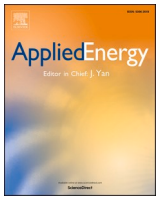
Applied Energy 364 (2024) 123130

Contents lists available at ScienceDirect

Applied Energy

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



Neural network-based surrogate modeling and optimization of a multigeneration system

Parviz Ghafariasl [a], Alireza Mahmoudan [b], Mahmoud Mohammadi [c], Aria Nazarparvar [d], Siamak Hoseinzadeh [e,*], Mani Fathali [c], Shing Chang [a], Masoomeh Zeinalnezhad [f], Davide Astiaso Garcia [e]

[a] Department of Industrial and Manufacturing Systems Engineering, Kansas State University, Manhattan, KS 66506, USA
[b] Energy Efficiency Group, Institute for Environmental Sciences and Forel Institute, University of Geneva, Boulevard Carl-Vogt 66, 1205 Genève, Switzerland
[c] Department of Aerospace Engineering, K.N. Toosi University of Technology, Tehran, Iran
[d] Flight faculty, Georgian Aviation University (GAU), 16 Ketevan Dedopli Ave., Tbilisi 0103, Georgia
[e] Department of Planning, Design, and Technology of Architecture, Sapienza University of Rome, Via Flaminia 72, 00196 Rome, Italy
[f] Department of Industrial Engineering, West Tehran Branch, Islamic Azad University, Tehran, Iran

HIGHLIGHTS

- ANN, CNN, and LSTM models were tuned for a multigeneration system.
- A composition of ANN and LSTM formed the best surrogate model.
- Six-objective optimization was performed at four solar irradiation values.
- All objectives were improved compared to the simulation-based optimization.

ARTICLE INFO

Keywords:
Surrogate modeling
Artificial neural network (ANN)
Long-short term memory (LSTM)
Convolutional neural network (CNN)
Multi-objective optimization
Multigeneration system

ABSTRACT

Multi-Objective Optimization (MOO) poses a computational challenge, particularly when applied to physics-based models. As a result, only up to three objectives are typically involved in simulation-based optimization. To go beyond this number, Surrogate Models (SMs) need to replace such high-fidelity models. In this exploratory study, the objectives are to perform comprehensive regression surrogate modeling and to conduct MOO for a Multi-Generation System (MGS). The most suitable SM was chosen among four neural-network models: Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM), and an ensemble model developed through brute-force search using the three aforementioned models. The final model was found to be superior to others, achieving R^2 values ranging from 0.9830 to 0.9999. Next, an optimization problem with six conflicting objectives was defined and performed at four distinct values of Direct Normal Irradiation (*DNI*), a time-dependent feature. This aimed to provide multi-criteria decision-making information based on atmospheric transparency. As a result, new understandings were gained: (I) exergy efficiency, production cost, and freshwater production rate were found to be highly influenced by *DNI*, and (II) the critical range of operation was observed within the *DNI* interval of 100 to 400 W/m^2. Furthermore, we compared the result of the six-objective optimization with that of the bi-objective optimization obtained in our simulation-based study and found that all objectives showed improvements ranging from 1.9% to 12.7%. Finally, based on the findings obtained in the present study, some practical recommendations were put forward for applying the proposed methodology to similar MGSs.

1. Introduction

Computer simulation of engineering processes is critical in the design

* Corresponding author.
E-mail addresses: siamak.hosseinzadeh@uniroma1.it, hoseinzadeh.siamak@gmail.com (S. Hoseinzadeh).

https://doi.org/10.1016/j.apenergy.2024.123130
Received 22 January 2024; Received in revised form 13 March 2024; Accepted 27 March 2024

stage of such problems. Taking the energy systems field as an example, this step can shed light on the weaknesses and strengths of the system in question in terms of performance, economic, and environmental efficiencies [1–5]. Simulations based on mathematical models, however, might require high computational resources depending on the degree of complexity associated with the problem. This issue can particularly manifest itself when multi-objective optimization (MOO), another key step in the design phase, has to be undertaken [6]. While new generations of computers have become fast and powerful, they are still incapable of perfectly meeting the demands of growing CPU-intensive calculations.

Alternatively, surrogate models (SMs) have been introduced to replace high-fidelity models with much less computationally expensive ones. Otherwise known as metamodeling or reduced-order modeling, surrogate modeling is a process that approximates underlying mathematical calculations by mapping the relationship between input-output data extracted from the original simulation. It is also a simple tool for exploring the solution space. In addition to polynomial-based SMs that are effective for linear/first-order problems, there are other surrogate models that rely on supervised machine learning techniques. These machine learning-based SMs can handle more complex data and are flexible enough to capture non-linear relationships, making them suitable for a wider range of problems. In this category, the common methods are as follows: Kriging, Radial Basis Function (RBF), inverse distance weighting, least squares, Support Vector Regression (SVR), and Artificial Neural Networks (ANN) [7]. There are various computer packages available for surrogate modeling. The Python libraries Scikit-learn and TensorFlow have been developed for constructing SVR and ANN models, respectively, while the Surrogate Modeling Toolbox developed by [8] can be used to construct the remaining models.

There are advantages and disadvantages to each of the above models. Regarding Kriging, its development is less time-consuming and the resulting surrogate model is highly flexible and suitable for problems with high dimensionality. However, on the downside, this model does not perform properly when the distance among sample points is small. Even though this issue can also be seen in RBF models, they are simple and fast for small datasets. The inverse distance weighting method requires no training process, which makes it easy-to-use, but it is at the expense of having low efficiency. As for the least squares, they are fast, simple, and very accurate for linear problems, but not a good choice for nonlinear ones. On the other hand, both linear and nonlinear problems even with high dimensionality can be successfully modeled by SVR. This technique requires a time-consuming training process, making it a compromise between accuracy and high dimensionality. Concerning the final method, ANN is the most commonly utilized model in a wide variety of research fields for its high accuracy and ability to treat problems with nonlinear structures. The significant shortcoming of this technique is, however, the need for a large-size dataset. Besides, while a high number of neurons is required to secure high accuracy, it must be executed with extra caution to prevent overfitting in solution [9–11].

A few essential steps for regression and time-series black-box (BB) modeling have been detected by Rätz et al. [12] through reviewing the relevant literature thoroughly. Here, we will exclusively focus on the steps that are pertinent to regression surrogate modeling. The following items provide a framework that the surrogate modeling process presented in this study is based upon:

- **Data Preprocessing**: this concept can be divided into three major steps: (I) scaling and normalizing, (II) handling missing data, and (III) identifying and treating outliers. In order to mitigate the effects of bias in results, the former process should be applied to features. In other words, the impact of a single feature with a large value may outweigh the collective effect of multiple features with much smaller values. The incident of missing data usually occurs when the data is

Nomenclature

DNI	Direct Normal Irradiation (W/m^2)
Eff_{exe}	Exergy efficiency (%)
$\dot{m}$	Mass flowrate (kg/s)
MAE	Mean absolute error (−)
MSE	Mean squared error (−)
n	Number of observations (−)
N_{hel}	Number of heliostats (−)
P	Pressure (kPa)
$\dot{Q}$	Heat transfer rate (kW)
R^2	R-squared (−)
$RMSE$	Root mean squared error (−)
RSS	Residual sum of squares (−)
T	Temperature (°C)
TSS	Total sum of squares (−)
$TUCP$	Total unit cost of products ($/GJ)
$\dot{W}_{net}$	Net electricity production (kW)
x	Data point
x'	Z-score value (−)
y	Actual values
$\hat{y}$	Predicted values
Z_{TM}	Figure of merit

Subscripts

air	Air
c	Cooling
fw	Freshwater
geo	Geothermal
h	Heating
j	Numerator
ORC	Organic Rankine Cycle
$Rank$	Rankine

Greek

η_{exe}	Exergy efficiency (%)
μ	Mean value
σ	Standard deviation

Abbreviation

1D	One dimensional
ANN	Artificial neural network
BB	Black box
CNN	Convolutional neural network
EES	Engineering equation solver
FGP	Forecasting Gaussian process
LSTM	Long-short term memory
MGS	Multi-generation system
MOO	Multi-objective optimization
NSGA	Non-dominated sorting genetic algorithm
ORC	Organic Rankine Cycle
RBF	Radial basis function
RNN	Recurrent neural network
SM	Surrogate model
SVR	Support vector regression

collected manually or by sensors. Since this does not occur in software simulation, it cannot be considered a challenge in surrogate modeling. As for outliers, the presence of extreme values in a dataset can lead to a skewed distribution, possibly undermining the accuracy of the model. However, the definition of an outlier in a particular problem can significantly affect the number of data points, potentially leading to an inaccurate model [13]. Thus, outliers must be treated properly.

- **Feature selection**: some features are of redundant/irrelevant nature or have minimal effect on labels. Thus, eliminating these features not only produces no adverse impact on the final results but may also accelerate the modeling process and reduce runtime. In this regard, three common methods can be used: filter, embedded, and wrapper. Feature selection is rather implemented to less sophisticated BB models, such as random forest, decision tree, SVR, etc. However, for more sophisticated methods, such as neural networks, where feature weighing is already embedded, reducing the number of features would inevitably compromise the system's overall performance [14].
- **Hyperparameter optimization**: this is a search for tailoring machine learning algorithms to the specific problem being addressed [15]. All BB models have hyperparameters and they can be set either manually (by trial and error), automatically (using grid search), randomly, or through more advanced methods (such as the Bayesian optimization algorithm). The objective is to tune hyperparameters to secure high-quality SM [16].
- **Model selection**: depending on the inherent characteristics of a problem and its associated data, different BB models can be selected [17]. Choosing an appropriate SM for a given problem requires considering a few key questions including: does the dataset have a linear or nonlinear characteristic? Is the problem regression or time-series? How much control over hyperparameters is needed? What degree of accuracy is demanded? How computationally efficient does the model have to be? And so on.
- **Model evaluation**: After training a model, model evaluation has to be conducted to figure out its fidelity. This can also help in prioritizing surrogate models in terms of performance. There are metrics available for measuring the accuracy of models that are suitable for regression problems, such as Mean Absolute Error (*MAE*), Mean Squared Error (*MSE*), Root Mean Squared Error (*RMSE*), R-squared (R^2), and so forth [18].

As mentioned earlier, one strong motive for developing surrogate models is to make the multi-objective optimization process feasible. Due to the complexity and nonlinearity inherent in real-world problems, the mathematical models representing them may also exhibit intricacies, possibly causing increased computational costs for MOO. There are various challenges in the energy sector that MOO must be employed to address. To name a few examples: balancing the energy consumption of systems that utilize different renewable energy sources; determining optimal design points for systems with conflicting objectives, such as cost, efficiency, and environmental impact; and, optimally scheduling systems with various objectives and multiple operating processes [19,20]. In such cases, using white-box models would exacerbate the computational process. As a solution, SMs can come into play to surmount this hurdle by offering computationally cheap approximations. Consequently, surrogate-based optimization has become increasingly popular among several engineering disciplines in recent years. This is owing to two facts: (I) locating global optimal solutions is feasible via SMs, and (II) coding machine learning/black-box algorithms has become quite manageable through open-source programming [21].

It can be anticipated that optimizing energy systems poses a computational dilemma, especially when several production rates should be taken into account as objectives along with other performance criteria. Hence, if the configuration of an energy system transitions from cogeneration to trigeneration or from trigeneration to multigeneration, one would expect a corresponding transformation in the level of difficulty associated with the optimization process [22]. The present study seeks to investigate the surrogate modeling of a Multi-Generation System (MGS) according to the common framework discussed earlier, as well as to perform surrogate-based optimization. To delve deeply into the chosen topic, the following literature review exclusively focuses on studies related to energy conversion systems that yield more than one product.

1.1. Literature review

Since energy systems can be assessed from various perspectives (such as energy, exergy, emergy, economic, environmental, and production size), different objective functions would be available to choose from for carrying out a MOO process. The presence of numerous objectives inevitably compels researchers to employ reduced-order models to ensure optimization feasibility [23]. In recent years, countless studies have utilized SMs merely for the optimization of cogeneration/trigeneration/multigeneration energy systems. In these studies, a machine learning algorithm, mainly an ANN, is trained to construct a meta-model with the intention of performing MOO efficiently. A number of such investigations are presented in the following paragraph:

Bahlawan et al. [24] employed life cycle energy and economic methods and developed an RBF-based surrogate model to optimize a MGS fueled by solar energy and natural gas. It was reported that the optimal results obtained from the MOO saved life cycle energy and costs by 17% and 18%, respectively. Also, they showed that making use of a surrogate model reduced the computation time by 78%. To optimize two cogeneration systems producing net electricity and cooling load, Zhou et al. [25] developed two SMs using Forecasting Gaussian Process (FGP) and SVR algorithms. In more detail, the SVR was developed to model the total cost rate, while the FGP was utilized for energy efficiency. The results demonstrated that through the MOO, a 65% increase in energy efficiency and a 16.32 $/h reduction in the total cost rate could be observed. Assareh et al. [26] evaluated an MGS from thermodynamic and economic standpoints and defined a two-objective optimization problem. They generated 700 random data points using EES to develop an ANN model. To analyze the surrogate model, the predictive accuracy and an error histogram were reported. The outcome of the MOO exhibited that the proposed system could attain 32.39% exergy efficiency, a 34.32 $/GJ cost rate, and an 1140 kW net electricity production rate.

Note that these examples represent only a fraction of the extensive body of research available in the literature. However, these investigations mainly focused on thermodynamic, economic, and/or environmental modeling and hardly ever presented any details about the process of surrogate modeling, including model selection, preprocessing, data collection, hyperparameters tuning, feature importance, etc. The only step followed carefully in such studies was model evaluation, through which authors obtained the accuracy of their model.

On the other hand, there are some studies on energy systems in which the process of surrogate modeling is elaborated in greater detail. It is worth mentioning that while our focus is on regression problems, some efforts have been concentrated on time-series problems [27–29]. Regarding regression surrogate modeling, a number of papers have dealt with evaluating single-product systems [30–32], but we aim to discuss only those studies on MGSs, the number of which is actually very limited:

Jiang et al. [33] employed a Convolutional Neural Network (CNN) algorithm to model a district heating, cooling, and net electricity generation system fed by natural gas, solar energy, and an external electrical grid. This study primarily focused on effective data generation techniques. It suggested a combination of Monte Carlo stratified sampling and the incorporation of various time-based scenarios into the simulation software to capture all possible inputs and outputs. They also

carefully evaluated the performance of the SM with different sample sizes. Various strategies for MOO were introduced and the results were presented based on a scoring system that allowed a direct comparison between objectives (i.e., system efficiency, renewable energy penetration, and operation cost) to discover the best strategy. The results showed that the generated SM had a 94.34% overall accuracy. Moreover, the process of MOO using the SM took only 2.30 s compared to 4942.94 s when the software simulation was employed instead. Tariq et al. [34] examined a multi-energy-source system aiming to store electricity and produce freshwater for a remote community located in Yucatan, Mexico. In more detail, different energy sources (such as wind, solar, and diesel fuel) along with three types of battery energy storage technologies were considered to create six scenarios of system configurations. They used the HOMER software for mathematical simulation and MATLAB for both surrogate modeling and MOO. A thorough investigation of hyperparameter tuning based on the number of hidden layers and epochs was given. This further elucidated the process of constructing independent ANN models for each objective/label (i.e., renewable fraction, cost of potable water, and CO_2 emissions). Model accuracy was presented by using R^2, *RMSE*, scatter index, and *MSE* as well as standard-error and absolute-frequency distributions. Furthermore, the authors obtained correlation coefficients and implemented a sensitivity analysis to rank the four features used in the modeling based on their impact on each objective. The best trade-off solution of the MOO process using Non-Dominated Sorting Genetic Algorithm II (NSGA-II) exhibited 98.5% for renewable fraction, 0.0447 $\$/m^3$ for the potable water cost, and 36,970 kg for CO_2 emission.

1.2. Motivation and novelty

As discussed above, only two studies have provided comprehensive details of regression surrogate modeling for energy systems that produce more than one commodity. The systems evaluated in these two papers were, in fact, cogeneration and trigeneration, and not precisely multigeneration, which typically involves the production of at least four products. This indicates the scarcity of such studies on multigeneration systems. Once this research gap was identified, we became inspired to implement detailed surrogate modeling to one of our previously studied MGSs (i.e., Mohammadi et al. [35]) and optimize it while incorporating every possible objective. In this respect, the present study is quite unprecedented. Moreover, since using SMs overcomes the hurdle of time-consuming calculations, there is an opportunity to explore the impact of time-dependent parameters within the MOO process. In our proposed system, Direct Normal Irradiation (*DNI*), which exerts a notable influence on the total performance, is a time-varying input. With the groundwork provided in this research, we are able to optimize the system for various *DNI*s. Thus, in addition to performing MOO with as many objective functions as possible, the idea of attaining optimal solutions at different *DNI* values also became a driving factor for us.

1.3. Research objectives

In the present study, various research objectives associated with the given framework of surrogate modeling have been identified by the authors. First, outlier removal will be taken into account within the data preprocessing stage to improve the accuracy of data-driven modeling. Here, the detection of outliers is carried out statistically by utilizing the *Z*-score method. Next, the refined data will be normalized using the same approach (Z-score) to eliminate bias in the modeling process. Along with ANN (the most common BB model used for regression problems), we try to examine two other well-known neural-network-based algorithms for the model selection step: one-dimensional (1D) CNN and Long Short-Term Memory (LSTM) which are particularly suitable for visual data analysis and time-series problems, respectively. Although the type of our problem is not actually image-processing or time-series, it is our objective to inquisitively explore the applicability of these two approaches to our case. We have to note that the reason why machine learning algorithms such as SVR are not taken into account in this study is that these methods have been extensively investigated in the literature and cannot compete well with neural network-based methods in terms of accuracy. Moreover, hyperparameter optimization is executed for all three models via a grid search. To attain even higher accuracy, we aim to create an ensemble model using a brute-force search method. Then, all these four models will be compared with each other to find the best one in terms of precision. Regarding feature selection, since neural network-based methods are employed, removing even one inconsequential feature would deteriorate the accuracy of these models. Thus, no special treatment is given to this step. Later, when the best model is determined, a sensitivity analysis will be performed to assess the importance of features based on their impact on each objective.

The other major purpose of this study is multi-objective optimization, for which a few goals are defined. Because of high computational complexity, merely two objective functions (exergy efficiency and total unit cost of products) were considered for the MOO process in our previous study [35]. With the convenience that SMs offer, performing MOO becomes undemanding. Having such an advantage, we aim to execute MOO with the highest number of objectives that can be defined for the investigated MGS. Hence, along with the two aforementioned objectives, the production rates of commodities are also included, making the total number of objectives six. Furthermore, since *DNI* is basically a time-dependent feature and cannot be scheduled or taken into account as a decision variable, the MOO is carried out at four distinct *DNI* values to encompass the entire range of this feature. This is owing to the fact that *DNI* plays a pivotal role in the performance of our MGS. Through this investigation, deeper insights into the optimal performance of the system will be gained. Finally, the results of two-objective optimization and six-objective optimization will be compared to see if any improvement is made in the current investigation.

2. Methodology

This section explains the methodology adopted for the present study. The entire procedure is illustrated in Fig. 1. Since the simulation was performed and elaborated on in our previous study [35], this step is excluded here. Data sampling is the first step of the procedure through which two datasets are generated for the day and night (not shown in the figure) operation of the MGS. Next, surrogate modeling is performed using the existing framework that includes data preprocessing, hyperparameter optimization, model selection, and model evaluation, as explained in the introduction section. Sensitivity analysis is also conducted to better understand how the outputs depend on the inputs within the best surrogate model. In the final step, multi-objective optimization is carried out utilizing all six performance indicators. This process is repeated four times to cover the feasible range of direct normal irradiation, a time-dependent feature. All these steps were programmed in Python, for which various libraries were used, such as TensorFlow, Keras, and Scikit-learn for black-box modeling and Platypus for the optimization process.

The schematic of the MGS is shown in Fig. 2. The system is driven by geothermal and solar energy sources to produce electricity, cooling load, heating load, and freshwater. The MGS is able to operate sustainably by harnessing clean energy sources, making it attractive for areas with good solar and geothermal energy potential. The essential requirements of the location where the system should be installed include:

- Good geothermal potential for drilling a deep borehole to attain geothermal water temperatures up to 165 °C.
- The consistency of solar energy throughout all seasons.
- A large space for the installation of a solar tower/heliostat field.
- A coastal area facing water scarcity or a remote coastal area without any potable water distribution system to reduce the cost of seawater

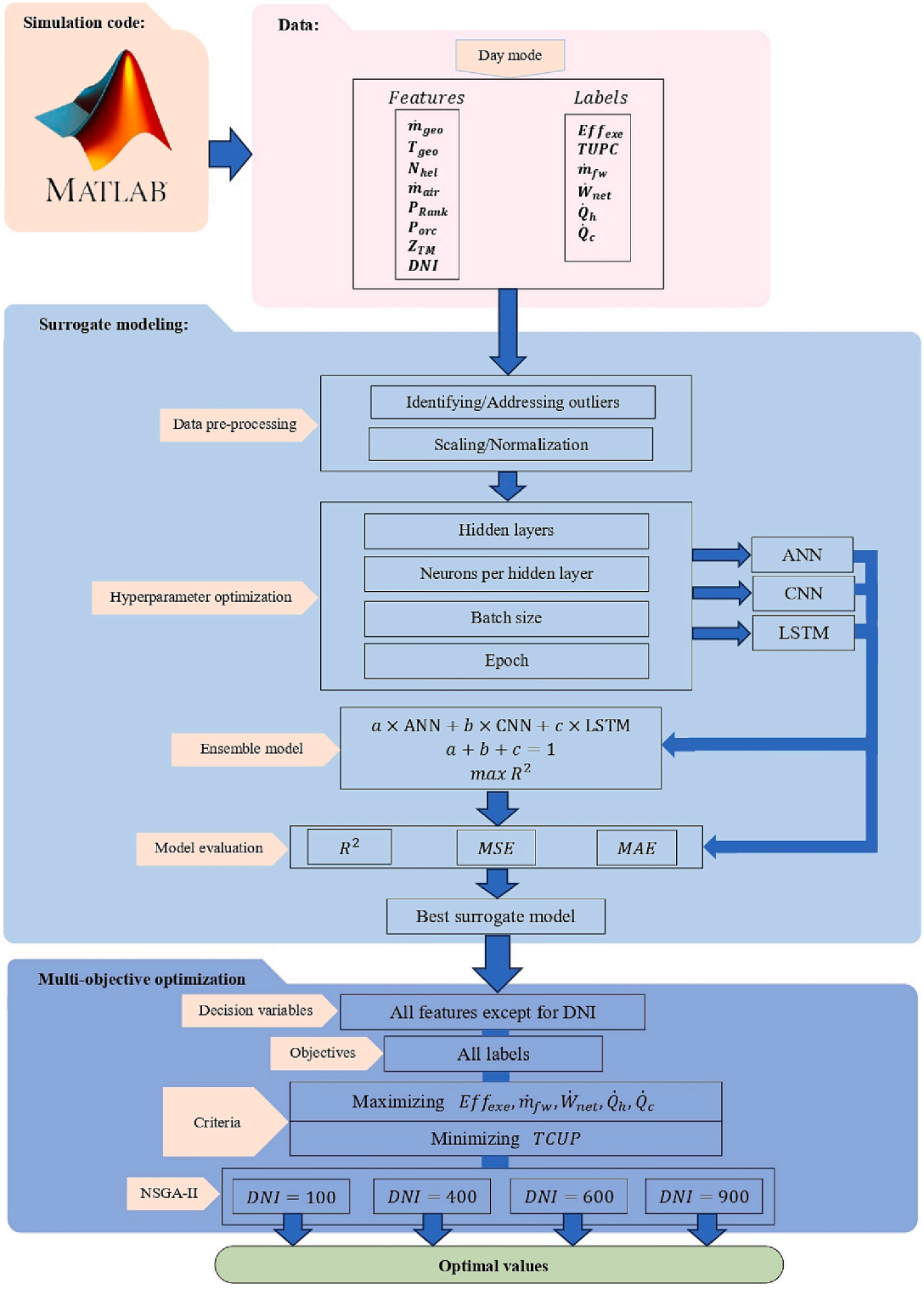


Fig. 1. The methodology implemented in this study.

piping to the desalination system. Alternatively, locations with poor water quality are also suitable because the reverse osmosis unit can remove impurities and contaminants from water, making it drinkable. Not to mention that the previous conditions must also be met.

It is worth noting that, based on a case study conducted using the meteorological data of Minab, a coastal city in the south of Iran, the system was shown to be capable of providing freshwater for an average of over 1200 individuals during the daytime all year long [35]. This example serves as an illustration of the practical benefits that the current MGS can provide in real-life scenarios.

2.1. Data sampling

Data sampling is the process of data selection from a larger dataset. For surrogate modeling, data sampling is defined as the process of

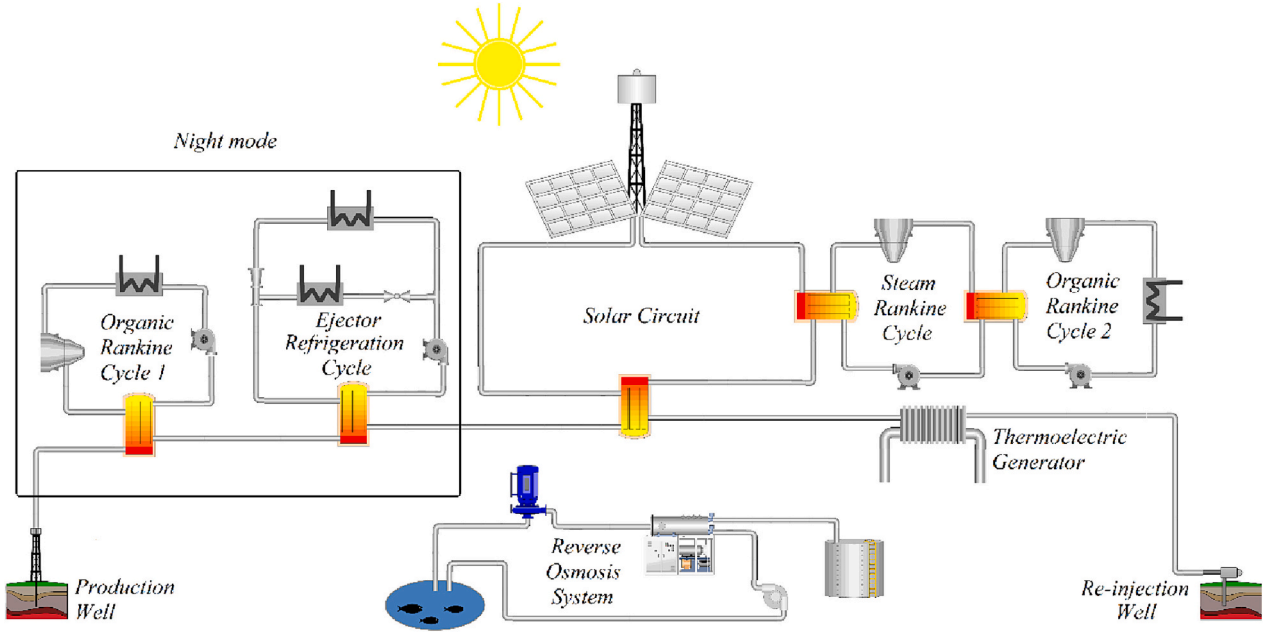


Fig. 2. A schematic of the MGS.

obtaining a dataset representing the behavior of a simulation experiment [36]. In addition, data sampling can play a crucial role in hyperparameter optimization. Many techniques are being used commonly: uniform random sampling, grid sampling, Latin hyperbolic sampling, Poisson disk sampling, farthest point sampling, best candidate sampling, and hybrid BC-GreedyFP sampling [37]. In this study, we adopted uniform random sampling, the most straightforward and unbiased approach, which led to the generation of a large dataset to ensure comprehensiveness. Eight features (inputs/decision variables) were considered, the ranges of which are listed in Table 1. The MATLAB code was run based on the random selection of features within their respective ranges to generate six labels (outputs/objectives): exergy efficiency (Eff_{exe} or η_{exe}), Total Unit Cost of Products ($TUCP$), freshwater mass flowrate ($\dot{m}_{fw}$), net electricity generation ($\dot{W}_{net}$), hot water generation ($\dot{Q}_h$), and cooling load ($\dot{Q}_c$). These features and labels collectively formed a dataset of 23,000 data points.

2.2. Data preprocessing

Data preprocessing involves preparing raw datasets into a format that is suitable for training black-box algorithms. This step involves three techniques: (I) scaling/normalization, (II) handling missing data, and (III) identifying and addressing outliers. The second approach is not commonly necessary for surrogate modeling, thanks to data extraction from software simulations. Unlike the outputs of sensors [38,39], this type of data does not face malfunctions or interruptions. In this study, the Z-score technique is employed to handle the data preprocessing phase, i.e., treating outliers and normalization. This technique uses the statistical measures of mean and standard deviation to normalize unstructured data [40]. The following formula calculates the Z-score value ($x^{'}$) of a given data point (x) using its mean (μ) and standard deviation (σ) values:

$$x^{'} = (x-\mu)/\sigma \tag{1}$$

Carrying out this transformation would ensure that all data points take a mean of zero and a standard deviation of one [41].

Since the presence of outliers (out-of-range values) causes inefficiency during the training phase, outlier detection and removal have to be undertaken. The Z-score technique embraces this concept as well. This can be done by assigning a specific value to the standard deviation as a threshold to eliminate any remaining data points beyond that threshold. While removing more outliers may appear to offer an opportunity for accuracy improvement, it is not necessarily guaranteed, as it can lead to a lower number of data points, jeopardizing the overall accuracy of the trained model. Therefore, only an optimal value must be chosen as the threshold.

Table 1
The range of the features included in the data sampling process [35].

Feature	Unit	Range
Geothermal mass flowrate ($\dot{m}_{geo}$)	kg/s	[2 15]
Geothermal temperature (T_{geo})	°C	[125165]
Number of heliostats (N_{hel})	–	[25 65]
Air mass flowrate ($\dot{m}_{air}$)	kg/s	[4 10]
Pressure of the Ranking cycle (P_{Rank})	kPa	[2000 5000]
Pressure of ORCs (P_{ORC})	kPa	[15,900 3500]
Figure of merit (Z_{TM})	–	[0.2 2.0]
Direct normal irradiation (DNI)	W/m^2	[11100]

2.3. Artificial neural network (ANN)

An ANN is an algorithmic model, inspired by the human nervous system, that can build empirical and nonlinear relationships between the inputs and outputs of a dataset [42–44]. This supervised learning technique operates at high speed and demonstrates high adaptability, generalization, robustness, and mapping ability [45]. An ANN consists of three distinct layers that are sequentially connected by a number of neurons: an input layer, one or more hidden layers, and an output layer (Fig. 3). The number of hidden layers and neurons per layer can entirely affect the ability of the ANN in modeling intricate relationships between input and output data. However, an excessive number of neurons may end up causing overfitting, a situation where the model struggles to generalize well. Another important hyperparameter of ANNs, as well as all other types of neural networks, is the batch size. This parameter determines the number of training samples employed in one iteration within the training process. The choice of the batch size can also impact the learning process of a model and therefore must be tuned to secure high predictive power. The final hyperparameter that we deal with in this study is epoch. An epoch refers to the number of iterations that the algorithm processes through the entire dataset and updates its internal

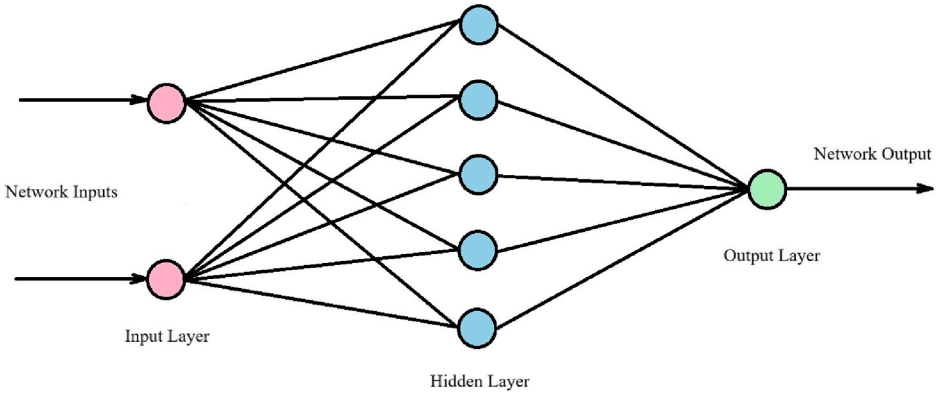


Fig. 3. Architecture of an artificial neural network (ANN).

parameters (i.e., weights and biases). An optimal choice of epoch would avoid underfitting and overfitting. The optimization of the aforementioned hyperparameters is elaborated on in the results section.

2.4. Convolutional neural network (CNN)

The CNN is another type of neural network that uses convolutional layers instead of fully connected layers (a characteristic of ANNs) to process grid-structured data, such as images [46]. As shown in Fig. 4, a CNN consists of five main layers: the input layer, convolutional layer, pooling layer, and fully-connected (dense) layers, and output layer. The convolutional layer applies filters (kernels) to the input data to extract features and patterns, the pooling layer is responsible for reducing the spatial dimension of the input data using down sampling, the fully-connected layers are similar to the hidden layers of ANNs in which the learning process occurs, and the output layer generates the final prediction [47]. Since the present study deals with a one-dimensional (1D) problem, a 1D-CNN model is developed. The same hyperparameters as those used for the ANN are also tuned for the CNN, which will be presented later.

2.5. Long-short term memory (LSTM)

An LSTM model belongs to the Recurrent Neural Networks (RNNs) family, which are bi-directional neural networks designed to handle sequential data or time-series [48]. The advantage of LSTMs over RNNs is their ability to store information for a longer term. However, due to limited memory, the algorithm must decide which information is crucial to be stored in the long-term memory; otherwise, it would be placed in the short-term memory. Fig. 5 shows the architecture of an LSTM model, which has three gates to control features: the input gate, the forget gate, and the output gate. The input gate controls new information flowing into the cell state, representing the long-term memory,the forget gate discards unimportant previous information already saved in the cell state, and the output gate decides the output information from the cell state and then determines the next hidden state, in which the information is stored for next iterations. An LSTM model can automatically save

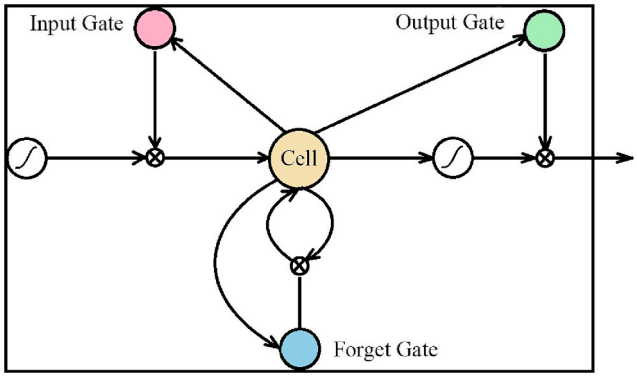


Fig. 5. Architecture of a long-short term memory (LSTM) algorithm.

or discard stored information using these gates [49].

2.6. Ensemble model

The ensemble model is developed using a brute-force search based on the results of the ANN, CNN, and LSTM. The brute-force search (or exhaustive search) is a straightforward technique for solving a problem by systematically considering all possible solutions [50]. In this context, the following formulation is considered by assigning a coefficient (a, b, and c) to each method:

$$\text{Ensemble} = a \times \text{ANN} + b \times \text{CNN} + c \times \text{LSTM} \quad (2)$$

The brute-force search goes through every combination of coefficients to maximize the R-squared (R^2) score (see Section 2.7.1). Note that the sum of all three coefficients must equal one. At each iteration, the R^2 score is computed, and, if it surpasses the current best, the coefficients and the R^2 score will be updated. The final result is the combination of coefficients yielding the highest R^2 score.

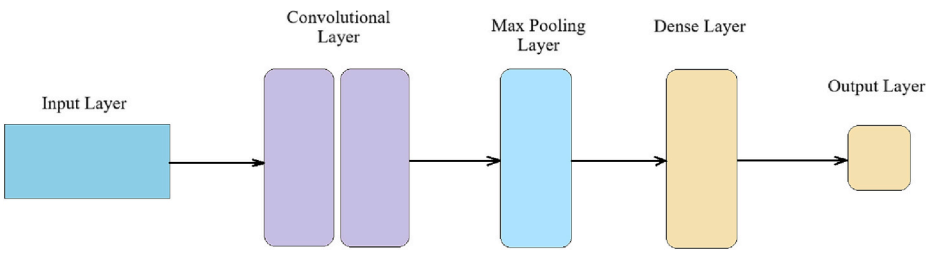


Fig. 4. Architecture of a convolutional neural network (CNN).

2.7. Model accuracy evaluation

Different measures are available to evaluate the accuracy of SMs in predicting values or classes as close as possible to the observed ones. In this study, since the type of the problem is regression, three commonly used regression measures are utilized: R-Squared (R^2), Mean Absolute Error (*MAE*), and Mean Squared Error (*MSE*).

2.7.1. R-squared (R^2)

R-squared (R^2), also referred to as the coefficient of determination, indicates how much variance in a label can be explained by features in a regression problem. The value of this measure ranges from zero to one. The higher the R-squared score, the better the model fits to the data. The typical expression for R^2 is given as follows [51]:

$$R^2 = 1 - \frac{RSS}{TSS} \tag{3}$$

Where *RSS* is the residual sum of the squared differences between the actual and predicted values, and *TSS* is the total sum of the squared differences between the actual values and the mean of labels.

2.7.2. Mean absolute error (MAE)

This metric measures the average absolute differences between the actual and predicted values regardless of whether the error is over-predicted or underpredicted. It can be mathematically presented as [52]:

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - \hat{y}_j\right| \tag{4}$$

Here, n is the number of observations, y_j and $\hat{y}_j$ are the actual and predicted values in the j-th observation, respectively.

2.7.3. Mean squared error (MSE)

The *MSE* quantifies the performance of a model based on the average squared differences between the actual and predicted values. In this way, bigger errors gain larger weights as the differences are squared. The formulation is as follows [53]:

$$MSE = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2 \tag{5}$$

2.8. Multi-objective optimization (MOO)

As illustrated in Fig. 1, the MOO process involves setting decision variables, objectives, and criteria as well as applying the NSGA-II algorithm. In this problem, all features can be regarded as decision variables, except for *DNI*, as this feature is inherently time-dependent and cannot be set manually. However, all labels can be taken into account as objectives, implying that up to six objectives are allowed to be embedded into the optimization process. These six objectives can be broken down into three major groups: the exergetic performance, total production cost, and production rates. Regarding optimization criteria, the exergetic performance and production rates must be maximized while the total production cost needs to be minimized. Although *DNI* cannot be regarded as a decision variable, it would be preferable to perform the six-objective optimization at different *DNI* values to cover the entire feasible range of this parameter. The optimal values obtained would assist in establishing an effective control system for the MGS to readjust the inputs at the selected *DNI* values, namely, 100, 400, 600, and 900 W/m^2.

The optimization method in this study is NSGA-II, which was also employed for the simulation-based optimization in our simulation study [35]. Employing this method in the current study allows us to compare the results of surrogate-based and simulation-based optimization with each other. The other reason is the robust capability of NSGA-II in addressing multi-objective optimization problems, as also emphasized by [34]. A flowchart of this process is shown in Fig. 6, where applying the TOPSIS method for discovering the best solution from Pareto frontiers is also a step [54,55]. NSGA II is an elitist non-dominated sorting genetic algorithm commonly used for MOO problems [56]. In the first step, a random initial population of potential solutions is generated, where each solution represents a set of decision variables. Next, the performance of the surrogate model outputs (objectives) is assessed to evaluate the suitability of the population. In the third step, the solutions are ranked using a non-dominated sorting technique according to their performance in improving the objectives. Now, a set of trade-off solutions for all objectives is identified. If the maximum number of generations is not reached, the genetic operators (i.e., selection, crossover, and mutation) come into play to create a new population. The same procedure is iterated over and over again until reaching the maximum number of generations. At this point, the final Pareto frontiers of non-dominated solutions are obtained. Finally, the TOPSIS method is utilized to locate the best trade-off solution among all the potential solutions acquired in the previous step.

3. Results and discussion

This section is divided into two major subsections: surrogate modeling and multi-objective optimization. In the first part, according to the framework, the steps for obtaining robust and accurate SMs are followed to emulate the physics-based model of the MGS at hand. In this stage, the best model among four neural network-based models, i.e., ANN, CNN, LSTM, and an ensemble model, is selected. Next, the performance of all models is studied through a case study. In the final step, a sensitivity analysis is performed for the chosen model. Regarding the second subsection, a six-objective optimization process using the best SM is performed at four *DNI* levels. The achieved results are then compared to the results of the simulation-based bi-objective optimization obtained in [35].

3.1. Surrogate modeling

3.1.1. Data preprocessing

In addition to using the technique of *Z*-score for outlier detection, some outliers can also be identified by domain knowledge and expertise [57]. Given the fact that the labels of our data represent performance indicators, a reasonable approach to define outliers would be to set thresholds for labels according to our understanding of the anticipated behavior of the given MGS. This way of treating outliers requires expert knowledge. Since economic performance can be regarded as one of the most important criteria in the present study, it would be reasonable to set a threshold based on *TUCP*. Our understanding of the system tells us that the values of *TUCP* predominantly range up to approximately 100 $/GJ and rarely exceed it. This value is, thus, taken as an outlier boundary, reflecting that any data value beyond 100 $/GJ will be eliminated from the dataset. As a result, out of 23,000 data points generated from the original simulation, 1152 data values were filtered in the first round before normalizing the dataset. After normalization, outlier detection through the *Z*-score technique was carried out on the tailored dataset containing 21,848 value points. The higher the *Z*-score value, the lower the number of outliers is. Multiplies of the standard deviation, such as 2 and 3, are commonly considered for the threshold value [58]. We started with five and came down by subtracting 0.5 each time. Through this trial and error process, 2.5 was recognized as the most affordable Z-score value, identifying 800 data points as outliers and leaving 21,048 for training. It should be noted that any values lower than 2.5 (e.g., 2) resulted in a relatively high number of outliers (e.g., 2922).

3.1.2. Pearson correlation coefficients

After refining the dataset, the Pearson correlation coefficients can provide a holistic view of the relationships among data variables. This

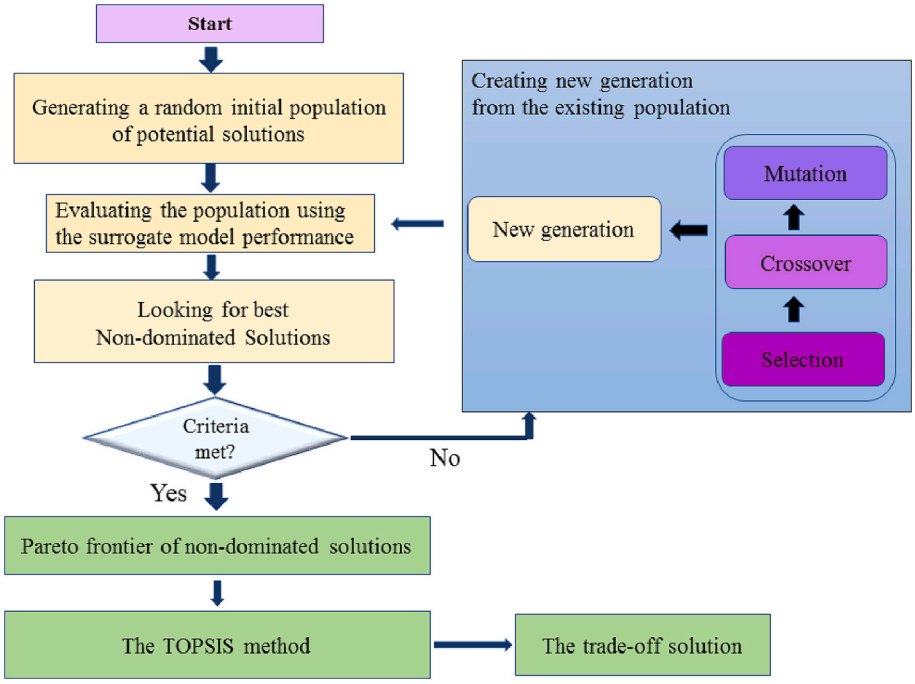


Fig. 6. Flowchart of the NSGA-II algorithm combined with TOPSIS.

method evaluates both the strength and direction of the linear relationships among data variables. As a result, a correlation heatmap generated from the Pearson correlation coefficients is illustrated in Fig. 7. As can be seen, positively correlated elements take a value between zero and 1 (shown in red), while a value between -1 and zero indicates that the two elements are negatively correlated (shown in blue). Moreover, the stronger the correlation, the more intense the color. Taking these factors into account, it can be noted that the hot water production rate exhibits the highest correlations with features. The geothermal temperature (with positive impact), pressure of the Rankine cycle (with negative impact), and geothermal mass flowrate (with positive impact) are the top three decisive features influencing this objective in a linear manner. The second objective with high correlation coefficients is net electricity production, on which the air mass flowrate

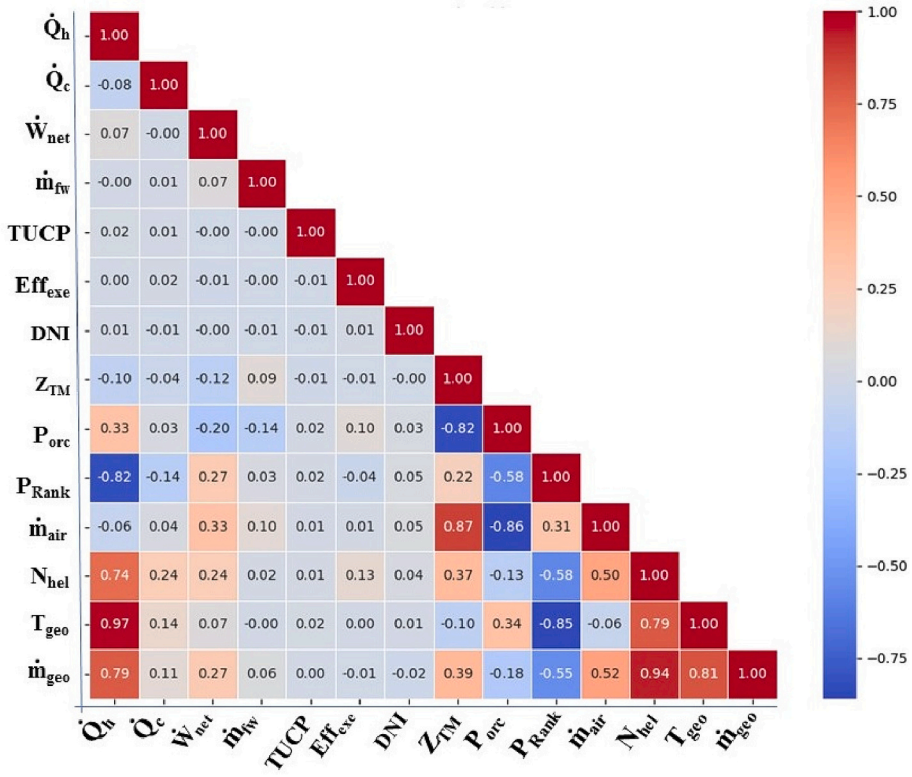


Fig. 7. A correlation heatmap obtained for all data variables.

of the solar circuit exerts the highest effect. In relation to the cooling load ($\dot{Q}_c$), which ranks as the third most correlated objective according to the heatmap, the number of heliostats exhibits the highest correlation. This variable, however, has no practical relationship with the cooling load since this objective is solely influenced by geothermal-based features. It is worth mentioning that, due to the linearity of this measure and the presence of nonlinearity in our dataset, the results may be somewhat limited or distorted, as observed in the case of cooling load. Along with this, the absence of relationships (i.e., almost zero correlations) for other objectives does not necessarily imply their nonexistence, as this method cannot detect complex and nonlinear relationships. After developing the surrogate model, a nonlinear sensitivity analysis will be conducted to address this concern.

3.1.3. Hyperparameter optimization

Three neural network-based methods are selected for comparison to determine the best one: ANN, CNN, and LSTM. However, before proceeding with this step, hyperparameter optimization has to be addressed to determine the optimal network structure of each method. This stage is of critical importance in data science, especially when it comes to neural network models, as they are difficult to configure, and a poor choice of hyperparameters may lead to slow training and inaccurate results. The grid search method is employed to search through a list of hyperparameter combinations and discover the best that returns the least error [59]. As can be seen in Table 2, the number of hidden layers, number of neurons per hidden layer, learning rate, batch size, and number of epochs are the considered hyperparameters, and optimal choices are shown in bold letters. Since high computational effort is required (as many times as the permutation of candidates) for the sequence of training to reach the best combination, only three choices are considered for each hyperparameter. The candidates listed in the table are commonly employed for such models. For more details on this topic, the reader is referred to the following sources: Jin et al. [60] for ANN, Bochinski et al. [61] for CNN, and Nakisa et al. [62] for LSTM.

As shown in Table 2, the highest number of hidden layers brings the best results in each method. However, a different pattern can be observed for the number of neurons per hidden layer; the ANN takes the highest number (128), followed by the CNN and LSTM, with 64 and 32, respectively. Regarding the batch size, the ANN takes 32, the value of which is usually considered a rule of thumb for neural networks, but the other two approaches take fewer samples (16). For the last hyperparameter, all three methods prefer to employ 100 epochs, which indicates the number of iterations an algorithm works through the dataset.

3.1.4. Model selection

After achieving the best network structure for each method, model evaluation is performed to figure out the best approach. As can be seen in Table 3, the results for all six objectives are presented. Among the three methods, the CNN performed poorly and failed to achieve any results better than 0.9560 in terms of R^2 scores. As was anticipated, this method is less suitable for one-dimensional regression problems. On the other hand, as shown in Fig. 8, for R^2 scores, the ANN and LSTM exhibited almost identical excellent results. This is somewhat surprising considering that the LSTM has been designed to address time-dependent problems rather than regression ones. Examining the obtained results in terms of *MAE* scores (Fig. 9), the ANN outperforms the LSTM in all objectives except for cooling load generation. Hence, the ANN can be selected as the most accurate method. This outcome reinforces the recognition of the ANN as one of the most preferred approaches for regression problems.

Table 2
Results of hyperparameters tuning via grid search.

Hyperparameters	ANN	CNN	LSTM
Hidden layers	1, 2, **3**	2, 3, **5**	2, 3, **4**
Neurons per hidden layer	32, 64, **128**	32, **64**, 128	16, **32**, 64
Batch size	16, **32**, 64	10, **16**, 32	8, **16**, 32
Epoch	50, **100**, 200	50, **100**, 200	50, **100**, 200

Table 3
Model evaluation of the four tuned models for all objectives using three common metrics.

Model	Objective	R^2	*MSE*	*MAE*
ANN	η_{exe}	0.9959	0.0040	0.0396
	TUCP	0.9847	0.0149	0.0624
	$\dot{m}_{fw}$	0.9982	0.0018	0.0254
	$\dot{W}_{net}$	0.9973	0.0027	0.0323
	$\dot{Q}_c$	0.9998	0.0002	0.0114
	$\dot{Q}_h$	0.9994	0.0006	0.0169
CNN	η_{exe}	0.9413	0.0576	0.1717
	TUCP	0.9087	0.0889	0.2199
	$\dot{m}_{fw}$	0.9560	0.0437	0.1662
	$\dot{W}_{net}$	0.9322	0.0665	0.1997
	$\dot{Q}_c$	0.9539	0.0447	0.1659
	$\dot{Q}_h$	0.9485	0.0507	0.1683
LSTM	η_{exe}	0.9954	0.0045	0.0478
	TUCP	0.9827	0.0167	0.0649
	$\dot{m}_{fw}$	0.9967	0.0033	0.0395
	$\dot{W}_{net}$	0.9961	0.0039	0.0429
	$\dot{Q}_c$	0.9988	0.0012	0.0265
	$\dot{Q}_h$	0.9979	0.0020	0.0327
Ensemble	η_{exe}	0.9971	0.0029	0.0294
	TUCP	0.9830	0.0166	0.0568
	$\dot{m}_{fw}$	0.9982	0.0018	0.0244
	$\dot{W}_{net}$	0.9976	0.0024	0.0281
	$\dot{Q}_c$	0.9999	0.0001	0.0080
	$\dot{Q}_h$	0.9996	0.0004	0.0129

In an effort to improve the performance of the ANN, an ensemble model is developed. As explained in Section 2.6, a brute-force search is utilized to evaluate the coefficients assigned to each neural network model, looking for the best combination leading to maximum R^2. The outcome shows that the weight of the CNN is zero, while those of the ANN and LSTM are 0.57 and 0.43, respectively. This would make sense when considering the undesirable effect of the CNN, along with a slightly higher accuracy of the ANN than LSTM. The scores of the four developed models for all objectives are also given in Table 3. When compared to the ANN, it is evident that the ensemble model yields even better results except for *TUCP*. Plotting the scores of R^2 and *MAE* can offer more clarification on the dominance of the ensemble model (Figs. 10–11). Specifically, from Fig. 11, it is noticeable that the *MAE* scores of the ensemble model are comparably smaller for all objectives, even for *TUCP*. Hence, this makes the ensemble model the most suitable surrogate model for the current study. It is worth mentioning that although running a model composed of two different models can be relatively time-consuming, obtaining the highest possible accuracy is our foremost concern.

3.1.5. Case study

The type of problem considered in this study is regression since almost all features are not time-dependent. In fact, the data was generated using random values for all features within their respective valid ranges to ensure the resulting dataset is comprehensive. However, *DNI* is an exception as it inherently changes over time. For this reason, an evaluation can be made based on the hourly changes of *DNI*. The hourly *DNI* values of the Reno city located in Nevada, USA, in a time interval between the 22nd and 28th of June 2021 are considered for this purpose [63]. In addition to receiving abundant solar energy all year long, Reno is among the best places in the US in terms of geothermal energy potential [64,65]. On the other hand, the poor quality of water from the Truckee River, which passes through Reno, has been an issue to the local municipality [66]. The installation of the current MGS can contribute to the removal of impurities and contaminants from the water

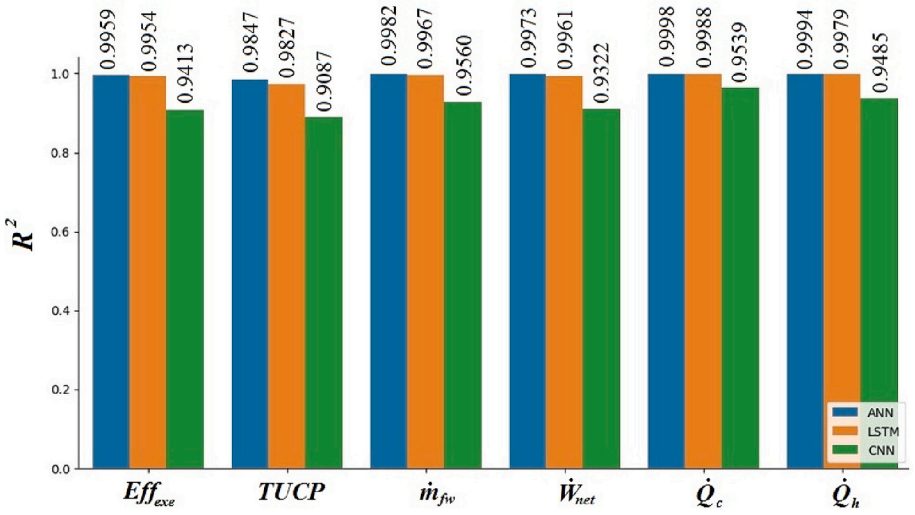


Fig. 8. R^2 scores of the ANN, CNN, and LSTM.

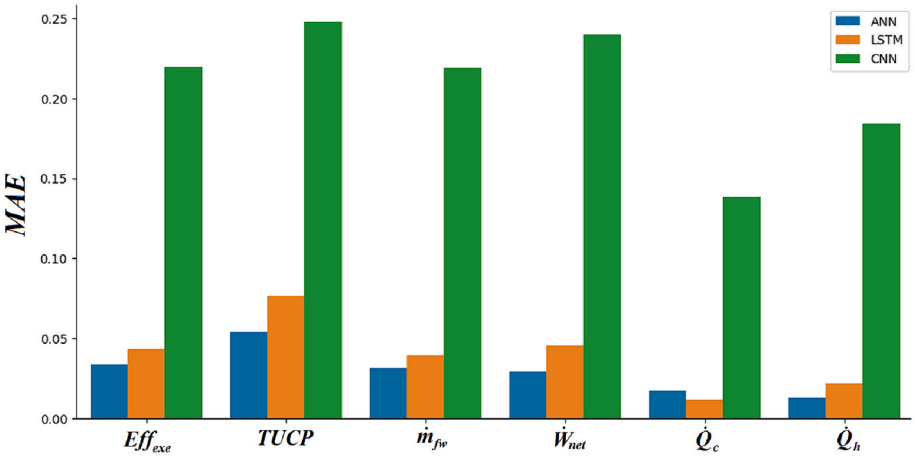


Fig. 9. *MAE* scores of the ANN, CNN, and LSTM.

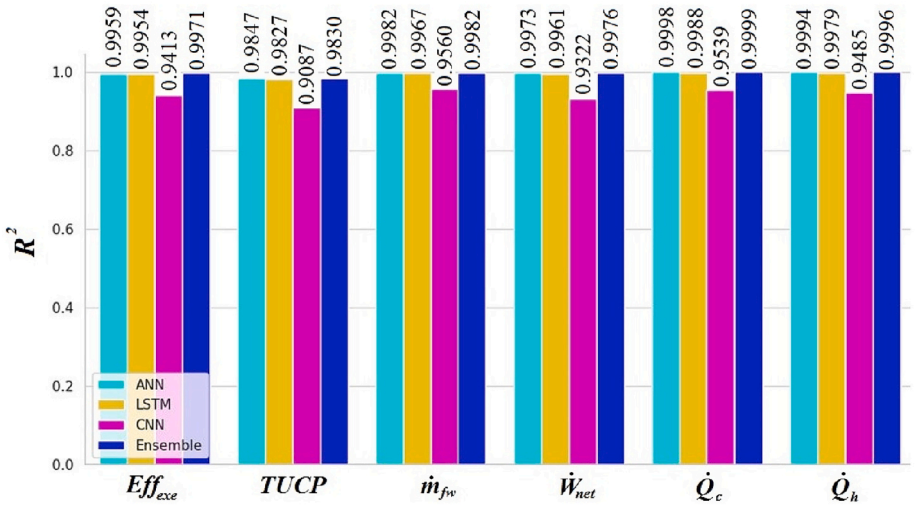


Fig. 10. R^2 scores of the ANN, CNN, LSTM, and the ensemble model.

in this area. Hence, this city can be an attractive location for conducting a case study.

The purpose of this subsection is to not only compare the results of the four surrogate models but also demonstrate the precision of the ensemble model by comparing it with the results obtained from the simulation. The hourly distributions of freshwater and net electricity

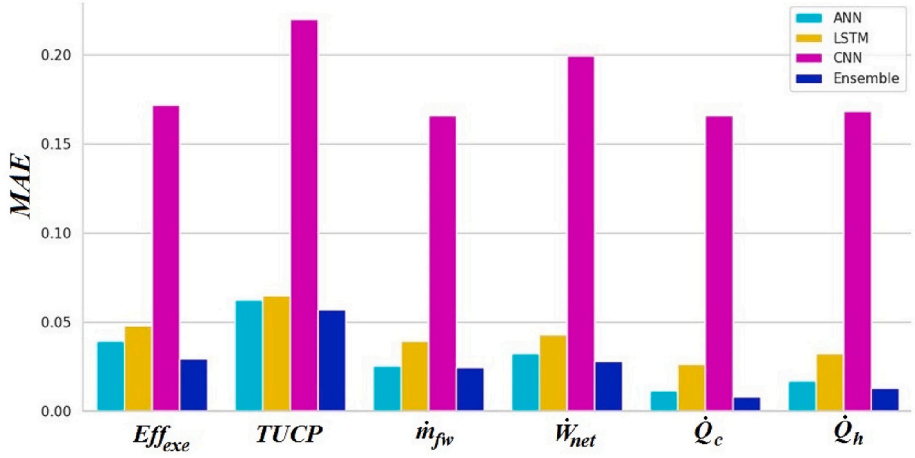


Fig. 11. *MAE* scores of the ANN, CNN, LSTM, and the ensemble model.

production rates are shown in Fig. 12. It must be mentioned that during times when no solar energy is available, the system switches to the night mode. Thus, where *DNI* is zero, no freshwater is produced and net electricity generation stays at a minimum level. Taking a glance at the graphs, it is evident that the CNN lags behind other models in terms of accuracy. Regarding the rest of the models, their results are clustered together near that of the simulation as their accuracy is high. However, looking closely, it is clear that the ensemble model outperforms other methods even at zero-*DNI* points.

3.1.6. Sensitivity analysis

In this subsection, further insights into the impact of features on objectives are presented through a sensitivity analysis, a crucial assessment for understanding the behavior of a data-driven model. This type of analysis can have the advantage of handling non-linearity between features and objectives. The results are shown in Fig. 13 through bar charts and pie charts. Here, instead of presenting the net effect, both positive and negative impacts (given in weights) are provided in the bar charts to offer a more detailed view. On the other hand, the absolute contribution of features on each objective is presented in the provided pie charts. As for total exergy efficiency, most features have a negative overall impact on this objective as shown in the respective bar chart. From the pie chart, it is evident that *DNI* makes the most significant impact, followed by geothermal mass flowrate and the number of heliostats, having identical effects. The dominance of *DNI* can be witnessed for all other objectives except for $\dot{Q}_h$ as it is fed merely by geothermal energy. Regarding the second most important feature, the number of heliostat comes into play for *TUCP* and $\dot{m}_{fw}$, while the geothermal mass flowrate takes effect on $\dot{W}_{net}$ and $\dot{Q}_h$. The third most effective feature varies for each objective: the geothermal mass flowrate for *TUCP*, the mass flowrate of the solar circuit for $\dot{m}_{fw}$, the geothermal temperature for $\dot{W}_{net}$, and the number of heliostats for $\dot{Q}_h$.

Through this analysis, we observed different general tendencies for each objective. Exergy efficiency is mainly affected by both solar and geothermal-related features while the weight of the former is slightly higher. More or less, a similar inclination can be seen for *TUCP*, $\dot{W}_{net}$, and $\dot{Q}_h$. However, the role of solar-driven features is more dominant in the case of $\dot{m}_{fw}$. Among operation features, the pressure of the ORC comparably plays a more important role, followed by the mass flowrate of the solar circuit and the pressure of the Ranking cycle. On the other hand, only features related to geothermal energy influenced the cooling load, with the geothermal mass flowrate being the most significant.

3.2. Multi-objective optimization

3.2.1. Six-objective optimization

As surrogate models are considerably more efficient to run than simulation codes, there would be almost no hurdle in carrying out optimization with many objectives. This study aims to incorporate as many objectives as possible to attain a trade-off solution for a multiple-criteria problem. Thus, six objectives are considered, including performance indicators of operation, cost, and production rates: overall exergy efficiency, total unit cost of products, freshwater mass flowrate, net electricity generation rate, cooling water production rate, and hot water production rate. Moreover, since *DNI* cannot be set manually despite being a feature in our dataset, the optimization problem is performed at four different *DNI* values (from 100 W/m^2 to 900 W/m^2) to cover the entire range of possible *DNI* values. The TOPSIS method is employed to figure out the best trade-off solution at each *DNI*, and the corresponding results for features and objectives are specified in Table 4. Before delving into the results of objectives, it is worth checking out the optimal values of decision variables (i.e., features). As can be seen, the geothermal water mass flowrate remains constant in all four cases at nearly the highest possible value (15 kg/s). The number of heliostats (N_{hel}) and the figure of merit (z_{TM}) also remain unaltered throughout the range at their respective lower boundaries. However, other features exhibit irregular patterns, having extremum values around either *DNI* = 400 W/m^2 or *DNI* = 600 W/m^2.

Optimal results for all objectives are plotted in Fig. 14. Regarding total exergy efficiency, an approximately 11% reduction can be noticed when *DNI* increases from 100 to 900 W/m^2. The reason is the growth in the exergy destruction rate of the solar tower as solar intensity rises. In contrast, *TUCP* increases in value over the range of *DNI*, at 78%. This is largely linked to the substantial growth of freshwater production, which increases more than 10 times throughout the four *DNI* cases. It should be noted that there is a considerable change in the values of the three aforementioned objectives between *DNI*s of 100 and 400 W/m^2. This may indicate a critical operating range for the system, wherein the solar input can induce a significant change. However, this pattern is not observed in the net electricity and hot water production rates, even though they also exhibit an increasing trend over the entire range of *DNI*. Therefore, it can be concluded that freshwater exerts the highest impact on both total exergy efficiency and *TUCP* among all products. It is worth mentioning that while cooling load production appears unrelated to *DNI*, slight variations in this objective can be observed at different *DNI* values. This can be justified by considering the changes occurring in the geothermal temperature (Table 4).

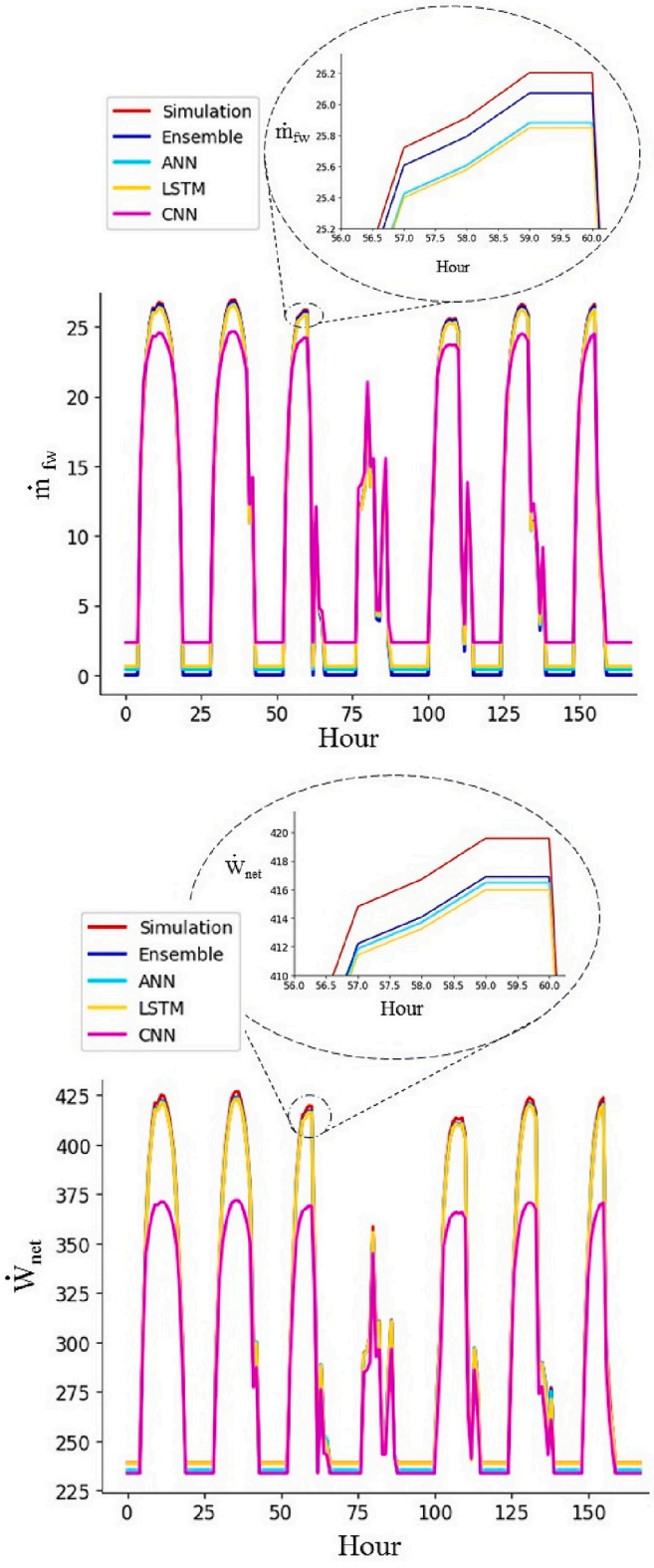


Fig. 12. Hourly distribution of freshwater and net electricity generation rates.

3.2.2. Six-objective vs bi-objective

After achieving the results of the six-objective optimization, it would be appealing to compare them with the results obtained by the bi-objective optimization using the simulation code [35]. For this purpose, only the optimal values corresponding to *DNI* = 600 W/m^2 are considered as the simulation-based optimization was performed for the exact same *DNI* value. As shown in Table 5, the range of improvements achieved in this study is from 1.9% to 12.7%, and the average improvement is 6.3%; exergy efficiency has the lowest, while *TUCP* shows the highest. In fact, the decrease of the total unit cost of products from 34.1 $/GJ to 29.77 $/GJ is of considerable value. This shows that the operation of the system at this optimal solution (obtained through the six-criteria scenario) would be notably more cost-effective. The second highest improvement is related to net electricity generation, showing a 9.5% increase from 423.7 kW to 463.90 kW. However, the enhancements of exergy efficiency and freshwater mass flowrate are insignificant. This implies that, in the six-objective optimization case, the optimal solution mostly favors *TUCP* and net electricity generation.

4. Limitations

Although this study made a pioneering effort in developing an accurate surrogate model for a multigeneration system and performing surrogate-based optimization with six objectives, the authors have to acknowledge several limitations inherent to the research:

- In this study, uniform random sampling was used for data collection. Considering the simplicity of this technique, a substantial amount of data needed to be collected to adequately capture the system's behavior. Implementing more sophisticated methods, such as Latin hyperbolic sampling, may result in a dataset of a smaller size while maintaining the same level of quality.
- When it comes to outlier detection, expert knowledge has to contribute along with a relevant numerical method. While general instructions can be suggested for numerical methods, expert knowledge heavily depends on the characteristics of the problem at hand. For the current MGS, the total unit cost of products was taken into account with a threshold of 100 $/GJ. However, choosing appropriate criteria varies from one problem to another, making generalization somewhat challenging.
- While NSGA-II has been endorsed by countless studies, it would be valuable to explore various other optimization methods in terms of both speed and accuracy, comparing them to NSGA-II.
- Since no environmental impact was analyzed in the simulation-based study, this crucial element was not considered in the current study as well, which might have yielded more insights to the MGS.
- As emphasized several times throughout the manuscript, the proposed methodology is only suited for regression analysis; thus, problems related to time-series analysis cannot be addressed. When an energy storage unit, such as flow batteries, pumped hydro storage, thermal energy storage, compressed air energy storage, etc., is part of a MGS, the problem would possess dynamic characteristics. Therefore, the processes of data sampling, surrogate modeling, and optimization must be adjusted to embrace the time-dependency of the problem. This also holds true when production rates fluctuate on an hourly, daily, or seasonal basis, requiring corresponding adjustments to the inputs of the system. This occurs when the demand pattern is changing and time-dependent. For systems exhibiting these two characteristics, the complexity will further increase as the interplay between these two cases comes into effect. While some efforts have been made in related research works [27–29], no specific study has been conducted for MGSs. This area remains open for investigation by future researchers.

5. Adaptability of the methodology for other MGSs

After pointing out the limitations of our study, exploring the adaptability of the introduced methodology for other multigeneration systems would be a valuable consideration. The present methodology is well-suited for addressing MGSs having a routine operation, or, in other words, exhibiting a regression characteristic. This implies that systems with time-dependent energy storage units and/or production rates

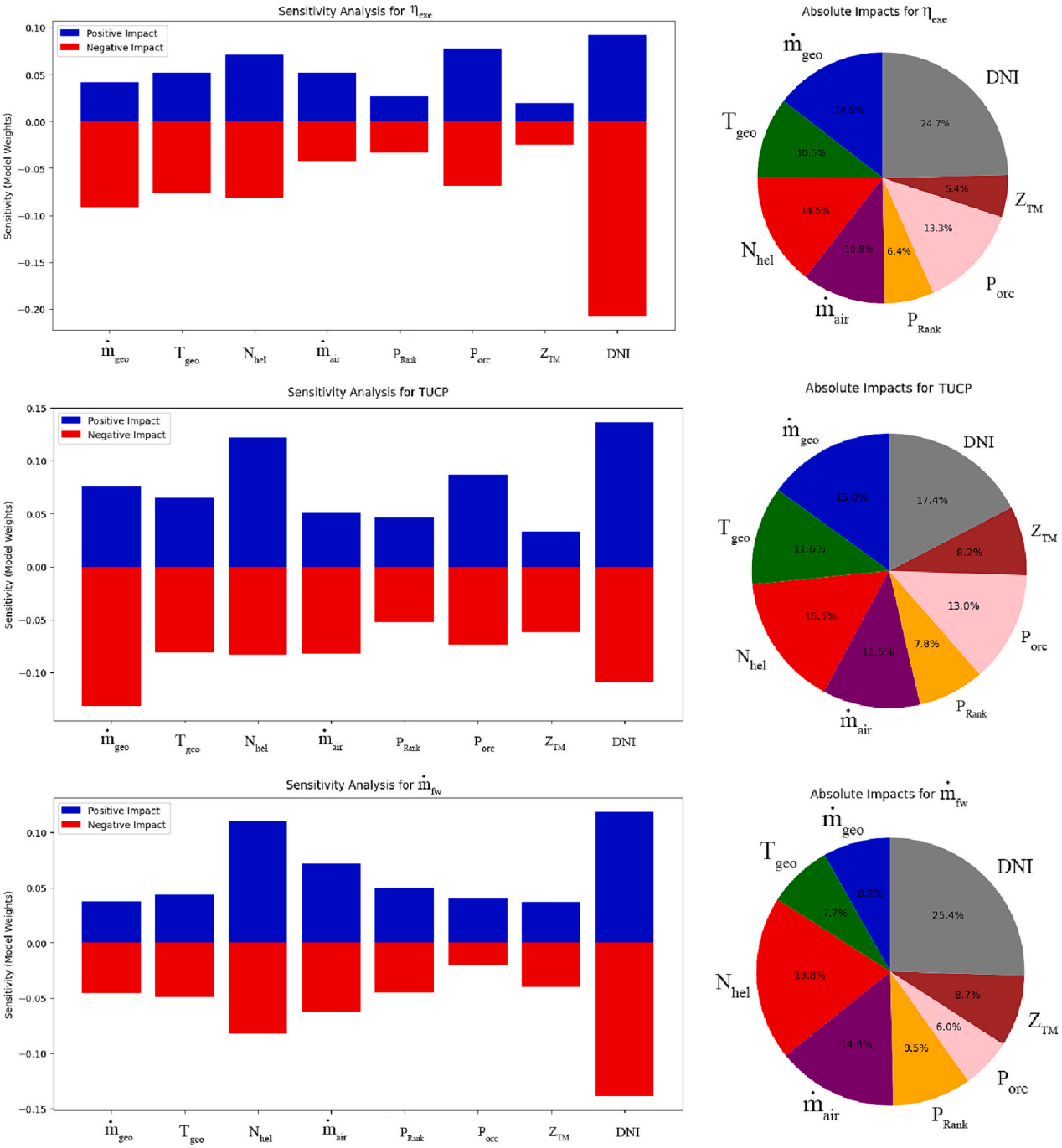


Fig. 13. Results of sensitivity analysis: (left) both positive and negative impacts of features, and (right) absolute effect of features on each objective.

characterized by fluctuating patterns, for which time-series analysis becomes the sole viable choice, cannot be handled by the current approach. For MGSs with integrated energy storage, a combination of static features and one of time-based, event-based, systematic, or scenario-based data must be organized for the data sampling step and an appropriate SM must be taken into account for surrogate modeling. As for pure time-series analysis, all features are time-dependent and data collection occurs at each time step. It is worth mentioning that there are various SMs available for addressing such problems, but they fall beyond the scope of the present study. On the other hand, while renewable energy sources, such as solar and wind, exhibit time-dependent characteristics, the approach to data sampling and surrogate modeling can remain consistent with the proposed methodology.

To carry out a similar procedure on any valid MGS, the following steps/guidance must be considered:

- For data sampling, uniform random sampling is a proper option. However, to be comprehensive, it is suggested to generate a large dataset. When the execution of the simulation code becomes computationally burdensome, alternative efficient methods can be employed to generate smaller datasets.
- While there are other methods for detecting/removing outliers, *Z*-score is a reliable choice for this purpose. According to our finding, a standard deviation of 2.5 can be suggested. However, when it comes to expert knowledge, the authors cannot offer a general suggestion,

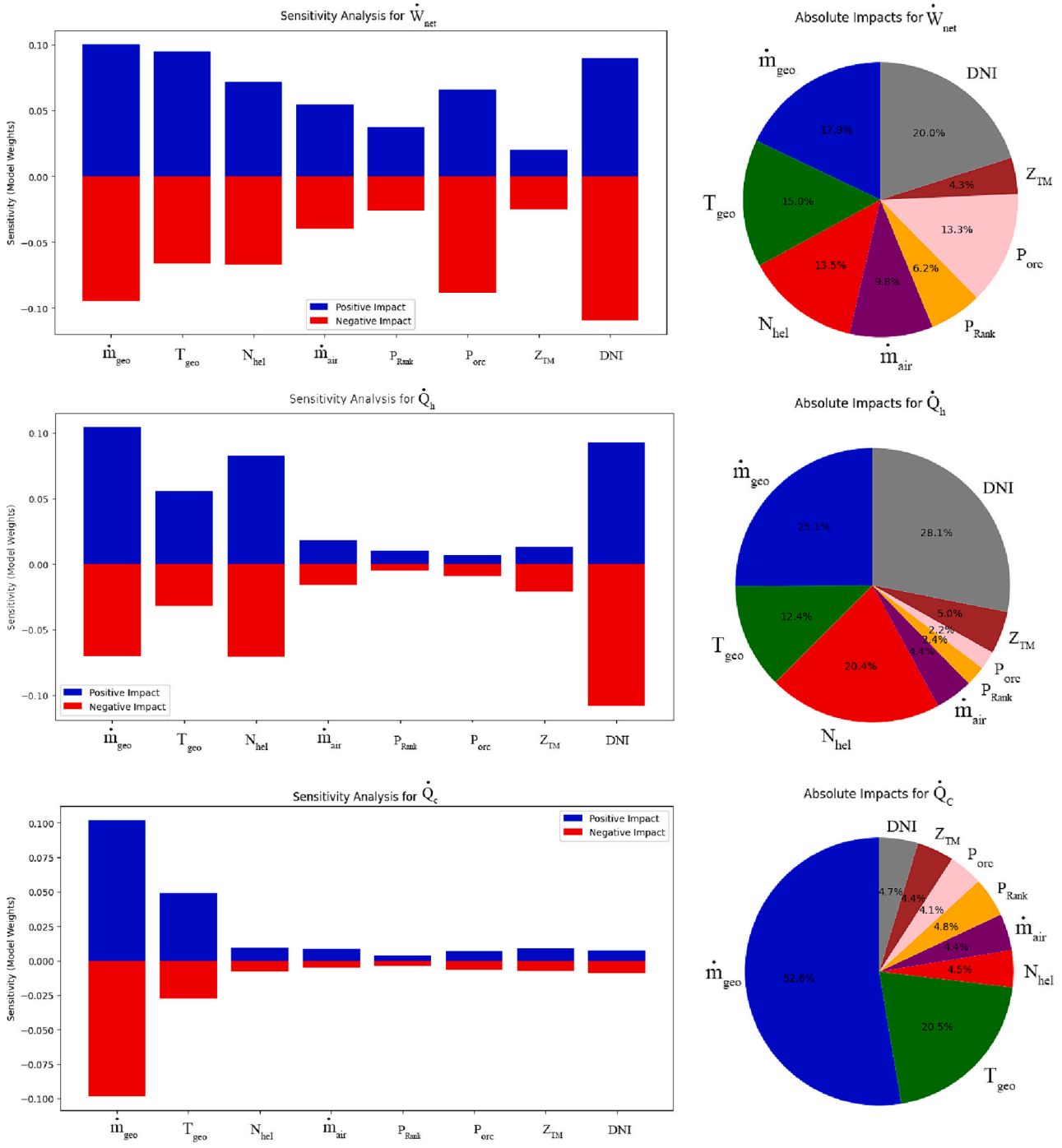


Fig. 13. (*continued*).

as this decision greatly depends on the characteristics of the given problem.

- The hyperparameters of three commonly used neural network methods (i.e., ANN, LSTM, and CNN) were tuned for the current MGS and provided in Table 2. These findings can be applied to other similar cases, though the use of a CNN is not recommended for such problems. Moreover, an ensemble model can be developed on top of the first two models, where the coefficients of the ANN and LSTM were calculated as 0.57 and 0.43, respectively.
- NSGA-II is a reliable optimization method, but the competence of other approaches can also be explored.
- For MOO, any time-dependent feature must be included with a fixed value. When more than one time-dependent feature is present, the number of times that optimization must be performed to cover the entire range of each feature would significantly increase.
- Indeed, there is no limitation on the number of objective functions for MOO; they can be as many as needed.

6. Conclusion

This study aimed to establish a pathway for developing a precise surrogate model tailored for multigeneration systems (MGSs) and to conduct surrogate-based optimization with as many objective functions as possible. For this purpose, a dataset containing 23,000 data points was generated from the simulation program we developed in our previous study [35]. Comprehensive surrogate modeling (including outlier

Table 4
Results of the six-objective optimization using eight features at four different *DNI*s.

Feature/ Objective	Unit	$DNI = 100$	$DNI = 400$	$DNI = 600$	$DNI = 900$
$\dot{m}_{geo}$	kg/s	14.99	14.94	14.99	14.98
T_{geo}	°C	162.77	158.21	164.94	164.89
N_{hel}	–	25	25	25	25
$\dot{m}_{air}$	kg/s	9.95	6.13	4.76	5.32
P_{Rank}	kPa	4976.07	2012.43	2295.65	2477.04
P_{ORC}	kPa	3459.66	3217.03	3195.70	3491.80
z_{TM}	–	0.22	0.20	0.20	0.22
Eff_{exe}	%	35.88	28.27	25.89	22.70
$TUCP$	$/GJ	17.90	27.93	29.77	31.83
$\dot{m}_{fw}$	kg/s	1.24	5.42	9.45	12.48
$\dot{W}_{net}$	kW	422.73	418.88	463.90	487.82
$\dot{Q}_c$	kW	134.25	128.88	132.50	133.62
$\dot{Q}_h$	kW	1754.79	1786.65	1953.30	2061.93

treatment, normalization, model development, and hyperparameter tuning of ANN, CNN, and LSTM as well as an ensemble model composed of the ANN and LSTM) was performed and model evaluation was carried out using R^2, *MSE*, and *MAE*. Through this step, specific suggestions were provided on how to replicate this process for other similar MGSs, as discussed in Section 5. The outcome showed that the ensemble model outperformed other models in accuracy, leading to its selection as the best surrogate model. This was also demonstrated in a case study using the meteorological data of the Reno city. In the final part of the simulation section, a sensitivity analysis was performed to deepen our understanding of the nonlinear impact of features on objectives. It was found that *DNI* was the most decisive feature for all objectives except for cooling load, which was solely driven by the mass flowrate and temperature of geothermal water. This was evident for exergy efficiency, *TUCP*, $\dot{W}_{net}$, and $\dot{Q}_h$ being mainly affected by the solar-based features, whereas the role of the geothermal-based features could not be overlooked. However, in the case of the freshwater mass flowrate, the importance of features related to geothermal energy was marginal. It is worth mentioning that the pressure of the ORC system (P_{orc}) was the most significant feature among the operation features (i.e., features unrelated to solar and geothermal energy), i.e., $\dot{m}_{air}$, P_{Rank}, and Z_{TM}.

In the next step, a six-objective optimization problem was defined and performed. Since *DNI* was a time-dependent feature, four scenarios of optimization were considered based on different levels of *DNI*, covering its entire range. The results demonstrated that the value of all objectives grew as *DNI* increased, except for exergy efficiency and $\dot{Q}_c$. This implied that as solar intensity increased, the performance and operating costs were negatively affected. Additionally, the production rates of all commodities improved, except for cooling load. The possible explanation lies in the increase of exergy destruction, which accounts for the low exergetic performance. This is coupled with the rise of production rates that, in turn, contributes to high operating costs (i.e., *TUCP*). When compared to the rate of production of other products, freshwater exhibited the most significant impact on exergy efficiency and *TUCP*. Moreover, the critical range of operation occurred between *DNI* values of 100 to 400 W/m^2, within which major changes in the performance, cost, and production rates of the MGS could be observed. At the final step, the optimal results of the present study were compared to those of the two-objective optimization obtained in our simulation-based study. It was found that the six-objective optimization could identify a better operating setting compared to the two-objective case. Consequently, every objective was improved: *TUCP* and net electricity generation achieved substantial improvements, with 12.7% and 9.5%,

Table 5
A comparison between the optimization results obtained by simulation [35] and that of the ensemble model developed in the present study.

Objective	Unit	Bi-objective optimization	Six-objective optimization	Improvement (%)
Eff_{exe}	%	25.4	25.89	1.9
$TUCP$	$/GJ	34.1	29.77	12.7
$\dot{m}_{fw}$	kg/s	9.1	9.45	3.8
$\dot{W}_{net}$	kW	423.7	463.90	9.5
$\dot{Q}_c$	kW	127.1	132.50	4.2
$\dot{Q}_h$	kW	1846	1953.30	5.8

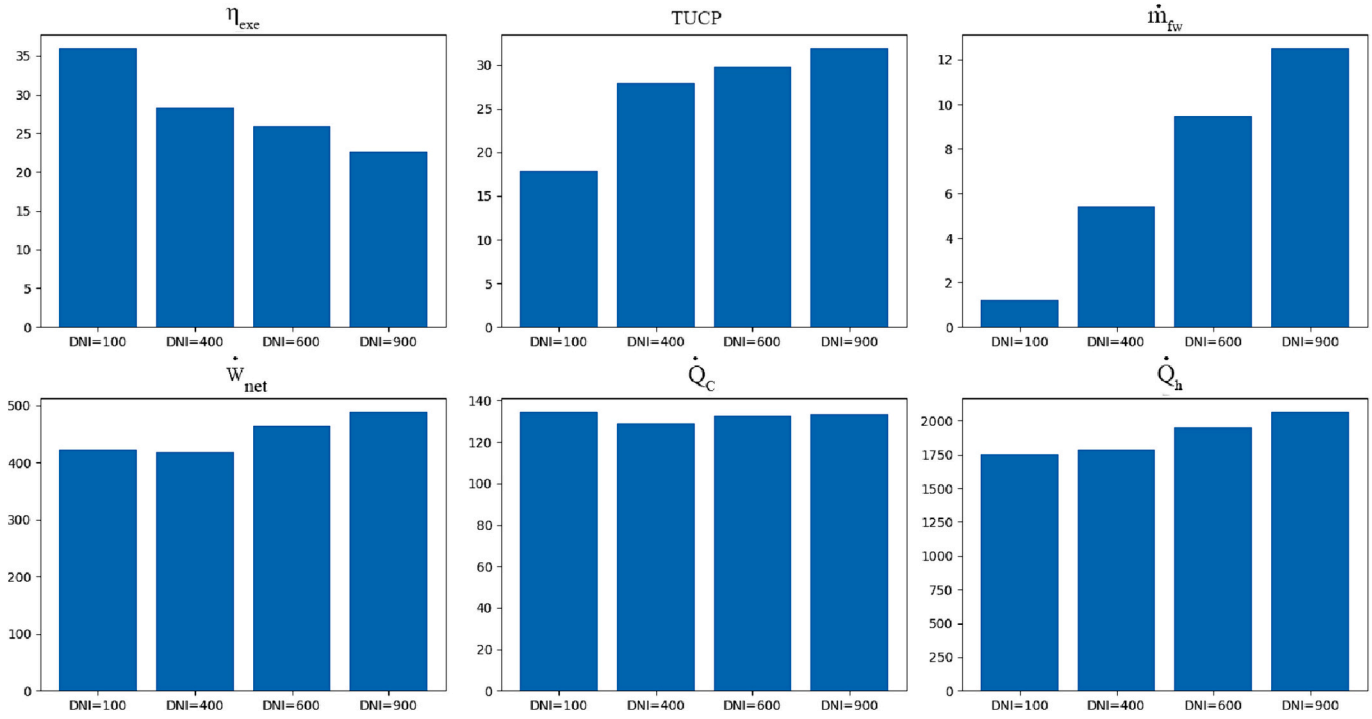


Fig. 14. Optimal objective values at four different *DNI* values.

respectively, while exergy efficiency and freshwater mass flowrate showed least enhancement, with 1.9% and 3.8%, respectively.

CRediT authorship contribution statement

Parviz Ghafariasl: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alireza Mahmoudan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mahmoud Mohammadi:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Aria Nazarparvar:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Siamak Hoseinzadeh:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Mani Fathali:** Writing – original draft, Visualization, Validation, Investigation, Formal analysis, Data curation, Conceptualization. **Shing Chang:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Masoomeh Zeinalnezhad:** Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Davide Astiaso Garcia:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition.

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.

References

[1] Yoro KO, Daramola MO, Sekoai PT, Wilson UN, Eterigho-Ikelegbe O. Update on current approaches, challenges, and prospects of modeling and simulation in renewable and sustainable energy systems. Renew Sustain Energy Rev 2021;150:111506. https://doi.org/10.1016/j.rser.2021.111506.

[2] Esmaeilion F, Soltani M, Dusseault MB, Rosen MA. Performance investigation of a novel polygeneration system based on liquid air energy storage. Energ Conver Manage 2023;277:116615. https://doi.org/10.1016/j.enconman.2022.116615.

[3] Nojedehi P, Heidari M, Ataei A, Nedaei M, Kurdestani E. Environmental assessment of energy production from landfill gas plants by using long-range energy alternative planning (LEAP) and IPCC methane estimation methods: a case study of Tehran. Sustain Energy Technol Assess 2016;16:33–42. https://doi.org/10.1016/j.seta.2016.04.001.

[4] Esmaeilion F, Soltani M. Technical and economic assessments of a novel multigeneration system based on desalination and liquid air energy storage. Desalination 2024:117497. https://doi.org/10.1016/j.desal.2024.117497.

[5] Hoseinzadeh S, Assareh E, Riaz A, Lee M, Astiaso Garcia D. Ocean thermal energy conversion (OTEC) system driven with solar-wind energy and thermoelectric based on thermo-economic analysis using multi-objective optimization technique. Energy Rep 2023;10:2982–3000. https://doi.org/10.1016/j.egyr.2023.09.131.

[6] Subramanian ASR, Gundersen T, Adams TA. Modeling and simulation of energy systems: a review. Processes 2018;6(12):238. https://doi.org/10.3390/pr6120238.

[7] Koziel S, Leifsson L. Surrogate-based modeling and optimization. Springer 2013. https://doi.org/10.1007/978-1-4614-7551-4.

[8] Bouhlel MA, Hwang JT, Bartoli N, Lafage R, Morlier J, Martins JRRA. A Python surrogate modeling framework with derivatives. Adv Eng Softw 2019;135:102662. https://doi.org/10.1016/j.advengsoft.2019.03.005.

[9] Khan AH, Omar S, Mushtary N, Verma R, Kumar D, Alam S. Digital twin and artificial intelligence incorporated with surrogate modeling for hybrid and sustainable energy systems. Handb Smart Energy Syst 2023:2837–59. https://doi.org/10.1007/978-3-030-97940-9_147.

[10] Alizadeh R, Allen JK, Mistree F. Managing computational complexity using surrogate models: a critical review. Res Eng Des 2020;31:275–98. https://doi.org/10.1007/s00163-020-00336-7.

[11] Eason J, Cremaschi S. Adaptive sequential sampling for surrogate model generation with artificial neural networks. Comput Chem Eng 2014;68:220–32. https://doi.org/10.1016/j.compchemeng.2014.05.021.

[12] Rätz M, Javadi AP, Baranski M, Finkbeiner K, Müller D. Automated data-driven modeling of building energy systems via machine learning algorithms. Energ Buildings 2019;202:109384. https://doi.org/10.1016/j.enbuild.2019.109384.

[13] Nyitrai T, Virág M. The effects of handling outliers on the performance of bankruptcy prediction models. Socioecon Plann Sci 2019;67:34–42. https://doi.org/10.1016/j.seps.2018.08.004.

[14] Salcedo-Sanz S, Cornejo-Bueno L, Prieto L, Paredes D, García-Herrera R. Feature selection in machine learning prediction systems for renewable energy applications. Renew Sustain Energy Rev 2018;90:728–41. https://doi.org/10.1016/j.rser.2018.04.008.

[15] Alruqi M, Sharma P. Biomethane production from the mixture of sugarcane vinasse, solid waste and spent tea waste: a Bayesian approach for hyperparameter optimization for Gaussian process regression. Fermentation 2023;9:120. https://doi.org/10.3390/fermentation9020120.

[16] Feurer M, Hutter F. Hyperparameter optimization. In: Automated machine learning: Methods, systems, challenges; 2019. p. 3–33. https://doi.org/10.1007/978-3-030-05318-5_1.

[17] Sharma P, Jain A, Bora BJ, Balakrishnan D, Show PL, Ramaraj R, et al. Application of modern approaches to the synthesis of biohydrogen from organic waste. Int J Hydrogen Energy 2023;48(55):21189–213. https://doi.org/10.1016/j.ijhydene.2023.03.029. 30.

[18] Botchkarev A. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. ArXiv Prepr ArXiv:1809.03006. 2018. https://doi.org/10.48550/arXiv.1809.03006.

[19] Gong YJ, Chen WN, Zhan ZH, Zhang J, Li Y, Zhang Q, et al. Distributed evolutionary algorithms and their models: a survey of the state-of-the-art. Appl Soft Comput 2015;34:286–300. https://doi.org/10.1016/j.asoc.2015.04.061.

[20] Allmendinger R, Jaszkiewicz A, Liefooghe A, Tammer C. What if we increase the number of objectives? Theoretical and empirical implications for many-objective combinatorial optimization. Comput Oper Res 2022;145:105857. https://doi.org/10.1016/j.cor.2022.105857.

[21] Kim SH, Boukouvala F. Surrogate-based optimization for mixed-integer nonlinear problems. Comput Chem Eng 2020;140:106847. https://doi.org/10.1016/j.compchemeng.2020.106847.

[22] Curry DM, Dagli CH. Computational complexity measures for many-objective optimization problems. Procedia Comput Sci 2014;36:185–91. https://doi.org/10.1016/j.procs.2014.09.077.

[23] Amsallem D, Zahr M, Choi Y, Farhat C. Design optimization using hyper-reduced-order models. Struct Multidiscip Optim 2015;51:919–40. https://doi.org/10.1007/s00158-014-1183-y.

[24] Bahlawan H, Morini M, Pinelli M, Spina PR, Venturini M. Simultaneous optimization of the design and operation of multi-generation energy systems based on life cycle energy and economic assessment. Energ Conver Manage 2021;249:114883. https://doi.org/10.1016/j.enconman.2021.114883.

[25] Zhou J, Ali MA, Sharma K, Almojil SF, Alizadeh AA, Almohana AI, et al. Using machine learning to predict performance of two cogeneration plants from energy, economic, and environmental perspectives. Int J Hydrogen Energy 2022. https://doi.org/10.1016/j.ijhydene.2022.12.018.

[26] Assareh E, Hoseinzadeh S, Agarwal N, Delpisheh M, Dezhdar A, Feyzi M, et al. A transient simulation for a novel solar-geothermal cogeneration system with a selection of heat transfer fluids using thermodynamics analysis and ANN intelligent (AI) modeling. Appl Therm Eng 2023;231:120698. https://doi.org/10.1016/j.applthermaleng.2023.120698.

[27] Pinto G, Deltetto D, Capozzoli A. Data-driven district energy management with surrogate models and deep reinforcement learning. Appl Energy 2021;304:117642. https://doi.org/10.1016/j.apenergy.2021.117642.

[28] Riboldi L, Nord LO. Offshore power plants integrating a wind farm: design optimisation and techno-economic assessment based on surrogate modelling. Processes 2018;6:249. https://doi.org/10.3390/pr6120249.

[29] Wang X, Liu Y, Zhao J, Liu C, Liu J, Yan J. Surrogate model enabled deep reinforcement learning for hybrid energy community operation. Appl Energy 2021;289:116722. https://doi.org/10.1016/j.apenergy.2021.116722.

[30] Mazzeo D, Herdem MS, Matera N, Bonini M, Wen JZ, Nathwani J, et al. Artificial intelligence application for the performance prediction of a clean energy community. Energy 2021;232:120999. https://doi.org/10.1016/j.energy.2021.120999.

[31] Abokersh MH, Vallès M, Cabeza LF, Boer D. A framework for the optimal integration of solar assisted district heating in different urban sized communities: a robust machine learning approach incorporating global sensitivity analysis. Appl Energy 2020;267:114903. https://doi.org/10.1016/j.apenergy.2020.114903.

[32] Bornatico R, Hüssy J, Witzig A, Guzzella L. Surrogate modeling for the fast optimization of energy systems. Energy 2013;57:653–62. https://doi.org/10.1016/j.energy.2013.05.044.

[33] Jiang J, Yu H, Song G, Zhao J, Zhao K, Ji H, et al. Surrogate model assisted multi-criteria operation evaluation of community integrated energy systems. Sustain Energy Technol Assess 2022;53:102656. https://doi.org/10.1016/j.seta.2022.102656.

[34] Tariq R, Cetina-Quiñones AJ, Cardoso-Fernández V, Daniela-Abigail HL, Soberanis ME, Bassam A, et al. Artificial intelligence assisted technoeconomic optimization scenarios of hybrid energy systems for water management of an isolated community. Sustain Energy Technol Assess 2021;48:101561. https://doi.org/10.1016/j.seta.2021.101561.

[35] Mohammadi M, Mahmoudan A, Nojedehi P, Hoseinzadeh S, Fathali M, Garcia DA. Thermo-economic assessment and optimization of a multigeneration system powered by geothermal and solar energy. Appl Therm Eng 2023;230:120656. https://doi.org/10.1016/j.applthermaleng.2023.120656.
[36] Garud SS, Karimi IA, Kraft M. Smart sampling algorithm for surrogate model development. Comput Chem Eng 2017;96:103–14. https://doi.org/10.1016/j.compchemeng.2016.10.006.
[37] Kamath C. Intelligent sampling for surrogate modeling, hyperparameter optimization, and data analysis. Mach Learn Appl 2022;9:100373. https://doi.org/10.1016/j.mlwa.2022.100373.
[38] Moradi H, Zhoulideh M, Ghafariasl M. Tunable and ultrasensitive sensor covering terahertz to telecommunication range based on a Fabry–Perot interference of graphene plasmonic waves. Opt Commun 2023;542:129592. https://doi.org/10.1016/j.optcom.2023.129592.
[39] Jafari B, Gholizadeh E, Jafari B, Zhoulideh M, Adibnia E, Ghafariasl M, et al. Highly sensitive label-free biosensor: graphene/CaF2 multilayer for gas, cancer, virus, and diabetes detection with enhanced quality factor and figure of merit. Sci Rep 2023;13:16184. https://doi.org/10.1038/s41598-023-43480-5.
[40] Fei N, Gao Y, Lu Z, Xiang T. Z-score normalization, hubness, and few-shot learning. In: InProceedings of the IEEE/CVF international conference on computer vision; 2021. p. 142–51.
[41] Patro SGK, sahu KK.. Normalization: a preprocessing stage. IARJSET 2015:20–2. https://doi.org/10.17148/iarjset.2015.2305.
[42] Panchal S, Mathew M, Dincer I, Agelin-Chaab M, Fraser R, Fowler M. Thermal and electrical performance assessments of lithium-ion battery modules for an electric vehicle under actual drive cycles. Electr Pow Syst Res 2018;163:18–27. https://doi.org/10.1016/j.epsr.2018.05.020.
[43] Panchal S, Dincer I, Agelin-Chaab M, Fraser R, Fowler M. Design and simulation of a lithium-ion battery at large C-rates and varying boundary conditions through heat flux distributions. Measurement 2018;116:382–90. https://doi.org/10.1016/j.measurement.2017.11.038.
[44] Panchal S, Dincer I, Agelin-Chaab M, Fraser R, Fowler M. Experimental investigation and simulation of temperature distributions in a 16Ah-LiMnNiCoO 2 battery during rapid discharge rates. Heat Mass Transf 2017;53:937–46. https://doi.org/10.1007/s00231-016-1870-x.
[45] Lai W, Kuang M, Wang X, Ghafariasl P, Sabzalian MH, Lee S. Skin cancer diagnosis (SCD) using artificial neural network (ANN) and improved gray wolf optimization (IGWO). Sci Rep 2023;13:19377. https://doi.org/10.1038/s41598-023-45039-w.
[46] Chen J, Manivanan M, Duque J, Kollmeyer P, Panchal S, Gross O, et al. A convolutional neural network for estimation of lithium-ion battery state-of-health during constant current operation. In: In2023 IEEE transportation electrification conference & expo (ITEC); 2023. p. 1–6. https://doi.org/10.1109/ITEC55900.2023.10186914.
[47] Haar LV, Elvira T, Ochoa O. An analysis of explainability methods for convolutional neural networks. Eng Appl Artif Intel 2023;117:105606. https://doi.org/10.1016/j.engappai.2022.105606.
[48] Chen J, Kollmeyer P, Panchal S, Masoudi Y, Gross O, Emadi A. Sequence training and data shuffling to enhance the accuracy of recurrent neural network based battery voltage models. SAE Technical Paper; 2024.
[49] Van Houdt G, Mosquera C, Nápoles G. A review on the long short-term memory model. Artif Intell Rev 2020;53:5929–55. https://doi.org/10.1007/s10462-020-09838-1.
[50] Irlbeck T, Mayer F. Brute forcevol. 6. Springer; 2018.
[51] Karch J. Improving on adjusted R-squared. Collabra Psychol 2020:6. https://doi.org/10.1525/collabra.343.
[52] Robeson SM, Willmott CJ. Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. PloS One 2023;18:e0279774. https://doi.org/10.1371/journal.pone.0279774.
[53] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj Comput Sci 2021;7:1–24. https://doi.org/10.7717/PEERJ-CS.623.
[54] Mahmoudan A, Esmaeilion F, Hoseinzadeh S, Soltani M, Ahmadi P, Rosen M. A geothermal and solar-based multigeneration system integrated with a TEG unit: development, 3E analyses, and multi-objective optimization. Appl Energy 2022;308:118399. https://doi.org/10.1016/j.apenergy.2021.118399.
[55] Mahmoudan A, Samadof P, Hosseinzadeh S, Garcia DA. A multigeneration cascade system using ground-source energy with cold recovery: 3E analyses and multi-objective optimization. Energy 2021;233:121185. https://doi.org/10.1016/j.energy.2021.121185.
[56] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 2002;6:182–97. https://doi.org/10.1109/4235.996017.
[57] Sejr JH, Schneider-Kamp A. Explainable outlier detection: what, for whom and why? Mach Learn Appl 2021;6:100172. https://doi.org/10.1016/j.mlwa.2021.100172.
[58] Aggarwal CC. Outlier analysis second edition. IBM T J Watson Res Cent 2016;24:379–84. https://doi.org/10.1007/978-3-319-47578-3.
[59] Priyadarshini I, Cotton C. A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis. J Supercomput 2021;77:13911–32. https://doi.org/10.1007/s11227-021-03838-w.
[60] Jin M, Liao Q, Patil S, Abdulraheem A, Al-Shehri D, Glatz G. Hyperparameter tuning of artificial neural networks for well production estimation considering the uncertainty in initialized parameters. ACS Omega 2022;7:24145–56. https://doi.org/10.1021/acsomega.2c00498.
[61] Bochinski E, Senst T, Sikora T. Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms. In: 2017 IEEE international conference on image processing (ICIP); 2017. p. 3924–8. https://doi.org/10.1109/ICIP.2017.8297018.
[62] Nakisa B, Rastgoo MN, Rakotonirainy A, Maire F, Chandran V. Long short term memory hyperparameter optimization for a neural network based emotion recognition framework. IEEE Access 2018;6:49325–38. https://doi.org/10.1109/ACCESS.2018.2868361.
[63] NSRDB. n.d, https://nsrdb.nrel.gov/data-viewer (accessed January 3, 2024).
[64] Atkinson DJ, Huxtable DD. Geothermal district heating for Renovol. 13. Nevada, U.S.A: Geothermics; 1984. p. 265–79. https://doi.org/10.1016/0375-6505(84)90020-8.
[65] Ayling BF. 35 Years of geothermal power generation in Nevada, USA: A review of field development, generation, and production histories. In: InProceedings of the 45t h workshop on geothermal reservoir engineering; 2020. p. 1–12.
[66] Rivord J, Saito L, Miller G, Stoddard SS. Modeling contaminant spills in the Truckee River in the Western United States. J Water Resour Plan Manag 2014;140:343–54. https://doi.org/10.1061/(asce)wr.1943-5452.0000338.