



Data driven insights for parabolic trough solar collectors: Artificial intelligence-based energy and exergy performance analysis



Hai Tao^{a,b,c}, Omer A. Alawi^d, Raad Z. Homod^e, Mustafa KA. Mohammed^f, Leonardo Goliatt^g, Hussein Togun^h, Shafik S. Shafikⁱ, Salim Heddam^j, Zaher Mundher Yaseen^{k,l,*}

^a School of Computer and Information, Qiannan Normal University for Nationalities, Duyun, 558000, Guizhou, China

^b Institute of Big Data Application and Artificial Intelligence, Qiannan Normal University for Nationalities, Duyun, 558000, Guizhou, China

^c Institute for Big Data Analytics and Artificial Intelligence (IBDAAD), Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia

^d Department of Thermofluids, School of Mechanical Engineering, Universiti Teknologi Malaysia, 81310, UTM Skudai, Johor Bahru, Malaysia

^e Department of Oil and Gas Engineering, Basrah University for Oil and Gas, Basra, Iraq

^f College of Remote Sensing and Geophysics, Al-Karkh University of Science, Al-Karkh Side, Haifa St. Hamada Palace, Baghdad, 10011, Iraq

^g Computational Modeling Program, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil

^h Department of Mechanical Engineering, College of Engineering, University of Baghdad, Baghdad, Iraq

ⁱ Experimental Nuclear Radiation Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Iraq

^j Faculty of Science, Agronomy Department, University, 20 Août 1955 Skikda, Route El Hadaïk, BP 26, Skikda, Algeria

^k Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia

^l Interdisciplinary Research Centre for Membranes and Water Security, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia

ARTICLE INFO

Handling Editor: Panos Seferlis

Keywords:

Parabolic trough solar collectors
Energy efficiency
Exergy efficiency
Artificial intelligence
Synthetic oils
Nanofluids

ABSTRACT

Artificial intelligence (AI) algorithms can potentially contribute to optimizing energy and exergy outputs in renewable resources to increase efficiencies and reduce environmental risk. This study utilized tree-based, linear, and non-linear regression techniques to predict the energy and exergy efficiency of Parabolic Trough Solar Collectors (PTSCs) using oil-based nanofluids. The cooling fluids were prepared from three main oil types, namely Therminol VP-1, Syltherm 800, and Dowtherm Q mixed with three metallic oxides, including Al_2O_3 , CuO , and SiO_2 , in various volume fractions. The two outputs were predicted according to a range of input parameters, namely Volume Fraction (%), Reynolds Number (Re), Inlet Fluid Temperature, Direct Solar Irradiance, Nusselt Number (Nu), and Friction Factor (f). Ensemble approaches such as Extra Trees Regressor (ETR), Extreme Gradient Boosting (XGBoost), Random Forest Regressor (RFR), Classification and Regression Trees (CART), and Adaptive Boosting (AdaBoost) were the top-performing models in the model selection process out of nine. The modeling results showed that, CART was the top model in predicting the energy efficiency using Syltherm 800- SiO_2 nanofluid with $R^2 = 0.9999$. Meanwhile, ETR was the top model in predicting the exergy efficiency using Dowtherm Q- SiO_2 nanofluid with $R^2 = 0.9988$. Moreover, in the business insights, the maximum errors in the energy and exergy models were observed (1.43 % and 1.97 %) using Therminol VP-1, (1.3 % and 2.44 %) using Syltherm 800 and Syltherm 800-CuO and (1.15 % and 2 %) using Dowtherm Q and Dowtherm Q-CuO, respectively.

1. Introduction

1.1. General background of study

The parabolic trough solar collector (PTSC) is one of the outstanding technologies of renewable energy resources because of its low operating

budget relative to other technologies in the same field. The principal mechanism of the PTSC is the use of parabolic troughs to collect and concentrate direct sun radiation (I). The receiver tube, a metal tube with a black selective coating on its surface (Daviran et al., 2019; Manikandan et al., 2019) and a vacuum glass cover [3], is placed in the focus of the parabola (Ajbar et al., 2022).

The heat transfer fluid (HTF), such as water, molten salt, and thermal

* Corresponding author. Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia.

E-mail addresses: haitao@sgmtu.edu.cn (H. Tao), omer.adnan85@gmail.com (O.A. Alawi), raadahmood@yahoo.com (R.Z. Homod), mustafa_kareem97@yahoo.com (M.KA. Mohammed), leonardo.goliatt@ufjf.br (L. Goliatt), hussein-tokan@coeng.uobaghdad.edu.iq (H. Togun), dr.shafik@alayen.edu.iq (S.S. Shafik), s.heddam@univ-skikda.dz (S. Heddam), z.yaseen@kfupm.edu.sa (Z.M. Yaseen).

oil (Bellos et al., 2016) plays a crucial role in the system's operation by absorbing the concentrated sunlight and transferring the heat to a power cycle for electricity generation. Furthermore, depending on the latitude,

Carboxymethyl cellulose (CMC) (99.5 % DW and 0.5 % CMC) incorporating Iron Oxide/Carbon nanotubes ($\text{Fe}_3\text{O}_4/\text{CNTs}$) was studied within a double-fluid parabolic trough solar collector (DFPTSC) utilizing a deep

Nomenclature

Adaboost	Adaptive Boosting Regression	\dot{m}	Mass flow rate (kg/s)
<i>Adjusted-R</i> ²	Adjusted-Coefficient of determination	MLR	Multiple linear regression
Al_2O_3	Aluminium Oxide	MOO	Multi-objective optimization
ANNs	Artificial Neural Networks	MSE	Mean Squared Error
A_{PTSC}	Area of solar collector (m ²)	Nu	Nusselt Number
CART	Classification and Regression Trees	PR	Polynomial regression
CMC	Carboxymethyl cellulose	PSO	Particle-swarm-optimization
CNTs	Carbon nanotubes	PTSC	Parabolic Trough Solar Collector
Cp_{nf}	Specific heat capacity (J/kg-K)	Re	Reynolds Number
CuO	Copper(II) oxide	RFR	Random Forest Regressor
DT	Decision tree	RMSE	Root Mean Squared Error
ETR	Extra Trees Regressor	SHAP	SHapley Additive exPlanations
f	Friction Factor	SiO_2	Silicon Dioxide
Fe_3O_4	Iron(II,III) oxide	SOE	Second-Order Equation
FFBP	Feedforward Backpropagation	SVR	Support Vector Regression
GA	Genetic algorithm	Tamb	Ambient temperature (K)
GBM	Gradient boosting machine	T_{fi}	Inlet temperature (K)
GII	Gini Index of Impurity	T_{fo}	Outlet temperature (K)
GPR	Gaussian process regression	T_{sun}	Sun temperature (K)
HTF	Heat Transfer Fluid	W_p	Pumping power (W)
I	Direct Solar Irradiance (W/m ²)	XGBoost	Extreme Gradient Boosting
KNR	K-Nearest Neighbours Regression	ΔP	Pressure drop (Pa)
LASSO	Least Absolute Shrinkage and Selection Operator	η_{en}	Energy efficiency
LR	Linear regression	η_{ex}	Exergy efficiency
MAE	Mean Absolute Error	η_{th-e}	Thermal-electrical conversion efficiency
		ρ_{nf}	Density of nanofluids (kg/m ³)

the PTSC should be installed in an E-W or N-S direction to get the lowest cosine loss. (Ajbar et al., 2022). Furthermore, placing it in sun-rich areas is advised because the net annual heat gain in these areas is nearly four times larger than in non-sun-rich areas for the identical solar tracking system (Liang et al., 2017). Simultaneously, the solar energy gathered on the PTSC's moving surface utilizing the sun tracking system is approximately 46.46 % greater than that collected on the fixed surface (Ajbar et al., 2022).

Artificial intelligence (AI) algorithms are among the most fascinating prospective modeling tools. It can be believed to be a profound tool for solving challenging problems on multiple topics (Vakili and Salehi, 2023). In general, it is a subset of AI models that comprises techniques that let computers find and make sense of data on their own and deliver machine learning applications (Elsheikh et al., 2019). Many academics prefer AI models to the traditional techniques for forecasting and modeling thermal system performance due to its high processing speed, large capacity, and ease of use. (Ghritlahre and Prasad, 2018).

1.2. Literature review on AI-based energy and exergy for PTC

This section is focused on reviewing the previous studies that applied various ML and DL algorithms to predict energy and exergy efficiency in parabolic trough solar collectors. The application of Artificial Neural Networks (ANNs) was proposed in order to detect three major objectives, namely faults in the optical efficiency, flow rate, and thermal losses in a model of a 50 MW parabolic-trough solar plant working on Therminol VP-1 as a heat transfer fluid (Ruiz-Moreno et al., 2023). The fault detection accuracy was found to be 71.72 %, 83.96 %, and 90.62 % for the first, second, and third approaches, respectively. A thermal efficiency of a non-Newtonian base fluid consisting of water and

learning approach (ANN) (Mustafa et al., 2022). The ANN demonstrated a high level of accuracy in predicting the thermal efficiency, exceeding 99 %. In another study, the thermal efficiency of a non-Newtonian DW-CMC/ Al_2O_3 nanofluid inside a PTSC was predicted using an ANN model, yielding an R^2 -value of 0.9998 (Al-Rashed et al., 2021). Moreover, the error between the numerical and ANN results was <0.2 %. A previous study (reference (Abubakr et al., 2020)) prepared a mixture of three individual synthetic oils, specifically Therminol VP-1, Syltherm-800, and Dowtherm-Q. These oils were combined with three different nanomaterials, namely CuO , Al_2O_3 , and SiO_2 , within a PTSC system. Multi-layered Feedforward Backpropagation (FFBP) was utilized to predict energy efficiency and exergy efficiency, resulting in an accuracy of >99 %. Furthermore, the study included two optimization techniques: genetic algorithm (GA) and multi-objective optimization (MOO). Two ANN models were developed using (TANSIG and LOGSIG) functions in order to predict the energy efficiency of solar parabolic trough collector according to six input parameters such as rim-angle, inlet-temperature, ambient-temperature, water volumetric flow rate, direct-solar-radiation and wind-speed (Ajbar et al., 2021a). The results showed that both ANN models achieved satisfactory results with a coefficient of determination of 0.9511 and a root mean square error of 0.0193. Then, ANN was optimized with genetic algorithms (GA) and particle swarm optimization (PSO) in order to find the optimum input variable. Moreover, the outlet temperature of PTSC was predicted using Six artificial neural networks (ANNs) and four multiple linear regression (MLR) models according to different inputs like Temperature, wind speed, rim angle, flow rate, and solar radiation (Ajbar et al., 2021b). The simulation showed that ANN-1 and MLR with Second-Order Equation (SOE) are the models that yielded the best results with $R^2 = 0.9984$ and $R^2 = 0.9958$ and with an RMSE = 0.7708 and 1.6031, respectively. The

entropy generation of PTSC using Al_2O_3 -60:40 % EG/DW nanofluid was minimized using ANN and quadratic algorithm (Ebrahimi-Moghadam et al., 2018). Their optimization results showed that, the rate of the entropy generation decreases by decreasing volume fraction, increasing particle diameter, and increasing average flow temperature. In another study, the entropy generation of PTSC using SiO_2 -Water Nanofluids was minimized using ANNs and the Genetic Algorithm (Okonkwo et al., 2020). The optimal mean square error used as a performance validation of the model was 0.02288 for training and 0.0282 for testing with an R^2 value of 0.9999. It was concluded that AI techniques can be an efficient tool for predicting the rate of entropy generation in a collector within the constraint of the defined parameters. ANN and superposition principle were used to predict the exit fluid temperature of the PTSC for one day of operation under fluctuated solar radiation (Heng et al., 2019). The predicted results can be used for initial system planning, heat balance analysis, and system design. The hourly electric production of the power plant in Morocco working on PTSC technology was estimated using three analytical and ANN models (Zaaoumi et al., 2021). The 1st analytical model was based on calculating the heat losses, while the 2nd analytical model was based on the energy efficiency. The 3rd model was an ANN. The simulation results indicated that the ANN model performed much better than the analytical models. The instantaneous energy efficiency of phase change material-based receiver tube for naturally circulated PTSC was predicted using models that include linear regression (LR), polynomial regression (PR), decision tree (DT), random forest regressor (RFR), artificial neural network (ANN), KNN regression, and Gaussian process regression (GPR) (Kottala et al., 2023). Among all these, the GPR model showed higher prediction performance (i.e., RMSE = 0.0049, R^2 = 0.9977) than the remaining developed ML models. Support Vector Regressor (SVR) was developed to predict energy efficiency (R^2 = 0.9998) and exergy efficiency (R^2 = 0.9996) of PTSC working with molten salt-based nanofluids (Solar Salt, Hitec, and Hitec XL and three nanoparticle types (Al_2O_3 , CuO, and SiO_2)) (Kaood et al., 2021).

1.3. Research objectives and novelty

In summary, few attempts in the literature review utilized AI models for predicting the energy efficiency (1st law efficiency) and exergy efficiency (2nd law efficiency) of PTSCs working on nanofluids as heat transfer fluids. However, there is still a significant necessity for reliable and cost-efficient approaches in order to estimate the energy and exergy performance of PTSC technologies using nanofluids. This study aims to fill this research gap by developing a complete AI methodology that utilizes tree-based, linear-based, and non-linear-based regression algorithms for predicting the 1st law efficiency and 2nd law efficiency in PTSCs. In the current study, three synthetic-based oils, namely (Therminol VP-1, Syltherm 800, and Dowtherm Q) were mixed with three different nanoscale metallic oxides, namely (Al_2O_3 , CuO, and SiO_2) to prepare nanofluids, with varying percentages of volume. Nine individual AI algorithms were implemented to predict the energy and exergy outputs based on multiple input features, such as Volume Fraction Percentage, Reynolds Number (Re), Inlet Fluid Temperature, Direct Solar Irradiance, Nusselt Number (Nu), and Friction Factor (f). The methodology given here is a novel and low-cost computational technique that aids decision-making to reduce the environmental impact of fossil fuels. Furthermore, it can also be employed in various industrial processes to create cleaner and sustainable production using established CSP technologies.

2. Experimental work and artificial intelligence models

2.1. Experimental data collection

In order to train and test the current AI computational models, authors have collected a dataset from a previous study was published in the

Renewable Energy journal (Abubakr et al., 2020). This study examined the 1st and 2nd law efficiencies of PTSCs using three oil-based fluids, including Therminol VP-1, Syltherm 800, and Dowtherm Q. In order to prepare nano-oil fluids, three metallic oxides, including Al_2O_3 , CuO, and SiO_2 , were mixed with the base fluids at varied volume percentages. It is important to note that the three oils and three nanoscale materials were chosen for the current application for two reasons: (i) their wide availability in industrial and CSP technologies and (ii) their comparatively inexpensive costs, respectively. The Energy Efficiency (%) and Exergy Efficiency (%) of PTSC were predicted based on several input parameters, including Volume percentage (%), Reynolds Number (-), Inlet Fluid Temperature (°C), Direct Solar Irradiance (W/m²), Nusselt Number (-), and Friction Factor (-). Equations (1) and (2) show how the 1st law and 2nd law efficiencies were experimentally calculated (Abubakr et al., 2020). The descriptive statistics of 12 scenarios were displayed in Tables (see Appendix A), providing information on several statistical measures. These measures include the total number of observations (N-total) for each scenario, the number of missing data (N-missing), the mean, standard deviation, skewness, kurtosis, minimum value, first quartile (Q1), median, third quartile (Q3), and maximum value. SHAP values (SHapley Additive exPlanations) were used to evaluate the significance and influence of specific features on the model's prediction, as shown in Figs. 1–3. It is important to note that, although SHAP provides insights into the contribution and significance of each feature in the model's prediction, it does not assess the accuracy or reliability of the regression itself.

$$\eta_{En} = \frac{\dot{m}C_{p,nf}(T_{fo} - T_{fi}) - \dot{W}_p/\eta_{th-e}}{IA_{PTSC}} \quad (1)$$

$$\eta_{Ex} = \frac{\dot{m}C_{p,nf}(T_{fo} - T_{fi}) - \dot{m}C_{p,nf}T_{amb} \ln\left(\frac{T_{fo}}{T_{fi}}\right) - \dot{m}T_{amb}\Delta P}{IA_{PTSC} \left[1 - \frac{4}{3} \left(\frac{T_{amb}}{T_{sun}} \right)^2 + \frac{1}{3} \left(\frac{T_{amb}}{T_{sun}} \right)^4 \right]} \quad (2)$$

In this regard, η_{En} and η_{Ex} are the energy and exergy efficiency. $C_{p,nf}$ and ρ_{nf} are the heat capacity (J/kg-K) and density (kg/m³) of nano-oils. \dot{m} is the mass flow rate (kg/s). A_{PTSC} is the area of solar collector (m²). \dot{W}_p is the pumping power ($\dot{m}\Delta P/\rho_{nf}$) (W). η_{th-e} is taken here as thermal-electrical conversion efficiency (33 %). T_{fo} , T_{fi} , T_{amb} and T_{sun} are the temperatures of outlet fluid, inlet fluid, ambient and sun surface (K). I is Direct Solar Irradiance (W/m²).

2.2. Applied AI models

2.2.1. SHAP values (SHapley additive exPlanations)

The SHAP (Shapley Additive exPlanations) method, which draws upon principles from game theory (Štrumbelj and Kononenko, 2014) and local explanations (Ribeiro et al., 2016), was introduced by Lundberg and Lee (2017) as a means to explain the model's output. This approach offers a means to assess and quantify the individual impact of each input feature (predictor). Consider a Random Forest model in which a group N , consisting of n characteristics, is utilized to make predictions for an output variable $v(N)$. In the SHAP construction, the allocation of the contribution of each feature to the model output $v(N)$ is written as (φ_i is contribution of feature i) and determined by their respective marginal contributions (Shapley, 1953). As per many previous sources, SHAP values are calculated using a set of established criteria to confirm a reasonable distribution of feature contributions (Parsa et al., 2020).

$$\varphi_i = \sum_{S \subseteq N(i)} \frac{|S|!(n - |S| - 1)!}{n!} [v(S \cup \{i\}) - v(S)] \quad (3)$$

The linear function g , can be calculated in the following equation. Also, g is dependent on binary features.

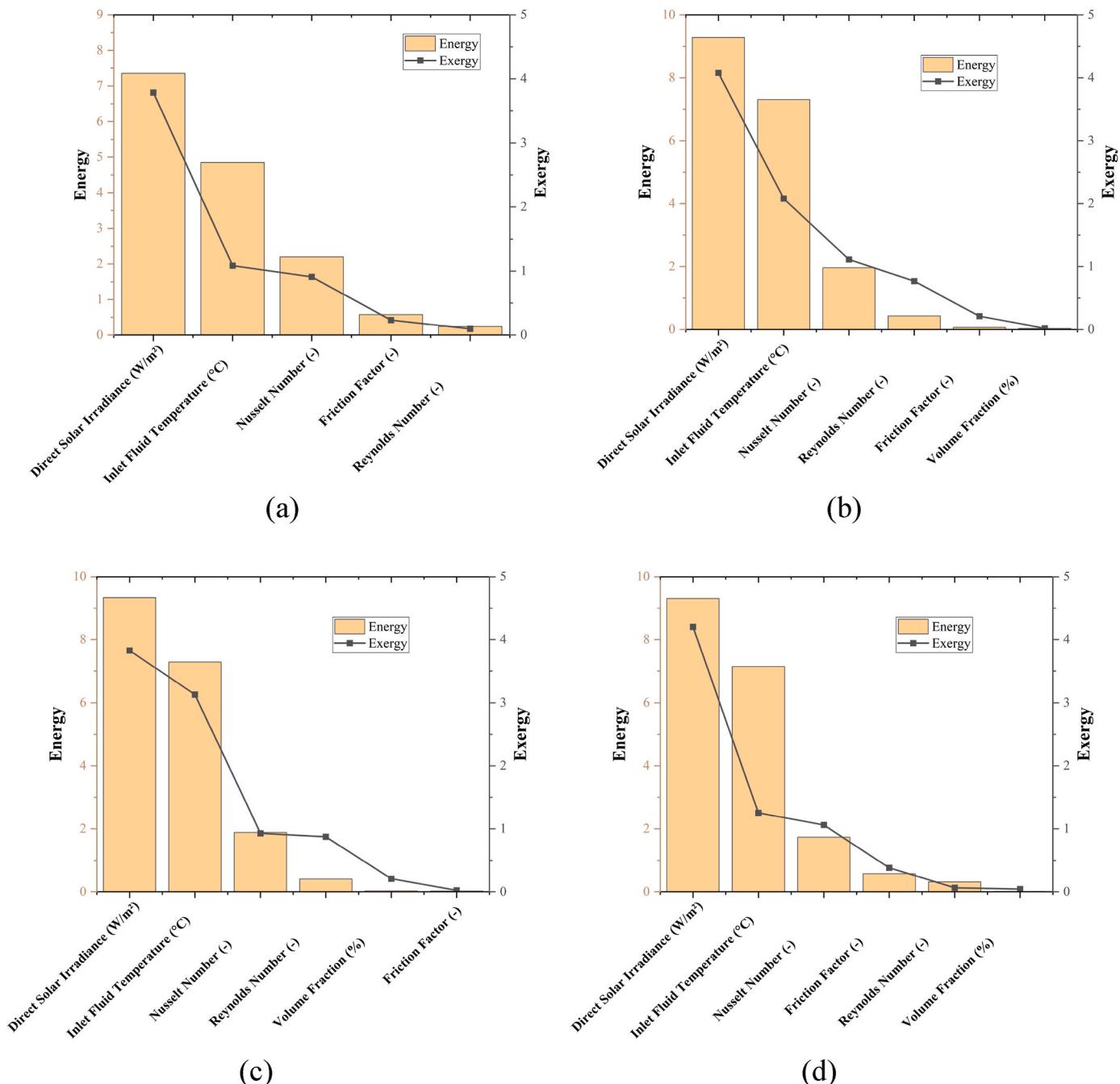


Fig. 1. SHAP values of Energy and Exergy; (a) Therminol VP-1, (b) Therminol VP-1-Al₂O₃, (c) Therminol VP-1-CuO, (d) Therminol VP-1-SiO₂.

$$g(\vec{z}) = \varphi_i + \sum_{i=1}^M \varphi_i z'_i \quad (4)$$

As above, \vec{z} is an element of the interval $(0, 1)^M$, where it takes the value of 1 when a feature is seen and 0 if not. Here, M corresponds to the total number of input parameters.

2.2.2. Extra tree regression (ETR)

The Extra Tree Regression (ETR) model is a type of ensemble ML algorithm that has been developed using decision trees (DTs) and bagging approaches (Geurts et al., 2006). This method includes the computation of the average outputs (energy and exergy) from many DTs, leading to an outstanding augmentation in the model's accuracy and minimized computational costs. The architectural design of the ETR

ensemble model follows a conventional top-down approach, as described in reference (Seyyedattar et al., 2020). One notable merit between the ETR approach and other ensemble methods, such as the Random Forest regression (RFR), is that ETR utilizes all the learning samples. In contrast, RF algorithm employs bootstrap replication and randomly divides nodes in the ETR algorithm using the cut-point of each feature (Jamei et al., 2021). Furthermore, the primary advantages of the ETR over classic DTs are as follows (Heddam et al., 2020): (i) It is more credible for individual decision trees to fix overfitting issues. (ii) Significant computational efficiency. (iii) The ETR model is a randomized ensemble model. The use of a single-averaging ensemble model results in lower accuracy than other models.

The number of random splits (K) and the minimum sample size required to divide a node (n_{min}) are the two key parameters to consider when implementing ETR algorithms for regression applications. The

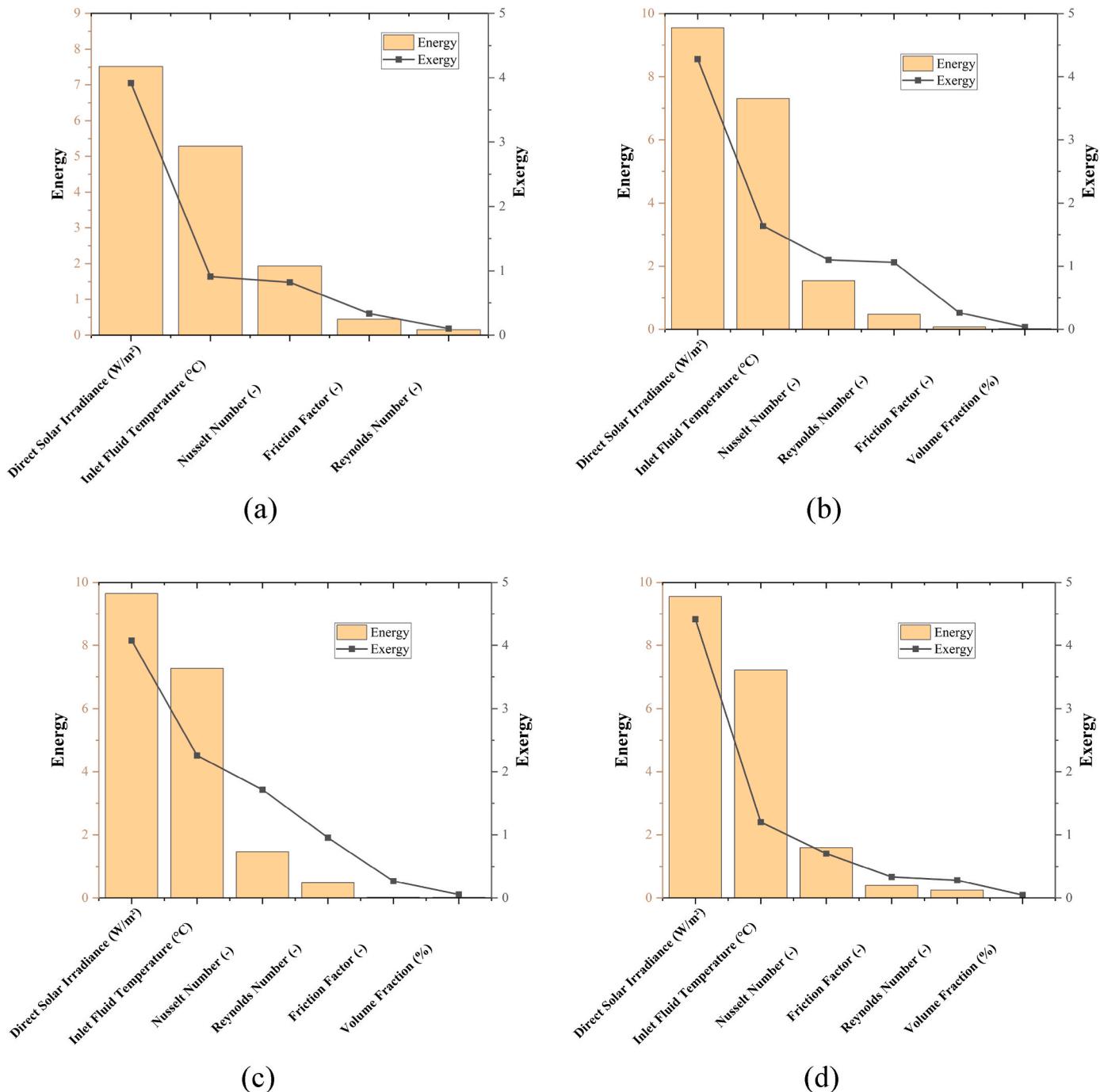


Fig. 2. SHAP values of Energy and Exergy; (a) Syltherm 800, (b) Syltherm 800-Al₂O₃, (c) Syltherm 800-CuO, (d) Syltherm 800-SiO₂.

process of tree growth involves the determination of K in each layer, which is repeated until the number of learning samples is less than n_{min} . The ETR strategy is employed in ensemble learning to reduce variation and mitigate bias. This is achieved through the explicit randomization of inputs and the averaging of processes (Geurts et al., 2006). The quantity of trees produced in the ETR model directly influences how variation is mitigated. The ETR demonstrates superior computational efficiency compared to a single decision tree while still preserving its robustness against irrelevant characteristics, particularly when the number of input parameters much exceeds the number of random splits (K) (Wehenkel et al., 2006). Figure (B-1) depicts the architecture of the ETR model. However, in order to enhance the regression efficiency of the ETR model, it is required to specify the maximum number of trees ($n_{Estimator}$),

maximum number of features ($max_{features}$), and maximum depth of the trees (max_{depth}) (Seyyedattar et al., 2020).

2.2.3. Extreme gradient boosting (XGBoost)

One notable application of the gradient boosting machine (GBM) method is the extreme gradient boosting (XGBoost) algorithm (Chen and Guestrin, 2016). The XGBoost model is executed using the “boosting” notion, wherein a collection of weak learners’ estimations are combined to construct a robust learner employing additive training techniques (Fan et al., 2018). The concept of XGBoost has been widely accepted across several sectors worldwide (Feigl et al., 2021; Ferreira and da Cunha, 2020; Jamei et al., 2022a; Ni et al., 2020; Sikorska-Senoner and Quilty, 2021; Tao et al., 2022; Xu et al., 2021; Zhang et al., 2021). In

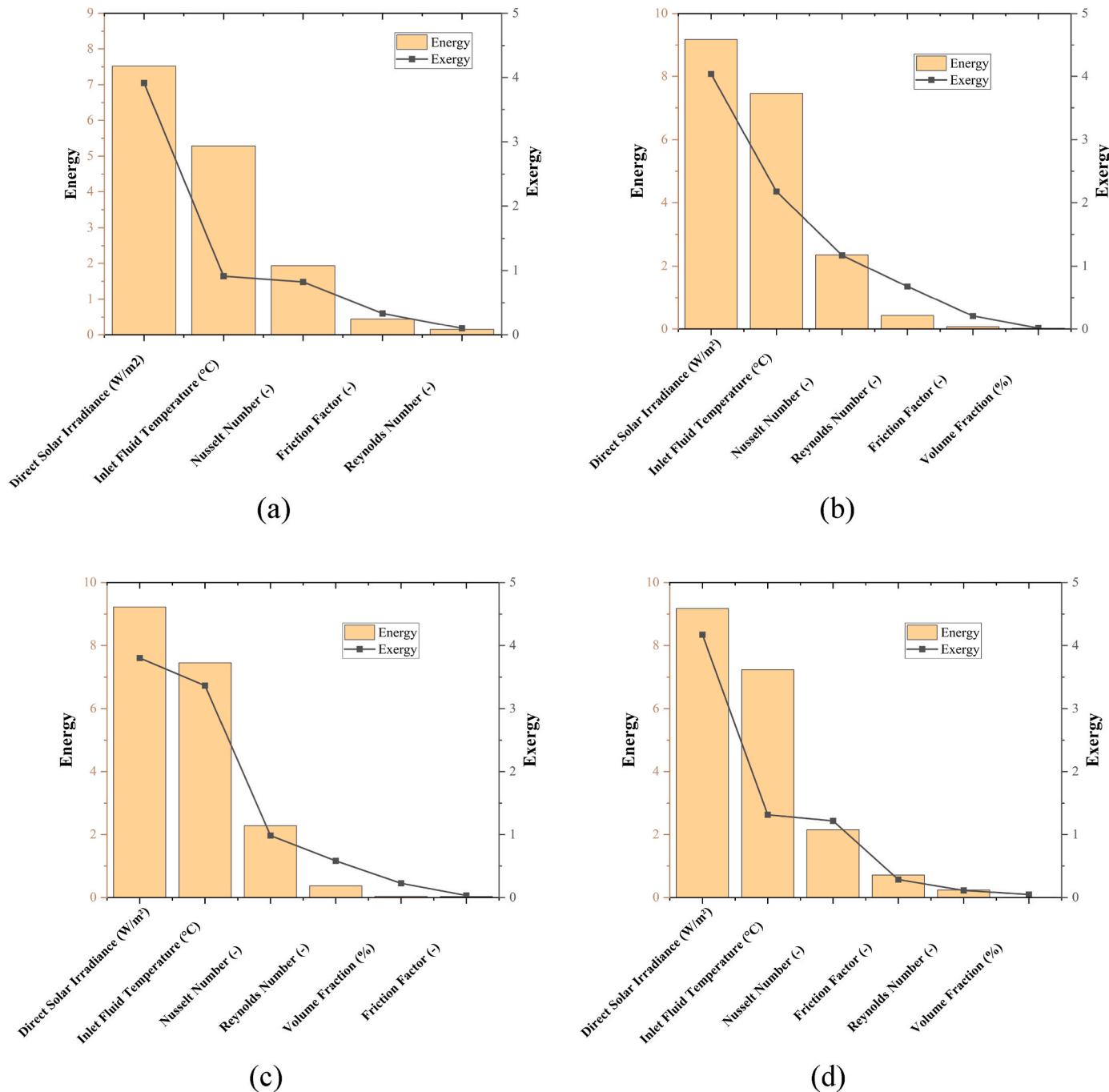


Fig. 3. SHAP values of Energy and Exergy; (a) Dowtherm Q, (b) Dowtherm Q-Al₂O₃, (c) Dowtherm Q-CuO, (d) Dowtherm Q-SiO₂.

XGBoost algorithms, the growth of trees follows a specific procedure. Firstly, a subset of the training set data is used to fit decision trees (DTs). Secondly, the residuals of the previous tree are fitted using a loss function, which varies depending on whether the task is regression or classification. This loss function quantifies the discrepancy between the output of the algorithms and the desired target value. Lastly, the next tree is constructed by incorporating the loss function and the previous tree (Brédy et al., 2020). The final output of the XGBoost model is determined by the weighted aggregate contribution of all the utilized decision trees. The XGBoost procedure is depicted in Figure (B-2), where it operates on input vectors (x) and employs independently and identically distributed random vectors (β_k). Furthermore, the predictive efficiency of XGBoost can be improved with the appropriate adjustment of its hyperparameters, specifically the learning_rate, n_estimators,

max_depth, subsample, and colsample_bytree.

2.2.4. Random Forest Regression (RFR)

The concept of Random Forest Regression (RFR) was initially introduced by Breiman in 2001 (Breiman, 2001). The author integrated decision forests with the Bagging framework as the foundational learner, and subsequently incorporated random property selection into the training process of the decision trees (Cui et al., 2019). In order to enhance accuracy, the RFR algorithm selects the optimal partition attribute from a subset that is randomly sampled from the entire collection of features. This occurs after the normal DTs select the best feature from the dataset at the current node (Gholizadeh et al., 2020; Jamei et al., 2021). The RFR model is a flexible and effective technique with numerous advantages. (i), it has a high level of accuracy in the

regression outputs. (ii) it can be applied to a large and complex database. (iii), it is efficient in handling high-dimensional datasets. (iv), it can maintain a high accuracy in estimating missing data even when the missing values are huge. (v), it can discover the impact between input parameters while selectively disregarding redundant information. It utilizes its own ensemble structure, as depicted in Figure (B-3).

2.2.5. Classification and Regression Tree (CART)

Classification and Regression Tree (CART) method was developed firstly by Breiman et al. (1984) in order to remove excessive variables and tackle complex datasets (Singh et al., 2013). The CART predictive model employs a recursive splitting algorithm to iteratively separate a given dataset into progressively smaller sub-groups (Ture et al., 2005). It is considered as a robust, non-parametric approach that applies recursive splitting to examine classification targets (categorical dependent variables) and regression (continuous dependent variables) outputs (Dastorani et al., 2018; Tsai et al., 2012). Additionally, it has been extensively used in various disciplines, leading to many successful applications and objectives (Alizamir et al., 2020; Jamei et al., 2022a; Phyo and Jeevananta, 2021; Torres-Barrán et al., 2019). Figure (B-4) demonstrates the workflow of CART model. After all, the Gini Index of Impurity (GII) is utilized to assess the efficiency of the divisions in classification trees and the overall error among nodes in regression trees. The expression for GII is stated as follows (Smetti et al., 2009):

$$GII = 1 - \sum_j p^2 \left(\frac{j}{t} \right) \quad (5)$$

In this context, the symbol p denotes the probability of class j occurring at a given node, while j represents the total number of classes present in node t .

2.2.6. Adaptive Boosting Regression (AdaBoost)

Freund and Schapire (1997) introduced the Adaptive Boosting Regression (AdaBoost) approach to enhance the regression accuracy of AI algorithms (Freund and Schapire, 1997). The AdaBoost regressor is a specific variant of the AdaBoost algorithm specifically designed to tackle

regression problems. The AdaBoost regressor's fundamental concept is the iterative adjustment of the weights assigned to training instances based on the prediction error observed at each boosting iteration. This process aims to train a sequence of weak regressors that exhibit a substantial bias error but a relatively low variance error. This underscores instances that the subsequent regressor inaccurately estimated in the initial phase. The final prediction is obtained by aggregating the coefficients of the weak regressors, resulting in a model that exhibits low bias and variance error (Pedregosa et al., 2011). In this study, the decision tree regressor, enhanced by AdaBoost, was employed as the primary regressor due to its efficiency in prediction (Bilali et al., 2021; Xiao et al., 2019). The mathematical formulations of the Adaboost algorithm are presented in the references (El Bilali et al., 2021; Jamei et al., 2022b).

Please take into account the provided training dataset: The dataset D is represented as a collection of m ordered pairs, denoted as $[(x_1, y_1), \dots, (x_m, y_m)]$. In this context, the target vector is denoted as y_m , the input vector is denoted as x_m , and m represents the total number of training data points. The subsequent method is employed to initialize the weight vector:

$$w_1(x_i) = \frac{1}{m}, \text{ for } i = (1, 2, \dots, m) \quad (6)$$

As an illustrative instance, the error rate can be computed from the next equation:

$$\varepsilon_t = \sum P_t(x_i)[h_t(x_i) - y_i] \quad (7)$$

The expression $P_t(x_i)$ is defined as the ratio of the weight $w_t(x_i)$ to the sum of all weights $\sum_{i=1}^m w_t(x_i)$. If the value of ε is > 0.5 and the condition T equals t minus 1 is met, the loop will be terminated. The weight confidence can be determined by employing the subsequent mathematical expression:

$$\alpha_t = \log \left(\frac{\varepsilon_t}{1 - \varepsilon_t} \right) \quad (8)$$

The subsequent phase facilitates the alteration of the weights

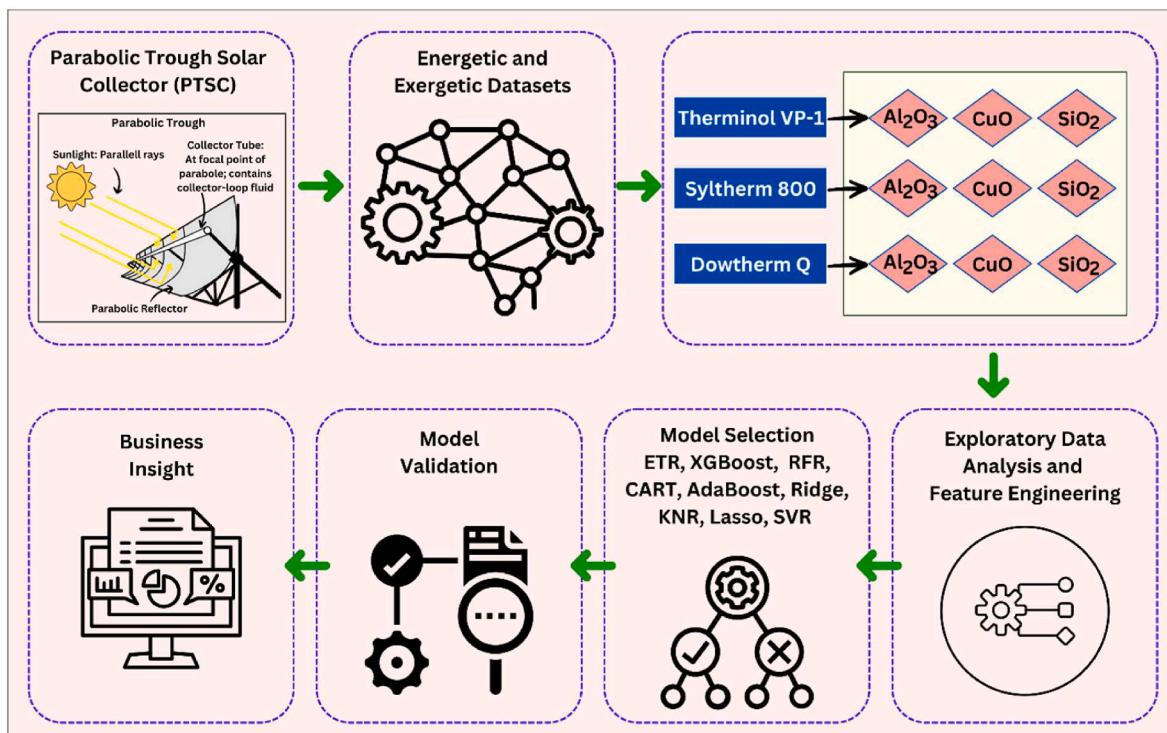


Fig. 4. Flowchart of the applied Artificial intelligence methodology.

associated with all training data:

$$w_{t+1}(x_i) = w_t(x_i) \cdot e^{-y_i h_t(x_i) \alpha_t} \quad (9)$$

If the value of ϵ_t is > 0.001 and the value of $t < T$, the task is to find the value of $P_t(x_i)$. This process should be repeated until all the necessary conditions are satisfied. The combination of weak learners results in the emergence of a strong learner.

$$H(x) = \text{sign} \left[\sum_j^T \alpha_t [h_t(x) = y] \right] \quad (10)$$

2.2.7. Kernel Ridge Regression

The Kernel Ridge Regression model, known as Kernel Regularized Least Squares, is an AI technique incorporating kernels and a ridge regression method. It aims to mitigate overfitting in regression problems by employing regularization and the kernel approach to identify nonlinear relationships (An et al., 2007; You et al., 2018). The mathematical equations of the Kernel Ridge Regression model are presented in the following equations (Ali et al., 2019; Jamei et al., 2022a):

$$\text{argmin}_q \frac{1}{q} \sum_{o=1}^q f_o - y_o^2 + \lambda f_H^2 \quad (11)$$

$$f_o = \sum_{p=1}^q \alpha_p \Phi(x_p, x_o) \quad (12)$$

$$K_{p,o} = \Phi(x_p, x_o) \quad (13)$$

$$(K + \lambda q I) = y \quad (14)$$

$$\tilde{y} = \sum_{o=1}^q \alpha_o \Phi(x_o, \tilde{x}) \quad (15)$$

$$\Phi(x_p, x_o) = x_p^T x_o \quad (16)$$

$$\Phi(x_p, x_o) = (x_p^T x_o + r)^d \quad (17)$$

$$\Phi(x_p, x_o) = \exp \left(\frac{-x_p - x_o^2}{2\sigma^2} \right) \quad (18)$$

Let H denote the Hilbert norm space. Kernel matrices are generated for K . The variable y represents the input vector. In parameter optimization, the Ridge method aims to determine the optimal values of α and λ based on a given set of parameters. The variable \tilde{x} represents the set of unknown sample points. Equations 16–18 represent the linear, polynomial, and Gaussian kernels, viable options for utilization inside the Ridge algorithm.

2.2.8. K-Nearest Neighbours Regression (KNR)

The K-Nearest Neighbours Regression (KNR) approach is a non-parametric method that does not rely on any assumptions about the data distribution. The underlying principle of this approach is to utilize benchmarks for similar data points (Ho and Yu, 2022). The KNR algorithm calculates the distance between a new data point and all the points in the training dataset. It subsequently selects the k-nearest neighbours and makes predictions based on their corresponding target values (Li et al., 2017). As equation (19) demonstrated, the Euclidean distance is a frequently employed metric for measuring distance.

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (19)$$

Let p represent the number of input features. Let x_{ik} denote the value of the k^{th} input feature for the i^{th} observation, and x_{jk} represent the value of the k^{th} feature for the j^{th} observation.

2.2.9. Least Absolute Shrinkage and Selection Operator (LASSO)

The “Lasso” (Least Absolute Shrinkage and Selection Operator), sometimes referred to as L1 regularization, is a method employed to mitigate overfitting by including the absolute values of the weights into the regression objective function (Zhang et al., 2019; Shi et al., 2020). The LASSO regression analysis was conducted utilizing descriptors obtained by implementing the genetic algorithm (GA). The LASSO method seeks to minimize the coefficients of certain model variables while simultaneously forcing other coefficients to be precisely zero. Moreover, the LASSO method combines the benefits of subset selection and ridge regression. As Equation (20) indicates, LASSO aims to minimize the Mean Squared Error (MSE), thereby reducing the number of predictors necessary for model development. The user has provided a numerical reference (Datta et al., 2017). The determination of the final set of coefficients (b^{LASSO}) occurs at the point of minimum mean squared error (MSE), as indicated by Equation (21). The user provided a numerical reference (Datta et al., 2017). In this context, the user-defined constant λ (lambda) is employed. The frequency of descriptors with a coefficient of zero exhibits a positive correlation with the magnitude of λ .

$$\min \sum_{i=1}^N (y_i - \sum b_j x_{ij})^2 + \lambda \sum |b_j| \quad (20)$$

$$b^{\text{LASSO}} = \text{argmin} \|y - xb\|_2^2 + \lambda \|b\|_1 \quad (21)$$

2.2.10. Support vector regression (SVR)

The Support Vector Regression (SVR) model identifies the optimal value within a predefined threshold, representing the distance between the hyperplane and the boundary line. In contrast, conventional regression algorithms minimize the discrepancy between the observed and predicted values (Rezaei et al., 2023). The modification described below enhances the SVR algorithm, rendering it a powerful regularization technique that relies on kernel methods. This adjustment allows users to adjust the acceptable error margin and tolerance rate, thus increasing the flexibility of the SVR algorithm. It has been observed that this phenomenon leads to a decrease in both the confidence interval and the empirical risk. The computational complexity of SVR is not affected by the size of the input space. Previous studies have shown evidence of the extraordinary performance of SVR, especially when dealing with limited sample sizes and extensive feature spaces. Moreover, it exhibits a high regression accuracy level and exceptional generalizability. The SVR can be used to find the best line or hyperplane in higher dimensions that best fits the data and decide the acceptable amount of model error. The proposed approach facilitates the preservation of relationships while moving data from an n -dimensional space to an $n + 1$ -dimensional space, utilizing a kernel trick as an assisting mechanism. Within this context, the hyperplane that encompasses the maximum quantity of data points is identified, and afterward, decision boundaries are constructed while allowing for a tolerable margin of error.

The model aims to minimize the mathematical function denoted by Equation (22).

$$\text{minimize} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n |\xi_i|, \text{subject to } \|R_{\text{train}} - \hat{r}\|_2^2 \leq \varepsilon + |\xi_i| \quad (22)$$

$$\hat{r} = X\beta \quad (23)$$

As mentioned previously, X represents the matrix of input variables, while β is the parameter vector of the regression model. The symbols ε and ξ represent the margin error and slack variables, respectively. Another hyperparameter, denoted as C , determines the tolerance for points that fall outside the range of ε .

2.3. Model development and configuration

As previously stated, the energy efficiency and exergy efficiency of

PTSC technology were predicted according to the experimental dataset collected from the literature. The dataset included three different synthetic oils, including Therminol VP-1, Syltherm 800, and Dowtherm Q, combined with three different nanomaterials, namely Al_2O_3 , CuO , and SiO_2 , in different volume fractions. Fig. 4 shows the overall workflow of the AI process.

All features (input parameters) were standardized during the Exploratory Data Analysis (EDA) and Feature Engineering stages. This process involves centering the features around their mean value and scaling them to have a standard variance using the StandardScaler technique. The dataset was divided into a training set with 70 % and a testing set with 30 %. After that, nine models were developed, namely ETR, XGBoost, RFR, CART, AdaBoost, Ridge, KNR, Lasso, and SVR in order to predict the outputs. In order to select the optimum models, the Leave-One-Out (LOO) cross-validation technique was implemented employing four different regression metrics: Adjusted-R², Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

Following that, four models were selected for the next step, namely (ETR, XGBoost, RFR, and CART) due to their regression accuracy. In the next step, the actual regression performance in predicting the two outputs was evaluated through the best four models using untrained data. Meanwhile, in the last stage, Business Insights were discussed because sustainable technologies leveraging AI for predictive modeling in PTSC can gain a competitive edge by offering more reliable and efficient solutions.

2.4. Models performance evaluation metrics

The assessment of prediction error rates and model efficiency in regression analysis mainly depends on the use of various performance metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (Coefficient of Determination) (Ibrahim et al., 2023; Jamei et al., 2023; Karbasi et al., 2023; Yaseen, 2023). The regression measurements indicated above can be expressed mathematically in the following equations.

$$MAE = \frac{1}{N} \times \sum_{i=1}^N |y_i - \hat{y}| \quad (24)$$

$$MSE = \frac{1}{N} \times \sum_{i=1}^N (y_i - \hat{y})^2 \quad (25)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \times \sum_{i=1}^N (y_i - \hat{y})^2} \quad (26)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (27)$$

In the given equations, y_i is the predicted value of energy and exergy and \hat{y} is the experimental value of energy and exergy.

3. Results and discussion

3.1. Model selection process

In this part, a comparative analysis will be conducted to evaluate the performance of nine algorithms in predicting the energy and exergy efficiency of PTSC. Four regression metrics will be used to determine the superior algorithm. According to Figs. 5–7, the ensemble methods such as ETR, XGBoost, RFR, CART, and AdaBoost demonstrate excellent performance with high Adj_R² scores, suggesting robust fits to the dataset. Meanwhile, linear regression-based techniques such as (Ridge and Lasso) and non-linear methods such as (KNR and SVR) display varying goodness of fit. Notably, the SVR model demonstrates the least promising performance among the models mentioned.

The study analyzes prominent models in predicting energy and exergy efficiency, explicitly focusing on Therminol VP-1-based nanofluids, as depicted in Fig. 5(a and b). The top models for energy efficiency in the case of base fluid (Therminol VP-1) are ETR (0.9846), XGBoost (0.9833), RFR (0.9818), and CART (0.9783). Moreover, the top models in predicting exergy efficiency include ETR with a score of 0.9853, XGBoost with a score of 0.9838, RFR with a score of 0.9796, and CART with a score of 0.9748. In the case of Therminol VP-1-Al₂O₃, outstanding models include ETR (0.9998), XGBoost (0.999), RFR (0.9989), and CART (0.9976), demonstrating high accuracy in predicting the energy efficiency of PTSC. Meanwhile, the most efficient models in predicting the exergy efficiency are XGBoost (0.9948), ETR (0.994), RFR (0.9888), and CART (0.9814). In the case of Therminol VP-1-CuO, the top models for energy efficiency include ETR with a coefficient of determination (R²) of 0.9996, XGBoost with a R² of 0.9987, RFR with a R² of 0.9985, and CART with a R² of 0.9966. Also, the exergy efficiency prediction demonstrates a similar trend in the ETR model (0.9944), XGBoost model (0.9936), RFR model (0.9892), and CART model

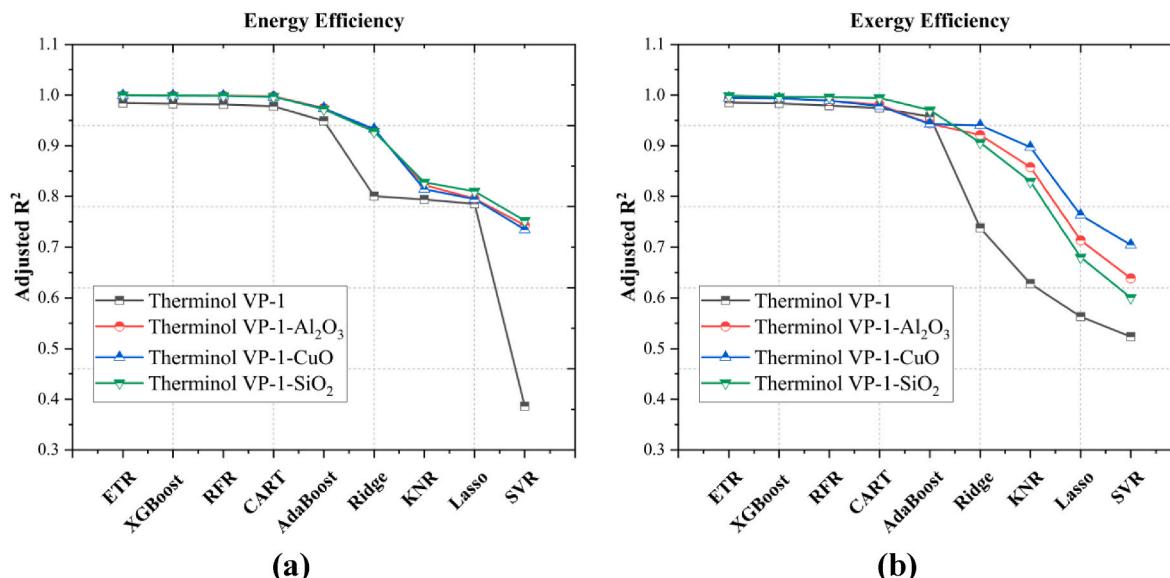


Fig. 5. Adjusted R² of Therminol VP-1-based nanofluids: (a) Energy efficiency, (b) Exergy efficiency.

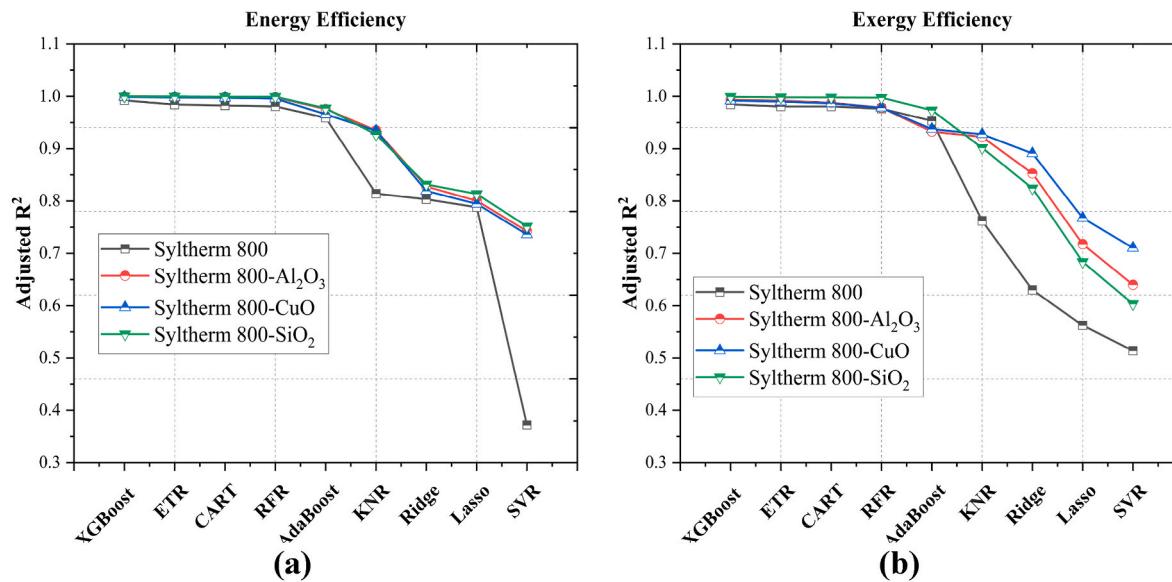


Fig. 6. Adjusted R² of Syltherm 800-based nanofluids: (a) Energy efficiency, (b) Exergy efficiency.

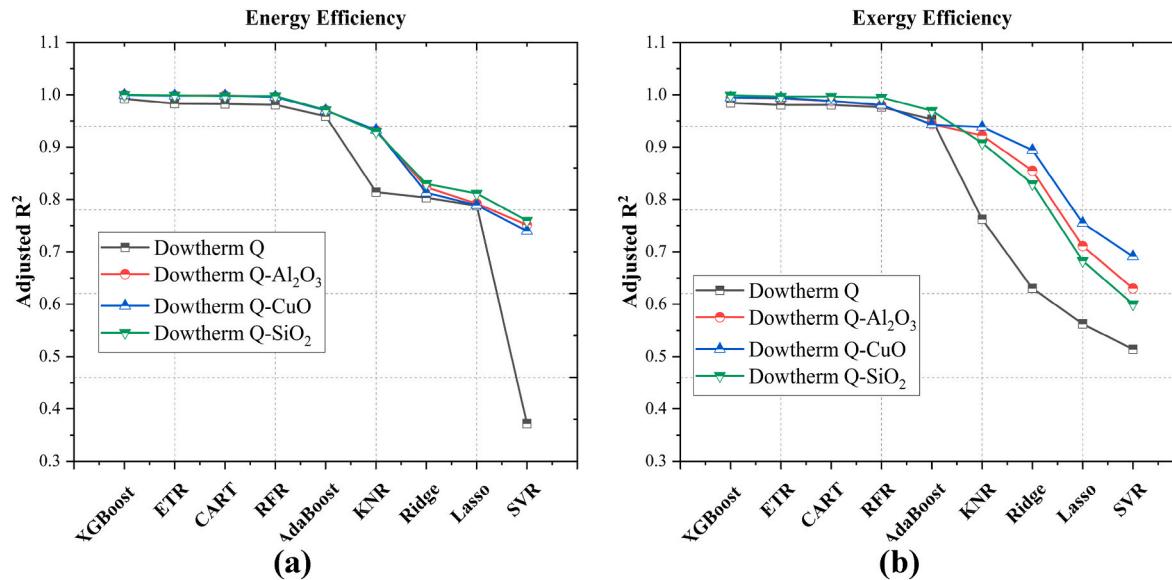


Fig. 7. Adjusted R² of Dowtherm Q-based nanofluids: (a) Energy efficiency, (b) Exergy efficiency.

(0.9786). Finally, the employment of Therminol VP-1-SiO₂ has resulted in the development of significant energy efficiency models, namely ETR (0.9999), RFR (0.999), XGBoost (0.9985), and CART (0.997). Furthermore, the exergy efficiency exhibits a similar trend to the ETR model (0.9988), RFR model (0.9965), XGBoost model (0.9961), and CART model (0.9944).

The investigation examines prominent models for predicting the energy and exergy efficiency of PTSC using Syltherm 800-based nanofluids, as depicted in Fig. 6(a and b). The superior models for energy efficiency in the case of base fluid (Syltherm 800) are XGBoost (0.9923), ETR (0.9844), CART (0.9824), and RFR (0.9809). The regression of exergy efficiency indicates that, ETR (0.9845), RFR (0.9807), XGBoost (0.9807), and CART (0.9763) exhibit the highest predicting performance. In the case of Syltherm 800-Al₂O₃, notable models that exhibit superior accuracy in energy efficiency include ETR (0.9997), RFR (0.9991), CART (0.9988), and XGBoost (0.9984). Meanwhile, XGBoost (0.9932), ETR (0.9918), RFR (0.9877), and CART (0.9785) exhibit superior performance in predicting the exergy efficiency. In the case of

Syltherm 800-CuO, the top models for predicting the energy efficiency include ETR with a coefficient of determination (R-squared) value of 0.9989, XGBoost with a value of 0.9977, RFR with a value of 0.9972, and CART with a value of 0.9957. In the meantime, the exergy efficiency regression demonstrates a similar trend, such as XGBoost (0.9918), ETR (0.99), RFR (0.9866), and CART (0.9782). Finally, in the case of Syltherm 800-SiO₂, superior models for predicting energy efficiency are ETR (1), RFR (0.9998), XGBoost (0.9993), and CART (0.9993). At this point, the regression accuracy in exergy efficiency is observed to be consistent with the values of ETR (0.9992), XGBoost (0.9981), RFR (0.9979), and CART (0.9974).

The study examines prominent approaches for predicting the energy and exergy efficiency of PTSC, as in Fig. 7(a and b), employing Dowtherm Q-based nanofluids. In the case of Dowtherm Q, the top models for predicting energy efficiency are XGBoost (0.9923), ETR (0.9833), CART (0.9824), and RFR (0.9809). At the same time, the top models for exergy efficiency prediction are ETR with a score of 0.9845, RFR with a score of 0.9807, XGBoost with a score of 0.9807, and CART with a score

of 0.9763. In the case of Dowtherm Q-Al₂O₃, superior models that exhibit high accuracy in energy efficiency include ETR (with a coefficient of determination of 0.9998), XGBoost (with a coefficient of determination of 0.9986), RFR (with a coefficient of determination of 0.9986), and CART (with a coefficient of determination of 0.9963). Regarding exergy efficiency, the superior models are ETR with a value of 0.9945, XGBoost with 0.9931, RFR with 0.9878, and CART with 0.9801. In the case of Dowtherm Q-CuO, the superior models predicting energy efficiency include ETR with a coefficient of determination (R^2) value of 0.9995, XGBoost with R^2 value of 0.9982, RFR with R^2 value of 0.9981, and CART with R^2 value of 0.9958. Then, the prediction of exergy efficiency demonstrates a similar trend in the regression metrics of ETR (0.9946), XGBoost (0.994), RFR (0.9881), and CART (0.9807). Finally, in the case of Dowtherm Q-SiO₂, the top models for predicting energy efficiency include ETR (0.9999), RFR (0.9989), XGBoost (0.9975), and CART (0.9974). Also, the accuracy in exergy efficiency is observed to be consistent with the values of ETR (0.999), XGBoost (0.9966), RFR (0.9965), and CART (0.9947).

Figures 8–10 display the Regression Error in predicting the energy and exergy efficiency of PTSC using synthetic oils named (Therminol VP-1, Syltherm 800, and Dowtherm Q) mixed with different nanoparticles (Al₂O₃, CuO, and SiO₂). As per Fig. 8(a and b), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Therminol VP-1 were as (ETR: MSE = 3.2753, RMSE = 0.9274, MAE = 1.8098), (SVR: MSE = 130.3882, RMSE = 6.9635, MAE = 11.4188) and (ETR: MSE = 0.6286, RMSE = 0.5019, MAE = 0.7928), (SVR: MSE = 20.3707, RMSE = 2.4923, MAE = 4.5134). As per Fig. 8(c and d), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Therminol VP-1-Al₂O₃ were as (ETR: MSE = 0.0461, RMSE = 0.0954, MAE = 0.2148), (SVR: MSE = 57.0604, RMSE = 4.1065, MAE = 7.5538) and (XGBoost: MSE = 0.2014, RMSE = 0.3542, MAE = 0.4488), (Lasso: MSE = 13.9209, RMSE = 2.6918, MAE = 3.7311). As per Fig. 8(e and f), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Therminol VP-1-CuO were as (ETR: MSE = 0.09, RMSE = 0.1434, MAE = 0.3), (SVR: MSE = 58.9152, RMSE = 4.272, MAE = 7.6756) and (ETR: MSE = 0.2591, RMSE = 0.3234, MAE = 0.509), (Lasso: MSE = 13.588, RMSE = 2.6242, MAE = 3.6862). As per Fig. 8(g and h), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Therminol VP-1-SiO₂ were as (ETR: MSE = 0.0153, RMSE = 0.0477, MAE = 0.1239), (SVR: MSE = 55.0929, RMSE = 3.9083, MAE = 7.4225) and (ETR: MSE = 0.0404, RMSE = 0.1409, MAE = 0.201), (Lasso: MSE = 13.834, RMSE = 2.7349, MAE = 3.7194).

As per Fig. 9(a and b), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Syltherm 800 were as (XGBoost: MSE = 1.6633, RMSE = 0.8912, MAE = 1.2897), (SVR: MSE = 134.8796, RMSE = 7.0418, MAE = 11.6138) and (ETR: MSE = 0.6677, RMSE = 0.5185, MAE = 0.8171), (SVR: MSE = 20.8682, RMSE = 2.5211, MAE = 4.5682). As per Fig. 9(c and d), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Syltherm 800-Al₂O₃ were as (ETR: MSE = 0.0637, RMSE = 0.1218, MAE = 0.2525), (SVR: MSE = 57.8321, RMSE = 4.1332, MAE = 7.6047) and (XGBoost: MSE = 0.2692, RMSE = 0.3805, MAE = 0.5188), (Lasso: MSE = 14.2863, RMSE = 2.7975, MAE = 3.7797). As per Fig. 9(e and f), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Syltherm 800-CuO were as (ETR: MSE = 0.2513, RMSE = 0.2047, MAE = 0.5013), (SVR: MSE = 59.2233, RMSE = 4.2703, MAE = 7.6957) and (XGBoost: MSE = 0.3914, RMSE = 0.4522, MAE = 0.6256), (Lasso: MSE = 13.8728, RMSE = 2.7487, MAE = 3.7246). As per Fig. 9(g and h), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Syltherm 800-SiO₂ were as (ETR: MSE = 0.0069, RMSE = 0.0349, MAE = 0.0832), (SVR: MSE = 55.8206, RMSE = 3.9754, MAE = 7.4713) and (ETR: MSE = 0.0305, RMSE = 0.1297, MAE = 0.1746), (Lasso: MSE = 14.3449, RMSE = 2.8337, MAE = 3.7875).

As per Fig. 10(a and b), the lowest and highest Regression Errors in

predicting the energy and exergy efficiency using Dowtherm Q were as (XGBoost: MSE = 1.6633, RMSE = 0.8912, MAE = 1.2897), (SVR: MSE = 134.8796, RMSE = 7.0418, MAE = 11.6138) and (ETR: MSE = 0.6649, RMSE = 0.5183, MAE = 0.8154), (SVR: MSE = 20.8682, RMSE = 2.5211, MAE = 4.5682). As per Fig. 10(c and d), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Dowtherm Q-Al₂O₃ were as (ETR: MSE = 0.0532, RMSE = 0.1119, MAE = 0.2307), (SVR: MSE = 55.8433, RMSE = 4.0632, MAE = 7.4728) and (ETR: MSE = 0.2112, RMSE = 0.324, MAE = 0.4596), (Lasso: MSE = 14.2922, RMSE = 2.7366, MAE = 3.7805). As per Fig. 10(e and f), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Dowtherm Q-CuO were as (ETR: MSE = 0.1131, RMSE = 0.1606, MAE = 0.3363), (SVR: MSE = 58.4406, RMSE = 4.2879, MAE = 7.6446) and (ETR: MSE = 0.2474, RMSE = 0.3299, MAE = 0.4974), (Lasso: MSE = 14.0878, RMSE = 2.6887, MAE = 3.7534). As per Fig. 10(g and h), the lowest and highest Regression Errors in predicting the energy and exergy efficiency using Dowtherm Q-SiO₂ were as (ETR: MSE = 0.0122, RMSE = 0.0476, MAE = 0.1103), (SVR: MSE = 54.0752, RMSE = 3.8471, MAE = 7.3536) and (ETR: MSE = 0.034, RMSE = 0.1345, MAE = 0.1844), (Lasso: MSE = 14.0842, RMSE = 2.7613, MAE = 3.7529).

3.2. Model validation

Consistent with the information provided in the previous section (3.1), this section includes the four most well-known algorithms recognized for their ability to balance high accuracy and minor regression errors. The algorithms in this context determine the most efficient performance on a dataset that has not been seen before.

Tables (1–3) presented the outcomes of model validation of energy and exergy efficiency. These validations were conducted by employing various oil-based nanofluids produced by combining three different oils (Therminol VP-1, Syltherm 800, and Dowtherm Q) with three different metallic oxides nanomaterials (Al₂O₃, CuO, SiO₂). The validation models employed in this study involve (ETR), (CART), (XGBoost), and (RFR) models, which showed superior regression efficiency in the previous part. Performance examination utilizes the following metrics: Test_Adj_R², Test_MSE, Test_RMSE, and Test_MAE.

The models provide robust prediction abilities for both energy and exergy efficiency, as demonstrated by high Adjusted R-squared values approaching one and low error measurements such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The models' effectiveness depends on the specific nano-fluid composition, the underlying fluid medium, and the nanoparticle additive employed. Particular combinations may yield superior performance for specific models compared to others. It has been observed that, the ETR and XGBoost models exhibit a strong performance across various nanofluids and additives. High Adjusted R-squared values and low error metrics support this. Furthermore, the RFR model exhibits excellent performance, although perhaps displaying marginally higher error metrics than ETR and XGBoost in specific scenarios. CART models typically demonstrate relatively poorer performance than others, characterized by slightly lower Adjusted R-squared values and higher error metrics. The selection of metallic oxides and base fluid (oil) influences the model's performance, as certain combinations result in superior predicted accuracy. In conclusion, the findings of this study indicate that the use of AI models provides a means of accurately predicting the energy and exergy efficiencies of PTSC using different types of nano oil additives as cooling mediums.

3.3. Business insights

Tables (4–6) report business insights into energy and exergy efficiencies of PTSC using different types of nano oil additives as cooling mediums. These tables exhibit the average energy and exergy values, prediction errors, and the corresponding percentage errors for each

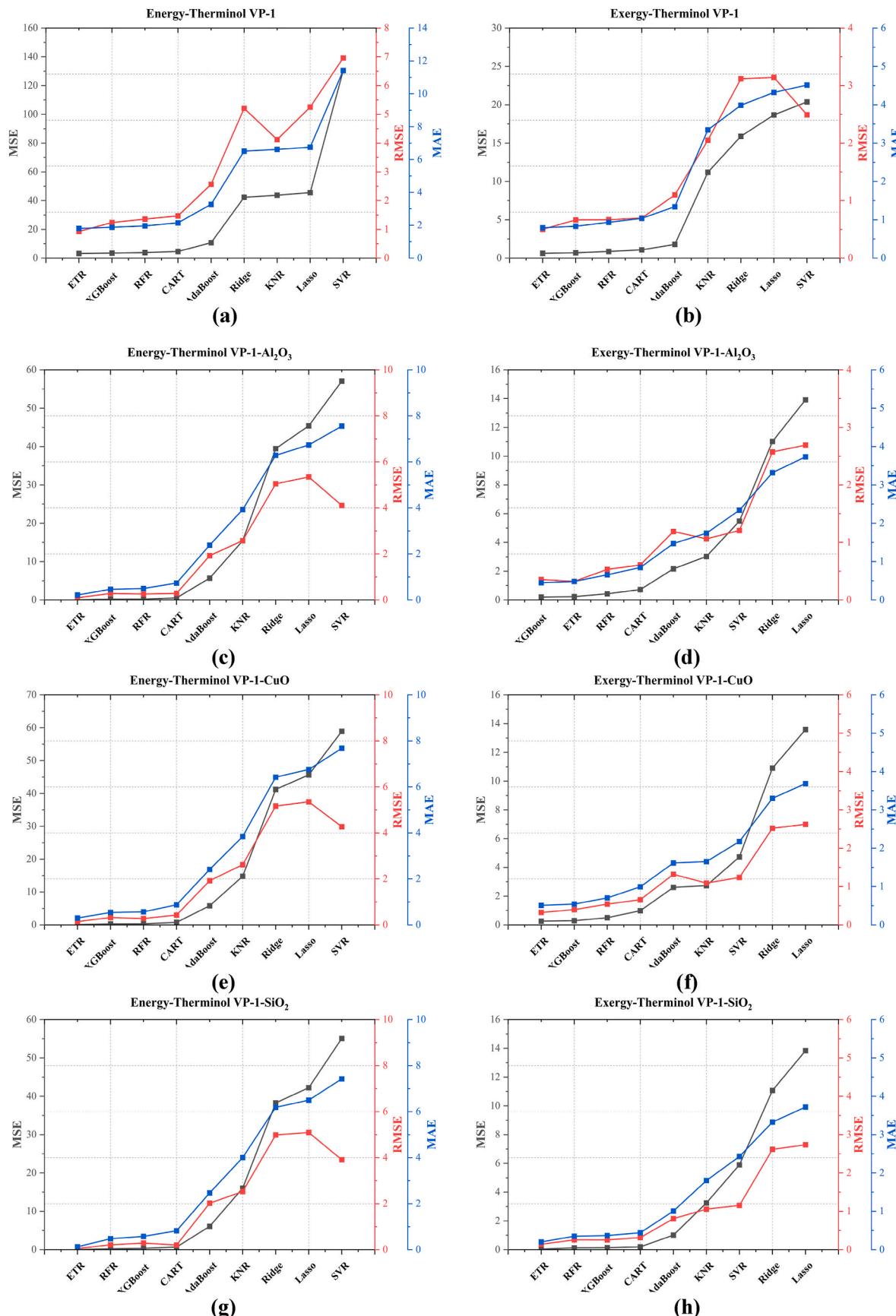


Fig. 8. MSE, RMSE and MAE of Therminol VP-1-based nanofluids; (a) Energy-Therminol VP-1, (b) Exergy-Therminol VP-1, (c) Energy-Therminol VP-1-Al₂O₃, (d) Exergy-Therminol VP-1-Al₂O₃, (e) Energy-Therminol VP-1-CuO, (f) Exergy-Therminol VP-1-CuO, (g) Energy-Therminol VP-1-SiO₂, (h) Exergy-Therminol VP-1-SiO₂.

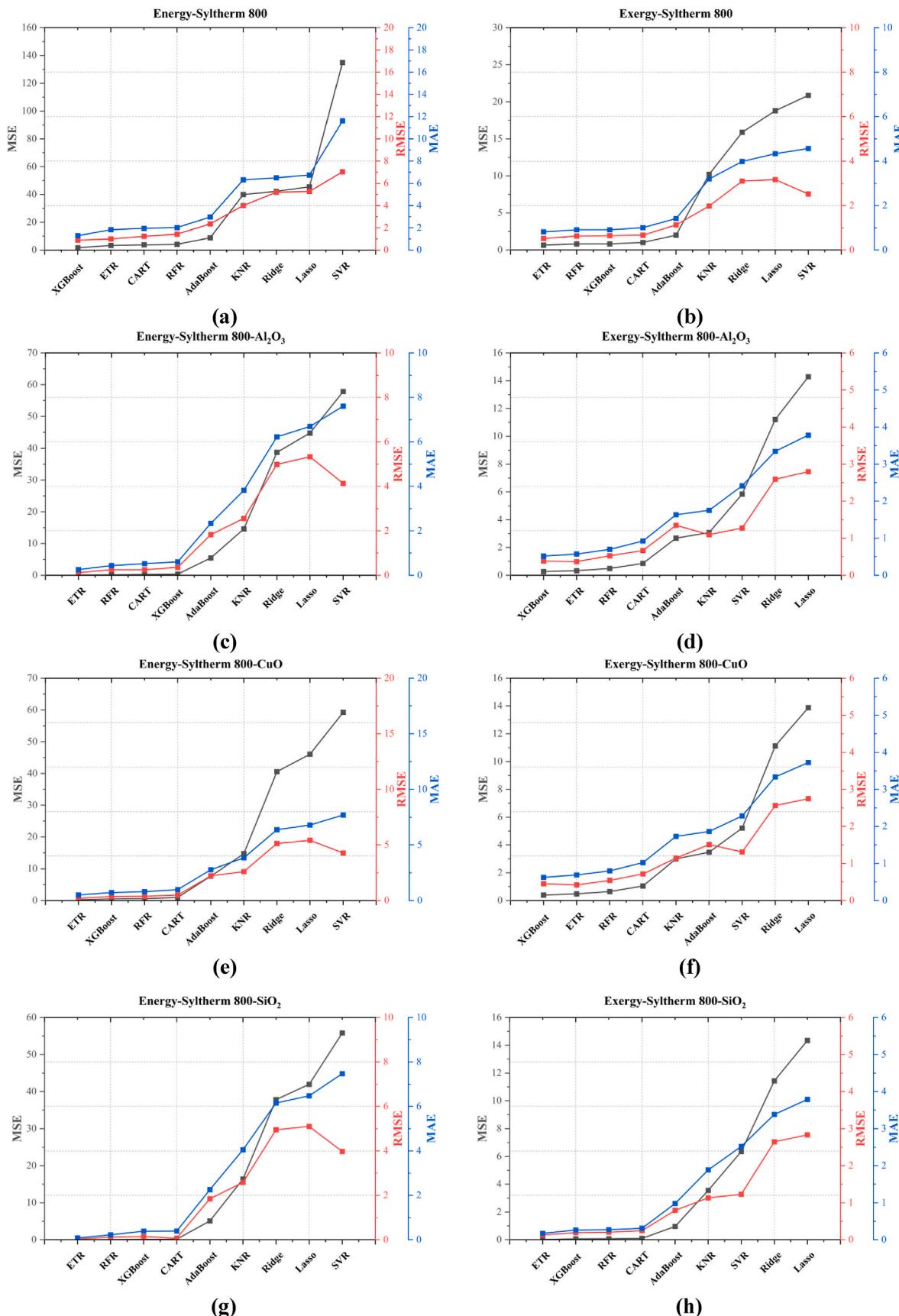


Fig. 9. MSE, RMSE and MAE of Syltherm 800-based nanofluids; (a) Energy- Syltherm 800, (b) Exergy- Syltherm 800, (c) Energy-Syltherm 800-Al₂O₃, (d) Exergy- Syltherm 800-Al₂O₃, (e) Energy- Syltherm 800-CuO, (f) Exergy- Syltherm 800-CuO, (g) Energy- Syltherm 800-SiO₂, (h) Exergy- Syltherm 800-SiO₂.

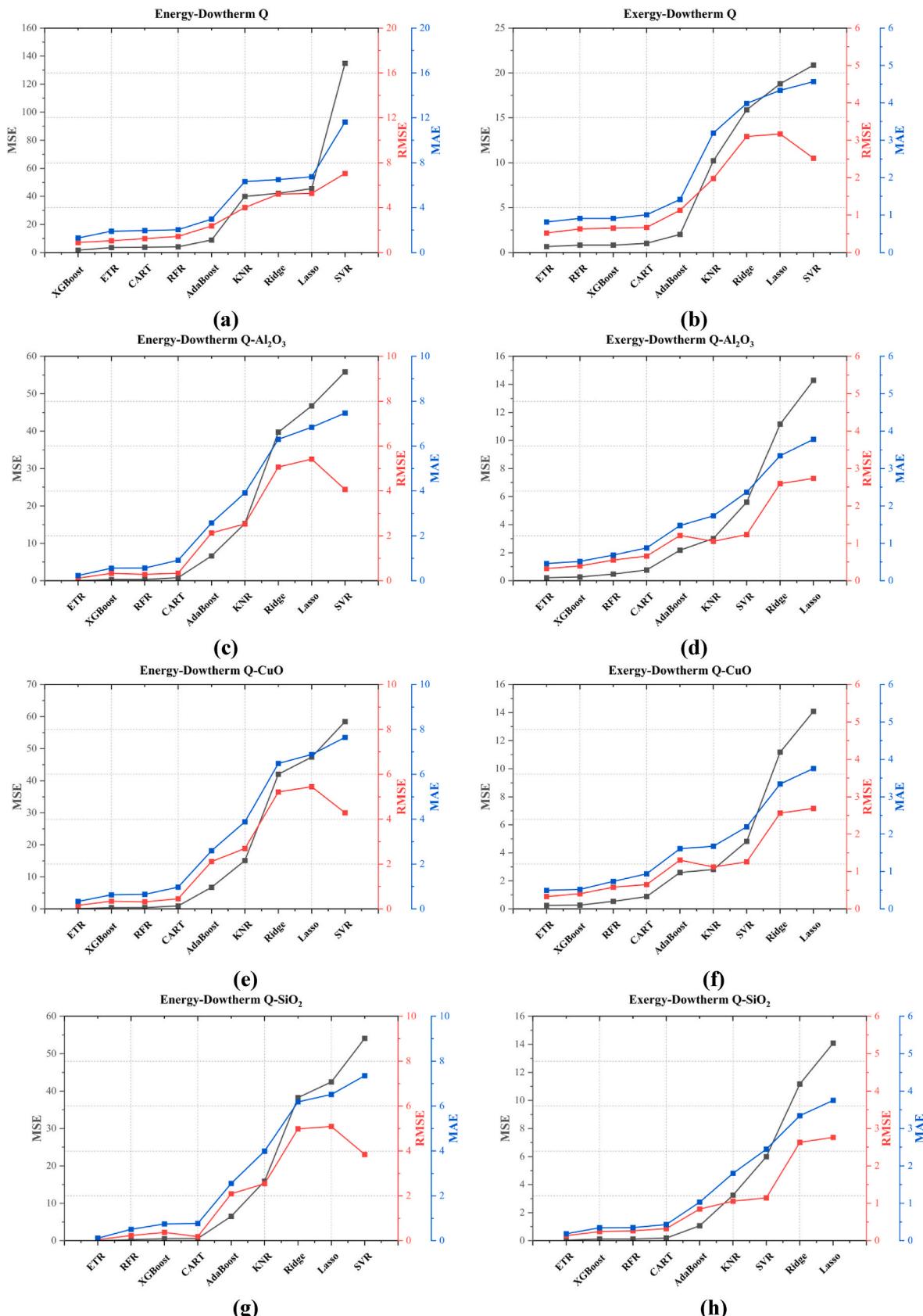


Fig. 10. MSE, RMSE and MAE of Dowtherm Q-based nanofluids; (a) Energy- Dowtherm Q, (b) Exergy- Dowtherm Q, (c) Energy- Dowtherm Q-Al₂O₃, (d) Exergy- Dowtherm Q-Al₂O₃, (e) Energy- Dowtherm Q-CuO, (f) Exergy- Dowtherm Q-CuO, (g) Energy- Dowtherm Q-SiO₂, (h) Exergy-Dowtherm Q-SiO₂.

Table 1

Model validation of energy and exergy efficiencies using Therminol VP-1-based nanofluids.

Energy-Therminol VP-1	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.982	1.6472	1.2834	0.7666
CART	0.9749	2.2912	1.5137	1.1546
XGBoost	0.9586	3.7853	1.9456	1.2862
RFR	0.9453	5.0057	2.2373	1.865
Exergy-Therminol VP-1	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
XGBoost	0.9746	0.4084	0.6391	0.5109
ETR	0.9619	0.6123	0.7825	0.5516
RFR	0.9463	0.8633	0.9291	0.6688
CART	0.941	0.9489	0.9741	0.723
Energy-Therminol VP-1-Al ₂ O ₃	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9989	0.2388	0.4886	0.1856
XGBoost	0.9984	0.3526	0.5938	0.3142
RFR	0.9973	0.5775	0.76	0.376
CART	0.9953	1.0121	1.006	0.4061
Exergy-Therminol VP-1-Al ₂ O ₃	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
XGBoost	0.9911	0.2816	0.5307	0.4003
ETR	0.989	0.3472	0.5893	0.3775
RFR	0.983	0.5368	0.7327	0.588
CART	0.9684	0.998	0.999	0.693
Energy-Therminol VP-1-CuO	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9986	0.3021	0.5496	0.2376
RFR	0.9976	0.5114	0.7151	0.3364
CART	0.9976	0.5173	0.7192	0.4028
XGBoost	0.9975	0.5507	0.7421	0.3904
Exergy-Therminol VP-1-CuO	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9895	0.3716	0.6096	0.3755
XGBoost	0.9879	0.4271	0.6535	0.4977
RFR	0.9825	0.6155	0.7845	0.5745
CART	0.9556	1.5642	1.2507	0.7704
Energy-Therminol VP-1-SiO ₂	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
CART	0.9999	0.0179	0.1337	0.071
ETR	0.9998	0.0398	0.1994	0.081
RFR	0.9987	0.283	0.532	0.2925
XGBoost	0.9978	0.4624	0.68	0.3683
Exergy-Therminol VP-1-SiO ₂	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9983	0.0522	0.2285	0.1662
XGBoost	0.9961	0.1246	0.353	0.2848
RFR	0.9933	0.2118	0.4603	0.326
CART	0.9893	0.3374	0.5809	0.4168

working fluid combination. According to the results of different nano-oils, the prediction errors for energy and exergy efficiencies are generally minor and vary between 0.05 and 0.76. Also, the error fraction is relatively low, most falling from 0.11 % to 2.44 %. This indicates that, the predictive AI models efficiently track business developments, emphasizing cleaner and sustainable production processes. These models provide crucial insights for data-driven decision-making, enhancing overall environmental responsibility and operational efficiency by pinpointing opportunities for energy reduction, emission control, and waste management. Metallic oxide nanomaterials, such as Al₂O₃, CuO, and SiO₂, exhibit varying impacts on efficiency predictions, as evidenced by the differences in prediction errors among different working fluid-additive combinations.

Regarding absolute error values, SiO₂ tends to lead to slightly smaller prediction errors than other additives in several cases. The average energy and exergy values obtained from the models closely align with the

Table 2

Model validation of energy and exergy efficiencies using Syltherm 800-based nanofluids.

Energy-Syltherm 800	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9858	1.3117	1.1453	0.6992
CART	0.9799	1.8532	1.3613	0.9248
XGBoost	0.9704	2.7282	1.6517	1.2356
RFR	0.9558	4.0722	2.018	1.6424
Exergy-Syltherm 800	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
XGBoost	0.9813	0.2866	0.5354	0.405
ETR	0.9753	0.3778	0.6147	0.3973
CART	0.9677	0.4938	0.7027	0.5158
RFR	0.9547	0.6929	0.8324	0.5811
Energy-Syltherm 800-Al ₂ O ₃	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9988	0.2719	0.5215	0.2248
RFR	0.9981	0.4141	0.6435	0.3131
XGBoost	0.9972	0.6119	0.7822	0.3895
CART	0.997	0.6643	0.815	0.3452
Exergy-Syltherm 800-Al ₂ O ₃	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
XGBoost	0.9908	0.3045	0.5518	0.3983
ETR	0.9835	0.5486	0.7407	0.4472
RFR	0.9809	0.6335	0.7959	0.5816
CART	0.9707	0.9731	0.9865	0.6951
Energy-Syltherm 800-CuO	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
XGBoost	0.9966	0.7421	0.8614	0.4356
ETR	0.9964	0.8041	0.8967	0.3536
CART	0.9948	1.1461	1.0706	0.5192
RFR	0.9941	1.3066	1.1431	0.4764
Exergy-Syltherm 800-CuO	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9798	0.7652	0.8748	0.5057
XGBoost	0.9791	0.7899	0.8888	0.6322
RFR	0.9714	1.0846	1.0414	0.6667
CART	0.9571	1.6244	1.2745	0.7797
Energy-Syltherm 800-SiO ₂	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
CART	0.9999	0.0145	0.1203	0.0569
XGBoost	0.9999	0.0248	0.1576	0.1081
ETR	0.9998	0.0336	0.1832	0.0711
RFR	0.9994	0.1212	0.3482	0.1903
Exergy-Syltherm 800-SiO ₂	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9983	0.0566	0.2379	0.1571
XGBoost	0.9974	0.0858	0.2929	0.2273
RFR	0.9971	0.098	0.3131	0.2234
CART	0.9942	0.1951	0.4417	0.3355

actual average values, indicating the models' ability to generalize well to new data. These insights can be valuable for industries aiming to optimize nanofluid compositions for enhanced energy and exergy efficiencies, as the models' accurate predictions offer practical guidance.

4. Conclusion

The researchers utilized tree-based, linear, and non-linear regression models to predict the energy and exergy efficiency of PTSCs. Three synthetic oils commonly were used as base fluids, namely Therminol VP-1, Syltherm 800, and Dowtherm Q. Then, the oils were mixed with different metallic oxide nanomaterials, namely Al₂O₃, CuO, and SiO₂, in different volume fractions. A total of nine algorithms were trained to choose the superior model. After that, four algorithms were selected for evaluation on a previously unseen dataset. Moreover, business evaluation was applied for optimizing the oil-based nanofluids to improve

Table 3

Model validation of energy and exergy efficiencies using Dowtherm Q-based nanofluids.

Energy-Dowtherm Q	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9871	1.1902	1.091	0.6205
CART	0.9799	1.8532	1.3613	0.9248
XGBoost	0.9704	2.7282	1.6517	1.2356
RFR	0.9558	4.0722	2.018	1.6424
Exergy-Dowtherm Q	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
XGBoost	0.9813	0.2866	0.5354	0.405
ETR	0.9757	0.3722	0.6101	0.3767
CART	0.9677	0.4938	0.7027	0.5158
RFR	0.9547	0.6929	0.8324	0.5811
Energy-Dowtherm Q-Al ₂ O ₃	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9987	0.2814	0.5305	0.2045
XGBoost	0.9971	0.6266	0.7916	0.4206
RFR	0.9971	0.6331	0.7957	0.4324
CART	0.9957	0.9267	0.9626	0.3496
Exergy-Dowtherm Q-Al ₂ O ₃	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
XGBoost	0.9903	0.3094	0.5562	0.4354
ETR	0.986	0.4471	0.6686	0.4566
RFR	0.9804	0.627	0.7918	0.6375
CART	0.9751	0.7964	0.8924	0.6633
Energy-Dowtherm Q-CuO	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
XGBoost	0.9988	0.267	0.5168	0.34
ETR	0.9981	0.4197	0.6478	0.2668
RFR	0.997	0.6553	0.8095	0.4249
CART	0.9958	0.9175	0.9579	0.4867
Exergy-Dowtherm Q-CuO	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9878	0.4289	0.6549	0.4095
XGBoost	0.9839	0.5682	0.7538	0.5613
RFR	0.9814	0.657	0.8106	0.6034
CART	0.9684	1.1135	1.0552	0.7253
Energy-Dowtherm Q-SiO ₂	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
CART	0.9998	0.049	0.2215	0.0963
ETR	0.9995	0.0993	0.3152	0.1176
RFR	0.9984	0.3394	0.5826	0.3179
XGBoost	0.9951	1.0724	1.0356	0.4526
Exergy-Dowtherm Q-SiO ₂	Test_Adj_R ²	Test_MSE	Test_RMSE	Test_MAE
ETR	0.9988	0.0386	0.1965	0.1442
XGBoost	0.996	0.1289	0.359	0.2686
RFR	0.9936	0.2087	0.4569	0.3414
CART	0.9934	0.2153	0.464	0.3733

energy and exergy efficiency. The primary findings of this investigation are summarized as follows:

- Direct Solar Irradiance always showed the higher SHAP value for both energy and exergy outputs such as (7.35 and 3.79), (9.33 and 3.83), (9.28 and 4.08) and (9.30 and 4.20) for Therminol VP-1, Therminol VP-1-CuO, Therminol VP-1-Al₂O₃, and Therminol VP-1-SiO₂, respectively. (7.52 and 3.92), (9.65 and 4.08), (9.55

and 4.28) and (9.56 and 4.42) for Syltherm 800, Syltherm 800-CuO, Syltherm 800-Al₂O₃ and Syltherm 800-SiO₂, respectively. (7.52 and 3.92), (9.22 and 3.80), (9.17 and 4.04) and (9.18 and 4.17), Dowtherm Q, Dowtherm Q-CuO, Dowtherm Q-Al₂O₃ and Dowtherm Q-SiO₂

- In the model selection process, Ensemble methods like ETR, XGBoost, RFR, CART, and AdaBoost consistently showed strong performance with high R² scores, indicating robust fits. In contrast, linear regression (Ridge and Lasso) and non-linear methods (KNN and SVR) exhibited varying goodness of fit. Notably, SVR performed less promisingly among these models.
- The maximum accuracy in predicting the energy and exergy outputs was CART = 0.9999 and ETR = 0.9983 using Therminol VP-1-SiO₂. Similarly, the maximum accuracy in predicting the energy and exergy outputs was CART = 0.9998 and ETR = 0.9983 using Syltherm 800-SiO₂. Additionally, the maximum accuracy in predicting the energy and exergy outputs was CART = 0.9999 and ETR = 0.9983 using Dowtherm Q-SiO₂.
- The lower regression error metrics in predicting the energy and exergy outputs were (CART: MSE = 0.0179, RMSE = 0.1337, MAE = 0.071) and (ETR: MSE = 0.0522, EMSE = 0.2285, MAE = 0.1662) using Therminol VP-1-SiO₂. Similarly, the lower regression error metrics in predicting the energy and exergy outputs were (CART: MSE = 0.0145, RMSE = 0.1203, MAE = 0.0569) and (ETR: MSE = 0.0566, EMSE = 0.2379, MAE = 0.1571) using Syltherm 800-SiO₂. Additionally, the lower regression error metrics in predicting the energy and exergy outputs were (CART: MSE = 0.049, RMSE = 0.2215, MAE = 0.0963) and (ETR: MSE = 0.0386, EMSE = 0.1965, MAE = 0.1442) using Dowtherm Q-SiO₂.
- In the business insights, the maximum errors in the energy and exergy models were observed like (1.43 % and 1.97 %) using Therminol VP-1, (1.3 % and 2.44 %) using Syltherm 800 and Syltherm 800-CuO and (1.15 % and 2 %) using Dowtherm Q and Dowtherm Q-CuO, respectively, reinforcing the reliability of these models for cleaner production, environmental, and sustainability research and applications.

5. Limitations and future recommendations

While the current study provides valuable insights into the problems of predicting the energy and exergy efficiency of PTSCs technologies through the utilization of various oil-based nanofluids, it is essential to acknowledge several limitations that may impact the interpretation and generalizability of the findings. i) the size of data collected was small, and the method was limited to a set time range, thus disregarding seasonal variations that could affect energy estimates. ii), the current study focused on various oil-based nanofluids, but the scope might be limited to specific nanofluid formulations. iii) additional influential factors that could affect the energy and exergy efficiency might have been omitted from the analysis. This could include external environmental conditions, variations in operational parameters, or specific design characteristics not considered in the current study. iv), the study might be confined to a specific domain, and the applicability of the findings to different PTSCs technologies or related fields might be unclear.

The following recommendations outline potential avenues for future research to address the identified limitations and propel our

Table 4

Business insights of energy and exergy efficiencies using Therminol VP-1-based nanofluids.

Working fluid	Energy average value	Energy prediction error	error %	Exergy average value	Exergy prediction error	error %
Therminol VP-1	53.5632	0.7666	1.43 %	25.9608	0.5109	1.97 %
Therminol VP-1-Al₂O₃	54.0042	0.1856	0.34 %	22.0033	0.4003	1.82 %
Therminol VP-1-CuO	54.1854	0.2376	0.44 %	20.3689	0.3755	1.84 %
Therminol VP-1-SiO₂	53.9177	0.071	0.13 %	22.7704	0.1662	0.73 %

Table 5

Business insights of energy and exergy efficiencies using Syltherm 800-based nanofluids.

Working fluid	Energy average value	Energy prediction error	error %	Exergy average value	Exergy prediction error	error %
Syltherm 800	53.9031	0.6992	1.30 %	26.0452	0.405	1.55 %
Syltherm 800-Al ₂ O ₃	54.1191	0.2248	0.42 %	22.3485	0.3983	1.78 %
Syltherm 800-CuO	54.1198	0.4356	0.80 %	20.7421	0.5057	2.44 %
Syltherm 800-SiO ₂	54.0899	0.0569	0.11 %	23.1759	0.1571	0.68 %

Table 6

Business insights of energy and exergy efficiencies using Dowtherm Q-based nanofluids.

Working fluid	Energy average value	Energy prediction error	error %	Exergy average value	Exergy prediction error	error %
Dowtherm Q	53.9031	0.6205	1.15 %	26.045	0.405	1.55 %
Dowtherm Q-Al ₂ O ₃	53.7063	0.2045	0.38 %	22.0233	0.4354	1.98 %
Dowtherm Q-CuO	53.9147	0.34	0.63 %	20.4446	0.4095	2.00 %
Dowtherm Q-SiO ₂	53.6041	0.0963	0.18 %	22.7937	0.1442	0.63 %

understanding of integrating AI into renewable energy technologies. i) Future research should consider a broader range of nanofluid formulations, including different base oils and nanoparticles, ii) Recognizing the trade-off between model complexity and interpretability, future studies should explore more advanced modeling techniques, iii) Factors such as external environmental conditions, variations in operational parameters, and specific design characteristics should be incorporated to provide a more holistic analysis, iv) Future studies should conduct comprehensive sensitivity analyses and quantify uncertainties, v) Future research should explore the transferability of predictive models to different domains within PTSCs technologies or related fields.

CRediT authorship contribution statement

Hai Tao: Writing – original draft, Visualization, Validation, Supervision, Investigation, Formal analysis, Data curation, Conceptualization. **Omer A. Alawi:** Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Raad Z. Homod:** Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation, Formal analysis, Data curation. **Mustafa KA. Mohammed:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Leonardo Goliatt:** Writing – original draft, Visualization, Validation, Software, Resources, Data curation, Conceptualization. **Hussein Togun:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Shafik S. Shafik:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Salim Heddam:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Zaher Mundher Yaseen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Investigation, Formal analysis, Data curation, 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.

Data availability

Data will be made available on request.

Acknowledgment

The authors very much thankful for the reviewers and editors'

comments. In addition, the first author would like to acknowledge the support received by the Key Research and Scientific Research Capability Improvement Project in Qiannan Normal University for Nationalities (2024zdzk06). Further, Zaher Mundher Yaseen would like to thanks the Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Saudi Arabia for its support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclepro.2024.141069>.

References

- Abubakr, M., Amein, H., Akoush, B.M., El-Bakry, M.M., Hassan, M.A., 2020. An intuitive framework for optimizing energetic and exergetic performances of parabolic trough solar collectors operating with nanofluids. *Renew. Energy* 157, 130–149.
- Ajbar, W., Parrales, A., Cruz-Jacobo, U., Conde-Gutiérrez, R., Bassam, A., Jaramillo, O., Hernández, J., 2021a. The multivariable inverse artificial neural network combined with GA and PSO to improve the performance of solar parabolic trough collector. *Appl. Therm. Eng.* 189, 116651.
- Ajbar, W., Parrales, A., Huicochea, A., Hernández, J., 2022. Different ways to improve parabolic trough solar collectors' performance over the last four decades and their applications: a comprehensive review. *Renew. Sustain. Energy Rev.* 156, 111947.
- Ajbar, W., Parrales, A., Silva-Martínez, S., Bassam, A., Jaramillo, O., Hernández, J., 2021b. Identification of the relevant input variables for predicting the parabolic trough solar collector's outlet temperature using an artificial neural network and a multiple linear regression model. *J. Renew. Sustain. Energy* 13 (4).
- Al-Rashed, A.A., Alnaqi, A.A., Alsarraf, J., 2021. Numerical investigation and neural network modeling of the performance of a dual-fluid parabolic trough solar collector containing non-Newtonian water-CMC/Al2O3 nanofluid. *Sustain. Energy Technol. Assessments* 48, 101555.
- Ali, M., Deo, R.C., Maraseni, T., Downs, N.J., 2019. Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. *J. Hydrol.* 576, 164–184.
- Alizamir, M., Kim, S., Kisi, O., Zounemat-Kermani, M., 2020. A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: case studies of the USA and Turkey regions. *Energy* 197, 117239.
- An, S., Liu, W., Venkatesh, S., 2007. Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. *Pattern Recogn.* 40 (8), 2154–2162.
- Bellos, E., Tzivanidis, C., Antonopoulos, K., Gkinis, G., 2016. Thermal enhancement of solar parabolic trough collectors by using nanofluids and converging-diverging absorber tube. *Renew. Energy* 94, 213–222.
- Bilali, A.E., Taleb, A., Bahlaoui, M.A., Brouziye, Y., 2021. An integrated approach based on Gaussian noises-based data augmentation method and AdaBoost model to predict faecal coliforms in rivers with small dataset. *J. Hydrol.* 599, 126510.
- Brédy, J., Gallichand, J., Celicourt, P., Gumiere, S.J., 2020. Water table depth forecasting in cranberry fields using two decision-tree-modeling approaches. *Agric. Water Manag.* 233, 106090.
- Breiman, L., 2001. Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat. Sci.* 16 (3), 199–231.
- Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. *Cart. Classification and Regression Trees*.
- Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

- Cui, J., Li, W., Fang, C., Su, S., Luan, J., Gao, T., Hu, L., Lu, Y., Chen, G., 2019. Adaboost ensemble correction models for TDDFT calculated absorption energies. *IEEE Access* 7, 38397–38406.
- Dastorani, M.T., Mahjoobi, J., Talebi, A., Fakhar, F., 2018. Application of machine learning approaches in rainfall-runoff modeling (case study: zayandeh Rood basin in Iran). *Civ. Eng. Infrastruct. J.* 51 (2), 293–310.
- Datta, S., Dev, V.A., Eden, M.R., 2017. Developing QSPR for predicting DNA drug binding affinity of 9-Amininoacridine derivatives using correlation-based adaptive LASSO algorithm. *Computer Aided Chemical Engineering* 2767–2772. Elsevier.
- Daviran, S., Krammer, A., Schüller, A., 2019. In-situ and post annealing effect on the microstructure and the optical properties of black Cu-Co-Mn oxide spinel coating for Parabolic Trough Collector (PTC) applications. *J. Phys. Conf.*, 012200. IOP Publishing.
- Ebrahimi-Moghadam, A., Mohseni-Gharyehsafa, B., Farzaneh-Gord, M., 2018. Using artificial neural network and quadratic algorithm for minimizing entropy generation of Al2O3-EG/W nanofluid flow inside parabolic trough solar collector. *Renew. Energy* 129, 473–485.
- El Bilali, A., Taleb, A., Brouziyne, Y., 2021. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agric. Water Manag.* 245, 106625.
- Elsheikh, A.H., Sharshir, S.W., Abd Elaziz, M., Kabeel, A.E., Guilan, W., Haiou, Z., 2019. Modeling of solar energy systems using artificial neural network: a comprehensive review. *Sol. Energy* 180, 622–639.
- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., Xiang, Y., 2018. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agric. For. Meteorol.* 263, 225–241.
- Feigl, M., Lebiedzinski, K., Herrnegger, M., Schulz, K., 2021. Machine-learning methods for stream water temperature prediction. *Hydrol. Earth Syst. Sci.* 25 (5), 2951–2977.
- Ferreira, L.B., da Cunha, F.F., 2020. New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. *Agric. Water Manag.* 234, 106113.
- Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* 55 (1), 119–139.
- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Mach. Learn.* 63, 3–42.
- Gholizadeh, M., Jamei, M., Ahmadianfar, I., Pourrajab, R., 2020. Prediction of nanofluids viscosity using random forest (RF) approach. *Chemometr. Intell. Lab. Syst.* 201, 104010.
- Ghritlahre, H.K., Prasad, R.K., 2018. Application of ANN technique to predict the performance of solar collector systems-A review. *Renew. Sustain. Energy Rev.* 84, 75–88.
- Heddam, S., Ptak, M., Zhu, S., 2020. Modelling of daily lake surface water temperature from air temperature: extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN. *J. Hydrol.* 588, 125130.
- Heng, S.Y., Asako, Y., Suwa, T., Nagasaka, K., 2019. Transient thermal prediction methodology for parabolic trough solar collector tube using artificial neural network. *Renew. Energy* 131, 168–179.
- Ho, W., Yu, F., 2022. Measurement and verification of energy performance for chiller system retrofit with k nearest neighbour regression. *J. Build. Eng.* 46, 103845.
- Ibrahim, H., Yaseen, Z.M., Scholz, M., Ali, M., Gad, M., Elsayed, S., Khadr, M., Hussein, H., Ibrahim, H.H., Eid, M.H., 2023. Evaluation and prediction of groundwater quality for irrigation using an integrated water quality indices, machine learning models and GIS approaches: a representative case study. *Water* 15 (4), 694.
- Jamei, M., Ali, M., Malik, A., Karbasi, M., Sharma, E., Yaseen, Z.M., 2022a. Air quality monitoring based on chemical and meteorological drivers: application of a novel data filtering-based hybridized deep learning model. *J. Clean. Prod.* 374, 134011.
- Jamei, M., Karbasi, M., Alawi, O.A., Kamar, H.M., Khedher, K.M., Abba, S., Yaseen, Z.M., 2022b. Earth skin temperature long-term prediction using novel extended Kalman filter integrated with Artificial Intelligence models and information gain feature selection. *Sustainable Computing: Informatics and Systems* 35, 100721.
- Jamei, M., Karbasi, M., Ali, M., Malik, A., Chu, X., Yaseen, Z.M., 2023. A novel global solar exposure forecast model based on air temperature: designing a new multi-processing ensemble deep learning paradigm. *Expert Syst. Appl.*, 119811.
- Jamei, M., Karbasi, M., Olumegbon, I.A., Mosharaf-Dehkordi, M., Ahmadianfar, I., Asadi, A., 2021. Specific heat capacity of molten salt-based nanofluids in solar thermal applications: a paradigm of two modern ensemble machine learning methods. *J. Mol. Liq.* 335, 116434.
- Kaood, A., Abubakr, M., Al-Oran, O., Hassan, M.A., 2021. Performance analysis and particle swarm optimization of molten salt-based nanofluids in parabolic trough concentrators. *Renew. Energy* 177, 1045–1062.
- Karbasi, M., Jamei, M., Malik, A., Kisi, O., Yaseen, Z.M., 2023. Multi-steps drought forecasting in arid and humid climate environments: development of integrative machine learning model. *Agric. Water Manag.* 281, 108210.
- Kottala, R.K., Balasubramanian, K., Jinshan, B., Divakar, S., Chigilipalli, B.K., 2023. Experimental investigation and machine learning modelling of phase change material-based receiver tube for natural circulated solar parabolic trough system under various weather conditions. *J. Therm. Anal. Calorim.* 1–24.
- Li, W., Kong, D., Wu, J., 2017. A novel hybrid model based on extreme learning machine, k-nearest neighbor regression and wavelet denoising applied to short-term electric load forecasting. *Energies* 10 (5), 694.
- Liang, H., Zheng, C., Zheng, W., You, S., Zhang, H., 2017. Analysis of annual performance of a parabolic trough solar collector. *Energy Proc.* 105, 888–894.
- Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 30.
- Manikandan, G., Iniyar, S., Goic, R., 2019. Enhancing the optical and thermal efficiency of a parabolic trough collector-A review. *Appl. Energy* 235, 1524–1540.
- Mustafa, J., Alqaed, S., Sharifpur, M., 2022. Numerical study on performance of double-fluid parabolic trough solar collector occupied with hybrid non-Newtonian nanofluids: investigation of effects of helical absorber tube using deep learning. *Eng. Anal. Bound. Elem.* 140, 562–580.
- Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., Liu, J., 2020. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *J. Hydrol.* 586, 124901.
- Okonkwo, E.C., Adun, H., Babatunde, A.A., Abid, M., Ratlamwala, T.A., 2020. Entropy generation minimization in a parabolic trough collector operating with SiO₂-water nanofluids using the genetic algorithm and artificial neural network. *J. Therm. Sci. Eng. Appl.* 12 (3), 031007.
- Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., Mohammadian, A.K., 2020. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accid. Anal. Prev.* 136, 105405.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Phyo, P.P., Jeenanunta, C., 2021. Daily load forecasting based on a combination of classification and regression tree and deep belief network. *IEEE Access* 9, 152226–152242.
- Rezaei, I., Amirshahi, S.H., Mahbadi, A.A., 2023. Utilizing support vector and kernel ridge regression methods in spectral reconstruction. *Results in Optics* 11, 100405.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144.
- Ruiz-Moreno, S., Gallego, A.J., Camacho, E.F., 2023. Artificial neural network-based fault detection and isolation in a parabolic-trough solar plant with defocusing strategy. *Sol. Energy* 262, 111909.
- Seyyedattar, M., Ghiasi, M.M., Zendehboudi, S., Butt, S., 2020. Determination of bubble point pressure and oil formation volume factor: Extra trees compared with LSSVM-CSA hybrid and ANFIS models. *Fuel* 269, 116834.
- Shapley, L.S., 1953. A Value for N-Person Games.
- Shi, X., Wang, K., Cheong, T.S., Zhang, H., 2020. Prioritizing driving factors of household carbon emissions: an application of the LASSO model with survey data. *Energy Econ.* 92, 104942.
- Sikorska-Senor, A.E., Quilty, J.M., 2021. A novel ensemble-based conceptual-data-driven approach for improved streamflow simulations. *Environ. Model. Software* 143, 105094.
- Singh, K.P., Gupta, S., Rai, P., 2013. Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmos. Environ.* 80, 426–437.
- Smetti, E., Thanasioulias, N., Lytras, E., Tzoumerkas, P., Golfinopoulos, S., 2009. Treated water quality assurance and description of distribution networks by multivariate chemometrics. *Water Res.* 43 (18), 4676–4684.
- Strumbelj, E., Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* 41, 647–665.
- Tao, H., Salih, S., Oudah, A.Y., Abba, S., Ameen, A.M.S., Awadh, S.M., Alawi, O.A., Mostafa, R.R., Surendran, U.P., Yaseen, Z.M., 2022. Development of new computational machine learning models for longitudinal dispersion coefficient determination: case study of natural streams, United States. *Environ. Sci. Pollut. Control Ser.* 29 (24), 35841–35861.
- Torres-Barrán, A., Alonso, Á., Dorronsoro, J.R., 2019. Regression tree ensembles for wind energy and solar radiation prediction. *Neurocomputing* 326, 151–160.
- Tsai, C.-C., Lu, M.-C., Wei, C.-C., 2012. Decision tree-based classifier combined with neural-based predictor for water-stage forecasts in a river basin during typhoons: a case study in taiwan. *Environ. Eng. Sci.* 29 (2), 108–116.
- Ture, M., Kurt, I., Kurum, A.T., Ozdamar, K., 2005. Comparing classification techniques for predicting essential hypertension. *Expert Syst. Appl.* 29 (3), 583–588.
- Vakili, M., Salehi, S.A., 2023. A review of recent developments in the application of machine learning in solar thermal collector modelling. *Environ. Sci. Pollut. Control Ser.* 30 (2), 2406–2439.
- Wehenkel, L., Ernst, D., Geurts, P., 2006. Ensembles of Extremely Randomized Trees and Some Generic Applications, Robust Methods for Power System State Estimation and Load Forecasting.
- Xiao, C., Chen, N., Hu, C., Wang, K., Gong, J., Chen, Z., 2019. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Rem. Sens. Environ.* 233, 111358.
- Xu, C., Liu, X., Wang, H., Li, Y., Jia, W., Qian, W., Quan, Q., Zhang, H., Xue, F., 2021. A study of predicting irradiation-induced transition temperature shift for RPV steels with XGBoost modeling. *Nucl. Eng. Technol.* 53 (8), 2610–2615.
- Yaseen, Z.M., 2023. Machine learning models development for shear strength prediction of reinforced concrete beam: a comparative study. *Sci. Rep.* 13 (1), 1723.
- You, Y., Demmel, J., Hsieh, C.-J., Vuduc, R., 2018. Accurate, fast and scalable kernel ridge regression on parallel and distributed systems. In: Proceedings of the 2018 International Conference on Supercomputing, pp. 307–317.
- Zaaoumi, A., Bah, A., Ciocan, M., Sebastian, P., Balan, M.C., Mechaqra, A., Alaoui, M., 2021. Estimation of the energy production of a parabolic trough solar thermal power plant using analytical and artificial neural networks models. *Renew. Energy* 170, 620–638.
- Zhang, T., Shi, X., Zhang, D., Xiao, J., 2019. Socio-economic development and electricity access in developing economies: a long-run model averaging approach. *Energy Pol.* 132, 223–231.
- Zhang, W., Wu, C., Zhong, H., Li, Y., Wang, L., 2021. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geosci. Front.* 12 (1), 469–477.