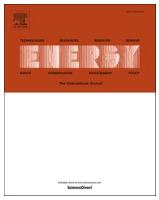
Energy 320 (2025) 135212

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy



Multi-criteria optimization of nanofluid-based solar collector for enhanced performance: An explainable machine learning-driven approach

Anjana Sankar [a,b], Kritesh Kumar Gupta [b,*], Vishal Bhalla [c], Daya Shankar Pandey [d,**]

[a] Department of Physics, Amrita School of Physical Sciences, Amrita Vishwa Vidyapeetham, Coimbatore, India
[b] Amrita School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore, India
[c] School of Energy and Environment, NICMAR University, Pune, India
[d] School of Mechanical Engineering, University of Leeds, Leeds, UK

ARTICLE INFO

Handling Editor: Dr A Mellit

Keywords:
Nanofluid-based solar collectors
Monte Carlo simulation
Gaussian process regression
Design optimization
Multi-criteria attainment
Explainable machine learning

ABSTRACT

This study presents a novel hybrid framework that leverages machine learning to enhance the performance of nanofluid-based solar collectors (NBSCs). The framework is designed to identify the optimal control variables required to meet multiple performance criteria (such as simultaneously maximizing outlet temperature, thermal efficiency, and optical efficiency). This study introduces an end-to-end multi-criteria optimization framework that combines numerical simulations with a Gaussian process regression (GPR) and genetic algorithm (GA) for designing optimized NBSCs. In this approach, a minimal number of random samples are selected using Monte-Carlo sampling to perform numerical simulations. The control variables of the system are varied within practical ranges, and key performance metrics such as outlet temperature [T_o (°C)], thermal efficiency (η_t), and optical efficiency (η_o) are recorded. The input and output data are utilized to develop a computationally efficient GPR model. The generalization capability of the developed explainable machine learning (xML) models allowed for various data-intensive analyses, including sensitivity analysis, uncertainty quantification, interactive influence of control variables, and multi-objective optimization. The proposed computational framework helped explore previously unknown territory, leading to the identification of optimal settings for simultaneously maximizing all the responses. The optimal parameters led to a simultaneous improvement in the responses, with a 23.44 °C rise in outlet temperature, a 37.48 % increase in thermal efficiency, and a 28.62 % boost in optical efficiency, compared to the base dataset. The developed framework is rigorously tested to ensure its robust generalization and its applicability to calibrate other physical systems. The results of this study offer valuable insights for designing optimal NBSCs with improved operational performance.

Nomenclature

A	Aperture area (m^2)
A_λ	Spectral absorption coefficient (m^{-1})
$A_{\lambda,water}$	Absorption coefficient of base fluid (water)
C_0	Speed of light in vacuum (3×10^8 m/s)
C_p	Specific heat (J/kg.K)
D	Diameter (nm)
E_λ	Spectral extinction coefficient
G_T	Total incident radiation (W/m^2)
H	Collector depth(m)
H	Planck's constant (6.6256×10^{-34} J s)
h_{conv}	Heat transfer coefficient for convection (W/m^2)
$I_{b\lambda}$	Spectral black body radiation (W/m^2 strμm)

(*continued on next column*)

(*continued*)

$I_{o\lambda}$	Spectral incident intensity (W/m^2 strμm)
K	Thermal conductivity (W/m^2)
k_B	Boltzmann constant (1.38×10^{-23} J/K)
L	Collector length (m)
M	Complex refractive index ($n+\iota\kappa$)
$\dot{m}$	Mass flow rate (kg/s)
$\dot{Q}$	Radiative heat flux
$Q_{e\lambda}$	Extinction efficiency
R^2	Correlation coefficient
S_{max}	Maximum value of the parameter
S_{min}	Minimum value of the parameter
S_r	Random sample generated within a parametric Range
S_λ	Spectral scattering coefficient

(*continued on next page*)

* Corresponding author.
** Corresponding author.
E-mail addresses: g_kriteshkumar@cb.amrita.edu (K.K. Gupta), d.pandey@leeds.ac.uk (D.S. Pandey).

https://doi.org/10.1016/j.energy.2025.135212
Received 9 September 2024; Received in revised form 6 February 2025; Accepted 21 February 2025
Available online 23 February 2025

(*continued*)

T	Temperature (°C)
$\overline{T}_{final}$	Average final temperature (°C)
T_s	Solar surface temperature (5526.85 °C)
V_f	Volume Fraction (%)
$\bar{y}$	Mean value of the response from the initial dataset
y_i	Individual true response corresponding to the ith sample
$\hat{y}_i$	Individual predicted response corresponding to the ith sample
% error	percentage relative error
Greek symbols	
Λ	Absorption coefficient
γ	Particle size ($\pi D/\lambda$)
H	Efficiency (%)
κ	Index of absorption
Λ	Wavelength
μ_x	Mean response corresponding to variations in x
P	Density (kg/m^3)
σ_x	Standard deviation of the response that arises from changes in x
Φ	Flux (W/m^2)
Ω	Solid angle (steradian)
Abbreviations	
2D	2-Dimensional
ANN	Artificial Neural Network
BIPVT	Building-Integrated Photovoltaic-Thermal
COV_x	The coefficient of variation related to the input feature 'x'(where x is an individual input)
DASC	Direct Absorption Solar Collector
FDM	Finite Difference Method
GA	Genetic Algorithm
GP	Gaussian Process
GPR	Gaussian Process Regressor
LINMAP	Linear Programming Technique for Multidimensional Analysis of Preference
MCCV	Monte-Carlo cross validation
MCS	Monte-Carlo sampling
ML	Machine Learning
ML-MOGA	Machine Learning-Multi Objective Genetic Algorithm
NBSC	Nanofluid-based Solar Collector
Pdf	Probability Density Function
PSO	Particle Swarm Optimization
RCV	Relative Coefficient of Variation
RTE	Radiative Transfer Equation
SHAP	Shapley Additive Explanation
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
UQ	Uncertainty Quantification
xML	Explainable Machine Learning
SUBSCRIPTS	
A	Ambient
Conv	Convection
F	Final
Min	Minimum
Max	Maximum
O	Optical
T	Thermal

1. Introduction

Energy is an essential need for modern society, as the quality of life heavily depends on it. Currently, the primary sources of energy (coal, oil, and natural gas) are depleting rapidly and causing long-term environmental damage. According to the IEA report for 2021, global consumption of coal, oil products, and natural gas was 38 million TJ, 160 million TJ, and 71 million TJ, respectively. This usage results in significant carbon dioxide emissions: 15,000 $MtCO_2$ from coal, 11,000 $MtCO_2$ from oil, and 7500 $MtCO_2$ from natural gas [1]. These emissions contribute to global warming and environmental changes, including desertification and increased frequency of severe weather events such as hurricanes, floods, and fires [2]. The unsustainable reliance on fossil fuels accelerates the depletion of fossil reserve and resulting in detrimental effects on the ecosystem. To address these issues, it becomes necessary to exploit renewable energy sources, such as solar energy, wind energy, geothermal energy, ocean thermal energy, and hydroelectricity, alongside existing fossil fuels. Renewable energy will not only help society meet its energy needs but also improve overall quality of life. Motivated by this, the present study focuses on harnessing solar energy, a fundamental source of energy. Other renewable sources like wind, biomass, and hydropower are ultimately driven by solar energy including fossil fuels, derived from solar energy captured millions of years ago and stored in the Earth's crust. Thus, effectively harnessing solar energy is essential for environmental sustainability.

1.1. Nanofluid based solar collectors (NBSC)

Most solar thermal technologies utilize an absorbing surface that captures solar radiation, heats up, and transfers the heat to a heat transfer fluid [3]. In power generation, solar thermal collectors are employed to run the Rankine cycle, which requires heating a fluid to high temperatures. These absorbing surfaces are specifically designed to have high absorptivity and low emissivity within the solar spectrum, optimizing their efficiency in capturing and transforming solar energy [4]. Despite the advanced optical properties of these absorbing surfaces, they still face inherent thermal barriers [5], which can reduce their efficiency below expectations. To improve thermal efficiency, one approach is to eliminate the absorber surface and allow the heat transfer fluid to directly absorb the solar radiation. However, a major challenge is that common heat transfer fluids, such as water, ethylene glycol, propylene glycol, toluene, silicone oil, and paraffin oil, are transparent to the solar spectrum (visible light) and thus have low solar absorptivity [6,7]. To enhance the absorptive properties of heat transfer fluids, it was suggested to develop nanofluids by incorporating nanomaterials. These nanofluids improve the efficiency of solar collectors by enhancing their ability to absorb solar radiation. Various approaches for direct absorption have been explored, including gas-particle suspensions [8–10] and liquid-film suspensions [11]. Solar collectors that utilize these nanofluids are known as nanofluid-based solar collector (NBSC). The performance of NBSCs are significantly influenced by the type of nanoparticles used. A range of materials, including metals, metal oxides, semiconductors, core-shell structures, and graphite-based nanoparticles, have been investigated for their ability to enhance solar energy absorption [12]. Metallic nanoparticles and carbon-based nanoparticles, in particular, are notable for their strong absorption in the visible spectrum, which improves both thermal and optical efficiency. Among these, specific metallic nanoparticles such as gold, silver, and copper have gained significant attention due to their plasmonic properties, which offer enhanced absorption and scattering capabilities.

Extensive numerical and experimental studies have demonstrated that nanofluid-based collectors, particularly those using metallic nanoparticles, exhibit higher thermal efficiency compared to traditional surface absorption solar thermal collectors [13–16]. For example, Taylor et al. [17] investigated the optical properties of various nanoparticles, including *Al*, *Ag*, TiO_2, *Cu*, and *C*, suspended in water. They found that silver (Ag) nanoparticles had a peak extinction coefficient around 400 nm, while aluminum (*Al*) and copper (*Cu*) nanoparticles demonstrated peaks around 300 nm. In contrast, carbon (*C*) nanoparticles displayed a broad and flat absorption spectrum. In experimental studies, Chen et al. [18,19] compared the efficiency of silver (*Ag*), zinc oxide (*ZnO*), and titanium dioxide (TiO_2) nanoparticles and found that *Ag* nanoparticles provided significantly higher thermal efficiency than *ZnO* and TiO_2. Their research indicated that variations in volume fraction, collector height, total exposure time, and solar irradiation improved the performance, outlet temperature, and thermal efficiency of direct absorption solar collectors (DASCs). Similarly, Bhalla et al. [20] evaluated the performance of nanofluid-based asymmetric parabolic concentrators using *Al*, *Ag*, and *C* nanoparticles and concluded that Ag nanoparticles delivered the highest thermal efficiency among the tested materials. Filho et al. [21] conducted experiments with *Ag* nanoparticle-based solar collectors under ambient conditions, testing three different concentrations (1.62, 3.25, and 6.5 ppm). The authors observed that these concentrations resulted in higher bulk temperatures compared to water, with thermal energy storage enhancements of 52 %, 93 %, and 144 %,

respectively. Several other contemporary studies also reported increases in thermal efficiency with higher concentrations of *Ag* nanoparticles [22–24].

1.2. Machine learning based predictive frameworks

The literature highlights that silver (*Ag*)-laden nanofluids can significantly improve the performance of solar collectors. However, the effectiveness of a solar collector depends on more than just the material and volume fraction of nanoparticles; various other factors also play a crucial role. For example, Baro et al. [25] investigated how different variables, such as channel dimensions (length, width, and height), volume fraction, mass flow rate, diameter, and inlet temperature, affect the performance of solar collectors using aluminum nanofluid. Similarly, Qin et al. [26] studied the impact of factors like the absorption coefficient (α), mass flow rate per unit width ($\dot{m}$), collector depth (H), and collector length (L) to optimize solar collector performance. Understanding the comprehensive influence of these variables can be challenging, either through large-scale experiments or numerical simulations. To address these complexities, researchers have increasingly turned to computational intelligence methods, such as meta-heuristic algorithms, for optimization [26,27]. Some studies have applied machine learning (ML) frameworks to predict the performance of NBSC [42,43,48,49]. For instance, Shahsavar et al. [28] compared the accuracy of various ML models to develop an effective predictive framework for Building-Integrated Photovoltaic-Thermal (BIPVT) collectors. The authors utilized geometrical parameters (length, width, and height) as input and energy gain as the output parameter. While several studies have focused on constructing predictive models using ML frameworks [29–34], few have explored methods for deploying these models in data-intensive investigations. Li et al. [35] proposed an artificial neural network (ANN)-based framework for high-throughput screening to identify optimal solutions for enhancing solar water heater performance. Similarly, Liu et al. [36] developed an ANN-based high-throughput screening framework to optimize intrinsic input features (such as tube length, number of tubes, tube center distance, tank volume, collector area, angle between tubes and ground, and final temperature) to improve heat collection rates. The proposed high-throughput screening methods [35,36] used trial and error-based sorting algorithms to find the best solutions, and hence lacked a systematic approach for design optimization. In contrast, Salari et al. [37] integrated numerical simulations with the ANN framework to develop a prediction mechanism that includes more robust optimal sample selection methods, such as TOPSIS, LINMAP, and Shannon entropy. Metaheuristic algorithms like Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) are effective at exploring search spaces to find global optimal solutions. However, while metaheuristics excel at exploring large search spaces nevertheless, they can be computationally expensive and may struggle with complex relationships without extensive parameter tuning.

1.3. Performance challenges

Despite the predictive capabilities of ML-based algorithms, most studies in the domain of solar-physical systems have primarily focused on constructing and validating ML models, rather than fully exploiting the computational power these models offer [50–53]. A key gap in the literature is the lack of robust mechanisms to leverage ML models for more comprehensive analyses. Traditional soft computing methods such as Design of Experiments, Response Surface Methods, and Taguchi techniques have been predominantly used for process optimization, and parametric sensitivity analysis [54–56,58–61]. These methods often struggle with generalizability, limiting their ability to accurately capture the full range of parametric variations. To address this limitation, we propose a computationally efficient ML-based framework. This framework effectively applies ML models to various downstream tasks, including uncertainty quantification, sensitivity analysis, multi-objective optimization, and model explainability. Specifically, we introduce a novel multi-criteria optimization framework that combines numerical simulations (algorithmically selecting a minimal set of simulations), an efficient Gaussian Process Regression (GPR)-driven ML model, and Genetic Algorithms (GA) for designing NBSC systems with enhanced performance. By integrating machine learning with metaheuristic optimization, this framework benefits from the strengths of both approaches, enabling faster, more accurate, and adaptable solutions. The proposed solution makes a significant contribution by presenting a novel, explainable, machine learning-driven closed-loop optimization framework designed to enhance the operational performance of NBSC systems. This framework not only identifies optimal solutions but also substantially reduces the computational costs associated with large-scale numerical simulations. It's simple, flexible design allows it to be applied to optimize a wide variety of physical systems and processes.

2. Modelling and simulation

2.1. Monte Carlo simulation

This investigation utilizes Monte Carlo simulation (MCS) based randomly sampled dataset (consisting the input features, i.e., volume fraction (V_f), flux (φ), height (H), mass flow rate ($\dot{m}$), ambient temperature (T_a), and heat transfer coefficient for convection (h)), and the desired quantities of interest (outlet temperature (T_o ($^\circ C$)), thermal efficiency (η_t), and optical efficiency (η_o) obtained from numerical simulations performed for MCS driven random samples). The investigation starts by randomly generating 125 nos. of samples by utilizing the MCS technique, wherein the random data instance (random set of 6 control variables presented in Table 1) is generated by using the following formulation [38]:

$$S_r = S_{\min} + X \times (S_{\max} - S_{\min}) \tag{1}$$

where, S_r denotes one of the random samples generated in between the range of parametric variation ($S_{\max} - S_{\min}$), $S_{\min}$ is the minimum value of the parameter, $S_{\max}$ is the maximum value of the parameter, and X denotes the randomly generated value in between 0 and 1. This generates 125 unique values of X in between 0 and 1. The unique feature of such sample-space generation is that it captures the complete continuous domain of parametric variation, which helps in constructing an efficient predictive model by using relatively lower number of samples. The randomly sampled 125 numbers of input samples are fed to the numerical simulation mechanism (refer to subsection *2.2*) for obtaining the desired quantities of interest such as outlet temperature [T_o($^\circ$ C)], thermal efficiency (η_t), and optical efficiency (η_o). Finally, the MCS driven finite difference method (FDM) simulations prepare the dataset to be utilized for the formation of computationally efficient ML model.

Table 1
The control variables and their parametric range of variation [21,46,47].

Control variables	Minimum value	Maximum value	Range of variation
Volume fraction, V_f (%)	0	0.004	0.004
Flux, φ (W/m^2)	500	3000	2500
Depth of the channel, H (m)	0.001	0.015	0.014
mass flow rate, $\dot{m}$ (kg/sec)	100	600	500
Ambient temperature, T_a (^{o}C)	25	30	5
Convective heat transfer coefficient, h (W/m^2)	1	10	9

2.2. Numerical model of the nanofluid-based solar thermal collector (NBSC)

This subsection explores the underlying physics of NBSC operation and performance, focusing on the key governing equations of radiative heat transfer. Fig. 1 illustrates the schematic of the NBSC system.

The following assumptions are made for the numerical modelling [5,7,13]:

1. The bottom surface of the collector is considered adiabatic to avoid the thermal losses from the collector to the environment [3].
2. The nanoparticles are evenly dispersed in the base fluid, making the nanofluid a single-phase fluid.
3. There is no scattering of radiation between the nanoparticles.

The numerical model is based on a 2-D rectangular channel with constant length ($L = 1$ m) and depth H, where H ranges from 0.001 to 0.015 m. The fluid flows from left to right through the channel. For simplification, it is assumed that the fluid velocity is uniform across the channel's cross-section and that no boundary layer forms within the pipe. Additionally, solar irradiation is assumed to be incident perpendicularly on the top surface of the collector. To simulate the incident solar intensity, the black body radiation relationship is applied using equation (2):

$$I_{b\lambda} = \frac{2.h.{C_o}^2}{\lambda^5} \frac{1}{\exp\left(\frac{h.C_o}{\lambda.K_B.T_s} - 1\right)} \tag{2}$$

where $I_{b\lambda}$ is spectral black body radiation, h is Planck constant (6.6256×10^{-34} J s), C_o is the speed of light in vacuum (3×10^8 m/s), λ is the wavelength in (nm), K_B is the Boltzmann constant (1.38×10^{-23} J/K), T_s is the solar surface temperature (5800 K).

During the numerical simulations, the attenuation of incident solar radiation in the outer atmosphere is disregarded. It is worth to mention that 96.3 % of the total radiation falls within the range of 300–2500 nm, and only this range is considered in the analysis. When the irradiation incidents on the nanofluid, the intensity of the radiation attenuates. This attenuation is assumed to vary in one dimension (along the depth), and the rate of intensity change along this depth is calculated using the Radiative Transfer Equation (RTE) [39], and is given in eq. (3).

$$\frac{dI_\lambda(\Omega)}{dy} = \underbrace{-(S_\lambda + A_\lambda)\, I_{o,\lambda}(\Omega)}_{A} + \underbrace{A_\lambda I_{b\lambda}}_{B} + \underbrace{\frac{S_\lambda}{4\pi} \int_{4\pi} I(\Omega')\, X(\Omega' \rightarrow \Omega) d\Omega}_{C} \tag{3}$$

The left-hand side of equation describes the rate of change of intensity along the depth of the collector, which depends on the sum of terms A, B, and C. Here, S_λ represents the spectral scattering coefficient, A_λ is the spectral absorption coefficient, $I_{o,\lambda}$ is the spectral incident intensity, Ω denotes the solid angle, $I_{b,\lambda}$ is the spectral black body radiation, and X ($\Omega' \rightarrow \Omega$) is the phase function. In Equation (3), terms A, B, and C represent the extinction term, scattering term, and in-scattering term, respectively. To solve the RTE, following assumptions are made:

(a) The size parameter ($\gamma = \pi D/\lambda$), $\gamma \ll 1$ and the base fluid used is water, having moderate refractive index ($\gamma|m\text{-}1|$), due to which Rayleigh scattering has been considered. Where, m is the complex refractive index ($m = n + i\kappa$). n is the index of refraction and κ is the index of absorption. The values of the n and κ have been taken from Ref. [39].
(b) The temperature rise of the fluid in the collector will be less than 427 °C, thus the emission from the fluid to the ambient is negligible.

By considering these assumptions, the term B and C becomes zero and equation (3) reduces to equation 4

$$\frac{dI_{o,\lambda}}{dy} = -(E_\lambda) I_{o,\lambda} \tag{4}$$

where E_λ is the spectral extinction coefficient and can be evaluated as $E_\lambda = S_\lambda + A_\lambda$. In the case of the nanofluid, the intensity decay results from both the base fluid and the nanoparticles. This intensity decay is represented by equation (5).

$$\frac{dI_{o,\lambda}}{dy} = -\left(A_{\lambda,\text{water}} + E_{\lambda,\text{nanoparticle}}\right) I_{o,\lambda} \tag{5}$$

For the pure fluid, the absorption coefficient is calculated by equation 6

$$A_{\lambda,\text{water}} = \frac{4\pi\kappa}{\lambda} \tag{6}$$

The values of κ for water is considered from Ref. [39]. The extinction of the irradiation in the nanofluid is calculated by equation 7

$$E_\lambda = \frac{1.5 \times V_f \times Q_{e\lambda}(\alpha, m)}{D} \tag{7}$$

where V_f is the volume fraction and $Q_{e\lambda}$ is the extinction efficiency, D is the hydrodynamic diameter of the nanoparticle. The extinction efficiency depends on the size parameter and the complex refractive index.

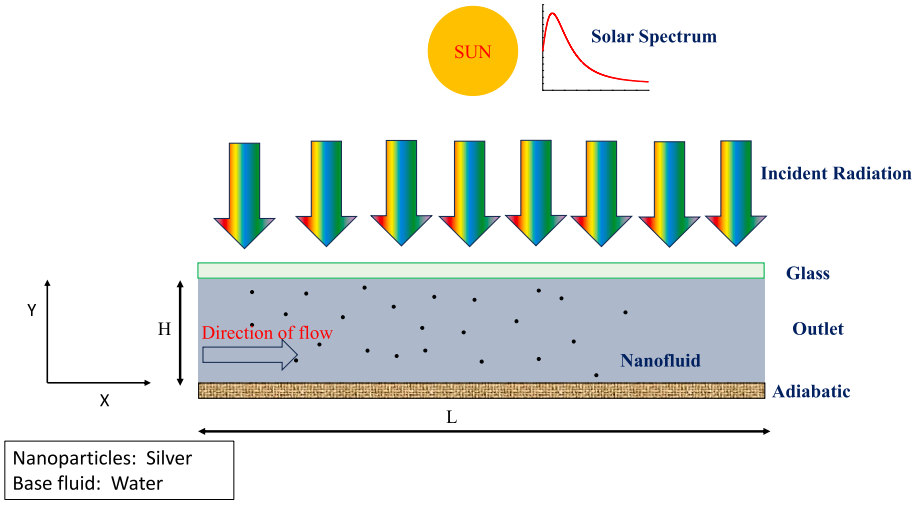


Fig. 1. Schematic of nanofluid-based solar thermal collector.

The extinction efficiency is calculated by equation 8

$$Q_{e,\lambda} = \overbrace{4\left(\frac{\pi D}{\lambda}\right)\mathrm{Im}\left\{\frac{m^2-1}{m^2+2}\left[1+\frac{(\pi D)^2}{15\lambda^2}\left(\frac{m^2-1}{m^2+2}\right)\left(\frac{m^4+27m^2+38}{2m^2+13}\right)\right]\right\}}^{(A)} + \underbrace{\frac{8}{3}\left(\frac{\pi D}{\lambda}\right)^4\left|\left(\frac{m^2-1}{m^2+2}\right)\right|^2}_{(B)} \tag{8}$$

where term A represents the absorption efficiency and term B represents the scattering efficiency. On substituting equation (8) in equation (7), the extinction coefficient can be calculated and given by equation 9

$$E_{e,\lambda} = \overbrace{6\left(\frac{\pi \times V_f}{\lambda}\right)\mathrm{Im}\left\{\frac{m^2-1}{m^2+2}\left[1+\frac{(\pi D)^2}{15\lambda^2}\left(\frac{m^2-1}{m^2+2}\right)\left(\frac{m^4+27m^2+38}{2m^2+13}\right)\right]\right\}}^{(C)} + \underbrace{\frac{8}{3}\left(\frac{\pi^4 D^3 \times V_f}{\lambda^4}\right)\left|\left(\frac{m^2-1}{m^2+2}\right)\right|^2}_{(D)} \tag{9}$$

where term C represents the absorption coefficient and term D represents the scattering coefficient. From equation (9), it can be observed that the extinction coefficient depends on the volume fraction, hydrodynamic diameter, and complex refractive index.

In the following stage, the temperature rise (due to absorption of the irradiation) of the nanofluid has been evaluated. For this evaluation, 1-D transient heat transfer equation has been used (refer to equation (10)) [40].

$$k\frac{\partial^2 T}{\partial x^2} - \frac{\partial Q}{\partial y} = \rho C_p u \frac{\partial T}{\partial x} \tag{10}$$

where k, ρ and C_p are the thermal conductivity, density, and specific heat of the nanofluid respectively, and Q is the radiative heat flux. The radiative heat flux is calculated by using equation (11).

$$Q = \int_{\lambda}\int_{\Omega} I_{\lambda}\, d\Omega\, d\lambda \tag{11}$$

The thermo-physical properties of the nanofluid (refer to Table 2) are calculated by using following set of equations.

$$\begin{aligned} (\rho)_{\text{nanofluid}} &= \rho_{\text{nanoparticle}}.V_f + (1-V_f).\rho_{\text{basefluid}} \\ (C_p)_{\text{nanofluid}} &= C_{p_{\text{nanoparticle}}}.V_f + (1-V_f).C_{p_{\text{basefluid}}} \\ \frac{k_{\text{nanofluid}}}{k_{\text{basefluid}}} &= \frac{k_{\text{nanoparticle}} + 2k_{\text{basefluid}} + 2\left(k_{\text{nanoparticle}} - k_{\text{basefluid}}\right)V_f}{k_{\text{nanoparticle}} + 2k_{\text{basefluid}} - \left(k_{\text{nanoparticle}} - k_{\text{basefluid}}\right)V_f} \end{aligned} \tag{12}$$

The optical efficiency (η_{ol}) is obtained by utilizing the fraction of solar irradiance intensity reaching the nanofluid (refer to equation 13) [12]:

$$\eta_{\text{ol}} = \frac{G_{\text{nanofluid}}}{G_{\text{aperture}}} \tag{13}$$

While, thermal efficiency (η_t) of the collector is calculated by using equation 14

$$\eta_t = \frac{\dot{m}C_{p_{nanofluid}}\left(\overline{T}_{final} - T_{initial}\right)}{A.G_T} \tag{14}$$

where $\dot{m}$ is the mass flow rate (kg/s), C_p is the specific heat of nanofluid, $\overline{T}_{\text{final}}$ is the average final temperature, A is the aperture area and G_T is the total incident radiation.

Table 2
Thermophysical properties of base fluid and nanoparticles [40].

Thermophysical property	Water (base fluid)	Silver (nanoparticle)
Density (kg/m^3)	1000	10,470
Thermal conductivity (W/m.K)	0.6	429
Specific heat (J/kg.K)	4180	235

2.3. Machine learning driven predictive framework

The dataset, consisting of 125 samples generated through Monte Carlo simulation (MCS) (refer for subsection 2.1), is used to develop the ML models. The prediction capability of different ML models including Decision Trees (DT), Linear Regression (LR), Single Layer Neural Network (SLNN), Support Vector Regression (SVR), and Gaussian Process Regression (GPR) is compared for selecting the best suited ML algorithm for modelling the FDM simulations derived NBSC dataset. The choice of the GPR model is based on its ability to effectively capture nonlinear interactions and uncertainties within the data. Unlike traditional regression methods that assume linear relationships between variables, GPR models are flexible and can accommodate complex interactions between input features and output predictions. Recent studies by Gupta et al. [37–39] have demonstrated that GPR models provide a high-fidelity predictive framework, requiring only a minimal number of initial samples, which can be applied to large-scale predictions. The detailed mathematical foundation of the ML models used in this study is presented in section SM1 of the supplementary material. This investigation developed and utilized three different models for individual output responses: outlet temperature (T_o (°C)), thermal efficiency (η_t), and optical efficiency (η_o). These models establish the relationships between the input features, i.e., volume fraction (V_f), flux (φ), height (H), mass flow rate ($\dot{m}$), ambient temperature (T_a), and heat transfer coefficient for convection (h) and predict their respective output responses. Initially, the dataset was randomly divided into a training set (90 %) and a test set (10 %). The training dataset was used to construct the machine learning models. The training and test performance measures (R^2 and *RMSE* values) are compared for obtaining the best suited for the NBSC dataset. The best model (GPR in this case) is further improved by employing Bayesian optimization for hyperparameter tuning. This process utilized the Bayesian framework to identify the optimal set of hyperparameters for the GPR models (refer to Table SM3 in the supplementary material), ensuring minimal mean square error in the predicted responses, while training [45]. The hyperparameter tuning process iteratively selected the combinations of different covariance (Kernel) functions and the prediction capability of the developed GPR models is assessed on the basis of correlation coefficient (R^2) and percentage relative error (%-error) in the prediction. The evaluation of R^2 and % error is carried out by using equation (15) [40]:

$$\begin{aligned} R^2 &= 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - \bar{y}\right)^2} \\ \%\ error &= \frac{\left(y_{true} - y_{predicted}\right)}{y_{true}} \times 100 \end{aligned} \tag{15}$$

where, y_i denotes the individual true response corresponding to the ith sample, $\hat{y}_i$ denotes the individual predicted response corresponding to the ith sample, $\bar{y}$ denotes the mean value of the response from the initial dataset. The evaluation of % error utilizes a difference in true response (y_{true}) and corresponding predicted response ($y_{predicted}$) for finding the error in prediction. Once satisfactory predictive accuracy was achieved during training, the models were validated using the out-of-fold test samples (16 samples, out-of-the-fold, separately generated by performing FDM simulations). It is important to note that the generalization capability of the developed models was ensured by performing Monte-Carlo cross validation using a separate (test) dataset that was not

included in the training process. To perform MCCV, randomly 8 samples are selected from the out-of-the-fold test samples (16 samples) in each iteration, and R^2 and *RMSE* is evaluated; this process is performed for 25 iterations. Such cross-validation mechanism establishes the generalization capability of the ML model by ensuring that developed model is capable in handling the variance in the unknown data.

2.4. Deployment of the machine learning model

With the sufficient confidence established in the predictive capabilities of the developed ML models, the models are further deployed for data-intensive investigations such as sensitivity analysis, uncertainty quantification, interactive parametric influence, and multi-objective optimization. In this regard, at first the large parametric space (with 10000 randomly generated samples) is generated using the MCS framework (refer to subsection 2.1). The generated (through MCS) large sample space is fed to developed ML models for obtaining the predictions for target variables (T_o, η_t, and η_o). For performing the sensitivity analysis, relative coefficient of variation (RCV) is evaluated as following [41,42]:

$$(RCV)_x = \frac{COV_x}{\sum_{x=1}^{6} COV_x} \qquad (16)$$

$$where, COV_x = \frac{\sigma_x}{\mu_x}$$

In this context, COV_x represents the coefficient of variation related to the input feature 'x' (where x is an individual input). Here, σ_x indicates the standard deviation of the response that arises from changes in x alone, while μ_x refers to the mean response corresponding to variations in x. For example, when the volume fraction (V_f) changes from 0 % to 0.004 %, with all other input features held constant at their mean values, σ_{Vf} and μ_{Vf} are determined for the output responses T_o, η_t, and η_o, which are then used to calculate COV_{Vf}. Similarly, the coefficients of variation COV_φ, COV_H, COV_m, COV_{Ta}, and COV_h are evaluated individually, allowing for the calculation of RCV_x.

The GPR models are used for multi-objective GA optimization to improve the performance of the NBSC system. For a comprehensive understanding of the GA-based multi-objective optimization, refer to section SM2 in the supplementary material. The optimization process is implemented using MATLAB with a population size of 100, a cross-over rate of 0.8, and an adaptive mutation function. This multi-criteria framework employs the ML models as fitness functions to predict solutions for selected generations from the search space. A cross-over rate of 0.8 ensures an 80 % probability of directional changes in the solutions, aiding in achieving optimal results.

Generally, the ML models act as black boxes for exploring the range of parameter variations needed for the desired properties. In this context, the models assess how variations in parameters affect the performance measures of the NBSC. Although the models' efficient prediction capabilities can be utilized directly, the exact mechanisms behind their functioning are not fully transparent. To address this inherent issue, the prediction capability of ML models developed in this study are analyzed using Shapley Additive exPlanations (SHAP) [41,43,44]. The SHAP analysis investigates the individual effects of input features on the prediction process, providing more interpretability to the ML results. The complete computational approach utilized to perform the present investigation is illustrated as the flow-diagram in Fig. 2.

3. Results and discussion

The desired quantities of interest such as outlet temperature [T_o (°C)], thermal efficiency (η_t), and optical efficiency (η_o) are evaluated by performing the numerical simulations (FDM) for the randomly sampled input space of 125 samples (refer to Fig. 3). The input space for

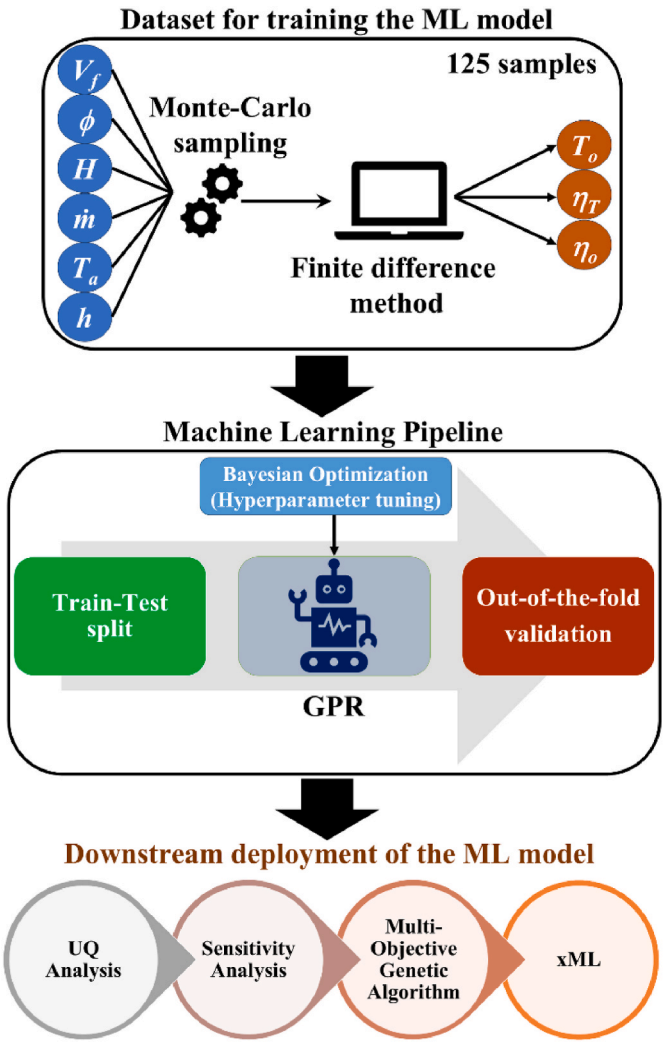


Fig. 2. Flow diagram of the presented computational framework.

numerical simulations is made up of randomly perturbed input features (volume fraction (V_f), flux (φ), height (H), mass flow rate ($\dot{m}$), ambient temperature (T_a), and heat transfer coefficient for convection (h)) that are within the specified parametric range of variation (see Table 1).

Prior to the data-driven modelling of the NBSC system is carried out, the results of the numerical simulation are validated with the experimental findings reported elsewhere [13]. The comparison of the collector's thermal efficiency (experimental *vs.* numerical simulation) demonstrates the close match (refer to Figure SM1 in the *supplementary file*). While performing the validation study, the same parameters were utilized to perform the numerical simulations.

To understand the one-on-one cross-correlation among input features and the responses (in the original FDM derived dataset), Pearson's correlation analysis is performed (refer to Fig. 4). It is evident from the Pearson's correlation that flux (φ) and mass flow rate ($\dot{m}$) have a moderately positive correlation with the T_o, whereas the depth (H) of the channel demonstrates a negative correlation with the T_o. When it comes to the efficiencies (thermal (η_t) and optical (η_o), the volume fraction (V_f) and the depth (H) of the channel exhibits moderately positive correlation. The statistical correlation among responses (T_o, η_t, and η_o) indicates, that there exists strong negative correlation between T_o and η_t, and T_o and η_o, whereas η_t and η_o shows mutually positive correlation. This preliminary understanding is obtained by the training data (without using the ML models) to assess the general trend and influence of the parametric variation. The complete dataset derived after

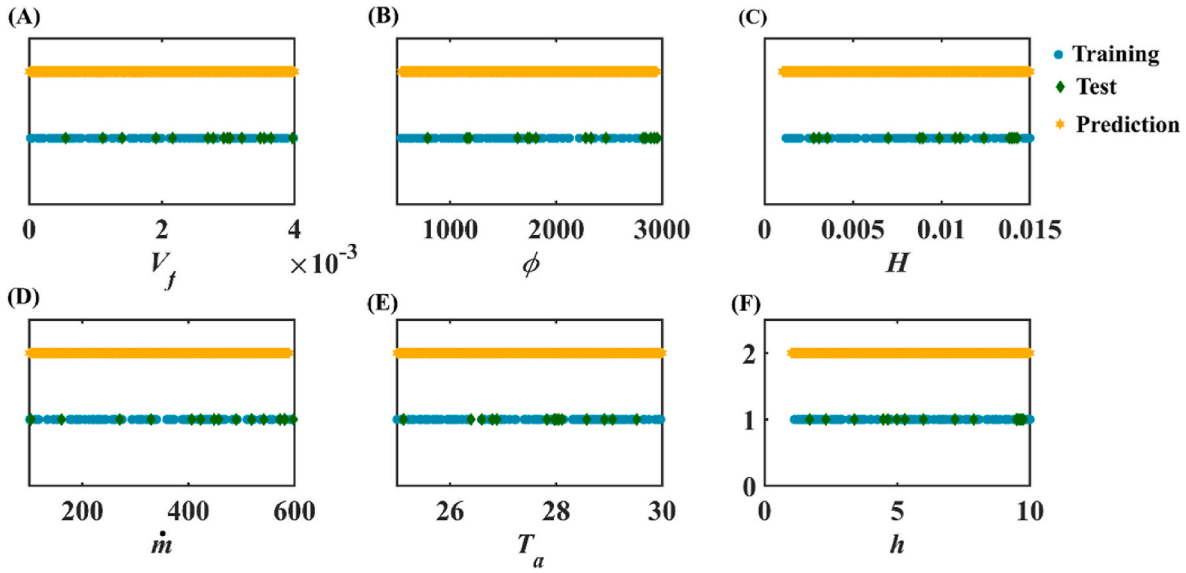


Fig. 3. Randomly (using Monte Carlo sampling) distributed input features considered in the present study for performing 125 numbers of numerical simulations (green points used for training and testing the ML models). The large-scale (10000) random samples (yellow points) are considered as virtual samples to get the predictions out of the ML models.

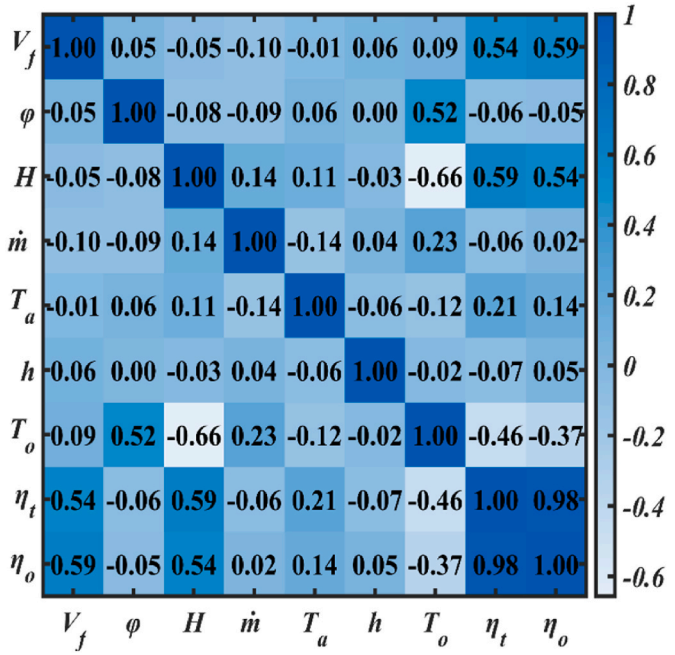


Fig. 4. Pearson's correlation analysis: Utilized to assess the individual parametric influence on the desired quantities of interest suggested by the initial (numerical simulations driven) dataset.

performing the numerical simulations (for 125 samples, 6 input features and 3 responses) is utilized to construct different ML models. The dataset utilized for training the models is provided in Table SM1 of the *Supplementary file*.

The ML models are trained by utilizing 90 % of the samples, while 10 % samples are utilized for testing the performance of the ML models. The prediction capability of different ML models including Decision Trees (DT), Linear Regression (LR), Single Layer Neural Network (SLNN), Support Vector Regression (SVR), and Gaussian process regression (GPR) is compared for selecting the best suited ML algorithm for modelling the FDM simulations derived NBSC dataset. The comparison of the evaluation metrics (R^2 and *RMSE*) is illustrated in Fig. 5, which indicates that among all the considered ML models the performance of GPR is consistent during training and testing regardless, of the target responses. Once the prediction capability of the GPR is ascertained, the performance of GPR is further improved by hyperparameter tuning, in this process, the sample-size to train the GPR models are varied in between 70 % and 100 %, by simultaneously varying the hyperparameters of the GPR model (presented in Table SM3 of the *supplementary file*). To understand the influence of individual hyperparameter on the predictive performance of developed GPR models, the most important hyperparameters (such as Kernel function, sigma values, and standardization status) [57] are individually varied. It is observed that squared exponential kernel function yielded the best R^2 and the lowest RMSE for predictions of T_o. However, for predicting efficiencies (η_t and η_o), the rational quadratic and Matérn 5/2 kernel functions provided the best performance, respectively (refer to Figure SM3 in the *supplementary file*). Apart from kernel function, the combination of sigma values and standardization status revealed to have a drastic influence on the performance of GPR models. The observations presented in Figure SM4 of the supplementary file revealed that performing standardization of the input features is essential to obtain best performance of GPR models. When the standardization status is considered false, the drastic reduction in the prediction performance of the GPR models is observed (refer to Figure SM4 in the *supplementary file*). The consideration of standardization of the input features makes the GPR models effective, irrespective of the sigma values considered.

The performance of these models (which are subjected to the variation in training samples size) is assessed by utilizing the out-of-the-fold 16 samples (refer to Table SM2 of *supplementary file*), which were separately derived from the FDM simulations (refer to Fig. 6). To ensure the proper generalization capability of the developed optimized GPR model, the variance is provided to the test dataset, by performing Monte-Carlo cross validation (MCCV) mechanism. To perform MCCV, randomly 8 samples are selected from the out-of-the-fold test samples (16 samples) in each iteration, and R^2 and *RMSE* is evaluated; this process is performed for 25 iterations. Such cross-validation mechanism establishes the generalization capability of the ML model by ensuring that developed model is capable in handling the variance in the unknown data. The observations of the MCCV are illustrated in Figure SM5 of the *supplementary file*, which indicates that the distribution of R^2 and *RMSE* values during MCCV remains conserved, *i.e.* for majority of the iterations

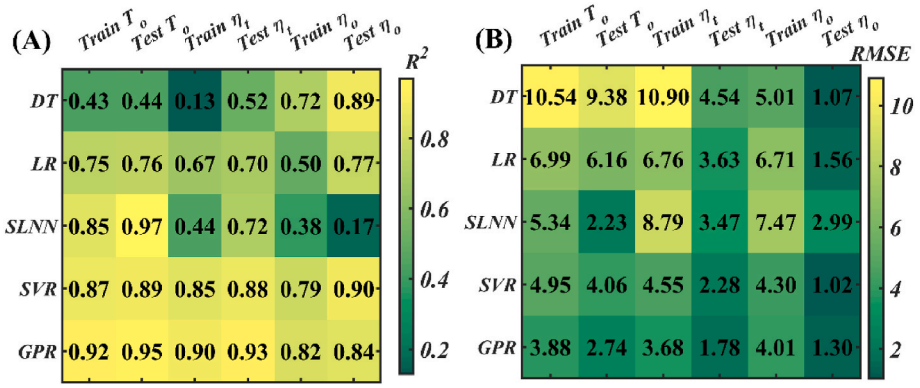


Fig. 5. Comparison of the performance of different ML models such as decision tree (DT), linear regression (LR), single layer neural network (SLNN), support vector regression (SVR), Gaussian process regression (GPR), **(A)** comparison of R^2 values, **(B)** comparison of *RMSE* values.

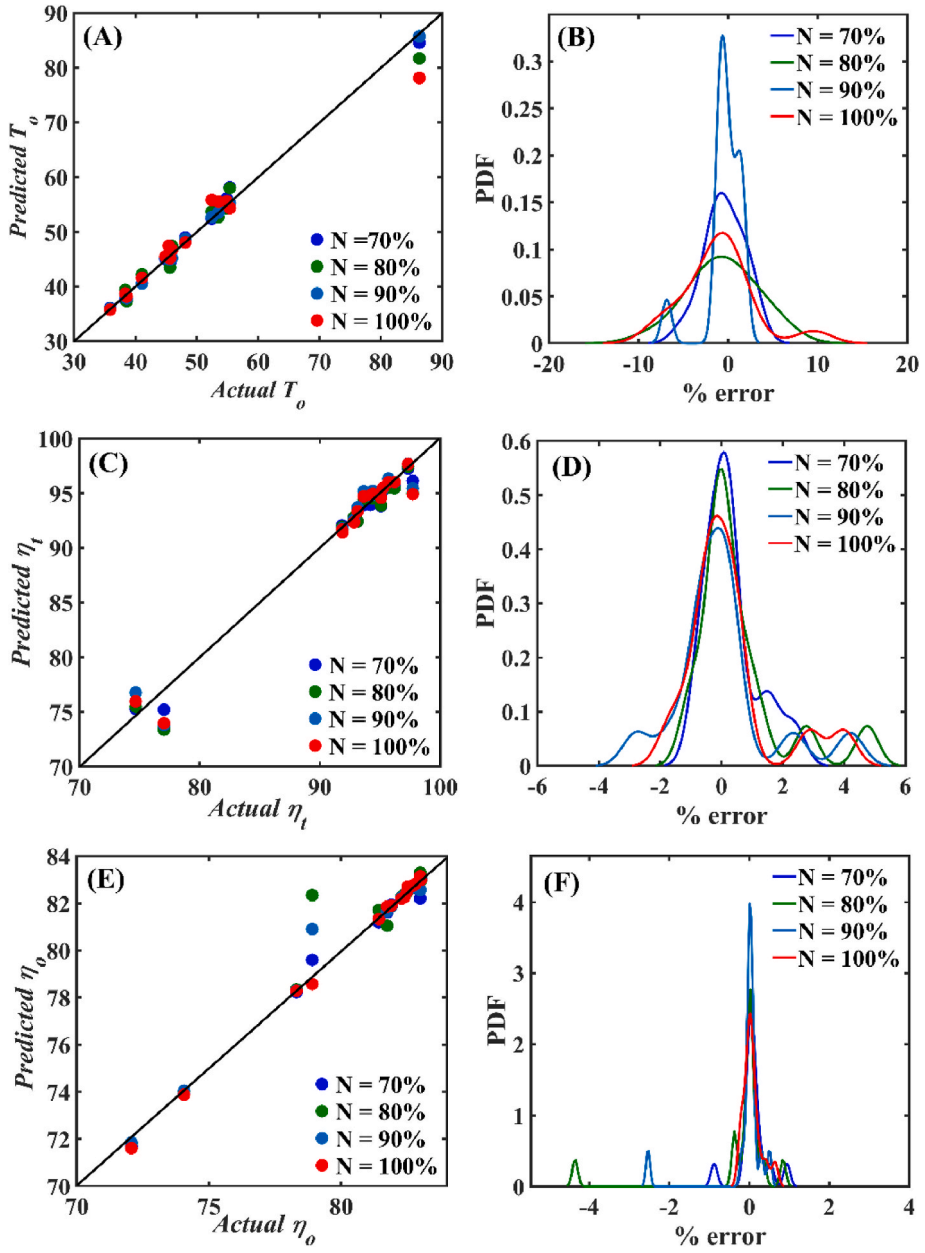


Fig. 6. Validation of developed GPR models: **(A, C, E)** Scatter plots between actual and predicted responses (T_o, η_t, and η_o) for training and test samples **(B, D, F)** probability distribution of sample wise percentage error in prediction while training and test the ML models for the responses (T_o, η_t, and η_o).

R^2 remains close to 1 and *RMSE* remains minimum (statistical description of the performance measures obtained from MCCV are provided in Table 3), regardless of the target response. This highlights the predictive accuracy and generalization capability of the developed GPR model.

The validation of the hyperparameter tuned GPR models is illustrated in Fig. 6 with the help of scatter plots (comparison between actual responses and GPR predicted responses) and sample-wise percentage error in the prediction with the help of probability density function (PDF) plots. It is evident from the scatter and error plots presented in Fig. 6, that regardless of the ML model (for T_o, η_t, and η_o) the GPR predicted observations are exceptionally close to the actual (FDM derived) responses, at the same time the high likelihood of the percentage error lies within ±5 %, it is to be noted that these observations are obtained for the out-of-the-fold validation samples, which were not involved during the training process of the model.

After establishing the generalization capability of the developed GPR models (GPR_T_o, GPR_η_t, and GPR_η_o), the models are deployed for large-scale data-intensive investigations including sensitivity and uncertainty quantification analysis, and are presented in Fig. 7. The bar plots presented in Fig. 7(A) highlights the statistical significance of the variation in an individual input feature with respect to the desired quantities of interest (T_o, η_t, and η_o), which reflects the RCV-based sensitivity of the responses over an individual input feature. The sensitivity analysis reveals that the variation in depth of the channel (H), flux (φ), and mass flow rate ($\dot{m}$) exhibits significant statistical influence over outlet temperature (T_o). Whereas thermal efficiency and optical efficiency are observed to be highly sensitive to the variation in depth of the channel (H) and volume fraction (V_f). With this understanding, the uncertainty analysis is performed by inducing the intentional uncertainty ($\varepsilon = \pm 2.5$ %, ±5 %, ±7.5 % and ±10 %) in the mean value of statistically important features (obtained through sensitivity analysis). For instance, the uncertainty is induced in the mean values of H (0.08), φ (1746), and $\dot{m}$ (346.4), using which the sample space of 10000 samples is constructed by similar random sampling approach (refer to Fig. 3). It is to be noted that rest of the input features are maintained at their corresponding mean values. The developed GPR model corresponding to outlet temperature (GPR_T_o) is utilized to make the predictions for such large-scale unknown samples. The predicted outlet temperature values of uncertainty induced (in H, φ, and $\dot{m}$) dataset are illustrated in terms of probability density function (PDF) plots in Fig. 7(B). It is evident from the 7(B) that with the increase in the magnitude of uncertainty, the spread of PDF plot increases which reveals the expected variation in the T_o corresponding to the uncertain H, φ, and $\dot{m}$. The same investigation is conducted for η_t, and η_o, wherein the intentional uncertainty is induced H and V_f, while considering mean value of the rest of the features. The observations associated with the uncertainty analysis conducted for η_t, and η_o are illustrated in Fig. 7(C) and (D), respectively. Similar to the uncertainty analysis of T_o, the increase in magnitude of uncertainty can result in a drastic variation in the desired thermal and optical efficiencies as well. It should be highlighted that the reported sensitivity and uncertainty analysis is carried out as part of the deployment of the developed GP models, highlighting the effective utilization of ML models in data-intensive studies that would otherwise go unexplored.

To understand the interactive influence of the simultaneous variations in the statistically important features (H, φ, and $\dot{m}$ for T_o, and H and V_f for η_t, and η_o (obtained from Fig. 8(A)), the simultaneous variation in the pair of features (two features at a time), while maintaining other features at their mean values is introduced to generate MCS driven 10000 samples, such unknown dataset is fed to the developed ML models and the prediction for desired output responses (T_o, η_t, and η_o) is made. The interactive influence of the simultaneous variations in the selected features is illustrated in Fig. 8, wherein Fig. 8(A–C) highlights the variation in outlet temperature subjected to simultaneous variation in H and φ, φ and $\dot{m}$, and $\dot{m}$ and H, respectively. The observations obtained from the original data based on the Pearson's correlation (presented in Fig. 4) are further fortified by the interactive influence of H, φ, and $\dot{m}$ on outlet temperature T_o (refer to Fig. 8(A–C)), wherein with the increase in H, the significant decline in T_o is evident. At the same time, the increase in φ and $\dot{m}$ demonstrated significant increase in outlet temperature. This behavior is directly correlated with flux (φ) and the input irradiation intensity. Concurrently, a decrease in channel depth results in a reduction of the nanofluid volume (assuming a constant heat transfer coefficient, h). With less nanofluid volume, the same amount of input power is concentrated in a smaller fluid volume, thereby increasing the outlet temperature. It is evident from Fig. 8(B) that simultaneous increase in φ and $\dot{m}$ leads to rapid increment in the outlet temperature. Further, increment in depth of the channel (H) at fixed flux (φ), and mass flow rate ($\dot{m}$) leads to drastic decline in the outlet temperature, which corroborates with the negative correlation of H over T_o illustrated in Fig. 4. The interpretation of this pattern explains that the fluid volume is directly correlated with the height of the channel. However, an increase in the mass flow rate results in a shorter exposure time of the fluid within the channel. Therefore, with a constant flux, increasing both the channel height and the mass flow rate leads to a decrease in the outlet temperature of the fluid. The influence of cross-interaction of variations in φ and $\dot{m}$ (refer to Fig. 8(B)) revealed that simultaneous increase in φ and $\dot{m}$ leads to significant increase in T_o. Since, H and V_f are observed to be having relatively higher statistical significance over thermal and optical efficiency (refer to Fig. 7(A)), the variation in η_t (refer to Fig. 8(D)), and η_o (refer to Fig. 8(E)) is recorded corresponding to the simultaneous variation in H and V_f. It is evident that, simultaneous increase in H and V_f leads to a rapid and drastic increment in thermal and optical efficiency. The reason for increase in thermal and optical efficiency is due to the volume fraction increase which prompts to increase the absorption capability of the nanofluid the optical efficiency. Furthermore, the increase in height and volume fraction leads to increase in outlet temperature of the fluid and the increase in outlet temperature is directly proportional to the thermal efficiency (eq. (14)). From eq. (14), it can be observed that the mass flow rate, specific heat of the nanofluid and the temperature difference between the outlet temperature and the inlet temperature are directly proportional to the thermal efficiency. Thus, the increase in mass flow rate led to higher outlet temperature, thus higher thermal efficiency. As the volume fraction of the nanoparticles increases the specific heat of the fluid decreases and the temperature difference depends on the height of the channel (volume of the fluid in the channel). Due to these reasons, the thermal efficiency of the DASC system is increasing.

So far, the constructed ML models are deployed to reveal the deep insights into the parametric perturbations and corresponding stochastic evaluation of the solar collector's thermal and optical performance. However, it is to be noted that the efficient design of silver nanoparticle based solar collector system should reflect appropriate trade-off between outlet temperature and efficiencies (since there exist a negative correlation between T_o and η_t, and T_o and η_o (Refer to Fig. 4)). Therefore, the developed models are utilized as the objective functions to perform the multi-objective GA-based optimization. In this regard, initially the combination of two responses (T_o and η_t, and T_o and η_o) are individually explored for obtaining the optimal input feature settings to ensure sufficiently maximized responses. Thereafter, all the three responses (T_o, η_t, and η_o) are considered together to obtain the optimal input features for ensuring the simultaneous maximization of all the considered responses. The Pareto solutions obtained from the multi-objective optimization are

Table 3
Statistical description of the models' performance measures obtained from MCCV for 25 iterations.

Model	R^2		RMSE	
	mean	Standard deviation	Mean	Standard deviation
GPR_T_o	0.9434	0.0681	2.02	1.222
GPR_η_t	0.9296	0.1019	1.03	0.4382
GPR_η_o	0.9862	0.0153	0.328	0.1291

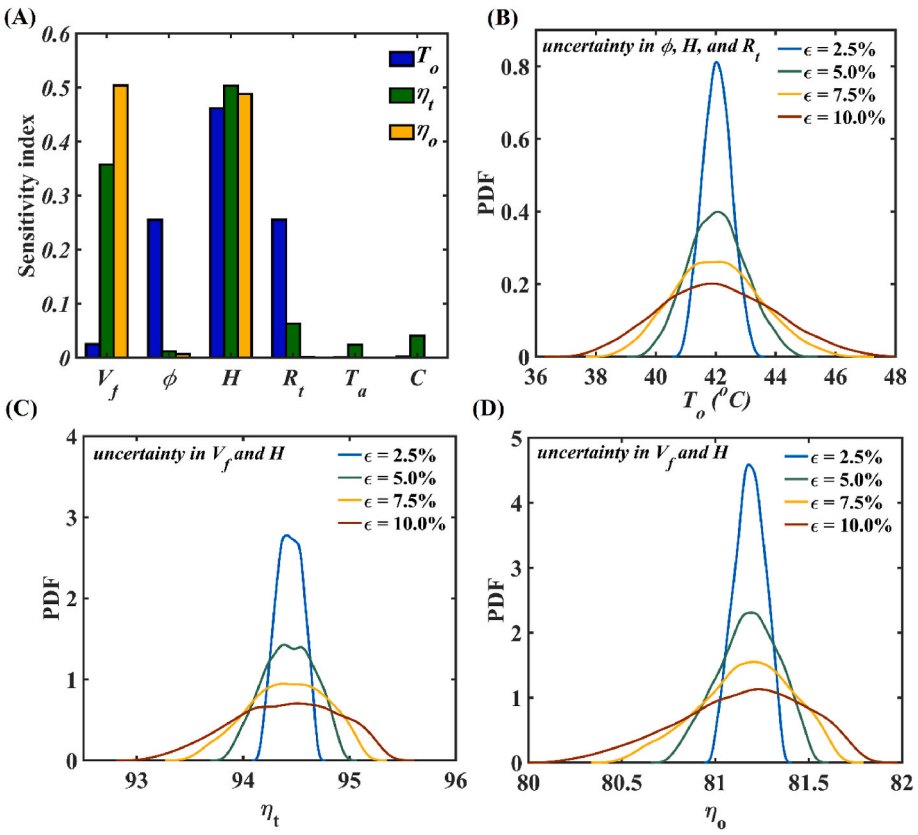


Fig. 7. Uncertainty analysis of the performance of NBSC. **(A)** ML driven sensitivity analysis for illustrating the statistically most significant parameters in terms of each target response (T_o, η_t, and η_o) **(B)** Intentional uncertainties (±2.5 %, ±5 %, ±7.5 %, ±10 %) introduced to mean values of φ, H, and $\dot{m}$ for their collective influence on T_o **(C)** Intentional uncertainties (±2.5 %, ±5 %, ±7.5 %, ±10 %) introduced to mean values of V_f, and H for their collective influence on η_t **(D)** Intentional uncertainties (±2.5 %, ±5 %, ±7.5 %, ±10 %) introduced to mean values of V_f, and H for their collective influence on η_o.

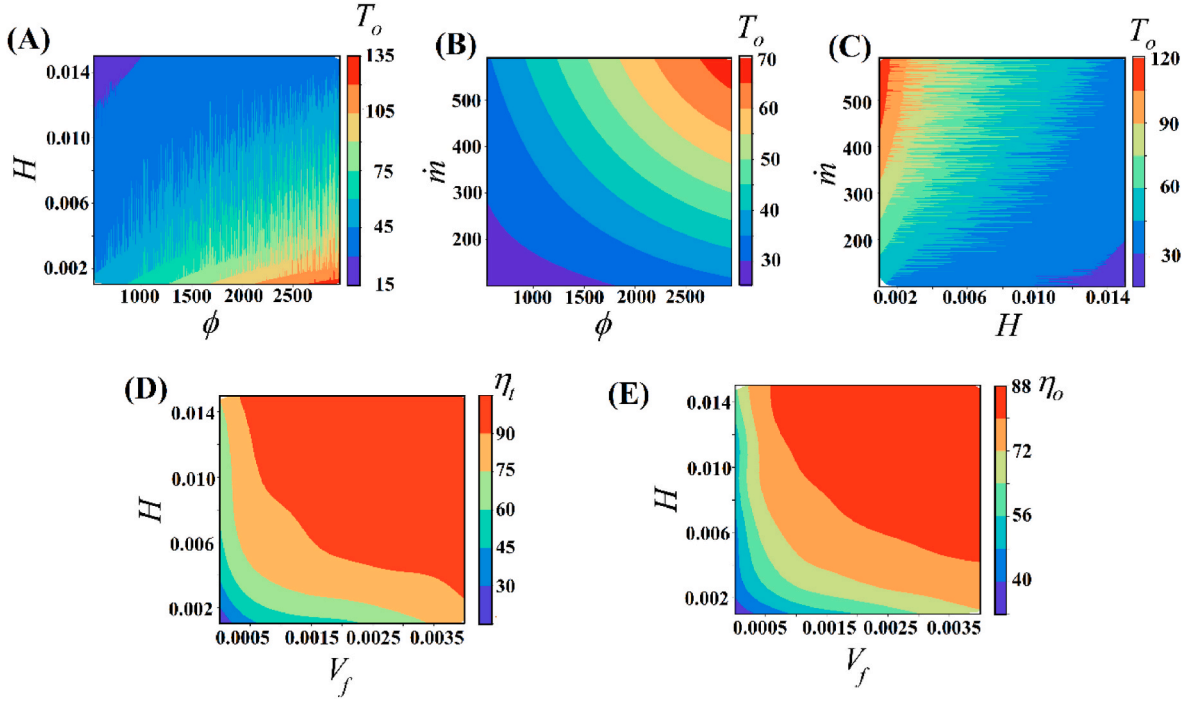


Fig. 8. Interactive influence of the most significant (suggested by sensitivity analysis) input variables on **(A–C)** outlet temperature (T_o), **(D)** thermal efficiency (η_t), **(E)** optical efficiency (η_o).

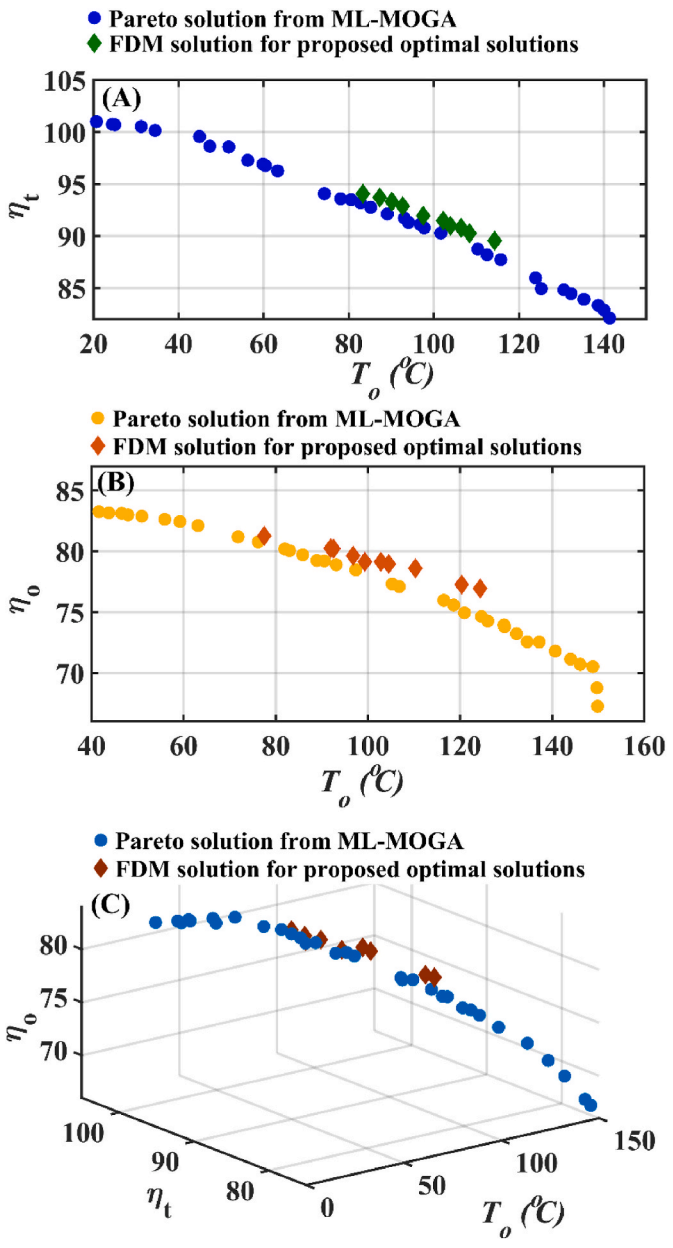


Fig. 9. Pareto analysis of machine learning driven multi-objective genetic algorithm (ML-MOGA) optimization framework. **(A)** Maximization of T_o, and η_t, suggested by ML-MOGA framework, proposed best solutions are validated by performing a set of numerical simulations (green points) **(B)** Maximization of T_o, and η_o, suggested by ML-MOGA framework, proposed best solutions are validated by performing a set of numerical simulations (orange points) **(C)** Simultaneous maximization of T_o, η_t and η_o suggested by ML-MOGA framework, proposed best solutions are validated by performing a set of numerical simulations (red points).

illustrated in Fig. 9, wherein Fig. 9(A) highlights the Pareto solution corresponding to the simultaneous maximization of T_o, and η_t. Similarly, Pareto solution corresponding to simultaneous maximization of T_o, and η_o is presented in Fig. 9(B). It is noted from Fig. 9(A) and (B) that all the solutions obtained from the ML-MOGA optimization does not qualify for the simultaneous maximization of the responses, but some of the points lying within the Pareto solution provides a suitable trade-off between T_o, and η_t (refer to Fig. 9(A)) and T_o, and η_o (refer to Fig. 9(B)).

These optimal input feature settings proposed by ML-MOGA are further validated by performing the numerical (FDM) simulations as depicted by a few points highlighted with different colour in Fig. 9. The similar understanding can be obtained from the Pareto solution presented in Fig. 9(C) corresponding to the simultaneous maximization of T_o, η_t and η_o. The optimal solutions along with the corresponding responses suggested in Fig. 9(A), (B) and 9(C) are presented in Table SM5 of the *supplementary file.* The Pareto solutions (in Fig. 9) proposed by the ML-MOGA framework represent the best solutions. However, to ensure the simultaneous maximization of T_o, η_t and η_o, the trade-offs between T_o, and η_t, as well as between T_o, and η_o, must be carefully considered. The optimal parameters identified were $V_f = 0.0037$ %, $\varphi = 2712.95$ W/m^2, $H = 0.003$ m, $\dot{m} = 542.79$ kg/s, $T_a = 27.62$ °C, and $h = 4.38$ W/m^2. Predictions from the ML-MOGA approach closely matched those from the numerical simulations (FDM), with slight deviations in temperature gain and efficiencies: ML-MOGA predicted $T_o = 117.27$ °C (FDM $T_o = 119.05$ °C), $\eta_t = 87.62$ % (FDM $\eta_t = 85.22$ %), and $\eta_o = 77.79$ % (FDM $\eta_o = 77.82$ %). It is important to note that the initial dataset used to construct the machine learning model, which was derived from the FDM simulations, recorded a maximum outlet temperature of 93.83 °C, along with a thermal efficiency of 50.14 % and an optical efficiency of 49.17 %. This highlights how the proposed ML-MOGA framework enabled the exploration of previously uncharted territory, facilitating the identification of optimal settings that simultaneously maximize all the key performance responses. Optimizing the design of a Nanofluid-Based Solar Collector (NBSC) requires careful consideration of the trade-offs between temperature, thermal efficiency, and optical efficiency, as represented by the presented Pareto solutions. This involves navigating a complex interplay between these factors, and the practical implications extend beyond the immediate performance metrics. The optical efficiency represents the fraction of incident radiation absorbed by the nanofluid and is defined by Equation (13). This equation indicates that optical efficiency is directly proportional to absorbed irradiation. A higher optical efficiency results in greater absorption of irradiation, leading to an increased temperature rise in the nanofluid. However, the temperature increase also intensifies re-radiation losses, which can reduce overall energy gain rather than enhancing thermal efficiency. The thermal efficiency as described in Equation (14), is directly proportional to the temperature difference ($T_{out} - T_{in}$). Although, when optical efficiency is high, the nanofluid temperature rises, subsequently improving thermal efficiency. Nevertheless, excessive temperature increase may necessitate the use of high temperature-resistant materials, potentially increasing material costs. Therefore, for practical implementation of NBSC system, several factors need to be considered, including the cost of nanofluids, the development of high-temperature-resistant designs, scalability, and nanoparticle stability. Currently, commercially available nanoparticles are expensive and often face stability challenges. Utilizing naturally occurring materials such as carbon soot or carbon dust presents a cost-effective and scalable alternative. Chemical treatment of these materials can further enhance their stability, making them more viable for solar energy applications. By leveraging these sustainable nanomaterials, solar energy harnessing can become more cost-effective and contribute to UN Sustainable Development Goal especially SDG 7 (Affordable and Clean Energy) and SDG13 (Climate Actions).

To properly understand the underlying mechanism for the predictions offered by the developed GPR models, the models are interpreted in terms of SHAP values. The agnostic approach of model's explanation, such as SHAP helps understanding the deep insights about the individual parametric influence on the prediction mechanism of the developed model. SHAP is a powerful tool for explaining complex machine learning models, providing both local and global interpretability, and can handle non-linear relationships. Such assessment on the interpretability of ML models is of great importance, to ensure a meaningful deployment of the ML models. The SHAP based summary plots for model explanation are illustrated in Fig. 10.

The SHAP analysis concerning the GP model for outlet temperature (refer to Fig. 10(A)) revealed that the model given prime importance to the variation in depth of channel (H), followed by flux (φ), and mass

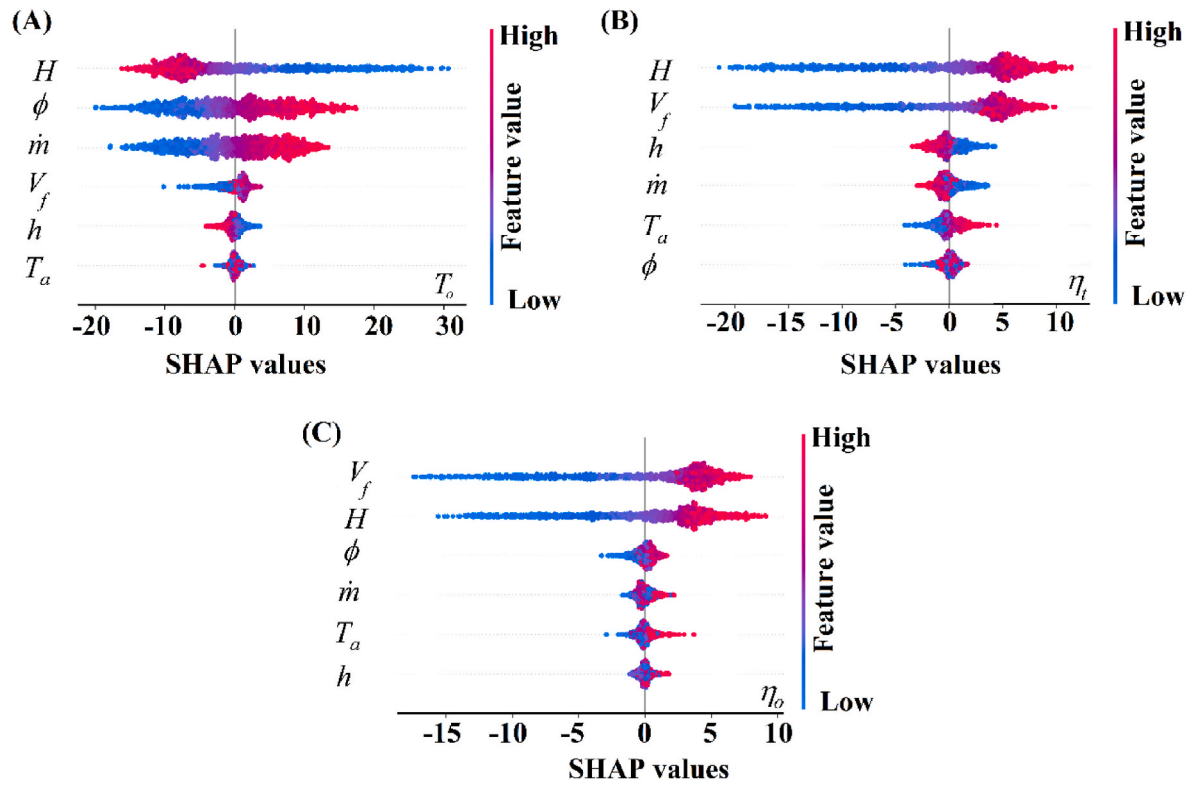


Fig. 10. SHAP based summary plots to explain the prediction mechanism of ML models corresponding to **(A)** T_o **(B)** η_t and **(C)** η_o.

flow rate ($\dot{m}$). The GPR model for T_o does not give much importance to the variation in volume fraction of nano particles (V_f), heat transfer coefficient for convection (h) and ambient temperature (T_a), while making the predictions. It is also revealed that while making the prediction for T_o, higher values of H, negatively impact the predicted response and is vice-versa for the lower value of H. In contrast, the lower values of φ and $\dot{m}$ negatively impact the prediction of T_o, and vice-versa. The SHAP analysis for the GPR model corresponding to the thermal efficiency is illustrated in Fig. 10(B), where it is revealed that higher values of depth of channel (H) and volume fraction (V_f) have a positive impact on the prediction of η_t, same observations can be drawn from the SHAP analysis for GPR model of η_o (refer to Fig. 10(C)). It is to be noted that the observations obtained from the SHAP analysis presents the corroboration of Pearson's correlation analysis (refer to Fig. 4) and data-driven sensitivity analysis (refer to Fig. 10(A)).

In summary, this study presents a machine learning (ML) model deployed for large-scale predictions to perform downstream tasks such as sensitivity analysis (illustrated in Fig. 7(A)), uncertainty quantification (illustrated in Fig. 7(B–D)), interactive influence of important features (illustrated in Fig. 8), and multi-objective optimization (illustrated in Fig. 9). These downstream tasks utilized randomly generated samples via Monte Carlo sampling (MCS), enabling large-scale predictions (for approximately 500,000 unknown samples) to be made instantaneously with the help of the computationally efficient Gaussian Process Regression (GPR) model. These tasks would have remained unexplored without the deployment of the computationally efficient ML-based framework. The comparison of the computational cost incurred in large-scale FDM simulations and large-scale predictions through ML model is illustrated in Figure SM2 of the *supplementary file*. The FDM simulations and ML modelling is performed using the laptop with the configurations of 6 core processor, 8 GB RAM, and 1.20 GHz clock-speed. It is evident from Figure SM2 that with the increase in number of simulations the computational time linearly increases, whereas for the increment in number of simulations the time for prediction through ML model remains a fraction of second. This shows that conducting conventional numerical simulations or lab experiments for such a large number of samples would have been time- and resource-intensive, demonstrating how the proposed framework is significantly faster than traditional approaches.

4. Conclusions

This paper introduces a data-driven multicriteria optimization approach for nanofluid-based solar thermal collectors (NBSCs) and utilized silver nanoparticles as a medium to enhance the performance metrics such as outlet temperature (T_o), thermal efficiency (η_t), and optical efficiency (η_o). The study systematically integrates three distinct modelling techniques: the Numerical Model (FDM), the Machine Learning Model (Gaussian Process Regression or GPR), and the Multi-objective Genetic Algorithm (GA) for optimal parameter setting. The important insights and findings from the study are as follows:

1. Pearson's correlation analysis shows that flux (φ) and mass flow rate ($\dot{m}$) have a moderate positive correlation with outlet temperature (T_o), while channel depth (H) has a negative correlation with T_o. Volume fraction (V_f) and channel depth (H) positively correlate with both thermal (η_t) and optical efficiency (η_o). There is a strong negative correlation between T_o and both η_t and η_o, with η_t and η_o being positively correlated.
2. Gaussian Process Regression (GPR) models were validated using out-of-the-fold samples, ensuring robust generalization capability. The models demonstrated high accuracy, with R^2 values ranging from 0.96 to 0.99 and prediction errors within ±5 %.
3. Sensitivity analysis identified key factors influencing the outlet temperature (T_o): depth of the channel (H), flux (φ), and mass flow rate ($\dot{m}$). Additionally, thermal efficiency (η_t) and optical efficiency (η_o) are most sensitive to depth (H) and volume fraction (V_f). The simultaneous increase in H and V_f lead to substantial improvements in η_t and η_o.

4. The optimal parameters identified were $V_f = 0.0037$ %, $\varphi = 2712.95$ W/m^2, $H = 0.003$ m, $\dot{m} = 542.79$ kg/s, $T_a = 27.62$ °C, and $h = 4.38$ W/m^2. Predictions from the ML-MOGA approach closely matched those from the numerical simulations (FDM), with slight deviations in temperature gain and efficiencies: ML-MOGA predicted $T_o = 117.27$ °C (FDM $T_o = 119.05$ °C), $\eta_t = 87.62$ % (FDM $\eta_t = 85.22$ %), and $\eta_o = 77.79$ % (FDM $\eta_o = 77.82$ %).
5. SHAP-based explanations of the developed ML models confirm the findings from Pearson's correlation and sensitivity analysis. They highlight the importance of channel depth (H), flux (φ), and mass flow rate ($\dot{m}$) in predicting outlet temperature (T_o), and show that higher values of H and volume fraction (V_f) positively impact predictions for thermal efficiency (η_t) and optical efficiency (η_o).

In summary, this study successfully demonstrates that the integration of ML and GA paradigm with numerical simulations can significantly enhance the design and performance of NBSCs by avoiding the preliminary computational cost of performing large scale simulations. The developed optimization framework provides a practical and effective method for developing highly efficient NBSCs, contributing valuable insights for future advancements in the field.

CRediT authorship contribution statement

Anjana Sankar: Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Kritesh Kumar Gupta:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Vishal Bhalla:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Conceptualization. **Daya Shankar Pandey:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.energy.2025.135212.

Data availability

Data will be made available on request.

References

[1] IEA Report" [Online]. Available: https://www.iea.org/data-and-statistics/data-tools/energy-statistics-data-browser?country=WORLD&fuel=Energysupply&indicator=TESbySource. [Accessed on : 25.August.2024].
[2] Tester JW, Drake EM, Driscoll MJ, Golay MW, Peters WA. Sustainable energy: choosing among options. MIT Press; 2012.
[3] Duffie J, Beckman W. Solar engineering of thermal processes. fourth ed. 2013.
[4] Khullar V, Tyagi H, Hordy N, Otanicar TP, Hewakuruppu Y, Modi P, Taylor RA. Harvesting solar thermal energy through nanofluid-based volumetric absorption systems. Int J Heat Mass Tran 2014;77:377–84.
[5] Lenert A, Wang EN. Optimization of nanofluid volumetric receivers for solar thermal energy conversion. Sol Energy 2012;86(1):253–65.
[6] Khullar V, Bhalla V, Tyagi H. Potential heat transfer fluids (nanofluids) for direct volumetric absorption-based solar thermal systems. J Therm Sci Eng Appl 2017;2017 [c)].
[7] Bhalla V, Khullar V, Tyagi H. Investigation of factors influencing the performance of nanofluid-based direct absorption solar collector using Taguchi method. J Therm Anal Calorim 2019;135(2).
[8] Bertocchi R, Karni J, Kribus A. Experimental evaluation of a non-isothermal high temperature solar particle receiver. Energy 2004;29(5–6):687–700.
[9] Miller F, Koenigsdorff R. Theoretical analysis of a high-temperature small-particle solar receiver. Sol Energy Mater 1991;24(1–4):210–21.
[10] Abdelrahman M, Fumeaux P, Suter P. Study of solid-gas-suspensions used for direct absorption of concentrated solar radiation. Sol Energy 1979;22(1):45–8.
[11] Bohn MS, Wang KY. Experiments and analysis on the molten salt direct absorption receiver concept. J Sol Energy Eng 1988;110(1):45–51.
[12] Bhalla V, Tyagi H. Parameters influencing the performance of nanoparticles-laden fluid-based solar thermal collectors: a review on optical properties. Renew Sustain Energy Rev 2018;84(December 2017):12–42.
[13] Tyagi H, Phelan P, Prasher R. Predicted efficiency of a low-temperature nanofluid-based direct absorption solar collector. J Sol Energy Eng 2009;131(4):041004.
[14] Khullar V, Tyagi H, Hordy N, Otanicar TP, Hewakuruppu Y, Modi P, Taylor Ra. Harvesting solar thermal energy through nanofluid-based volumetric absorption systems. Int J Heat Mass Tran 2014;77(November 2015):377–84.
[15] Taylor RA, Otanicar TP, Phelan PE. Technological advances to maximize solar collector energy output : a review. 2019. p. 140 (December 2018).
[16] Bhalla V, Tyagi H. Solar energy harvesting by cobalt oxide nanoparticles, a nanofluid absorption based system. Sustain Energy Technol Assessments 2017:1–10.
[17] Taylor RA, Phelan PE, Otanicar TP, Adrian R, Prasher R. Nanofluid optical property characterization: towards efficient direct absorption solar collectors. Nanoscale Res Lett 2011;6(1):1–11.
[18] Chen M, He Y, Huang J, Zhu J. Synthesis and solar photo-thermal conversion of Au, Ag, and Au-Ag blended plasmonic nanoparticles. Energy Convers Manag 2016;127:293–300.
[19] Chen M, He Y, Zhu J, Shuai Y, Jiang B, Huang Y. An experimental investigation on sunlight absorption characteristics of silver nanofluids. Sol Energy 2015;115:85–94.
[20] Bhalla V, Khullar V, Parupudi RV. Design and thermal analysis of nanofluid-based compound parabolic concentrator. Renew Energy 2022;185:348–62.
[21] Bandarra Filho EP, Mendoza OSH, Beicker CLL, Menezes A, Wen D. Experimental investigation of a silver nanoparticle-based direct absorption solar thermal system. Energy Convers Manag 2014;84:261–7.
[22] Moravej M, Doranehgard MH, Razeghizadeh A, Namdarnia F, Karimi N, Li LKB, Mozafari H, Ebrahimi Z. Experimental study of a hemispherical three-dimensional solar collector operating with silver-water nanofluid. Sustain Energy Technol Assessments 2021;44(January):101043.
[23] Singh P, Kumar S, Chander N, Bagha AK. Experimental investigation of an asymmetric compound parabolic concentrator–based direct absorption solar collector using plasmonic nanofluids. Environ Sci Pollut Control Ser 2023;30(21):60383–98.
[24] Kimpton H, Cristaldi DA, Stulz E, Zhang X. Thermal performance and physicochemical stability of silver nanoprism-based nanofluids for direct solar absorption. Sol Energy 2020;199(February):366–76.
[25] Baro RK, Kotecha P, Anandalakshmi R. Multi-objective optimization of nanofluid-based direct absorption solar collector for low-temperature applications. J Build Eng 2023;72(November 2022):106258.
[26] Qin C, Kang K, Lee I, Lee BJ. Optimization of a direct absorption solar collector with blended plasmonic nanofluids. Sol Energy Jul. 2017;150:512–20. https://doi.org/10.1016/j.solener.2017.05.007. Elsevier BV.
[27] Qin C, Kang K, Lee I, Lee BJ. Optimization of the spectral absorption coefficient of a plasmonic nanofluid for a direct absorption solar collector. Sol Energy Jul. 2018;169:231–6. https://doi.org/10.1016/j.solener.2018.04.056. Elsevier BV.
[28] Shahsavar A, Moayedi H, Al-Waeli AHA, Sopian K, Chelvanathan P. Machine learning predictive models for optimal design of building-integrated photovoltaic-thermal collectors. Int J Energy Res Mar. 19, 2020;44(7):5675–95. https://doi.org/10.1002/er.5323. Hindawi Limited.
[29] Sankar A., Gupta K.K., Bhalla V., "Development of data-driven predictive framework for nanofluid-based solar thermal collector: a machine learning approach", Paper no. CST 3056, International conference on sustainable technologies in civil and environmental engineering (ICSTCE 2024).
[30] Ataee S, Ameri M, Askari IB, Keshtegar B. Evaluation and intelligent forecasting of energy and exergy efficiencies of a nanofluid-based filled-type U-pipe solar ETC using three machine learning approaches. Energy 2024;298:131355.
[31] Chilambarasan L, Thangarasu V, Ramasamy P. Solar flat plate collector's heat transfer enhancement using grooved tube configuration with alumina nanofluids: prediction of outcomes through artificial neural network modeling. Energy 2024;289:129953.
[32] Mausam K, Singh S, Ghosh SK, Singh RP. Thermal performance modelling of solar flat plate parallel tube collector using ANN. Energy 2024:131940.
[33] Diwania S, Kumar M, Kumar R, Kumar A, Gupta V, Khetrapal P. Machine learning-based thermo-electrical performance improvement of nanofluid-cooled photovoltaic–thermal system. Energy Environ 2024;35(4):1793–817.
[34] Alhamayani A. Performance analysis and machine learning algorithms of parabolic trough solar collectors using Al2O3-MWCNT as a hybrid nanofluid. Case Stud Therm Eng 2024;57:104321.
[35] Li H, Liu Z, Liu K, Zhang Z. Predictive power of machine learning for optimizing solar water heater performance: the potential application of high-throughput screening. Int J Photoenergy 2017;2017:1–10. https://doi.org/10.1155/2017/4194251. Hindawi Limited.
[36] Liu Z, Li H, Liu K, Yu H, Cheng K. Design of high-performance water-in-glass evacuated tube solar water heaters by a high-throughput screening based on machine learning: a combined modeling and experimental study. Sol Energy Jan. 2017;142:61–7. https://doi.org/10.1016/j.solener.2016.12.015. Elsevier BV.
[37] Salari A, Shakibi H, Soltani S, Kazemian A, Ma T. Optimization assessment and performance analysis of an ingenious hybrid parabolic trough collector: a machine

learning approach. Appl Energy Jan. 2024;353:122062. https://doi.org/10.1016/j.apenergy.2023.122062. Elsevier BV.
[38] Mukhopadhyay T, Naskar S, Gupta KK, Kumar R, Dey S, Adhikari S. Probing the stochastic dynamics of coronaviruses: machine learning assisted deep computational insights with exploitable dimensions. Advanced Theory and Simulations 2021;4(7):2000291.
[39] Brewster MQ. Thermal radiative transfer and properties. Wiley; 1992.
[40] Cengel Y a. Heat transfer. second ed. 2002.
[41] Gupta KK, Bhowmik D. Exploring sustainable solutions for soil stabilization through explainable Gaussian process-assisted multi-objective optimization. Mater Today Commun 2024:110154.
[42] Gupta KK, Barman S, Dey S, Naskar S, Mukhopadhyay T. On exploiting nonparametric kernel-based probabilistic machine learning over the large compositional space of high entropy alloys for optimal nanoscale ballistics. Sci Rep 2024;14(1):16795.
[43] Gupta KK, Barman S, Dey S, Mukhopadhyay T. Explainable machine learning assisted molecular-level insights for enhanced specific stiffness exploiting the large compositional space of AlCoCrFeNi high entropy alloys. Mach Learn: Sci Technol 2024;5(2):025082.
[44] Pandey DS, Raza H, Bhattacharyya S. Development of explainable AI-based predictive models for bubbling fluidised bed gasification process. Fuel 2023;351:128971.
[45] Pan I, Pandey DS. Incorporating uncertainty in data driven regression models of fluidized bed gasification: a Bayesian approach. Fuel Process Technol 2016;142:305–14.
[46] Khullar V, Tyagi H. A study on environmental impact of nanofluid-based concentrating solar water heating system. Int J Environ Stud 2012;69(2):220–32.
[47] Han D, Meng Z, Wu D, Zhang C, Zhu H. Thermal properties of carbon black aqueous nanofluids for solar absorption. Nanoscale Res Lett 2011;6:1–7.
[48] Shafizadeh A, Shahbeik H, Nadian MH, Gupta VK, Nizami AS, Lam SS, Peng W, Pan J, Tabatabaei M, Aghbashlo M. Turning hazardous volatile matter compounds into fuel by catalytic steam reforming: an evolutionary machine learning approach. J Clean Prod 2023;413:137329.
[49] Chen X, Shafizadeh A, Shahbeik H, Rafiee S, Golvirdizadeh M, Moradi A, Peng W, Tabatabaei M, Aghbashlo M. Machine learning-based optimization of catalytic hydrodeoxygenation of biomass pyrolysis oil. J Clean Prod 2024;437:140738.
[50] Alawi OA, Kamar HM, Salih SQ, Abba SI, Ahmed W, Homod RZ, Jamei M, Shafik SS, Yaseen ZM. Development of optimized machine learning models for predicting flat plate solar collectors thermal efficiency associated with Al2O3-water nanofluids. Eng Appl Artif Intell 2024;133:108158.
[51] Alawi OA, Kamar HM, Abdelrazek AH, Mallah AR, Mohammed HA, Homod RZ, Yaseen ZM. Design optimization of solar collectors with hybrid nanofluids: an integrated ansys and machine learning study. Sol Energy Mater Sol Cell 2024;271:112822.
[52] Çolak AB, Bayrak M. A novel machine learning study: maximizing the efficiency of parabolic trough solar collectors with engine oil-based copper and silver nanofluids. Heat Tran Res 2024;55(16).
[53] Alhamayani A. Performance analysis and machine learning algorithms of parabolic trough solar collectors using Al2O3-MWCNT as a hybrid nanofluid. Case Stud Therm Eng 2024;57:104321.
[54] Ataee S, Ameri M, Askari IB, Keshtegar B. Evaluation and intelligent forecasting of energy and exergy efficiencies of a nanofluid-based filled-type U-pipe solar ETC using three machine learning approaches. Energy 2024;298:131355.
[55] Qin C, Seo J, Yoon S, Lee BJ, Zhu Q. Performance analysis and design optimization of a direct-absorption parabolic-trough solar collector with concentric nanofluid partitions. Sol Energy Mater Sol Cell 2025;282:113327.
[56] Tyagi PK, Kumar R. Thermodynamic modeling and performance optimization of nanofluid-based photovoltaic/thermal system using central composite design scheme of response surface methodology. Renew Energy 2024;225:120341.
[57] Hai T, Basem A, Alizadeh AA, Sharma K, Jasim DJ, Rajab H, Ahmed M, Kassim M, Singh NSS, Maleki H. Optimizing Gaussian process regression (GPR) hyperparameters with three metaheuristic algorithms for viscosity prediction of suspensions containing microencapsulated PCMs. Sci Rep 2024;14(1):20271.
[58] Zhou Y, Zheng S, Liu Z, Wen T, Ding Z, Yan J, Zhang G. Passive and active phase change materials integrated building energy systems with advanced machine-learning based climate-adaptive designs, intelligent operations, uncertainty-based analysis and optimisations: a state-of-the-art review. Renew Sustain Energy Rev 2020;130:109889.
[59] Zhou Y, Zheng S, Zhang G. Machine learning-based optimal design of a phase change material integrated renewable system with on-site PV, radiative cooling and hybrid ventilations—study of modelling and application in five climatic regions. Energy 2020;192:116608.
[60] Zhou Y, Zheng S. Multi-level uncertainty optimisation on phase change materials integrated renewable systems with hybrid ventilations and active cooling. Energy 2020;202:117747.
[61] Zhou Y, Zheng S, Zhang G. Machine-learning based study on the on-site renewable electrical performance of an optimal hybrid PCMs integrated renewable system with high-level parameters' uncertainties. Renew Energy 2020;151:403–18.