with the ideal transport properties, several studies are focused on the optimization of the TE material. From Figure 2, without the application of complex material engineering, any two of the transport properties can have ideal values with the third transport property having a nonideal value. This fact is used as a benchmark in this study, that is, nonlinear constraints are defined in the algorithm such that the optimal TE materials can only occur when any of the two transport properties have favorable values that may lead to a high figure of merit. For example, an optimal TE material is found when values of the Seebeck coefficient and electrical conductivity are large (favorable condition); however, the value of the thermal conductivity is relatively large (unfavorable condition).

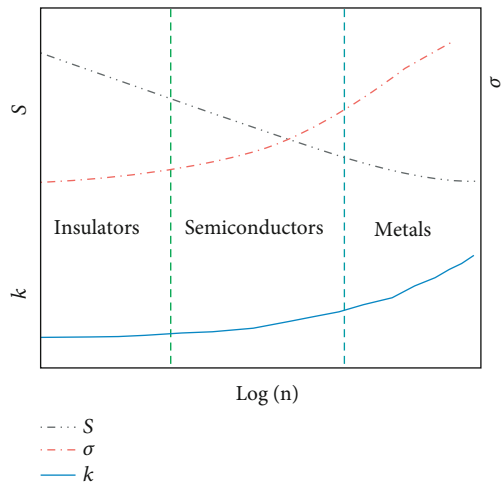


Figure 2: Relationship of transport properties on the carrier concentration.

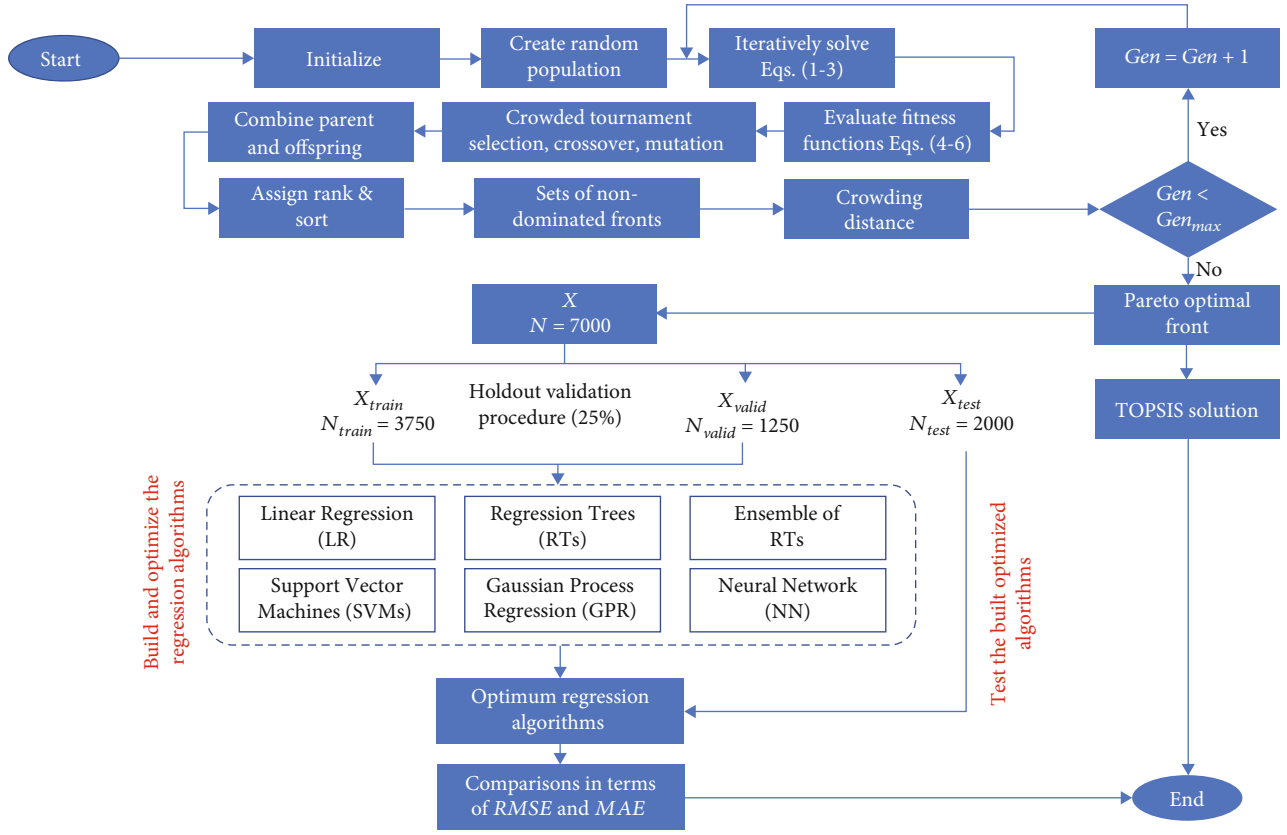


Figure 3: Flowchart of the MOGA, TOPSIS, and ML algorithms used to optimize the CPV-TE.

Parameters of the CPV-TE system is optimized using multiobjective genetic algorithm, and on termination of the algorithm, optimal Pareto front which has 7000 data points is revealed. The data points are used as input to the ML algorithms, and at the same time, TOPSIS decision method is used to determine the best compromise. The optimization process is completed once the TOPSIS decision has been made, while the ML process is completed once the ML algorithm with the best prediction performance has been identified. Herein, diverse ML algorithms (such as linear regression, regression trees, support vector machines, Gaussian process regression, ensembles of regression trees, and regressive neural network) have been built and optimized (in terms of their internal parameters) to predict the performance of the CPV-TE system for model deployment in industrial applications. For clarification purposes, a flowchart linking the MOGA, TOPSIS decision method, and the ML algorithms is provided and exemplified in Figure 3. Specifically, the flowchart summarizes the detailed steps of MOGA and TOPSIS solution (Figure 3 (upper)) (Section 2.1), and ML algorithms investigated to estimate the CPV-TE system's performance and the corresponding estimation evaluation process (Figure 3 (bottom)) (Section 2.3).

2.3. Machine Learning Dataset. The dataset generated in this work from the proposed CPV-TE system comprises the following features which significantly affect the performance of the hybrid system.

Table 2: Statistical characteristics of the CPV-TE's features.

Feature	Minimum	Maximum
Concentration ratio	10.0422	163.9594
Electrical conductivity p-type	5.0530×10^4	5.0532×10^4
Electrical conductivity n-type	5.0412×10^4	5.0413×10^4
Seebeck coefficient p-type	230.5651	232.9684
Seebeck coefficient n-type	-245.9402	-243.3020
Thermal conductivity p-type	0.5000	0.5047
Thermal conductivity n-type	0.5000	0.5078
External load resistance	1.0002	1.2297
Power output	3.7313	44.6872
Exergy	18.2930	24.9384
CO_2 saving	0.0142	0.1698

(1) Concentration ratio (C)

(2) Electrical conductivity p-type (S_p) (S/m)

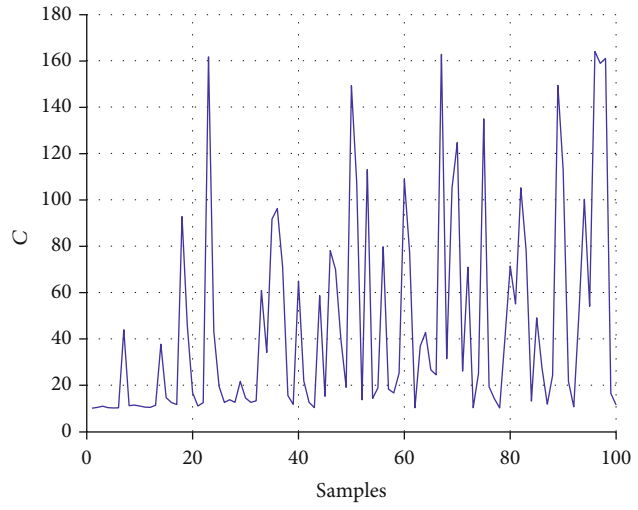
(3) Electrical conductivity n-type (S_n) (S/m)

(4) Seebeck Coefficient p-type (σ_p) (μV/K)

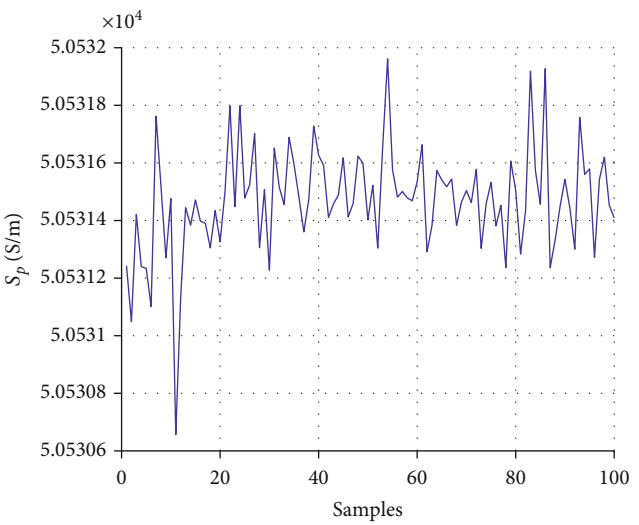
(5) Seebeck Coefficient n-type (σ_n) (μV/K)

(6) Thermal conductivity p-type (k_p) (W/(m.K))

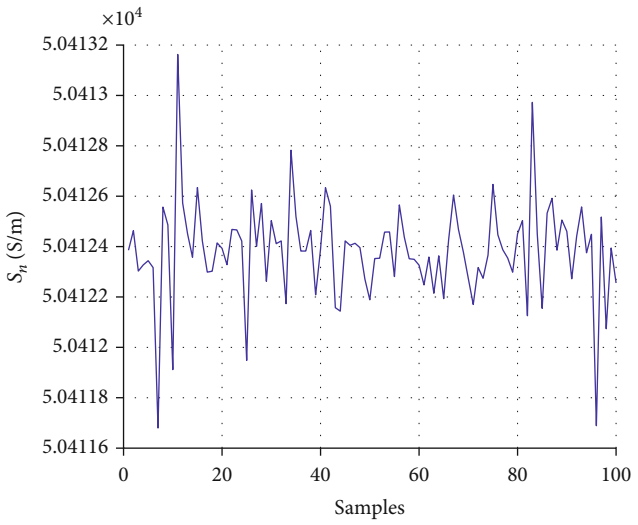
(7) Thermal conductivity n-type (k_n) (W/(m.K))



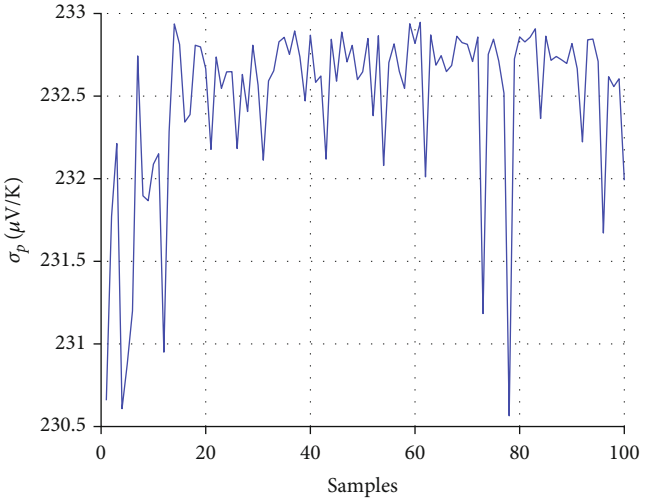
(a)



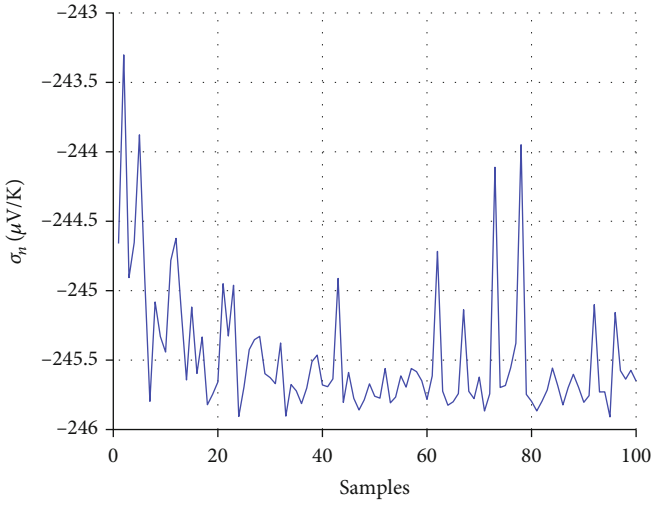
(b)



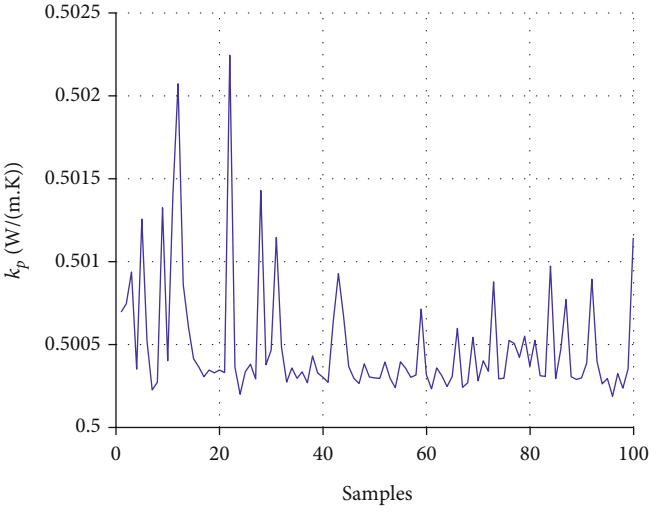
(c)



(d)



(e)



(f)

Figure 4: Continued.

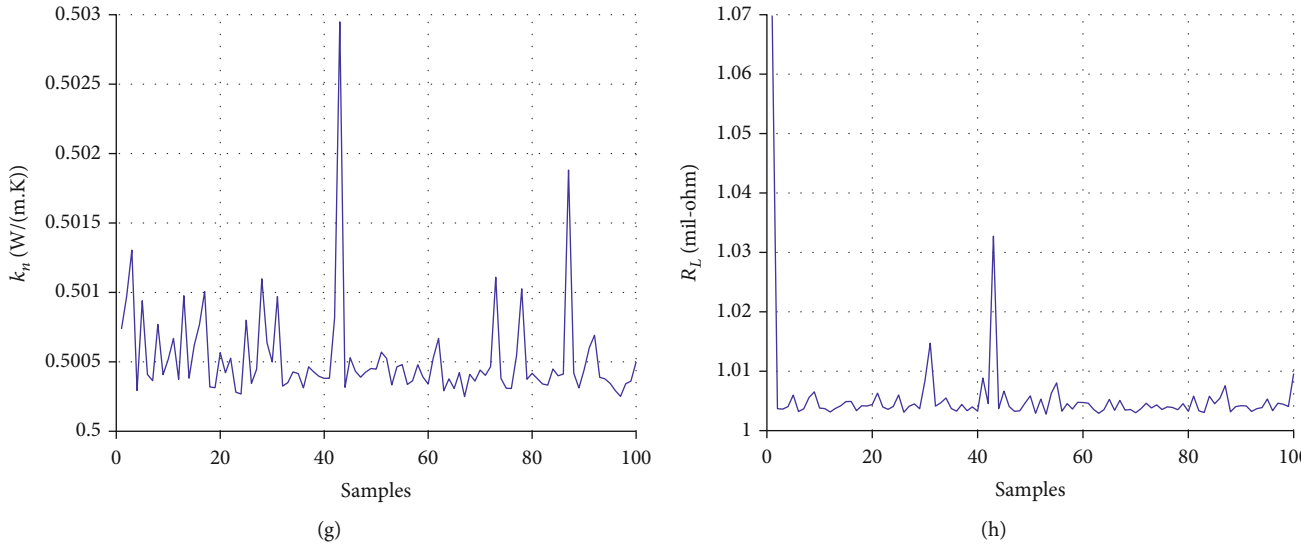


Figure 4: Input features' evolution. (a) Concentration ratio, (b) electrical conductivity p-type, (c) electrical conductivity n-type, (d) Seebeck coefficient p-type, (e) Seebeck coefficient n-type, (f) thermal conductivity *p*-type, (g) thermal conductivity *n*-type, and (h) external load resistance.

(8) External load resistance (R_L) (mΩ)

(9) System's power output (P_{sys}) (W)

(10) System's exergy (Ψ_{sys}) (%)

(11) CO_2 saving/day ($CO_{2saving/day}$) (g/day)

A total of $N = 7000$ data points have been generated, and the overall data points have been appended to the matrix $\mathbf{X}$. For estimation purposes, the former eight features (i.e., predictors) have been used as inputs to the investigated regression algorithms, whereas the latter three features (i.e., responses) have been used as individual targets (representative of the hybrid system's performance) while developing the regression algorithms.

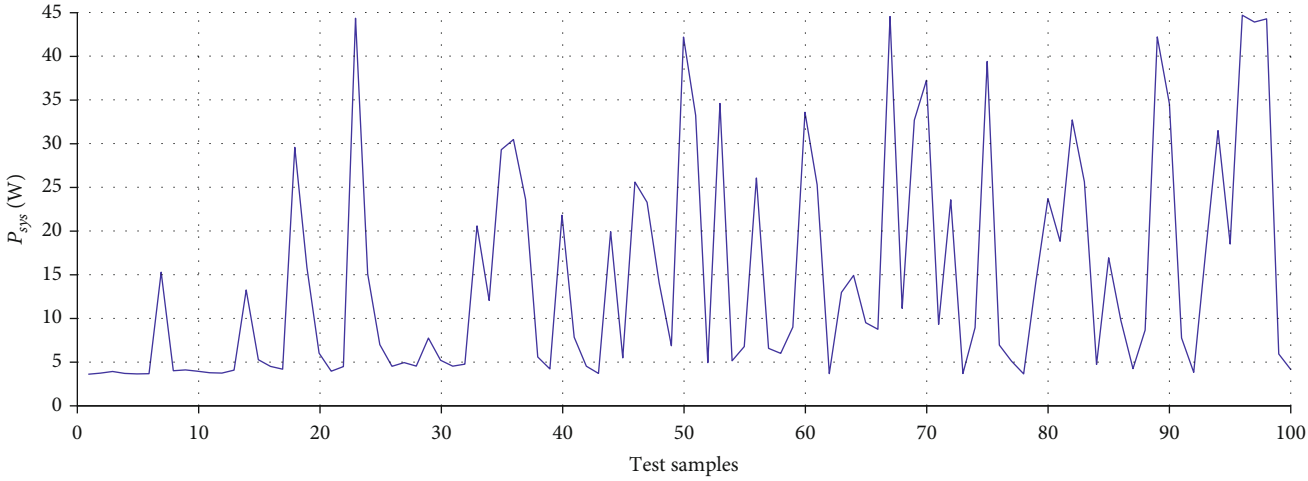
For clarification purposes, Table 2 reports the statistical characteristics of the features at hand, whereas Figures 4 and 5 show the features' evolutions of the inputs and outputs (the first 100 samples for illustration purposes), respectively. The inputs are the first 7 features while the targets are the last 3 features. The original range of the features in Table 2 were expressed in Equations (7)–(11). The statistical data from the features shown in Table 2 are selected based on the optimal values obtained from the MOGA process. The visualizations of these selections are discussed in the Results and Discussion. There is no specific distribution for each feature since a random distribution is assumed.

Furthermore, the initial optimization range for each parameter was provided in Equations (7)–(11) in which a wide range for each parameter was initially considered. The MOGA identified the optimum values for each parameter/feature as given in Table 2. Although the concentration ratio has the highest influence on the performance of the system, each parameter (Seebeck coefficient, electrical conductivity, thermal conductivity, and load resistance) can influence the overall performance of the system. It should be noted that small variations of the transport properties or the load resistance will not cause any significant change in the output performance of the system, but the optimization is needed to determine the optimum range of each parameter. Operating the hybrid system with optimum parameters can lead to a significant increase in the output performance.

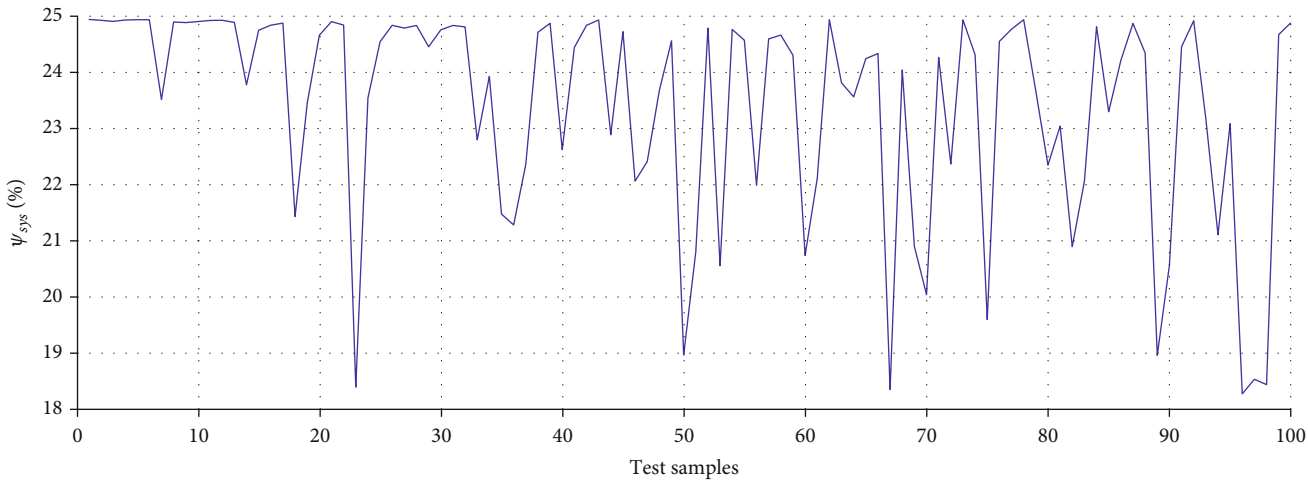
For the estimation task, the overall dataset ($\mathbf{X}$) has been divided into training, validation, and testing datasets/matrices used for building, optimizing, and evaluating the investigated regression algorithms, respectively. The former $N = 5000$ data points have been divided into training ($\mathbf{X}_{train}$) and validation ($\mathbf{X}_{valid}$) datasets—by resorting to the Holdout validation procedure with a fraction of 25%—to build and optimize the regression algorithms, respectively. In contrast, the remaining $N = 2000$ data points constitute the test dataset ($\mathbf{X}_{test}$) and used to evaluate the ultimately built algorithms. Specifically,

(1) $\mathbf{X}_{train}$. It comprises $N_{train} = 3750$ pairs of input-output selected randomly with a fraction of 75% from the first 6000 data points. The training dataset aims to develop/build/train the various regression algorithms being investigated in this work to capture the hidden mathematical relationship established between the inputs with each output

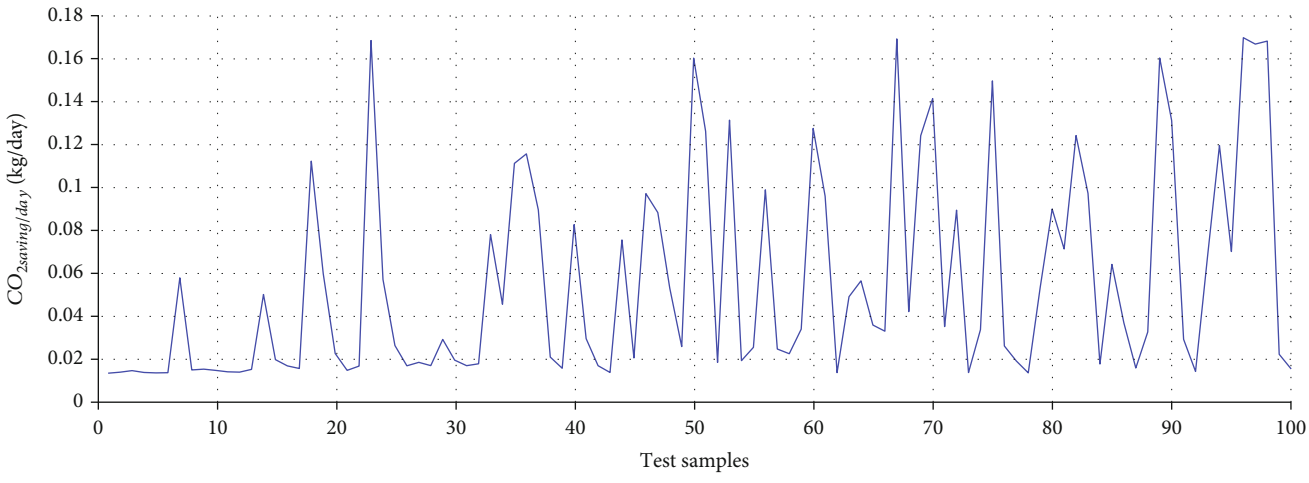
(2) $\mathbf{X}_{valid}$. It comprises $N_{valid} = 1250$ pairs of input-output remained from the first 6000 data points (a fraction of 25% from the first 6000 data points by resorting to the Holdout validation procedure). The validation dataset aims to optimize the investigated regression algorithms in terms of their configurations and hyperparameters to boost their predictability by resorting to various performance metrics from the literature



(a)



(b)



(c)

Figure 5: Output features' evolution.

(3) $\mathbf{X}_{test}$. It comprises unseen/new $N_{test} = 2000$ pairs of input-output selected initially to evaluate and compare the goodness of the built and optimized regression algorithms in terms of various performance metrics from the literature

2.4. Machine Learning Prediction Methodology. Various regression algorithms have been investigated in this work to accurately learn and predict the performance of the proposed hybrid CPV-TE system. Very specific, three parameters of the proposed hybrid system have been independently estimated (i.e., system's power output (W), system's exergy (%), and CO_2 saving (g/day)). The estimation results have been evaluated using the investigated regression algorithms in terms of various standard performance metrics from the literature.

To this aim, the Regression Learner App available in the Machine Learning and Deep Learning toolbox at 2021b MATLAB® [65] software has been utilized in this work. The investigated regression models range from linear regression (LR) models, regression trees (RTs), support vector machines (SVMs), Gaussian process regression (GPR) models, ensembles of RTs, and neural network (NN) models for regression. The investigated models have been effectively optimized in terms of their configurations and hyperparameters by using one of the inherent optimization algorithms available in the 2021b MATLAB® software toolbox [65]. The Bayesian optimization (BO) optimizer with 30 iterations (default settings) has been employed in this work because it is the most effective optimizer used in practice to optimize the regression models' configurations during the training/development stage [66].

The Holdout validation procedure has been employed to assess the goodness of the algorithms built on the training dataset and select its optimum configuration and hyperparameters. Once the optimum settings of each algorithm are defined, the algorithms have been used to estimate the proposed system's performance parameters of the "unseen" test dataset. The introduction of each algorithm, their detailed technical characteristics, and the employed performance metrics are hereafter reported in the following subsections for completeness.

2.4.1. Linear Regression (LR). Four linear regression models have been investigated, they are linear, interaction linear, robust linear, and stepwise learner. The model that provides minimum root mean square error (RMSE), minimum mean absolute error (MAE), and maximum coefficient of determination (R^2) will be selected for estimation and comparison purposes on the test dataset ($\mathbf{X}_{test}$). More details about the LR can be found in [67].

2.4.2. Regression Trees (RTs). The RT has been built and optimized in terms of minimum leaf size (e.g., 1-3400) by resorting to the BO optimizer. The best RT that satisfies the criteria mentioned above will be selected for estimation and comparison purposes on the test dataset ($\mathbf{X}_{test}$). More details about the RTs can be found in [68].

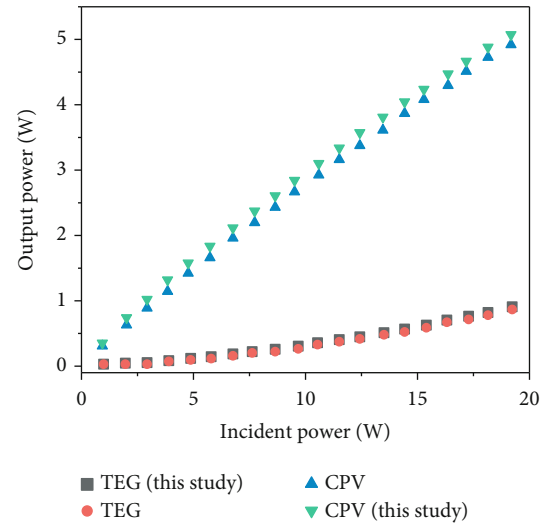


Figure 6: Validation of the present study.

2.4.3. Support Vector Machines (SVMs). The SVM has been built and optimized in terms of kernel function (e.g., linear, Gaussian, quadratic, and cubic), kernel scale (e.g., 0.001-1000), box constraint (e.g., 0.001-1000), and Epsilon value (e.g., 0.014111-1411.1379) by resorting to the BO optimizer. The best SVM that satisfies the criteria mentioned above will be selected for estimation and comparison purposes on the test dataset ($\mathbf{X}_{test}$). More details about the SVMs can be found in [69].

2.4.4. Gaussian Process Regression (GPR). The GPR has been built and optimized in terms of basic function (e.g., zero, constant, and linear), kernel function (e.g., Nonisotropic Exponential, Nonisotropic Matern 3/2, Nonisotropic Matern 5/2, Nonisotropic Rational Quadratic, Nonisotropic Squared Exponential, Isotropic Exponential, Isotropic Matern 3/2, Isotropic Matern 5/2, Isotropic Rational Quadratic, and Isotropic Squared Exponential), kernel scale (e.g., 0.15349-153.4916), and sigma value (e.g., 0.0001-111.7238) by resorting to the BO optimizer. The best GPR that satisfies the criteria mentioned above will be selected for estimation and comparison purposes on the test dataset ($\mathbf{X}_{test}$). More details about the GPR can be found in [70].

2.4.5. Ensemble of RTs. An ensemble of RTs has been built and optimized in terms of ensemble method (e.g., Bagging, LSBoost), minimum leaf size (e.g., 1-3400), number of learners (e.g., 10-500), learning rate (e.g., 0.001-1), and the number of predictors to sample (e.g., 1-8) by resorting to the BO optimizer. The best ensemble of RTs that satisfies the criteria mentioned above will be selected for estimation and comparison purposes on the test dataset ($\mathbf{X}_{test}$). More details about the ensemble of RTs can be found in [71–73].

2.4.6. Neural Network (NN). A NN regression model has been built and optimized in terms of the number of fully connected hidden layers (e.g., 1-3) and their number of neurons/nodes (e.g., 1-300 for each layer), hidden neuron/node activation function (e.g., Tanh, ReLU, sigmoid, and none),

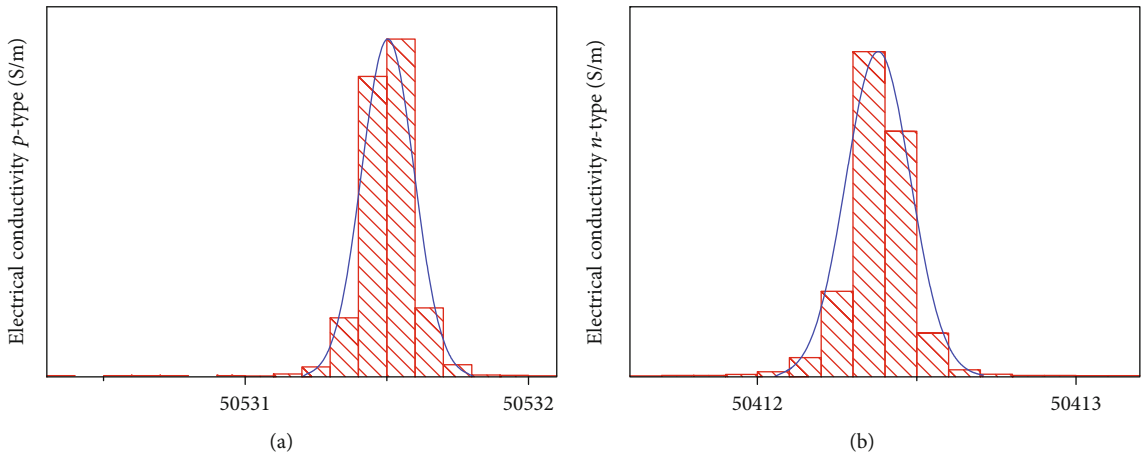


Figure 7: Search for the optimal electrical conductivity: (a) *p*-type; (b) *n*-type.

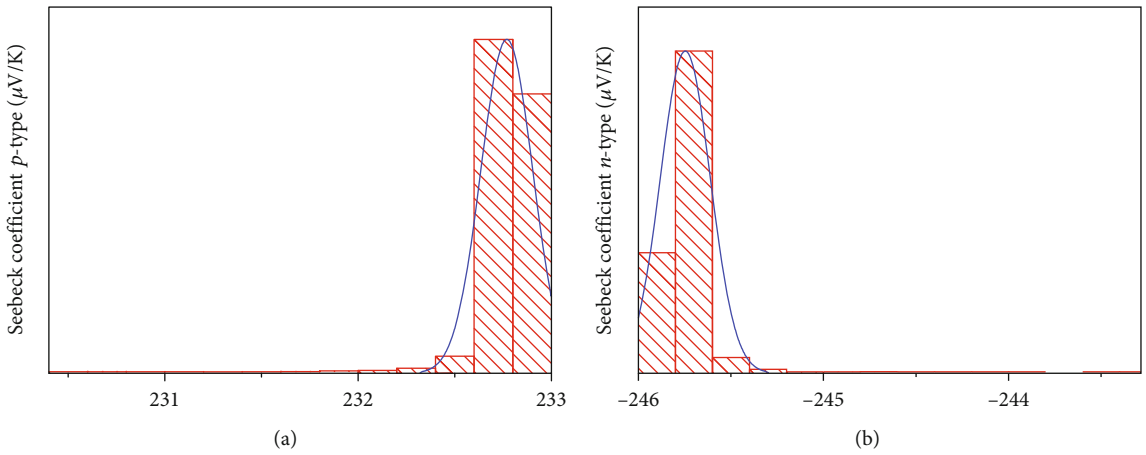


Figure 8: Search for the optimal Seebeck coefficient: (a) p-type; (b) n-type.

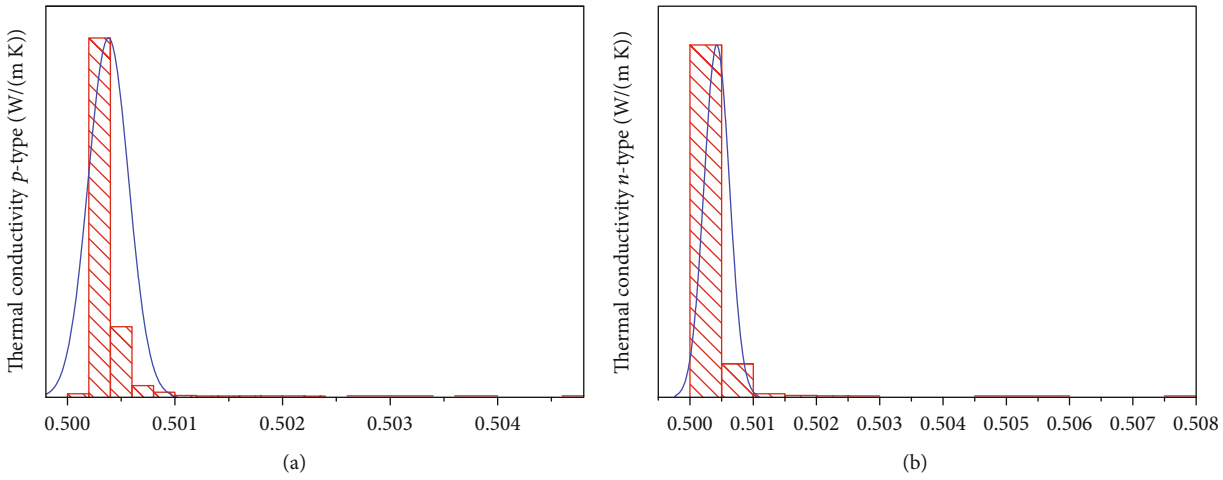


Figure 9: Search for the optimal thermal conductivity: (a) p-type; (b) n-type.

and regularization strength (lambda value—e.g., 1.4706*e*-09-14.7059) by resorting to the BO optimizer. The best NN regression model that satisfies the criteria mentioned above will be selected for estimation and comparison purposes on the test dataset ($\mathbf{X}_{\text{test}}$). More details about the NN can be found in [74].

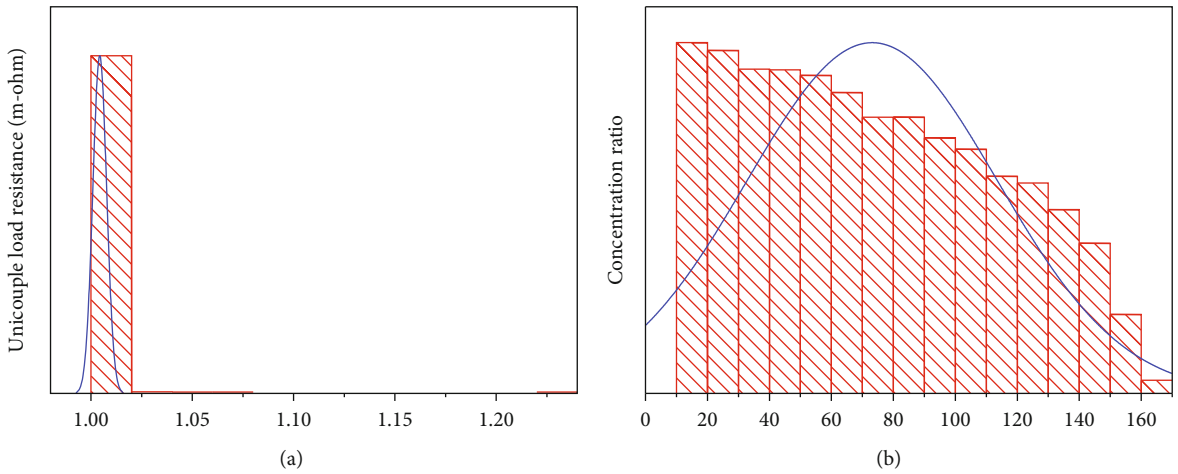


Figure 10: Search for the optimal (a) load resistance and (b) concentration ratio.

2.5. Machine Learning Performance Metrics. Two performance metrics are considered in this work to evaluate the predictability of the proposed hybrid system's performance of the built-regression models on the test dataset. They are (reported for the system's power output) the following:

(1) Root mean square error (RMSE). This metric computes the average mismatch between the actual (P_{sys}) and estimated ($\widehat{P}_{\text{sys}}$) system's power output on the test dataset. A small value of this metric (i.e., close to 0) entails the superiority of the investigated regression algorithm and vice versa [75]

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{N_{\text{test}}} \left(\widehat{P}_{\text{sys}_j} - P_{\text{sys}_j}\right)^2}{N_{\text{test}}}} \tag{14}$$

(2) Mean absolute error (MAE). This metric computes the average absolute mismatch between the actual (P_{sys}) and estimated ($\widehat{P}_{\text{sys}}$) system's power output on the test dataset. Like the RMSE, a small value of this metric (i.e., close to 0) entails the superiority of the investigated regression algorithm and vice versa [76]

$$\text{MAE} = \frac{\sum_{j=1}^{N_{\text{test}}} \left|\widehat{P}_{\text{sys}_j} - P_{\text{sys}_j}\right|}{N_{\text{test}}}, \tag{15}$$

where P_{sys_j} and $\widehat{P}_{\text{sys}_j}$ are the actual and estimated P_{sys} of the j th test pattern, $j = 1, \cdots, N_{\text{test}}$.

In addition to the above-mentioned accuracy metrics, the computational efforts required to develop and evaluate the regression algorithms have also been considered.

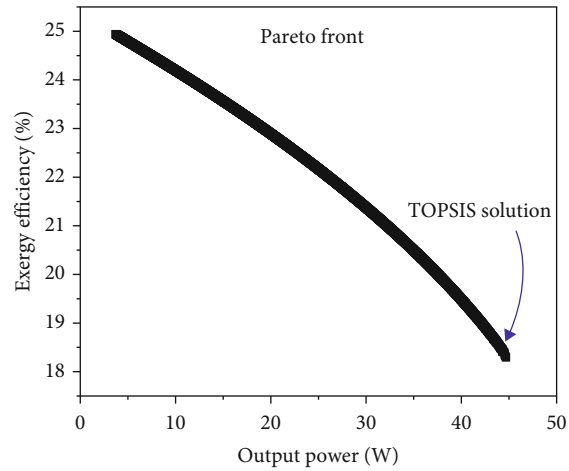


Figure 11: Optimal Pareto front.

Table 3: Optimized TE materials for the TEM.

Optimized TE materials	Symbol	p-type	n-type
Seebeck coefficient (μV/K)	3BS	231.67	-245.16
Electrical conductivity (S/m)	σ	50531	50411
Thermal conductivity (W/(m K))	k	0.5	0.5

3. Results and Discussion

3.1. CPV-TE Model Validation. The CPV-TE model developed in this study is validated against a 3-dimensional numerical study by Li et al. [18] as shown in Figure 6. In both studies, the output power of the PV operating in the CPV-TE system increases with an increase in the incident power; this will continue until maximum power is achieved. Similarly, as the incident power increases, the temperature gradient across the TEM increases, and this led to an increase in the output power of the TEM. Furthermore, some assumptions made by Li et al. [18] (blue line) were not reported in the paper. For example, the thermal resistances between the PV and TEG, TEG, and heat sink were not reported. In this case, we must compute the thermal

Table 4: Optimal settings of each regression algorithm acquired for the system's power output on the validation dataset.

Regression algorithm	Optimal configuration	RMSE [W]	MAE [W]	Training time [seconds]
LR	Interactions linear	0.70783	0.53914	**2.0319**
RT	Minimum leaf size is 4	0.096749	0.083006	13.523
SVM	Kernel function is quadratic Kernel scale is 0.001713 Box constraint is 0.0034653 Epsilon is 0.020246 Standardization is yes	0.06849	0.031594	335.15
GPR	Basis function is constant Kernel function is nonisotropic Matern 5/2 Kernel scale is 101.9278 Sigma value is 0.00015396 Standardization is no	**0.002334**	**0.0007575**	4329.8
Ensemble of RTs	Ensemble method: bag Minimum leaf size: 3 Number of learners: 10 Learning rate is 0.7737 Number of predictors to sample: 8	0.037564	0.028594	33.168
NN	Number of fully connected layers is 3 [number of neurons is 19, 18, 57] Activation is sigmoid Regularization strength (lambda) is 0.0004636 Standardization is yes	0.12684	0.02902	248.53

resistance using our model, and that may be one of the reasons why the output power of our model is consistently higher than that of the reference. Similarly, we assumed a convection heat transfer coefficient of 9.5 W/(m^2 K) which corresponds to the wind speed of 1 m/s (using Equation (12)). This value may be different from what Li et al. [18] considered because it was not explicitly stated. Since the PV is sensitive to both the thermal resistance and the convection coefficient, its output values may differ if any of the parameters is different from that of the reference. Overall, the two models presented have an average disparity of 3%; in other words, they have good agreement. Hence, the model presented in this study can be used to accurately predict the performance of a direct-coupled CPV-TE system.

3.2. NSGA-II Optimization Results. The population size determines the size of the population at each generation. When the optimization variables are less than or equal to 7, a typical population size of 200 is sufficient to enable the NSGA-II to converge to a set of Pareto optimal solutions. The NSGA-II can search more points and obtain better solutions if the number of generations, as well as the population size, is increased. However, this increases the computational time. In this study, a single run of the NSGA-II is completed in 3 h 50 min due to the population size of 10000 and a maximum number of generations of 9000. This ensures a large search space and provides a dataset for machine learning. Figures 7(a) and 7(b) show the normal distribution of electrical conductivities of *p*- and *n*-type TE materials, respectively. The nondominated solutions have high densities of electrical conductivity in the range 500531–50532 (S/m) for the p-type and 50412–50413 (S/m) for the n-type materials. Since more than 90% of the optimal solutions are within this range, one can suggest that the optimal electrical conductivities are found within this range. Several runs of the algorithm revealed the same points indicating global optimum.

The negatively skewed and positively skewed distributions of Seebeck coefficients for p- and n-type TE materials are given in Figures 8(a) and 8(b), respectively. More than 95% of the points of the *p*-type Seebeck coefficient lie between 232 μV/K and 233 μV/K, indicating that the optimal solutions are within that range. On the other hand, the optimal solutions of the *n*-type Seebeck coefficient are found within the range –245.25 μV/K and –246 μV/K. The other points outside the given ranges demonstrate the ability of the NSGA-II to expand the search space. The algorithm would move to the new search space had it been the fitness score of the new search space dominates the search space with the crowded population.

Similarly, symmetrical distributions of the thermal conductivities of the *p*- and *n*-type TE materials are given in

Table 5: Optimal settings of each regression algorithm acquired for the system's exergy on the validation dataset.

Regression algorithm	Optimal configuration	RMSE [%]	MAE [%]	Time [seconds]
LR	Interactions linear	0.0089714	0.0070405	**1.9813**
RT	Minimum leaf size is 4	0.015066	0.012725	12.982
SVM	Kernel function is linear Kernel scale is 0.01335 Box constraint is 0.078003 Epsilon is 0.0021042 Standardization is no	0.015403	0.010762	642.24
GPR	Basis function is zero Kernel function is isotropic Exponential Kernel scale is 3.5542 Sigma value is 0.033124 Standardization is no	**0.00096808**	**0.00037508**	1139.2
Ensemble of RTs	Ensemble method: bag Minimum leaf size: 1 Number of learners: 51 Learning rate is 0.3351 Number of predictors to sample: 8	0.0047669	0.0030295	46.845
NN	Number of fully connected layers is 2 [number of neurons is 1, 21] Activation is none Regularization strength (lambda) is 0.0031254 Standardization is yes	0.011626	0.0089548	380.65

Figures 9(a) and 9(b), respectively. In both figures, the optimal solutions are found to be around 0.5 Wm^{-1} K^{-1}. It can be observed that the algorithm did not explore a large search space in a bid to find the optimal thermal conductivity. This confirms that the optimal solutions are at 0.5 Wm^{-1} K^{-1}. One important thing to observe is that the algorithm converges with high Seebeck coefficients, low electrical conductivities, and low thermal conductivities. This shows that under the given constraints, feasible solutions are found with this material.

Figures 10(a) and 10(b) show the optimum load resistance of a unicouple and the concentration ratio, respectively. After several runs, the high density of the optimal solutions is found at around 1 mΩ (Figure 10(a)); this indicates an optimal point, that is, the suitable load resistance that will enable high output performance. In Figure 10(b), the concentration ratio shows a positively skewed distribution with the nonuniform distribution. Since both objective functions depend on the concentration ratio, one can see that the algorithm explores a large search space in a bid to obtain the optimal solutions.

Figure 11 shows an optimal Pareto front obtained from the NSGA-II algorithm. Although the two objective functions are benefit functions, increasing one function reduces the other. In line with that, a multicriteria decision-making method such as TOPSIS, VIKOR, and LINMAP is needed to select an optimal solution among the compromised solutions. Herein, the TOPSIS decision-making method is used, and the optimal solution reveals 44.6 W and 18.3% as the optimal output power and exergy efficiency, respectively. At this point, carbon saving per day is found to be 0.17 g/day. These values are obtained at the concentration ratio of 164 and optimum load resistance of a unicouple as 1×10^{-3} Ω. Equally, the corresponding TE materials are termed optimized TE materials as given in Table 3.

From the TE material point of view, the performance of a TEM can be dictated by a dimensionless figure of merit of the material $zT = TS^2\sigma/k$. In the optimization, some nonlinear constraints are added such that the zT of each n- and p-type material lies between 1 and 2, at the absolute temperature of 300 K. These values are chosen because typical zT values of commercial TE materials fall within that range [77]. An interesting conclusion can be drawn when the transport properties are examined, due to the nonlinear constraints which limit the dimensionless figure of merit of the TE materials to be ≤2. The results of Figures 7–10 and Table 3 show that, at a fixed geometry and with an upper limit boundary of zT, the performance of a CPV-TE system can be optimal if TE materials with high Seebeck coefficients and low values of thermal and electrical conductivities are used; this is in comparison to when TE materials with either high values of Seebeck coefficients, electrical and thermal conductivities, or high electrical conductivities, and low values of Seebeck coefficients and thermal conductivities are used. The results also reveal that the n- and p-type TE materials should have similar transport properties to avoid

Table 6: Optimal settings of each regression algorithm acquired for the system's CO_2 saving on the validation dataset.

Regression algorithm	Optimal configuration	RMSE [g/day]	MAE [g/day]	Time [seconds]
LR	Interactions linear	0.0026985	0.0020439	**2.8143**
RT	Minimum leaf size is 2	0.00036491	0.00031423	17.46
SVM	Kernel function is linear Kernel scale is 1 Box constraint is 0.0012404 Epsilon is 6.1303*e*-05 Standardization is yes	0.0037698	0.0023712	380.53
GPR	Basis function is zero Kernel function is isotropic Exponential Kernel scale is 3.5542 Sigma value is 0.033124 Standardization is no	**1.0004*e*-05**	**7.7158*e*-06**	2113.9
Ensemble of RTs	Ensemble method: bag Minimum leaf size: 1 Number of learners: 474 Learning rate is 1 Number of predictors to sample: 8	9.6717*e*-05	7.2527*e*-05	58.647
NN	Number of fully connected layers is 2 [number of neurons is 15, 2] Activation is sigmoid Regularization strength (lambda) is 0.00029705 Standardization is yes	0.0041681	0.0032614	149.64

Table 7: Performance metrics obtained by the built and optimized regression models on the test dataset for the three outputs.

Regression algorithm	RMSE			MAE		
	Power	Exergy ($\times 10^{-4}$)	CO_2 ($\times 10^{-5}$)	Power	Exergy ($\times 10^{-4}$)	CO_2 ($\times 10^{-5}$)
LR	0.7664	1.2510	1.1066	0.5732	0.0073	0.0022
RT	0.0092	2.3350	0.013246	0.0820	0.0128	3.11×10^{-4}
SVM	7.95×10^{-4}	7.8892	3.7447	0.0198	0.0233	0.0029
GPR	$\mathbf{0.08 \times 10^{-4}}$	**0.017142**	$\mathbf{2.64 \times 10^{-5}}$	$\mathbf{7.82 \times 10^{-4}}$	$\mathbf{4.18 \times 10^{-4}}$	$\mathbf{0.085 \times 10^{-4}}$
Ensemble of RTs	0.0018	0.26282	9.84×10^{-4}	0.0318	0.0032	0.64032
NN	0.0052	1.9907	2.8328	0.0431	0.0096	0.0038

a mismatch of current and voltage that would be generated by the TE legs. All these conclusions are drawn from a single run of the NSGA-II; this confirms its superiority over a time-consuming parametric study. This analysis would be useful in deciding the TE materials that should be used for a CPV-TE system.

3.3. Machine Learning Prediction Results. The whole considered regression algorithms have been built (i.e., by using the training dataset), fully optimized (i.e., by using the validation dataset), and evaluated (i.e., by using the test dataset) for the estimation task of the proposed hybrid system's power output and its corresponding exergy, as well as the CO_2 saving. Specifically, a dedicated regression model has been independently developed for each output of interest. Each model has been fully optimized in terms of its configurations and hyperparameters by resorting to the BO optimizer and the Holdout validation procedure with a fraction of 25% [65].

It is worth mentioning that the standardization of the dataset (i.e., mean = 0 and standard deviation = 1 for each feature) used to develop the SVM, GP, and NN regression algorithms has also been considered an optimization parameter seeking better estimation performance.

Tables 4–6 report the optimal performance metrics' values obtained by the optimized regression algorithms on the validation dataset for each target/output (i.e., system's power output, exergy, and CO_2 saving). These tables also report the optimal settings of the regression algorithms

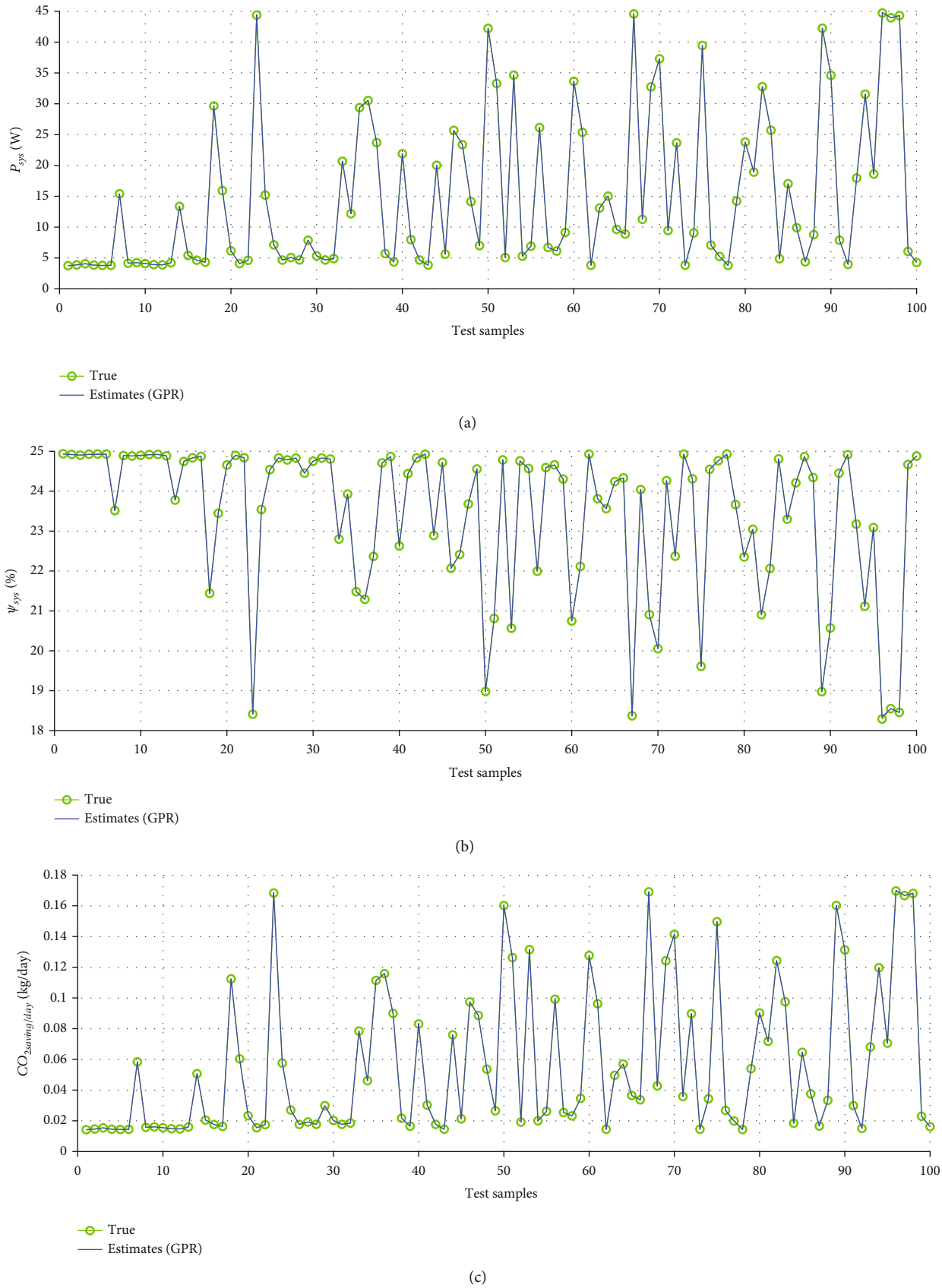


FIGURE 12: GPR's estimations of the test dataset for the (a) power output, (b) exergy, and (c) CO_2 saving hybrid system's performance parameters.

acquired by the BO optimizer. For example, the optimal settings of the GPR model are constant, nonisotropic exponential, 0.15663, and 0.00069583 for the basic function, kernel function, kernel scale, and sigma value, respectively.

Looking at Tables 4–6, one can observe the following:

(1) The GPR achieves accurate prediction results (lowest RMSE and MAE values) for the three hybrid system performance parameters amongst the other regression algorithms

(2) In contrast, the SVM exhibits poor training performance (highest RMSE and MAE values) for the three hybrid system's performance parameters amongst the other regression algorithms

(3) The GPR requires significant computational efforts (highest training time values) for the three hybrid system's performance parameters amongst the other regression algorithms

(4) In contrast, the LR model requires negligible computational efforts (lowest training time values), as expected, for the three hybrid system performance parameters amongst the other regression algorithms

Once the optimum regression models are obtained, they will be evaluated and compared using the test dataset (Table 7). Still, the performance of the GPR is superior to the other models on the "unseen" test dataset.

For completeness, Figure 12 shows the estimations provided by the best regression model (i.e., GPR) on the test dataset for the three hybrid system's performance parameters (i.e., power output, exergy, and CO_2 saving) and their true/actual values. Looking at Figure 12, one can clearly notice the capability of the GPR in accurately estimating the corresponding hybrid system's performance parameters with negligible mismatch between the predicted and true/actual values.

4. Conclusions

The aim of this work was to find the right combinations of thermoelectric material properties that will maximize the CPV-TE's 3E performance since no previous research has been conducted in this area. A multiobjective optimization and decision-making of a CPV-TE system via the NSGA-II and TOPSIS algorithms, respectively, were implemented to achieve this. Subsequently, diverse machine learning algorithms, namely, the linear regression (LR) models, regression trees (RTs), support vector machines (SVMs), Gaussian process regression (GPR) models, ensembles of RTs, and regressive neural network (NN), were built using the dataset generated from the optimization study for model deployment in industrial applications. The optimized parameters involved the concentration ratio, *n*- and *p*- Seebeck coefficient, electrical conductivity, thermal conductivity, and external load resistance. Meanwhile, the target outputs were the system's power output, exergy, and carbon dioxide savings, representing the exergo-environmental performance of the hybrid system. Based on the analysis, the following conclusions can be drawn:

(i) A high-output performance of a CPV-TE system can be achieved when TE materials with very low thermal conductivity are used

(ii) If any two of the transport properties of the TE materials are to be optimized, Seebeck coefficient and thermal conductivity should be chosen

(iii) To ensure an even performance across the TE legs, the transport properties of the *n*- and *p*-type TE materials should have similar values

(iv) The optimum TOPSIS solution generated a maximum output power of 44.6 W, maximum exergy efficiency of 18.3%, and minimum CO_2 saving of 0.17 g/day

(v) Various regression (linear and nonlinear) techniques were optimally built by a dataset generated by the optimization algorithm to predict the performance of the CPV-TE system

(vi) Compared to the other investigated regression algorithms, the GPR was superior in predicting the performance of the CPV-TE system with the lowest root mean square error and mean absolute error of 2.64×10^{-10} and 0.085×10^{-10}, respectively

(vii) However, the computational efforts required by the GPR was relatively expensive compared to those spent by, for instance, the LR, as expected, while compromising the prediction accuracy of the CPV-TE system

Nomenclature

A: Area, m^2
C: Concentration ratio
CO_2: Carbon dioxide
E: Daily energy generation, kWh
G: Global solar irradiance, Wm^{-2}
h: Heat transfer coefficient, W/m^2K
k: Thermal conductivity, $Wm^{-1}K^{-1}$
L: Length of the TE leg, mm
N: Number of data points
n: Number of thermocouples
P: Power output, W
R: Thermal resistance, K/W
R^2: Coefficient of determination
S: Seebeck coefficient, μV/K
T: Temperature, K
x: Coefficient of carbon dioxide savings, kg/kWh
ZT: Dimensionless figure of merit.

Greek Symbols

α: Absorptivity
β: Efficiency temperature coefficient, K^{-1}
γ: Stefan-Boltzmann constant, W/m^2K^4

ε: Emissivity
σ: Electrical conductivity, S/m
η: Efficiency, %
τ: Transmissivity
ψ: Exergy efficiency, %.

Abbreviations

CPV: Concentrated photovoltaic
GPR: Gaussian process regression
LR: Linear regression
ML: Machine learning
MAE: Mean absolute error
MOGA: Multiobjective genetic algorithm
NSGA: Nondominated sorting genetic algorithm
NN: Neural network
PV: Photovoltaic
RMSE: Root mean square error
RT: Regression trees
SPVM: Support vector machines
TE: Thermoelectric
TEM: Thermoelectric module
TOPSIS: Technique for order performance by similarity to ideal solution.

Subscripts

a: Ambient
c: Cold
c-hs: Cold junction and heat sink
conv: Convection
g: Glass
h: Hot
L: TE leg, electrical load
n: n-type TE material
p: p-type TE material
pv: Photovoltaic
pv-te: Photovoltaic-thermoelectric
r: Reference
sys: System
s: Sun
w: Water.

Data Availability

The MATLAB codes used to build the optimized ML algorithms on the test dataset for the three outputs of the proposed CPV-TE system are available at the GitHub repository https://github.com/chikap421/Forecasting-solar-panel-performance-using-machine-learning-algorithms. These codes can be used for validation and replication purposes. Further, a sample of the codes extracted from the Regression Learner App (MATLAB®) that could be used to recreate the built models is available on the GitHub repository for replication purposes. Finally, the data used in this work is available upon reasonable request from the corresponding author.

Conflicts of Interest

The author(s) declare(s) that they have no conflicts of interest.

Acknowledgments

The authors appreciate the following funding agencies: the Deanship of Scientific Research at Najran University for funding this work under the research collaboration funding program grant code (NU/RC/SERC/11/9), the Manson Benedict (1932) Fellowship at Massachusetts Institute of Technology with the cost object number: 3292100, the funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444), the Deanship of Scientific Research at King Faisal University for funding this work under the research collaboration funding program grant number 2783, the financial sponsorship provided by the Deanship of Scientific Research at King Faisal University, and the Office of Research and Sponsored Programs (ORSP) at Abu Dhabi University for funding this paper through a research fund.

References

[1] C. G. Ozoegwu, "The solar energy assessment methods for Nigeria: the current status, the future directions and a neural time series method," *Renewable and Sustainable Energy Reviews*, vol. 92, pp. 146–159, 2018.

[2] "World's most polluted city by air is in ... Nigeria, CNN," 2022, October 2022, https://www.cnn.com/2016/05/31/africa/nigeria-cities-pollution.

[3] "Nigeria suffers widespread blackouts after electricity grid fails, Reuters," 2022, October 2022, https://www.reuters.com/world/africa/nigeria-suffers-widespread-blackouts-after-electricity-grid-fails-2022-09-26/.

[4] "Nigeria's electricity grid collapses for the second time in a month, Reuters," 2022, October 2022, https://www.reuters.com/world/africa/nigerias-electricity-grid-collapses-second-time-month-2022-04-09/.

[5] "Nigerian businesses turn to solar sources amid high diesel costs | Energy News | Al Jazeera," 2022, October 2022, https://www.aljazeera.com/economy/2022/7/12/nigerian-businesses-turn-to-solar-as-diesel-costs-bite.

[6] P. Mulyungi, *Ashama Solar Power Plant, West Africa's Largest, Coming up in Nigeria*, Construction Review Online, Nairobi, 2021.

[7] F. K. A. Nyarko, G. Takyi, E. H. Amalu, and M. S. Adaramola, "Generating temperature cycle profile from in-situ climatic condition for accurate prediction of thermo-mechanical degradation of c-Si photovoltaic module," *Engineering Science and Technology, an International Journal*, vol. 22, no. 2, pp. 502–514, 2019.

[8] M. T. Zarmai, N. N. Ekere, C. F. Oduoza, and E. H. Amalu, "Evaluation of thermo-mechanical damage and fatigue life of solar cell solder interconnections," *Robotics and Computer-Integrated Manufacturing*, vol. 47, pp. 37–43, 2017.

[9] D. Sato and N. Yamada, "Review of photovoltaic module cooling methods and performance evaluation of the radiative cooling method," *Renewable and Sustainable Energy Reviews*, vol. 104, pp. 151–166, 2019.

[10] A. E. Kabeel, M. Abdelgaied, R. Sathyamurthy, and A. Kabeel, "A comprehensive review of technologies used to improve the performance of PV systems in a view of cooling mediums, reflectors design, spectrum splitting, and economic analysis," *Environmental Science and Pollution Research*, vol. 28, no. 7, pp. 7955–7980, 2021.

[11] K. K. Dixit, I. Yadav, G. K. Gupta, and S. K. Maurya, "A review on cooling techniques used for photovoltaic panels," in *2020 International Conference on Power Electronics and IoT Applications in Renewable Energy and Its Control, PARC*, pp. 360–364, Mathura, India, 2020.

[12] M. Sharaf, M. S. Yousef, and A. S. Huzayyin, "Review of cooling techniques used to enhance the efficiency of photovoltaic power systems," *Environmental Science and Pollution Research*, vol. 29, pp. 26131–26159, 2022.

[13] E. Yin, Q. Li, and Y. Xuan, "One-day performance evaluation of photovoltaic-thermoelectric hybrid system," *Energy*, vol. 143, pp. 337–346, 2018.

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

[15] A. Gupta, S. Agrawal, and Y. Pal, "Performance analysis of the photovoltaic-thermoelectric system combined with serpentine type water collector," *International Journal of Energy Research*, vol. 45, no. 13, pp. 19147–19164, 2021.

[16] M. A. I. Khan, M. I. Khan, A. H. Kazim et al., "An experimental and comparative performance evaluation of a hybrid photovoltaic-thermoelectric system," *Frontiers in Energy Research*, vol. 9, p. 9, 2021.

[17] S. Mahmoudinezhad, S. A. Atouei, P. A. Cotfas, D. T. Cotfas, L. A. Rosendahl, and A. Rezania, "Experimental and numerical study on the transient behavior of multi-junction solar cell-thermoelectric generator hybrid system," *Energy Conversion and Management*, vol. 184, pp. 448–455, 2019.

[18] Y. Li, Z. Zhang, H. Zhang, Z. Xiao, L. I. Luming, and P. Jiang, "Numerical and experimental performance evaluation of a laser-concentrated photovoltaic-thermoelectric generator hybrid system," *Optics Express*, vol. 30, no. 11, pp. 19465–19478, 2022.

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

[20] M. Ge, Y. Zhao, Y. Li, W. He, L. Xie, and Y. Zhao, "Structural optimization of thermoelectric modules in a concentration photovoltaic-thermoelectric hybrid system," *Energy*, vol. 244, article 123202, 2022.

[21] G. Li, X. Chen, Y. Jin, and J. Ji, "Optimizing on thermoelectric elements footprint of the photovoltaic-thermoelectric for maximum power generation," *Energy Procedia*, vol. 142, pp. 730–735, 2017.

[22] 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, article 114065, 2021.

[23] B. Lorenzi, P. Mariani, A. Reale, A. Di Carlo, G. Chen, and D. Narducci, "Practical development of efficient thermoelectric-photovoltaic hybrid systems based on wide-gap solar cells," *Applied Energy*, vol. 300, article 117343, 2021.

[24] S. Shittu, G. Li, X. Tang, X. Zhao, X. Ma, and A. Badiei, "Analysis of thermoelectric geometry in a concentrated photovoltaic-thermoelectric under varying weather conditions," *Energy*, vol. 202, article 117742, 13 pages, 2020.

[25] P. Wang, B. Wang, K. Wang, R. Gao, and L. Xi, "An analytical model for performance prediction and optimization of thermoelectric generators with varied leg cross-sections," *International Journal of Heat and Mass Transfer*, vol. 174, article 121292, 2021.

[26] G. Han, Y. Sun, Y. Feng, G. Lin, and N. Lu, "Machine learning regression guided thermoelectric materials discovery – a review," *ES Materials & Manufacturing*, vol. 14, pp. 20–35, 2021.

[27] H. Sun, Y. Ge, W. Liu, and Z. Liu, "Geometric optimization of two-stage thermoelectric generator using genetic algorithms and thermodynamic analysis," *Energy*, vol. 171, pp. 37–48, 2019.

[28] G. Contento, B. Lorenzi, A. Rizzo, and D. Narducci, "Simultaneous materials and layout optimization of non-imaging optically concentrated solar thermoelectric generators," *Energy*, vol. 194, article 116867, 2020.

[29] A. Z. Sahin, K. G. Ismaila, B. S. Yilbas, and A. Al-Sharafi, "A review on the performance of photovoltaic/thermoelectric hybrid generators," *International Journal of Energy Research*, vol. 44, no. 5, pp. 3365–3394, 2020.

[30] K. G. Ismaila, A. Z. Sahin, B. S. Yilbas, and A. Al-Sharafi, "Thermo-economic optimization of a hybrid photovoltaic and thermoelectric power generator using overall performance index," *Journal of Thermal Analysis and Calorimetry*, vol. 144, no. 5, pp. 1815–1829, 2021.

[31] W. Chen and Y. Chiou, "Geometry design for maximizing output power of segmented skutterudite thermoelectric generator by evolutionary computation," *Applied Energy*, vol. 274, article 115296, 2020.

[32] X. Li, C. Xie, S. Quan, Y. Shi, and Z. Tang, "Optimization of thermoelectric modules' number and distribution pattern in an automotive exhaust thermoelectric generator," *IEEE Access*, vol. 7, pp. 72143–72157, 2019.

[33] R. Arora, S. C. Kaushik, and R. Arora, "Thermodynamic modeling and multi-objective optimization of two stage thermoelectric generator in electrically series and parallel configuration," *Applied Thermal Engineering*, vol. 103, pp. 1312–1323, 2016.

[34] A. M. A. Youssef, Z. J. Zhai, and R. M. Reffat, "Genetic algorithm based optimization for photovoltaics integrated building envelope," *Energy and Buildings*, vol. 127, pp. 627–636, 2016.

[35] R. Wongsathan and A. Nuangnit, "Optimal hybrid neuro-fuzzy based controller using MOGA for photovoltaic (PV) battery charging system," *International Journal of Control, Automation and Systems*, vol. 16, no. 6, pp. 3036–3046, 2018.

[36] M. M. M. El-Arini, A. M. Othman, and A. Fathy, "A new optimization approach for maximizing the photovoltaic panel power based on genetic algorithm and Lagrange multiplier algorithm," *International Journal of Photoenergy*, vol. 2013, Article ID 481468, 12 pages, 2013.

[37] L. Zhang, C. Wang, Y. Chen, and L. Zhang, "Multi-objective optimization method for the shape of large-space buildings dominated by solar energy gain in the early design stage," *Frontiers in Energy Research*, vol. 9, p. 9, 2021.

[38] A. Akbar, G. Najafi, S. Gorjian, A. Kasaeian, and M. Mazlan, "Performance enhancement of a hybrid photovoltaic-

thermal-thermoelectric (PVT-TE) module using nanofluid-based cooling: indoor experimental tests and multi- objective optimization," *Sustainable Energy Technologies and Assessments*, vol. 46, article 101276, 2021.

[39] 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, article 118943, 2022.

[40] M. Świechowski, "Deep Learning and Artificial General Intelligence: Still a Long Way to Go," 2022, http://arxiv.org/abs/2203.14963v2.

[41] Z. He, M. Yang, L. Wang, E. Bao, and H. Zhang, "Concentrated photovoltaic thermoelectric hybrid system: an experimental and machine learning Study," *Science*, vol. 15, no. 47–56, 2021.

[42] 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.

[43] 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, article 101297, 2021.

[44] T. Y. Kim, "Prediction of system-level energy harvesting characteristics of a thermoelectric generator operating in a diesel engine using artificial neural networks," *Energies*, vol. 14, no. 9, p. 2426, 2021.

[45] Y. Zhu, D. W. Newbrook, P. Dai, C. H. K. de Groot, and R. Huang, "Artificial neural network enabled accurate geometrical design and optimisation of thermoelectric generator," *Applied Energy*, vol. 305, article 117800, 2022.

[46] W. Demeke, Y. Kim, J. Jung, J. Chung, B. Ryu, and S. Ryu, "Neural network-assisted optimization of segmented thermoelectric power generators using active learning based on a genetic optimization algorithm," *Energy Reports*, vol. 8, pp. 6633–6644, 2022.

[47] R. Kishore, R. 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.

[48] 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.

[49] 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, article 2100011, 2021.

[50] D. Mazzeo, M. S. Herdem, N. Matera et al., "Artificial intelligence application for the performance prediction of a clean energy community," *Energy*, vol. 232, article 120999, 2021.

[51] A. Jamali, P. Ahmadi, and M. N. Mohd Jaafar, "Optimization of a novel carbon dioxide cogeneration system using artificial neural network and multi-objective genetic algorithm," *Applied Thermal Engineering*, vol. 64, no. 1-2, pp. 293–306, 2014.

[52] Z. Pang, F. Niu, and Z. O'Neill, "Solar radiation prediction using recurrent neural network and artificial neural network: a case study with comparisons," *Renewable Energy*, vol. 156, pp. 279–289, 2020.

[53] A. Selamat, J. C. Lin, M. A. Eds, and R. Goebel, *Advances and Trends in Artificial Intelligence. Artificial Intelligence Practices. Vol. 12798*, Springer International Publishing, Cham, 2021.

[54] C.-L. Hwang, Y.-J. Lai, and T.-Y. Liu, "A new approach for multiple objective decision making," *Computers & Operations Research*, vol. 20, no. 8, pp. 889–899, 1993.

[55] A. Yusuf and S. Ballikaya, "Thermal resistance analysis of trapezoidal concentrated photovoltaic - thermoelectric systems," *Energy Conversion and Management*, vol. 250, article 114908, 2021.

[56] A. Duran Sahin, I. Dincer, and M. A. Rosen, "Thermodynamic analysis of solar photovoltaic cell systems," *Solar Energy Materials and Solar Cells*, vol. 91, no. 2-3, pp. 153–159, 2007.

[57] D. Li, Y. Xuan, Q. Li, and H. Hong, "Exergy and energy analysis of photovoltaic-thermoelectric hybrid systems," *Energy*, vol. 126, pp. 343–351, 2017.

[58] IEA, "Global Energy & CO2 Status Report 2019 – Analysis," 2019, https://www.iea.org/reports/global-energy-co2-status-report-2019.

[59] B. Hamawandi, H. Mansouri, S. Ballikaya et al., "A comparative study on the thermoelectric properties of bismuth chalcogenide alloys synthesized through mechanochemical alloying and microwave-assisted solution synthesis routes," *Frontiers in Materials*, vol. 7, pp. 1–13, 2020.

[60] B. Hamawandi, S. Ballikaya, H. Batili et al., "Facile solution synthesis, processing and characterization of n-and p-type binary and ternary Bi-Sb tellurides," *Applied Sciences*, vol. 10, no. 3, p. 1178, 2020.

[61] E. Yin, Q. Li, and Y. Xuan, "Optimal design method for concentrating photovoltaic-thermoelectric hybrid system," *Applied Energy*, vol. 226, pp. 320–329, 2018.

[62] S. Shittu, G. Li, X. Zhao, and X. Ma, "Series of detail comparison and optimization of thermoelectric element geometry considering the PV effect," *Renewable Energy*, vol. 130, pp. 930–942, 2019.

[63] W. Dong, Z. Zhou, L. Zhang et al., "Effects of Y, GdCu, and Al addition on the thermoelectric behavior of CoCrFeNi high entropy alloys," *Metals*, vol. 8, no. 10, p. 781, 2018.

[64] M. K. Bin, P. Kumar, R. A. Singh, and J. Singh, "Odyssey of thermoelectric materials: foundation of the complex structure," *Communications*, vol. 2, no. 6, article 062001, 2018.

[65] "Regression Learner App - MATLAB & Simulink," 2022, July 2022, https://www.mathworks.com/help/stats/regression-learner-app.html.

[66] A. S. Elamary and I. B. M. Taha, "Determining the shear capacity of steel beams with corrugated webs by using optimised regression learner techniques," *Materials*, vol. 14, no. 9, p. 2364, 2021.

[67] P. J. Huber and E. M. Ronchetti, "Regression," in *Robust Statistics*, pp. 149–198, John Wiley & Sons, 2009.

[68] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, vol. 19, First. Taylor & Francis, 1984.

[69] V. Kecman, T.-M. Huang, and M. Vogt, "Iterative single data algorithm for training kernel machines from huge data sets: theory and performance," in *Support Vector Machines: Theory and Applications*, L. Wang, Ed., pp. 255–274, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[70] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, Massachusetts, 2005.

[71] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, R. Tibshirani, and J. Friedman, "Ensemble learning," in *The Elements of Statistical Learning*, T. Hastie, R. Tibshirani, and J. Friedman, Eds., pp. 605–624, Springer New York, New York, NY, 2009.

[72] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[73] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[74] P. J. Braspenning, F. Thuijsman, and A. Weijters, *Artificial Neural Networks: An Introduction to ANN Theory and Practice*, vol. 931, Springer-Verlag, Berlin Heidelberg, 1995.

[75] G. Huang, H. G. Bin, S. Song, and K. You, "Trends in extreme learning machines: a review," *Neural Networks*, vol. 61, pp. 32–48, 2015.

[76] A. H. A. Al-Waeli, K. Sopian, J. H. Yousif, H. A. Kazem, J. Boland, and M. T. Chaichan, "Artificial neural network modeling and analysis of photovoltaic/thermal system based on the experimental study," *Energy Conversion and Management*, vol. 186, pp. 368–379, 2019.

[77] 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.