Environmental Challenges 13 (2023) 100779

Contents lists available at ScienceDirect

Environmental Challenges

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

Machine learning approaches for real-time forecasting of solar still distillate output

Deepak Kumar Murugan[a,*], Zafar Said[b,h], Hitesh Panchal[c], Naveen Kumar Gupta[d,i], Sekar Subramani[e], Abhinav Kumar[f], Kishor Kumar Sadasivuni[g,j,*]

[a] Department of Mechanical Engineering, Velammal Engineering College, Chennai, India
[b] Department of Sustainable and Renewable Energy Engineering, University of Sharjah, United Arab Emirates
[c] Department of Mechanical Engineering, Government Engineering College Patan, Gujarat, India
[d] Department of Mechanical Engineering, GLA University, Mathura, India.
[e] Department of Mechanical Engineering, Rajalakshmi Engineering College, Chennai, India
[f] Department of Nuclear and Renewable Energy, Ural Federal University Named after the First President of Russia Boris Yeltsin, Ekaterinburg 620002, Russia
[g] Centre for Advanced Materials, Qatar University, Qatar
[h] Department of Industrial and Mechanical Engineering, Lebanese American University (LAU), Byblos, Lebanon
[i] Department of Mechanical Engineering, Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India
[j] Department of Mechanical and Industrial Engineering, Qatar University, PO Box 2713, Doha, Qatar

ARTICLE INFO

Keywords:
Solar still
Machine learning techniques
Decision tree modelling
Predictive models
Productivity estimation

ABSTRACT

Solar stills provide a promising avenue for freshwater production in regions grappling with water scarcity, especially remote locales. However, their efficiency is often constrained by the variable climatic conditions. Conventional prediction methods fall short in consistently forecasting the yield, leaving a significant gap in optimizing solar still operations. Recognizing this, the introduction of machine learning becomes pivotal. With a robust predictive model, operators can avoid inefficiencies, inconsistent outputs, and sub-optimal resource utilization. The primary objective of this research is to determine the most suitable machine learning model tailored for predicting solar still output under specific environmental conditions. This research work assessed various machine learning models, including linear regression, decision trees, random forest, support vector machines, and multilayer perceptron. Evaluation metrics encompassed Mean Absolute Error (MAE), cross-validation, grid search, and randomized search techniques. Our results identified the Decision Tree model, registering a MAE of 5.43 and 5.74 through random and grid search methods, respectively, as the preeminent predictor for our dataset. This machine learning-centric methodology elevates the precision of solar still output predictions and paves the way for enhanced solar still designs and superior optimization of solar energy conversion mechanisms.

1. Introduction

Water scarcity, increasingly exacerbated by factors like droughts, over-exploitation of aquifers, and burgeoning population demands, has emerged as a critical global concern (Atteya & Abbas, 2023; Panchal et al. 2019). This crisis transcends mere water shortages, manifesting in catastrophic ramifications like agricultural downturns, economic setbacks, and heightened resource competitions that occasionally escalate into conflicts (Panchal, 2017). A crucial aspect of this challenge lies in the health sector: inadequate access to potable water can be a breeding ground for waterborne ailments such as cholera, hepatitis, and typhoid, with studies revealing diseases like these leading to fatalities, notably in children (Khatod et al., 2022). The global narrative is grim, with over a billion people deprived of clean water, culminating in an alarming annual death toll of 3.4 million due to preventable water-associated diseases. Although there are numerous desalination methods to tackle the water issue by converting seawater to freshwater, solar stills are an excellent alternative because they are cheap and easy to maintain. (Mahmoud et al., 2018, Panchal et al., 2021a, Elgendi et al., 2023, Sibagariang et al., 2022). They do not require any external energy source, making them suitable for remote and off-grid communities (Panchal et al. 2017; Pansal et al. 2020; Panchal et al. 2020). In addition to being more eco-friendly than alternative desalination methods, solar stills are also more cost-effective. (Ghandourah et al., 2022; Panchal and Shah, 2014; Panchal et al. 2020). In contrast, several other desalination

* Corresponding authors.
E-mail addresses: deepu.energy@gmail.com (D.K. Murugan), Kishorkumars@qu.edu.qa (K.K. Sadasivuni).

https://doi.org/10.1016/j.envc.2023.100779
Received 1 August 2023; Received in revised form 3 October 2023; Accepted 13 October 2023
Available online 14 October 2023

Nomenclature

ANN	Artificial Neural Network
GI	Galvanized Iron
DT	Decision Tree
LR	Linear Regression
MAE	Mean Absolute Error
MLP	Multi-layer Perceptron
RF	Random Forest
RMSE	Root Mean Square Erroe
SVM	Support Vector Machine

processes necessitate substantial energy input, i.e. RO requires high pressure to push water through the membrane, which uses energy and produces brine. (Shalaby et al., 2022). Safe drinking water can be produced with the help of solar still, which uses solar energy to distil water. (Aglan et al., 2021). A solar still works by heating water, collecting the resulting water vapour, and then condensing it into liquid form (Mevada et al. 2020). This process effectively removes any impurities present in the original water source, such as bacteria, salts, and other dissolved solids. Solar stills can be constructed in various designs, but the box-type solar still is the most common (Panchal and Thakkar, 2016). This design consists of a water-containing container covered with a transparent material, such as plastic or glass. Sunlight passing through the transparent material heats the water in the container, causing it to evaporate. When the water vapour rises, it condenses on the underside of the transparent material, which can subsequently be used to collect the water in its pure form (Panchal, 2011).

A solar still's efficiency relies on its type, transparent material quality, angle and orientation, and air temperature and humidity. (Zhao et al., 2022, Peng at al., 2022, Isah et al., 2022; Panchal 2010). With the right design and optimal operating conditions, a solar still can produce a significant amount of purified water per unit area. Off-grid and rural areas often have limited access to clean water, but solar stills provide a low-cost, long-term solution. The technology has been used to provide potable water in many regions worldwide and can be integrated with other water treatment methods, such as reverse osmosis, to improve water quality. The efficiency of this clean energy system could be improved by the limited availability of solar energy during the day. A multitude of studies have been conducted to tackle this critical issue (Wei et al., 2023; Lisboa et al., 2022; Chen & Xie, 2022; Modi & Gamit, 2022; Murugan et al., 2021; Sharshir et al., 2022). Despite their widespread use, conventional solar still typically only achieves an energy efficiency of around 30%, leaving considerable scope for further enhancements (Shatar et al., 2023; Rabishokr & Daghigh, 2023; Shoeibi et al., 2023; Panchal et al., 2021b). The output of a solar still is driven by the temperature differential between the evaporating and condensing surfaces. Recent advances in material science have led to the use of novel materials with superior solar absorption capabilities to elevate the temperature of the evaporating surface (Ebrahimpour & Shafi, 2022; Shoeibi et al., 2022; Gandhi et al., 2022; Chauhan et al., 2022; Peng et al., 2021). Additionally, innovative techniques have been proposed to decrease the temperature of the condensing surface (Amiri, 2022; Sambare et al., 2022; Lauvandy et al., 2022; Dubey et al., 2022; essa et al. 2021).

Machine learning is a subfield of artificial intelligence that develops algorithms and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed (Entezari et al., 2023). The growth of this field has been significant in recent years and holds tremendous potential for revolutionizing the optimization of energy systems (Liu et al., 2023; McLaughlin and Choi, 2023; Behzadi & Sadrizadeh, 2023; Kapp et al., 2023; Sohani et al., 2022; Alabi et al., 2022). The main types of machine learning algorithms are supervised learning, unsupervised learning, and reinforcement learning (Janiesch et al., 2021). In the context of solar stills, machine learning has the potential to enhance various aspects of the system's performance, efficiency, and maintenance (Wang et al., 2021; Zayed et al., 2022). For instance, machine learning models can be employed to predict the optimal inclination and orientation of the still based on weather conditions, which would result in improved efficiency.

Moreover, machine learning can be used to optimize the design of the still through the selection of materials and the determination of its size, thus improving performance (Rashidi et al., 2022; Maddah et al., 2020). In addition, machine learning algorithms can monitor and control the water quality and output of the solar still, ensuring that the produced water is potable. Furthermore, machine learning can also assist in the optimization of the cleaning and maintenance processes of the still, thereby extending its lifespan and reducing costs.

In this research study, we have a regression problem where the goal is to predict the distillate output of a solar still. The dataset has several features which may be used to predict the output, and several models have been considered. The Decision Tree model is a popular choice for regression problems because of its interpretability and ability to handle non-linear relationships. It works by recursively splitting the dataset into smaller subsets, and at each split, it selects the feature and threshold that results in the most homogeneous subsets. The final result is a tree of decisions that can be used to predict the output for new examples. This model is simple to understand, interpret and visualize, a major advantage of this problem. On the other hand, Random Forest is an ensemble method combining multiple decision trees to make predictions. The idea behind this method is to randomly select a subset of features and build a decision tree using them. This process is repeated multiple times, and the final predictions are made by averaging the predictions of all the trees. This model is more robust to overfitting, a common problem in decision trees. Multi-layer Perceptron (MLP) is an artificial neural network widely used for regression and classification problems. This model comprises multiple layers of perceptrons, which are simple feedforward neural networks. MLP can learn non-linear relationships and is good at handling high-dimensional datasets. Support Vector Machine (SVM) is a supervised learning algorithm that can be used for classification and regression problems. In this case, the goal is to find a hyperplane that maximizes the margin between the dataset and the hyperplane. This model is robust to overfitting and good at handling high-dimensional datasets. In summary, we have chosen these models based on their ability to handle non-linear relationships, robustness to overfitting, interpretability and suitability for high-dimensional datasets.

In this study, we harness machine learning models to accurately predict the distillate output of solar still, considering crucial environmental factors such as ambient temperature, humidity, wind speed, and solar radiation. Using a comprehensive dataset of these factors against actual distillate outputs, we trained and critically assessed models like Decision Trees, Random Forests, MLPs, and SVMs. Our systematic evaluation methods, including cross-validation and hyperparameter tuning, led us to identify the most effective model. The implications of our findings resonate beyond mere predictions, potentially guiding solar still operational enhancements and efficient freshwater resource management. This research not only underscores the competence of machine learning in this domain but also lays a foundation for future explorative endeavours.

2. Machine learning models

The models were chosen based on their capability to handle different types of datasets, complexity, and interpretability. These models were selected because they are commonly used in the literature and have proven successful in various applications. Linear regression is used as a simple base model, Decision tree and Random Forest are used for non-linear relationships, MLP is used for complex non-linear datasets, and

SVM is used for high dimensional datasets.

2.1. Linear regression

Linear Regression models the association between a dependent variable and one or more independent variables. Fig. 1 presents a diagram depicting the structure of the multiple linear regression model. It seeks the most optimal linear connection between the dependent and independent variables. The equation of a simple linear regression line is $Y = mX + b$ where Y is the target variable, X is the predictor variable, m is the slope, and b is the y-intercept. It is a simple and easy-to-use model suitable for problems with a linear relationship between the independent and dependent variables. It is a good starting point for solving regression problems and can provide a baseline for comparison with more complex models.

2.2. Decision tree

Decision Tree is a tree-based model that recursively splits the dataset into smaller subsets based on the values of the features. It starts with the root node, which represents the entire dataset. The root node is split into two or more child nodes based on the values of a feature. The process is repeated for each child node until it reaches the leaf nodes, representing the final predictions, as shown in Fig. 2. The decision tree algorithm builds a tree by repeatedly splitting the dataset to maximise the subsets' homogeneity. It is a simple and interpretable model that can handle linear and non-linear relationships. It is easy to visualize and understand the decision-making process of a decision tree, which makes it useful for understanding the relationship between the independent and dependent variables.

2.3. Random forest

Random Forest is an ensemble method that combines multiple decision trees to make predictions, as shown in Fig. 3. The idea behind this method is to randomly select a subset of features and build a decision tree using them. This process is repeated multiple times, and the final predictions are made by averaging the predictions of all the trees. It is an ensemble method that combines multiple decision trees to make predictions. It is more robust to overfitting than a single decision tree and often performs better than a single decision tree.

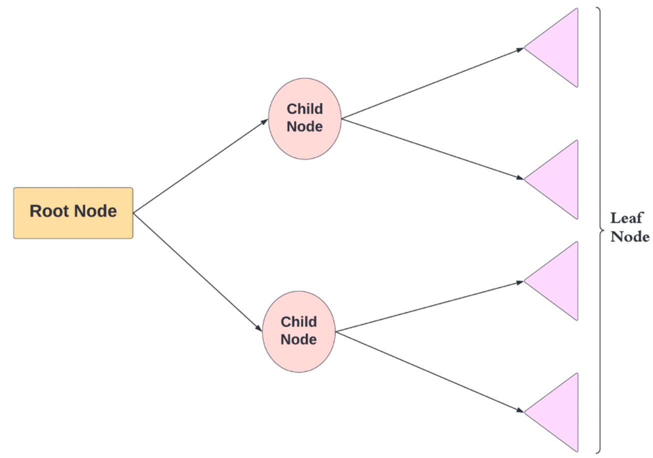

Fig. 2. Graphical representation of Decision Tree model (Otero et al., 2012).

2.4. Multi-layer perceptron

Fig. 4 illustrates the architecture of a Multi-layer Perceptron (MLP), a class of artificial neural network that has garnered significance in regression and classification problems. The depicted structure consists of three primary layers:

(a) Input Layer: This layer, composed of blue nodes marked "t," represents the initial data points or features fed into the neural network. Each node stands for a specific feature or input variable.

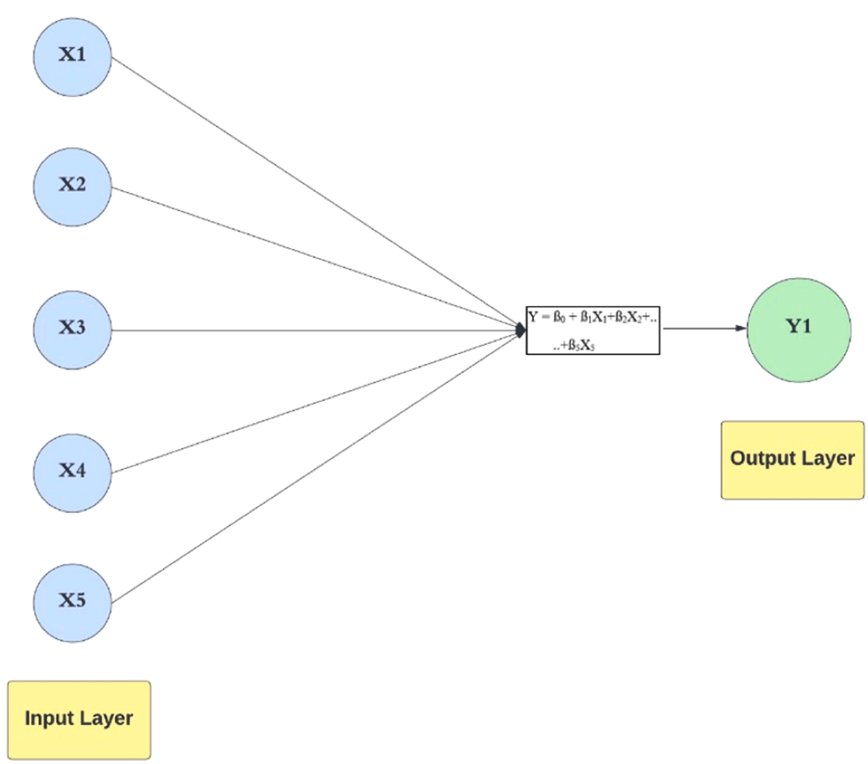

Fig. 1. Graphical representation of the multiple linear regression model (Yang et al., 2012).

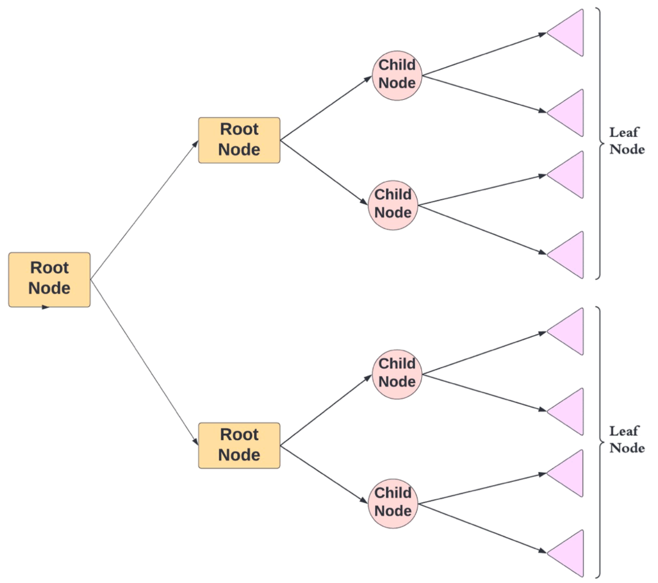

Fig. 3. Graphical representation of the Random Forest model (Cavusoglu, 2019).

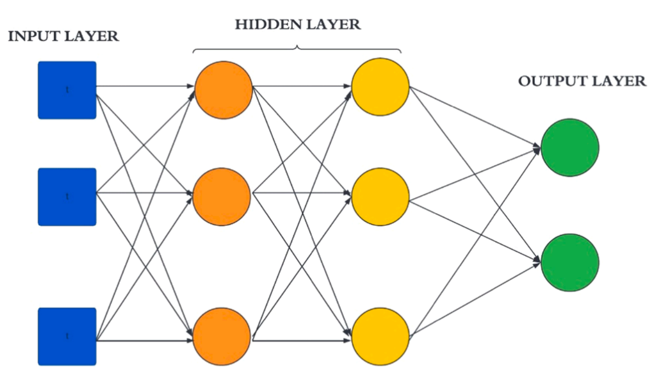

Fig. 4. Graphical representation of the Multi-Layer Perceptron model (Encisco & Zingaretti, 2019).

(b) Hidden Layer: Central to the MLP are the hidden layers, shown here as orange nodes. These intermediate layers, equipped with activation functions, enable the MLP to capture and learn intricate non-linear relationships within the data. They play a pivotal role in the MLP's generalisation ability from the training data.
(c) Output Layer: The green nodes signify the output layer. Depending on the application, this layer can produce a singular output (regression problems) or multiple outputs (classification tasks).

The interconnectedness of the nodes, signified by the lines, denotes the weighted relationships, with each connection having a specific weight. The strength and direction (positive or negative) of these weights are learned during training, enabling the MLP to make accurate predictions on new, unseen data. Notably, the ability of MLPs to handle high-dimensional datasets and discern non-linear patterns makes them an invaluable tool in machine learning applications.

2.5. Support vector machine

Fig. 5 presents the structural representation of a kernel-based Support Vector Machine (SVM). SVM, a prominent supervised learning technique, is adept at tackling both classification and regression challenges. In the depicted architecture:

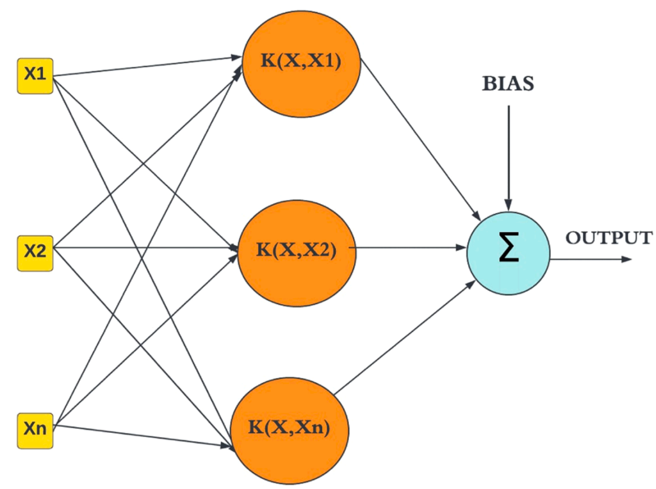

Fig. 5. Graphical representation of the Support Vector Machine model (Liu et al., 2019).

Input Nodes (X1, X2, ... Xn): These yellow nodes signify individual data points or features introduced into the SVM.

Kernel Functions (K(X, X1), K(X, X2), ... K(X, Xn)): The orange nodes represent the kernel transformations, a pivotal component of SVMs. Kernel functions enable SVM to operate in a transformed feature space, allowing it to handle non-linear relationships.

Summation (Σ) and Bias: Before arriving at the final output, the weighted sum of the kernel-transformed inputs, along with the bias, undergoes summation in the cyan node.

Output: The SVM's prediction or classification result emerges from this structure based on the weighted summation and bias.

A distinguishing feature of SVMs is their pursuit of the optimal hyperplane, which yields the maximal margin between different classes. This characteristic lends SVM its resistance to overfitting, particularly when navigating high-dimensional datasets.

3. System description & methodology

A conventional single basin single slope solar still is considered for the experiment in this research work. The schematic line diagram is shown in Fig. 6a. The still has a base area of 0.49 m^2. The solar still is fabricated using a galvanized iron (GI) basin with a thickness of 3 mm, a black painted surface to absorb more solar radiation, a top cover made of 5mm thick glass, an inlet for feedwater, a drain channel for washing away the feedwater, and a collector channel for collecting the freshwater. The use of a galvanized iron basin was chosen for its durability and cost-effectiveness. Galvanized iron has good corrosion resistance, which is important in outdoor experimental research work. It has a relatively low cost compared to other materials, such as stainless steel, making it a suitable choice for this experimental setup. The black paint applied on the basin's surface serves as an absorber of solar radiation. The black colour absorbs a higher amount of solar radiation than other colours, increasing the solar still's efficiency. The top cover of the still is made of 5 mm thick glass. Glass was chosen for its transparency and ability to allow solar radiation to pass through while keeping the inside of the still insulated. The inlet for feedwater is designed to introduce the feedwater into the still. The drain channel is placed at the bottom of the still to wash away the feedwater, which prevents contamination of the distillate. Finally, the collector channel is provided to collect the distilled water. In conclusion, the experimental setup described in this paper is designed to utilize solar radiation to distil water efficiently. The use of a galvanized iron basin, black paint, glass top cover, inlet, drain channel and collector channel were chosen based on their durability, cost-effectiveness and efficiency in increasing the solar still

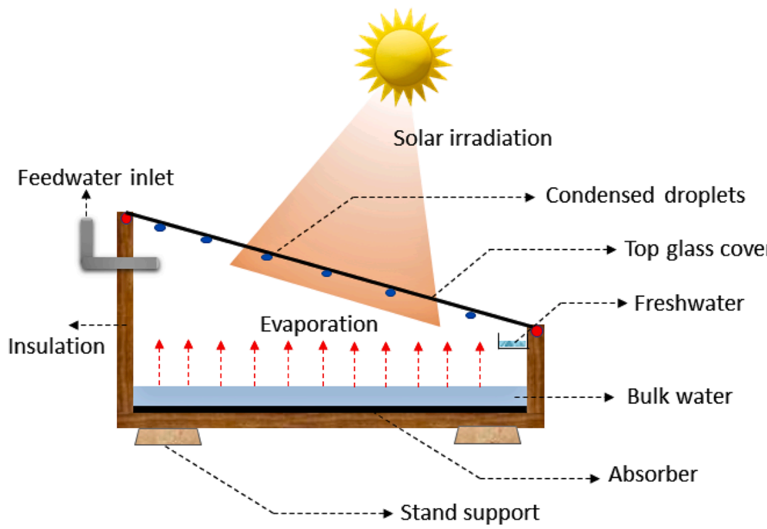

Fig. 6a. Schematic diagram of the experimental setup.

performance. The photograph of the experimental setup is shown in Fig. 6b.

The amount of distillate produced by a standard single basin single slope solar still in Chennai (13.0827°N, 80.2707°E), India, was the subject of a thorough experimental examination. The study was conducted in April, May, and June 2022 to construct a reliable machine-learning model to estimate the solar still's distillate yield. Over 90 days, sun irradiance, basin temperature, water temperature, glass cover temperature, and wind velocity were measured daily between 0800 and 1800 hours. The study collected a large amount of data meticulously filtered to eliminate the impact of exceptional events. Specifically, 6 days of data were excluded from the analysis due to significant variations in solar radiation intensity caused by clouds, which would have otherwise biased the results. The remaining data were utilised to train and test the machine learning model to increase the predictive accuracy. Summer months have been chosen to collect many data and improve the model's accuracy. The extensive data collection, filtering, and analysis, along with the careful selection of the study period, all contribute to the robustness and validity of the results, which will be discussed in detail in the following sections.

4. Algorithm implementation

Python is used for running code that uses multiple machine learning models to forecast solar still distillate output. The code uses several libraries such as pandas, numpy, sklearn, matplotlib, and seaborn to load, preprocess, and visualize the data. After loading the data, the code performs data preprocessing steps such as removing missing values, removing the date column, and normalizing the data. The data is then split into features and target variables. The features are the independent

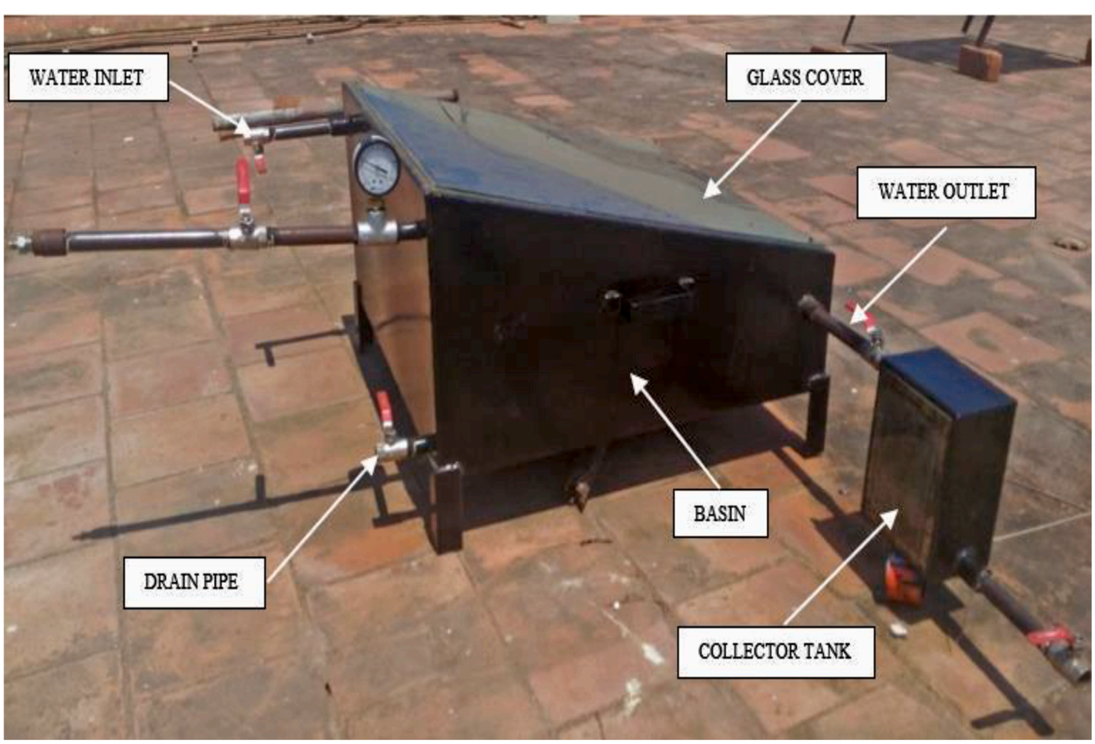

Fig. 6b. Photograph of the experimental setup.

variables used to predict the target variable, the distillate output. In this code, 80% of the data is used for training, and 20% is used for testing. The code then defines five machine learning models: Linear Regression, Decision Tree, Random Forest, Multi-layer Perceptron (MLP), and Support Vector Machine (SVM). These models are trained on the data, and their performance is evaluated using cross-validation. Cross-validation is used to evaluate the models using different subsets of the data, and the metric used to measure the performance of the models is the Mean Absolute Error (MAE). The code then uses Grid Search and Random Search to tune the parameters of the models. Grid Search is used to try different combinations of the parameters, while Random Search is used to select the parameters of the models randomly. The goal of tuning the parameters is to achieve the best performance of the models. After evaluating all the models, the code compares their performance and chooses the best model based on the lowest mean absolute error. The code also includes several data visualizations such as scatter plots, histograms, and heatmaps better to understand the data and the performance of the models. The flow chart of the code executed is as shown in Fig 7. In conclusion, this code provides a comprehensive and automated workflow for predicting the distillate output of solar still using machine learning techniques.

5. Results and discussion

5.1. Experimental results

The solar irradiance is a crucial factor in determining the output of a solar still. Solar irradiance variation is observed as a sinusoidal curve, as shown in Fig. 8a. At noon, the Sun's rays are perpendicular to the Earth's surface, leading to the maximum possible solar irradiance. As the day progresses, the angle of incidence of the Sun's rays decreases, causing the solar irradiance to decrease accordingly. The variation in solar irradiance throughout the day significantly affects the temperatures of various components in a solar still. The ambient temperature, basin temperature, water temperature, and glass cover temperature are all influenced by solar irradiance and follow a similar trend. The transient variation of various components such as basin, water and glass cover is shown in Fig. 8b. Throughout the trial, the average wind speed ranged from 1.9 to 4.6 metres per second. It is found that the effect of the wind velocity on distillate output would be greater above 2 meters per second. Convection cooling causes solar still output to drop as wind speeds rise.

The freshwater yield of a solar still is an important factor in its overall performance, and monitoring it to optimize its functioning regularly is essential. The temperature difference between the evaporating and condensing surfaces affects water evaporation and condensation. Fig. 9a depicts a clear diurnal pattern, with a noticeable increase in the morning hours and a gradual decrease towards the end of the day. This pattern is closely linked to the solar radiation levels, exhibiting a similar diurnal trend. The morning increase in freshwater yield can be attributed to the thermal storage capacity of the water in the solar still. The temperature of the water within the still and the glass cover gradually rises throughout the morning due to the gradual increase in the surrounding air temperature. This rise in temperature does not appreciably change the temperature difference between the evaporating and condensing surfaces, slowing water evaporation and lowering freshwater production. The temperature differential between the evaporating and condensing surfaces grows as the day advances, increasing water evaporation and freshwater output. The maximum yield is typically observed around 1400 hours when the temperature difference is at its highest, and the solar radiation levels are at their peak. As the ambient temperature drops at night, the glass cover temperature drops, reducing the temperature difference between the evaporating and condensing surfaces. This reduction in temperature difference results in a slower rate of water evaporation and a gradual decrease in the freshwater yield.

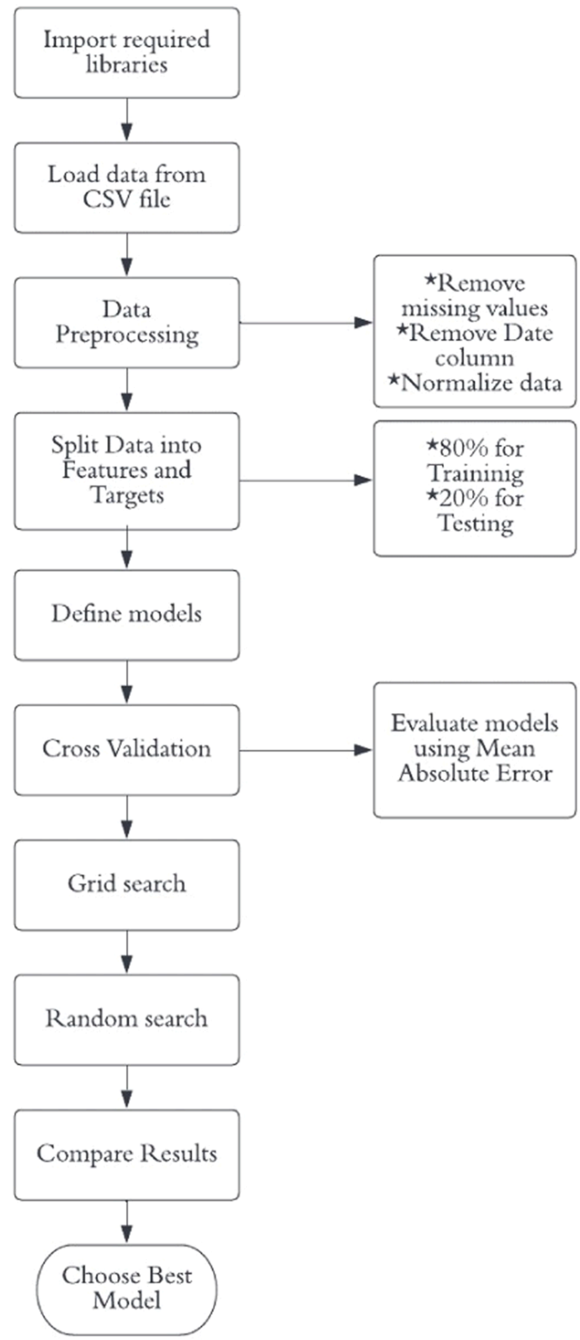

Fig. 7. Machine Learning workflow for predicting distillate output of a solar still.

5.2. Computational results

The code was executed using Python version 3.9 in the Spyder integrated development environment. The correlation heat map is used to visualize the relationship between different features in a dataset. It is a 2D representation of data where individual values are represented as colours, as shown in Fig. 9b. The values mapped with each feature represent the strength and direction of the relationship between the two variables. The computed correlation values between distillate output and basin temperature, water temperature, cover temperature, solar radiation, and wind velocity were analyzed to determine the strength of linear relationships between the variables. The correlation coefficients are 0.97, 0.96, and 0.96 for basin, water, and cover temperatures. It reveals a solid positive linear relation between distillate output and

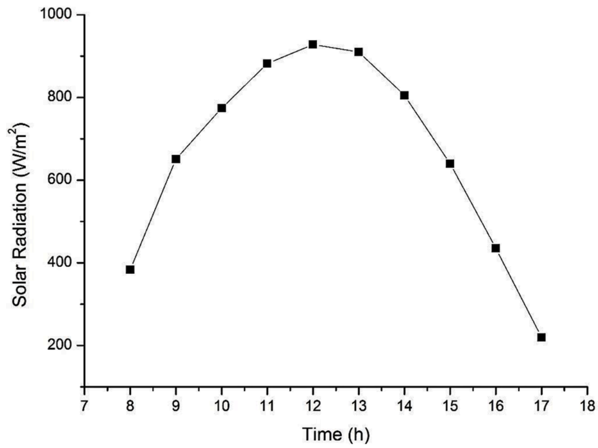

Fig. 8a. Variation of solar irradiance throughout the day.

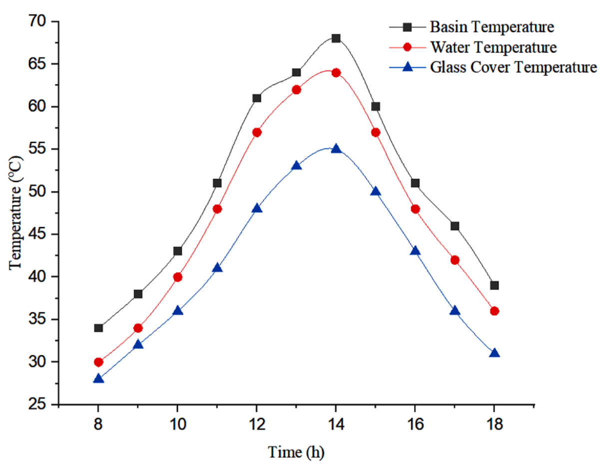

Fig. 8b. Transient temperature variations.

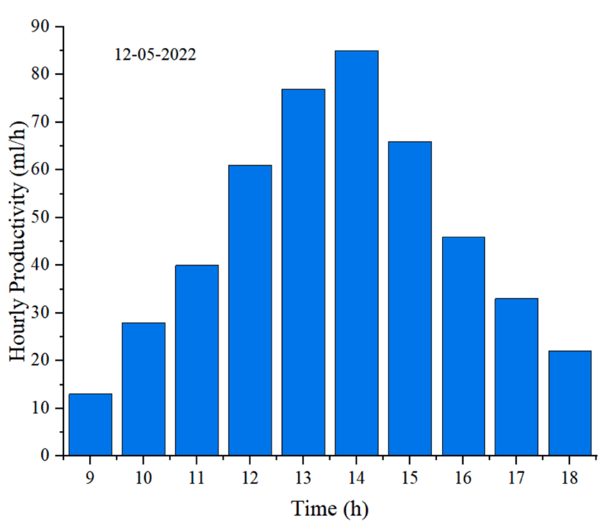

Fig. 9a. Variation of hourly productivity throughout the day.

these temperature variables, meaning that increasing these factors significantly increases distillate output. Conversely, the correlation coefficient of 0.53 between distillate output and solar radiation suggests a moderate positive linear relationship, implying that while solar radiation significantly impacts the distillate output, other factors may also play a role in determining the distillate output. Additionally, the correlation coefficient 0.065 between distillate output and wind velocity signifies a weak linear relationship, indicating that wind velocity has a negligible impact on the distillate output and that other factors likely dominate in determining the distillate output. The correlation heat map provides valuable information on the relationships between the variables and the impact each variable has on the distillate output, further enabling us to choose the best machine learning model for solar still optimization.

The scatter plot in Fig. 9c visualizes the comparison between the actual distillate output from the solar still and the predicted values generated by the machine learning model. The X-axis depicts the actual values, and the Y-axis represents the predicted values. The plot was generated utilizing the matplotlib.pyplot library in Python, which provides a comprehensive interface for plotting data. The scatter function was employed to create the scatter plot and required two inputs: the y_test and y_pred arrays, which contain the actual and predicted values, respectively. This plot offers a comprehensive evaluation of the performance of the machine learning model and its ability to predict the distillate output from the solar still accurately. It can be deduced from the plot that the actual and predicted values are closely aligned when modelled using a Decision Tree algorithm, with minimal deviation present.

Fig. 10 presents a residual scatter plot of the distillate production from the solar still. This plot represents the difference between the actual test set values (y_test) and the predicted values (y_pred) for each data point in the test set. The x-axis displays the actual values, and the y-axis displays the residuals, which are calculated as the difference between the actual and predicted values. Each scatter point in the plot represents a single data point from the test set, with its position on the y-axis indicating the magnitude of the error between the actual and predicted values for that point. Positive residuals indicate instances where the model predicted a higher value than the actual value, while negative residuals indicate instances where the model predicted a lower value. Observing the residual plot, it can be concluded that the residuals are distributed between both positive and negative values, and the range of residuals is lower than the actual values, implying that the machine learning model has demonstrated improved performance in predicting the distillate output of the solar still.

Numerous research initiatives have employed machine learning to optimise solar still outputs in the rapidly evolving renewable energy domain and sustainable water resource management. Table 1. presents a comparative summary of our current study's findings with recent works in the field.

6. Conclusion

In the realm of renewable energy and sustainable water management, machine learning has emerged as a potent tool, especially for predicting the outputs of systems like solar stills. This research underscores the pre-eminence of the Decision Tree model in forecasting distillate outputs from a single basin single slope solar still, advancing the domain of sustainable water sources. The important conclusions are summarized as follows,

- The Decision Tree model demonstrated superior performance over other evaluated models, including linear regression, MLP, SVM, and Random Forest.
- Through rigorous cross-validation evaluations and parameter tuning via Grid Search and Random Search, the Decision Tree model achieved the optimal cross-validation score of 5.57.
- The methodology introduced in this study provides valuable insights into optimising solar stills, vital instruments in offering clean water, especially in arid locales.

Future avenues of research that can build upon this study include

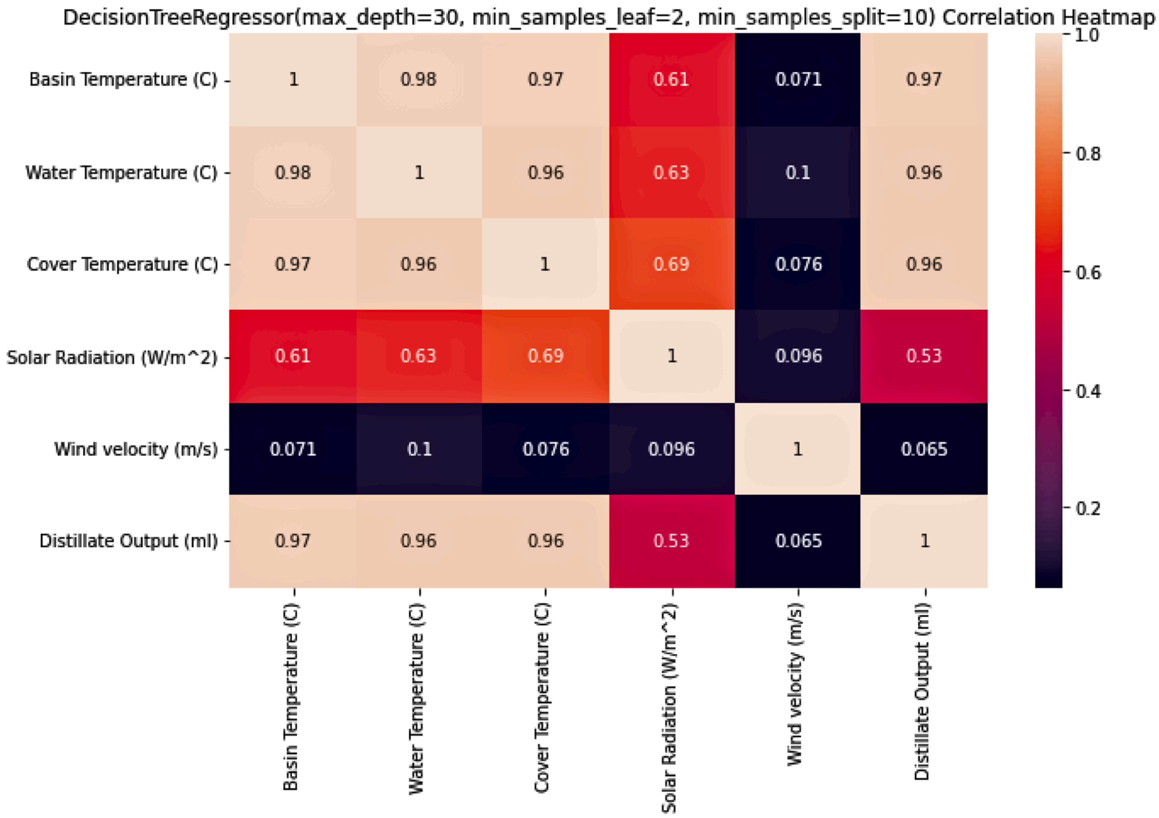

Fig. 9b. Correlation heat map of solar still performance parameters.

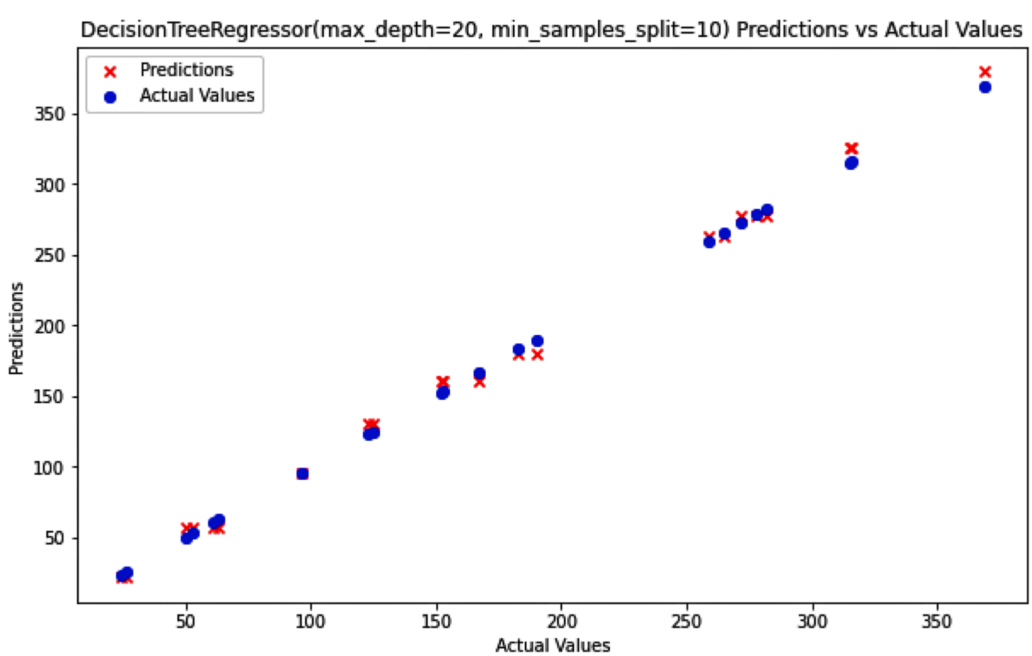

Fig. 9c. Scatter plot showing the predicted and actual values using the decision tree algorithm.

refinement of the models and incorporating additional data to increase the accuracy of predictions. Advanced machine learning techniques, such as deep learning models and considering climatic variables, can enhance the model's performance. The present study has certain limitations that future studies could address. The study's data was collected over three months; thus, it may need to reflect the solar still's performance over longer periods correctly. To address this, future studies could incorporate data from different seasons to better understand the solar still's performance. Additionally, this study was focused on a single type of solar still, and future studies could explore the performance of alternative designs, such as multi-basin and multi-slope configurations.

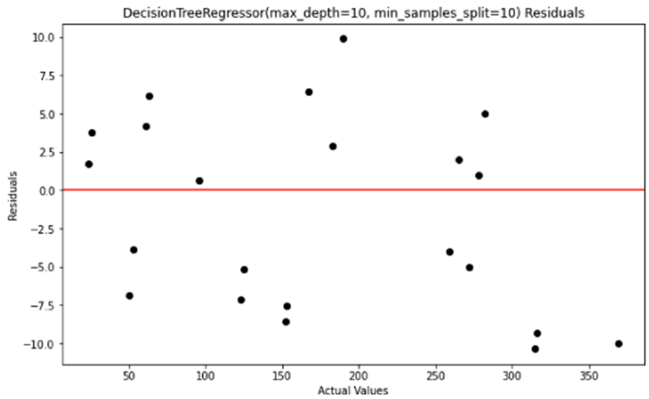

Fig. 10. Residual scatter plot of the distillate output of the solar still.

Table 1
Comparison of the current study with recent works.

Type of still & Location	Machine Learning Technique	Remarks	Ref.
Pyramid solar still, UAE	Linear Regression & Artificial Neural Network	ANN model performed better than LR in predicting the yield	Elgendi & Atef (2023)
Single slope solar still, Los Angeles	Stepwise Linear Regression	Exceptional prediction accuracy achieved; R^2 value of 1 and RMSE < 0.016 for distillate outputs F < 30 mL/day.	Maddah et. al. (2023)
Parabolic dish concentrator box solar still, Iran	Multivariate Adaptive Regression splines (MARS), Gene-Expression programming (GEP), Evolutionary Polynomial Regression (EPR), Model Tree (MT)	MARS outperformed other ML models in predicting the energy efficiency of the solar still	Nazari et. al. (2022)
Tubular Solar Still, Egypt	Random Forest, Artificial Neural Network, Bayesian Optimization	Random Forest model outperformed other models in predicting the hourly production rate of solar still.	Wang et al. (2021).
Single slope solar still, India	Linear regression, Decision Tree, Random Forest, Multi-layer Perceptron, Support Vector Machine	Decision Tree outperformed other models in predicting the hourly production rate of solar still.	Present study

Declaration of Competing Interest

Current paper has no conflict of interests.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported by Qatar National Research Fund under the grant no. MME03-1226-210042. The statements made herein are solely the responsibility of the authors.

References

Aglan, AM, Aklan, M, Momin, AE, 2021. Solar-powered desalination, a novel solar still directly connected to solar parabolic trough. Energy Rep. 7, 2245–2254. https://doi.org/10.1016/j.egyr.2021.04.041.

Alabi, TM, Aghimien, EI, Agbajor, FD, Yang, Z, Lu, L, Adeoye, AR, Gopaluni, B, 2022. A review on the integrated optimization techniques and machine learning approaches for modelling, prediction and decision making on integrated energy systems. Renew. Energy 194, 822–849. https://doi.org/10.1016/j.renene.2022.05.123.

Amiri, H, 2022. Enhancing the stepped solar still performance using a built-in passive condenser. Sol. Energy 248, 88–102. https://doi.org/10.1016/j.solener.2022.11.006.

Atteya, TEM, Abbas, F, 2023. Testing a stepped solar still with different sand beds and reflectors. Case Stud. Therm. Eng. 43, 102782 https://doi.org/10.1016/j.csite.2023.102782.

Behzadi, A, Sadrizadeh, S, Karimi, N, 2023. A rule-based energy management strategy for a low-temperature solar/wind-driven heating system optimized by the machine learning-assisted grey wolf approach. Energy Convers. Manage. 277, 116590 https://doi.org/10.1016/j.enconman.2022.116590.

Cavusoglu, U, 2019. A new hybrid approach for intrusion detection using machine learning methods. Appl. Intell. 49, 2735–2761. https://doi.org/10.1007/s10489-018-01408-x.

Chauhan, VK, Shukla, SK, Rathore, PKS, 2022. A systematic review for performance augmentation of solar still with heat storage materials: a state of the art. J. Energy Storage 47, 103578. https://doi.org/10.1016/j.est.2021.103578.

Chen, WL, Xie, G, 2022. Performance of multi-stage tubular solar still operating under vacuum. Renew. Energy 201, 34–46. https://doi.org/10.1016/j.renene.2022.11.023.

Dubey, A, Singh, SK, Tyagi, SK, 2022. Advances in design and performance of dual slope solar still: a review. Sol. Energy 244, 189–217. https://doi.org/10.1016/j.solener.2022.08.050.

Ebrahimpour, B, Shafi, MB, 2022. Experimental evaluation of the effect of boulders and fines in biodegradable organic materials on the improvement of solar stills. Sol. Energy 247, 453–467.

Elgendi, M, Atef, M, 2023. Calculating the impact of meteorological parameters on pyramid solar still yield using machine learning algorithms. Int. J. Thermofluids 18, 100341. https://doi.org/10.1016/j.ijft.2023.100341.

Elgendi, M, Kabeel, AE, Essa, FA, 2023. Improving the solar still productivity using thermoelectric materials: a review. Alexandr. Eng. J. 65, 963–982. https://doi.org/10.1016/j.aej.2022.10.011.

Enciso, M, Zingaretti, LM, 2019. A guide on deep learning for complex trait genomic prediction. Genes 10, 553. https://doi.org/10.3390/genes10070553.

Entezari, A, Aslani, A, Zahedi, R, Noorollahi, Y, 2023. Artificial intelligence and machine learning in energy systems: a bibliographic perspective. Energy State Rev. 45, 101017 https://doi.org/10.1016/j.esr.2022.101017.

Essa, F.A., Omara, Z., Abdullah, A., et al., 2021. Augmenting the productivity of stepped distiller by corrugated and curved liners, CuO/paraffin wax, wick, and vapor suctioning. Environ. Sci. Pollut. Res. 28, 56955–56965. https://doi.org/10.1007/s11356-021-14669-w.

Gandhi, AM, Shanmugan, S, Kumar, R, Elsheikh, AH, Sharifpur, M, Bewoor, AK, Bamisile, O, Hoang, AT, Ongar, B, 2022. SiO_2/TiO_2 nanolayer synergistically trigger thermal absorption inflammatory responses materials for performance improvement of stepped basin solar still natural distiller. Sustain. Energy Technol. Assess. 52, 101974 https://doi.org/10.1016/j.seta.2022.101974.

Ghandourah, E, Panchal, H, Fallatah, O, Ahmed, HM, Moustafa, EB, Elsheikh, AH, 2022. Performance enhancement and economic analysis of pyramid solar still with corrugated absorber plate and conventional solar still: a case study. Case Stud. Therm. Eng. 35, 101966 https://doi.org/10.1016/j.csite.2022.101966.

Isah, AS, Takaijudin, HB, Singh, BSM, Gilani, SIU, Yusof, KW, Abdurrasheed, AS, Abimbola, TO, Shoeb, MM, 2022. Solar energy desalination distillate yield and cost evolution, and statistical relationship between meteorological variables and distillate yield. Sol. Energy 246, 256–272. https://doi.org/10.1016/j.solener.2022.09.025.

Janiesch, C, Zschech, P, Heinrich, K, 2021. Machine learning and deep learning. Electron. Markets 31, 685–695. https://doi.org/10.1007/s12525-021-00475-2.

Kapp, S, Choi, J, Hong, T, 2023. Predicting industrial building energy consumption with statistical and machine-learning models informed by physical system parameters. Renew. Sustain. Energy Rev. 172, 113045 https://doi.org/10.1016/j.rser.2022.113045.

Khatod, KJ, Katekar, VP, Deshmukh, SS, 2022. An evaluation for the optimal sensible heat storage material for maximizing solar still productivity: a state-of-the-art review. J. Energy Storage 50, 104622. https://doi.org/10.1016/j.est.2022.104622.

Lauvandy, AF, Raihananda, FA, Philander, E, Lugiyana, NH, Budiman, BA, Fitriyanti, M, Juangsa, FB, Sambegoro, P, 2022. Improving condensing performance of a low-cost floating solar still by surface characteristic alteration. Sustain. Energy Technol. Assess. 54, 102835 https://doi.org/10.1016/j.seta.2022.102835.

Lisboa, AAV, Segurado, R, Mendes, MAA, 2022. Solar still performance for small-scale and low-cost seawater desalination: model-based analysis and water yield enhancement techniques. Sol. Energy 238, 341–362. https://doi.org/10.1016/j.solener.2022.04.007.

Liu, W, Shen, Y, Aungkulanon, P, Ghalandari, M, Le, BN, Alviz-Meza, A, Cardenas-Escrocia, Y, 2023. Machine learning applications for photovoltaic system optimization in zero green energy buildings. Energy Rep. 9, 2787–2796. https://doi.org/10.1016/j.egyr.2023.01.114.

Liu, Z, Wu, D, Liu, Y, Han, Z, Lun, L, Gao, J, Jin, G, Cao, G, 2019. Accuracy analyses and model comparison of machine learning adopted in building energy consumption production. Energy Explor. Exploit. 1–26. https://doi.org/10.1177/0144598718822400.
Mahmoud, A, Fath, H, Ahmed, M, 2018. Enhancing the performance of a solar driven hybrid solar still/humidification-dehumidification desalination system integrated with solar concentrator and photovoltaic panels. Desalination 430, 165–179. https://doi.org/10.1016/j.desal.2017.12.052.
McLaughlin, E, Choi, J, 2023. Utilizing machine learning models to estimate energy savings from an industrial energy system. Resour., Conserv. Recycl. 12, 100103 https://doi.org/10.1016/j.resenv.2022.100103.
Mevada, Dinesh, Panchal, H., Kumar Sadasivuni, K., Israr, M., Suresh, M., Dharaskar, S., Thakkar, H., 2020. Effect of fin configuration parameters on performance of solar still: a review. Groundw. Sustain. Dev. 10 https://doi.org/10.1016/j.gsd.2019.100289. Article 100289.
Modi, KV, Gamit, AR, 2022. Investigation on performance of square pyramid solar still using nanofluid and thermal energy storage material: An experimental and theoretical study. J. Cleaner Prod. 381, 135115 https://doi.org/10.1016/j.jclepro.2022.135115.
Murugan, DK, Elumalai, N, Subramani, S, 2021. Experimental investigation on the effect of basin pressure in a single basin solar still. Desalin. Water Treat. 229, 10–16. https://doi.org/10.5004/dwt.2021.27385.
Nazari, S, Najafzadeh, M, Daghigh, R, 2022. Techno-economic estimation of a non – cover box solar still with thermoelectric and antiseptic nanofluid using machine learning models. Appl. Therm. Eng. 212, 118584 https://doi.org/10.1016/j.applthermaleng.2022.118584.
Otero, FEB, Freitas, AA, Johnson, CG, 2012. Inducing decision trees with ant colony optimization algorithm. Appl. Soft Comput. 12, 3615–3626. https://doi.org/10.1016/j.asoc.2012.05.028.
Pansal, K., Ramani, B., Kumar Sadasivuni, K., Panchal, H., Manokar, M., Sathyamurthy, R., Suresh, M., Israr, M., 2020. Use of solar photovoltaic with active solar still to improve distillate output: a review. Groundw. Sustain. Dev. 10, 100341.
Panchal, H, Sadasivuni, KK, Essa, FA, Shanmugan, S, Sathyamurthy, R, 2021a. Enhancement of the yield of solar still with the use of solar pond: a review. Heat Transf. 50, 1392–1409. https://doi.org/10.1002/htj.21935.
Panchal, H, Sadasivuni, KK, Shanmugan, S, Pandya, N, 2021b. Performance analysis of waste brick magnesia as a storage material in a solar still. Heat Transf. 50, 1799–1811. https://doi.org/10.1002/htj.21956.
Panchal, H.N., Shah, P.K., 2014. Improvement of solar still productivity by energy absorbing plates. J. Renew. Energy Environ. 1 (1), 1–7. https://doi.org/10.30501/jree.2014.70052.
Panchal, H.N., Thakkar, H., 2016. Theoretical and experimental validation of evacuated tubes directly coupled with solar still. Therm. Eng. 63, 825–831. https://doi.org/10.1134/S0040601516110045.
Panchal, Hitesh, 2017. Life cycle cost analysis of a double-effect solar still. Int. J. Ambient Energy 38 (4), 395–399. https://doi.org/10.1080/01430750.2015.1132767.
Panchal, Hitesh Nikunj Patel, Thakkar, Hemin, 2017. Various techniques for improvement in distillate output from active solar still: a review. Int. J. Ambient Energy 38 (2), 209–222. https://doi.org/10.1080/01430750.2015.1076518.
Panchal, Hitesh, Hishan, S.S., Rahim, R., Sadasivuni, K.K., 2020. Solar still with evacuated tubes and calcium stones to enhance the yield: an experimental investigation. Process. Saf. Environ. Prot. 142, 150–155.
Panchal, Hitesh, Taamneh, Yazan, Sathyamurthy, Ravishankar, Kabeel, A.E., El-Agouz, S. A., Naveen Kumar, P, Muthu Manokar, A, Arunkumar, T., Mageshbabu, D., Bharathwaaj, R., 2019. Economic and exergy investigation of triangular pyramid solar still integrated to inclined solar still with baffles. Int. J. Ambient Energy 40 (6), 571–576.
Panchal, Hitesh, 2010. Experimental analysis of different absorber plates on performance of double slope solar still. Int. J. Eng. Sci. Technol. 2 (11), 6626–6629.
Panchal, Hitesh, 2011. Experimental investigation of varying parameters affecting on double slope single basin solar still. Int. J. Adv. Eng. Sci. 2 (1), 17–21.
Peng, G, Sharshir, SW, Wang, Y, An, M, Ma, D, Zang, J, Kabeel, AE, Yang, N, 2021. Potential and challenges of improving solar still by micro/nano-particles and porous materials - a review. J. Clean. Prod. 311, 127432 https://doi.org/10.1016/j.jclepro.2021.127432.
Peng, G, Xu, Z, Ji, J, Sun, S, Yang, N, 2022. A study on the upper limit efficiency of solar still by optimizing the mass transfer. Appl. Therm. Eng. 213, 118664 https://doi.org/10.1016/j.applthermaleng.2022.118664.
Rabishokr, K, Daghigh, R, 2023. A portable solar still's productivity boost combining a magnetic stirrer and thermoelectric. Desalination 549, 116340. https://doi.org/10.1016/j.desal.2022.116340.
Rashidi, S, Karimi, N, Yan, W, 2022. Applications of machine learning techniques in performance evaluation of solar desalination systems - a concise review. Eng. Anal. Boundary Elem. 144, 399–408. https://doi.org/10.1016/j.enganabound.2022.08.031.
Sambare, RK, Dewanga, SK, Gupta, PK, Joshi, S, 2022. Energy, exergy and economic analyses of tubular solar still with various transparent cover materials. Process Saf. Environ. Prot. 168, 1101–1108. https://doi.org/10.1016/j.psep.2022.10.064.
Shalaby, SM, Sharshir, SW, Kabeel, AE, Kandeal, AW, Abosheiasha, HF, Abdelgaied, M, Hamed, MH, Yang, N, 2022. Reverse osmosis desalination systems powered by solar energy: preheating techniques and brine disposal challenges – a detailed review. Energy Convers. Manage. 251, 114971 https://doi.org/10.1016/j.enconman.2021.114971.
Sharshir, SW, Omara, MA, Elsisi, G, Jospeh, A, Kandeal, AW, Ali, A, Bedair, G, 2022. Thermo-economic performance improvement of hemispherical solar still using wick material with V-corrugated basin and two different energy storage materials. Sol. Energy 249, 336–352. https://doi.org/10.1016/j.solener.2022.11.038.
Shatar, NM, Sabri, MFM, Salleh, MFM, Ani, MH, 2023. Investigation on the performance of solar still with thermoelectric cooling system for various cover material. Renew. Energy 202, 844–854. https://doi.org/10.1016/j.renene.2022.11.105.
Shoeibi, S, Saemian, M, Kargarsharifabad, H, Hosseinzade, S, Rahbar, N, Khiadani, M, Rashidi, MM, 2022. A review on evaporation improvement of solar still desalination using porous material. Int. Commun. Heat Mass Transf. 138, 106387 https://doi.org/10.1016/j.icheatmasstransfer.2022.106387.
Shoeibi, S, Saemian, M, Kargarsharifabad, H, Mirjalily, SAA, 2023. Influence of PV/T waste heat on water productivity and electricity generation of solar stills using heat pipes and thermoelectric generator: An experimental study and environmental analysis. Energy Convers. Manage. 276, 116504 https://doi.org/10.1016/j.enconman.2022.116504.
Sibagariang, YP, Napitupulu, FH, Kawai, H, Ambarita, H, 2022. Investigation of the effect of a solar collector, nozzle, and water cooling on solar still double slope. Case Stud. Therm. Eng. 40, 102489 https://doi.org/10.1016/j.csite.2022.102489.
Sohani, A, Sayyaadi, H, Cornaro, C, Shahverdian, MH, Pierro, M, Moser, D, Karimi, N, Doranehgard, MH, 2022. Using machine learning in photovoltaics to create smarter and cleaner energy generation systems: a comprehensive review. J. Clean. Prod. 364, 132701 https://doi.org/10.1016/j.jclepro.2022.132701.
Wang, Y, Kandeal, AW, Swidan, A, Sharshir, SW, Abdelaziz, GB, Halim, MA, Kabeel, AE, Yang, N, 2021. Prediction of tubular solar still performance by machine learning integrated with Bayesian optimization algorithm. Appl. Therm. Eng. 184, 116233 https://doi.org/10.1016/j.applthermaleng.2020.116233.
Wei, X, Liu, Y, Xie, G, Chen, W, Jiang, Z, 2023. A visualization study of vacuum enhancement on vapor flow and yield in tubular solar still. Sol. Energy 252, 145–155. https://doi.org/10.1016/j.solener.2023.01.045.
Yang, Z, Deng, Y, Wang, T, 2012. Multiple Linear Regression model based on neural network and its application in the MBR simulation. Abstract Appl. Anal. 703153 https://doi.org/10.1155/2012/703153.
Zayed, ME, Katekar, VP, Tripathy, RK, Deshmukh, SS, Elsheikh, AH, 2022. Predicting the yield of stepped corrugated solar distiller using kernel-based machine learning models. Appl. Therm. Eng. 213, 118759 https://doi.org/10.1016/j.applthermaleng.2022.118759.
Zhao, Y, Ramadan, O, Kong, H, Xue, X, Riffat, S, Zheng, H, 2022. Performance analysis and optimization of a novel high-efficiency flower-inspired solar still. Energy Convers. Manage. 251, 114878 https://doi.org/10.1016/j.enconman.2021.114878.