Journal of Materials Chemistry A

# Solar Stills: The Future Enable by Machine Learning



| | |
|---|---|
| Journal: | *Journal of Materials Chemistry A* |
| Manuscript ID | TA-REV-09-2024-006316.R1 |
| Article Type: | Review Article |
| Date Submitted by the Author: | 23-Oct-2024 |
| Complete List of Authors: | Li, Rui; Zhejiang A&F University<br>Wang, Chaohai; Henan University of Urban Construction<br>He, Chang; Zhejiang A&F University<br>Nam, Ho Ngoc; Nagoya University - Higashimaya Campus, Materials Process Engineering;<br>Wang, Junning; Henan University of Urban Construction<br>Mao, Yanli; Henan University of Urban Construction,<br>Zhu, Xinfeng; Henan University of Urban Construction<br>Liu, Wei; Zhejiang A and F University,<br>Kim, Minjun; University of Queensland<br>Yamauchi, Yusuke; University of Queensland, Chemical Engineering; Nagoya University, |
| | |

# Solar Stills: The Future Enable by Machine Learning

*Rui Li, Chaohai Wang*, Chang He, Ho Ngoc Nam, Junning Wang, Yanli Mao, Xinfeng Zhu, Wei Liu*, Minjun Kim*, Yusuke Yamauchi*

R. Li, C. He, W. Liu
College of Optical, Mechanical and Electrical Engineering, Zhejiang A & F University, Hangzhou 311300, China.
E-mail: liuwei@zafu.edu.cn

R. Li, C. Wang, C. He, J. Wang, Y. Mao, X. Zhu
Henan International Joint Laboratory for Green Low Carbon Water Treatment Technology and Water Resources Utilization, School of Municipal and Environmental Engineering, Henan University of Urban Construction, Pingdingshan 467036, China.
E-mail: chaohai@huuc.edu.cn

H. N. Nam, Y. Yamauchi
Department of Materials Process Engineering, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464–8603, Japan.

M. Kim, Y. Yamauchi
School of Chemical Engineering & Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, Brisbane, Queensland 4072, Australia.
E-mail: minjun.kim@uq.edu.au

**Keywords:** solar desalination, solar stills, machine learning, data-driven methods, optimization

## Abstract

Desalination is a highly energy-intensive process often requiring the consumption of costly fossil fuels, inevitably causing various environmental hazards. As a sustainable and renewable energy source, however, solar energy is anticipated to alleviate such environmental concerns associated with the energy-intensive desalination process. Recently, machine learning, a powerful data analysis method, has been employed for modeling and prediction to enhance the productivity of solar stills, an effective solution to water scarcity owing to low cost and simple operation. In this review, machine learning techniques are particularly emphasized, along with exploring distinctions between solar stills and other solar desalination technologies. Machine learning models can achieve further optimization through additional avenues such as model selection, hyperparameter tuning, feature selection, and dataset management. The findings specifically highlight the crucial role of machine learning in enhancing solar desalination through improved prediction and optimization. Furthermore, this paper discussed different machine-learning prediction techniques while offering suggestions for future research in the field.


**Nomenclature**

| | | | |
|---|---|---|---|
| SS | solar still | ANFIS | adaptive neuro-fuzzy inference systems |
| ML | machine learning | DT | decision trees |
| HDH | humidification-dehumidification | RF | random forests |
| RO | reverse osmosis | SVM | support vector machines |
| MD | membrane distillation | RM | regression models |
| MSF | multi-stage flash | MLP | multilayer perceptron |
| VC | vapor compression | WNN | wavelet neural networks |
| MVC | machine vapor compressor | RBF | radial basis function |
| TVC | thermal vapor compressor | ENN | elman neural network |
| SLeM | simulating learning methodology | KNN | k-nearest neighbors |
| | | GPR | gaussian process regression |
| MED | multi effect distillation | SP | solar pond |
| ETC | evacuated tube collector | FF | feedforward |
| PV-T | photovoltaic-thermal | NARX | nonlinear autoregressive exogenous |
| PV | photovoltaic | HBLS | heat localization bilayered structure |
| SLM | successive linearization method | CSS | conventional solar still |
| $R^2$ | R-squared | MSS | modified solar still |
| MAE | mean absolute error | SVR | support vector regression |
| RMSE | root mean square error | LVR | linear support vector regression |
| | | RVFL | random vector functional link |

| | | | |
|---|---|---|---|
| | | FA | firefly algorithm |
| MAPE | mean absolute percentage error | DSS | double slope solar still |
| DOE | Design of Experiment | BOA | bayesian optimization algorithm |
| RSM | response surface methodology | %ADD | average absolute deviation |
| TM | taguchi methodology | ICA | imperialist competitive algorithm |
| FD | factorial design | GA | genetic algorithm |
| AGMD | air-gap membrane distillation | PSO | particle swarm optimization |
| OI | overall index | LM | levenberg-marquardt |
| EC | efficiency coefficient | CRM | coefficient of residual mass |
| COV | coefficient of variation | MSSIE | modified solar still integrated with earth distillate |
| | | DPSS | developed pyramid solar still |
| DL | deep learning | IR | solar radiation intensity |
| | | HF | hourly freshwater |
| | | IEE | instantaneous energy efficiency |
| ANN | artificial neural networks | $V_w$ | wind speed |
| DSD | dish solar distiller | $T_b$ | basin slab temperature |
| PCM | phase change material | $T_{sw}$ | brine temperature |
| BP | back-propagation | $T_c$ | cover temperature |
| RBP | radial basis function | $T_\infty$ | ambient temperature |
| RNN | recurrent neural networks | $T_{w,basin}$ | basin nanofluid |
| CNN | conventional neural networks | $T_{vapor}$ | vapor |
| DRL | deep reinforcement learning | $T_{glass,in}$ | inner glass |
| LSTM | long-short-term memory neural networks | $T_{glass,out}$ | outer glass |
| RH | relative humidity | $T_{min}$ | minimum temperature |
| SR | radiation intensity | $T_{max}$ | maximum temperature |
| SSWESM | solar still with energy storage materials | $\eta$ | thermal efficiency |

# 1 Introduction

Recent economic development and population growth have led to massive global demand for water resources. However, approximately 97% of Earth's water found in the hydrosphere is made up of seawater, while only less than 1% of the freshwater resources are directly usable for humans.[1] Although it is indisputable that the fresh water is vital for the survival and development of human society, its resources are extremely limited.[2] Moreover, additional man-made factors such as irrational utilization, waste, and water pollution are rapidly exacerbating

the state of water scarcity, potentially impacting 4.7 billion people by 2025 according to the reports[3]. Therefore, an effective way to address the issue of water scarcity is essential and the use of advanced desalination method to gain access to seawater for freshwater has become a core of the research. A modern large-scale desalination plant can produce tens of thousands to a million tons of fresh water daily.[4,5] However, in the case of traditional desalination process, it is an energy-intensive industry which requires the consumption of non-renewable fossil energy sources (e.g., coal) to provide heat or power, therefore, increasing environmental risks.[6-9] Hence, the feasible usage of non-polluting energy sources for desalination is a pivotal step to alleviate the earth's condition of water scarcity.[10] Solar energy, the most abundant renewable resource, accounts for nearly 57% of the renewable energy in desalination market.[11] The ability to operate independently of steam and electricity, along with non-polluting and safe operation, renders it highly valuable in regions with energy scarcity and stringent environmental requirements.

In fact, desalination technologies such as multi-stage flash evaporation,[12,13] reverse osmosis,[14,15] vapor compression, and multi-effect distillation[16,17] in solar desalination often require high maintenance and installation costs. In contrast, solar stills (SS) offer low installation costs, simple design, and easy maintenance, making them an increasingly attractive technology,[18] particularly suitable for arid and semi-arid regions with low to medium water demand levels.[19-21] Multiple important variables such as weather conditions, sunlight intensity, temperature and system configuration make SS highly complex systems.[22,23] The use of accurate and convenient auxiliary analytical tools is therefore necessary to improve the desalination system's performance, as well as to save labor and material resources, especially when considering time-consuming and inefficient conventional experimental approaches. Traditional mathematical methods are no longer effective in adapting to and accurately accessing the intricate processes in modern desalination systems.[24,25] Most of the traditional methods take pre-determined steps in order to approach the possible optimal solution. Mostly, these algorithms start with a random speculation of the solution, and the exploration directions are obtained according to the specified migration rules. Then, only one direction is searched in order to find the possible optimal solution. In desalination and wastewater treatment, these traditional methods do not work.[26] Nevertheless, the evaluation and validation of these models require a substantial amount of data, which can limit the applicability of the predicted output and the model's effectiveness. In this regard, the emergence of machine learning (ML) techniques is particularly important (see Fig. 1) because it serves as a powerful data analysis tool that can accurately analyze the various aspects of solar desalination. In contrast to

traditional mathematical analysis methods, these methods do not require extensive professional experience, numerical or control equations, or explicit assumptions describing the underlying engineering processes. The judicious use of ML can create more efficient, environmentally friendly and economically viable solutions and facilitate the tuning of operating parameters to achieve maximum efficiency and minimise energy use.[27,28] Since the knowledge it provides is valuable in improving the design and functionality of solar stills, researchers have dedicated their efforts to this subject. For example, Xu et al.[29] proposed a machine learning automation method for the Simulating Learning Methodology (SLeM). Elsheikh et al.[30] reviewed the use of artificial neural networks in solar desalination, while Rashidi et al.[31] provided an overview of the application of ML methods in solar desalination. ML techniques have begun to show potential benefits in solar desalination systems.

By appropriately selecting training models, input-output pairs, and segmenting the dataset, ML methods can also be optimized to improve further their accuracy in predicting solar seawater desalination systems. The main goal of this study is to provide an overview of solar desalination technologies using ML, a powerful data analytics method for simplifying and accelerating the traditional desalination process. Finally, we discuss optimizing relevant ML algorithms to enhance their accuracy in predictive modeling.

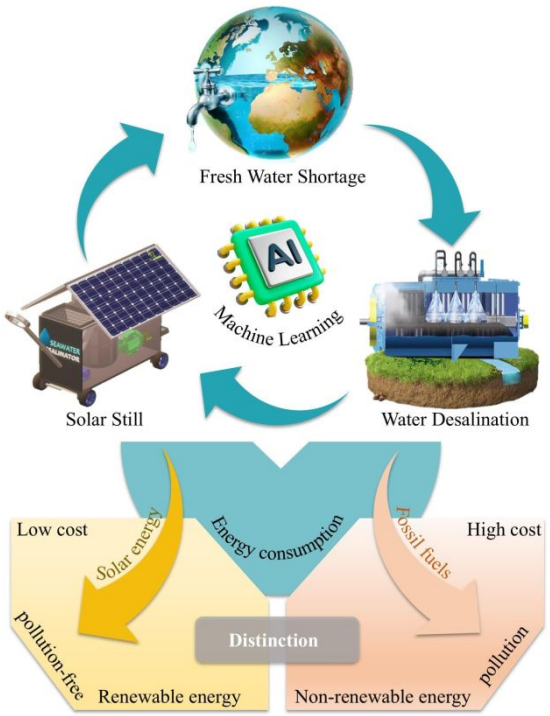


Fig. 1 Schematic of machine learning-assisted solar desalination.

## 2 Overview of solar desalination

With the continued development of solar seawater desalination technology, there are now many methods applicable for different systems, leading to the diversification of the available solar

desalination methods. Based on their operating principles, we can specifically categorize these methods into two types: direct and indirect methods.[32] This section begins with an overview of the indirect method of solar desalination technology, followed by a detailed description of the SS in the direct method.

**2.1 Indirect processes**

Indirect solar desalination systems use solar energy as an energy source to drive seawater desalination. Unlike the direct use of solar energy to evaporate seawater, an indirect solar desalination system uses solar energy to generate heat, which is then transferred to a still or other heat transfer device in the desalination system.

In humidification-dehumidification (HDH) desalination, brackish and saline water is heated to humidify the air, and freshwater is produced by condensation of the humid air generated at atmospheric pressure, as illustrated in Fig. 2a.[33] According to its circulation form, the HDH process can be divided into four main categories: closed air-open water circulation, closed air-closed water circulation, closed water-open air circulation, and open water-open air circulation. HDH is usually combined with external heaters such as solar collectors, flat plate solar collectors, vacuum tube solar collectors, and parabolic trough solar collectors. So far, very limited work has been done to identify the advantages and disadvantages of different configurations, whose impact on HDH performance is considerable.[34]

Reverse osmosis (RO) desalination technology utilizes the properties of a reverse osmosis membrane. As seawater passes through the RO membrane, micropores in the membrane allow water molecules to pass through and be collected. Salts and other dissolved substances are trapped on the other side of the RO membrane, forming concentrated water. The salt and other dissolved substances in the seawater are separated from the water molecules, thus realizing the desalination of seawater. The RO modules can also be connected in series or parallel configurations, as presented in Fig. 2b.[35,36] Currently, water pre-treatment is an obstacle to reverse osmosis systems, as water supplies often require extensive pre-treatment procedures.[37,38] In addition, improving the anti-fouling ability of the reverse osmosis membrane also has a beneficial effect on its treatment efficiency.

The difference in vapor pressure between the two sides of the membrane is used as the mass transfer driving force and thermally drives in membrane distillation (MD). Water vapor is then sent through the porous, hydrophobic membrane material to remove salt. A typical solar membrane distillation unit is introduced in Fig 2c.[35,39] There are four main types of membrane distillation processes: air gap membrane distillation, vacuum membrane distillation, sweeping gas membrane distillation, and direct contact membrane distillation.[39] Operational parameters

in membrane distillation, such as feed water temperature, flow rate, air gap thickness, membrane thickness, membrane thermal conductivity, porosity, curvature, and long-term operation, have an effect on distillate yield.[40] The heart of the membrane distillation process is considered to be a porous hydrophobic membrane. However, traditional polymer materials and membrane preparation methods are difficult to meet the current conditions. Therefore, developing new membrane materials and devising preparation methods for separation membrane are critical to advancing MD technology.[41]

Multi-stage flash (MSF) evaporation systems evaporate seawater to produce fresh water by using the method of reduced pressure and expansion flash evaporation. Accordingly, the seawater is first heated to a specific temperature and then introduced into the flash chamber. After multi-stage evaporation, the resulting water vapor enters the condenser and the collected liquid water is fresh water. Fig. 2d illustrates a schematic diagram of a solar MSF desalination system.[35,42] MSF accounts for about 21% of the worldwide desalination capacity, placing it to the second most common desalination method following reverse osmosis.[43] In particular, solar MSF integrates solar collectors to the conventional MSF desalination systems, making the selection of solar collectors, the correct design of the solar heating cycle, and the design and optimization of the MSF plant crucial for its successful operation.

In desalination by vapor compression (VC), seawater first undergoes heating to evaporation temperature, which then feeds the heated seawater into an evaporator to evaporate the water. A machine vapor compressor (MVC) or a thermal vapor compressor (TVC) then compresses the resulting vapor. A condenser finally cools it, transforming it into fresh water. In Fig. 2e, MVC desalination units are illustrated.[35,44] Photovoltaic modules, wind turbines, and water storage tanks can also be intergrated into MVC desalination plants, forming a complementary wind and solar system. This hybrid power unit provides the necessary energy, offering a more flexible energy supply. Currently, the vapor compression desalination equipment itself is expensive to manufacture and requires specialist maintenance and operating personnel to manage and maintain, increasing the difficulty and cost of operation.

In Fig. 2f, the principle of multi effect distillation (MED) is introduced.[45] By arranging several evaporators in parallel, seawater is heated in the previous evaporator, and water molecules begin to evaporate into water vapor. Pipes transfer the water vapor to the bottom of the next group of evaporators, where it enters the condenser. Through multiple uses of steam, MED achieves energy savings. Compared to MSF, a conventional MED desalination system uses about half the energy of an MSF system and produces almost the same amount of heat as MSF if both have the same gain ratio.[46] Solar MED can be a sustainable alternative option for

medium- to large-scale conventional desalination plants, although large-scale plants have not yet been built.[45]

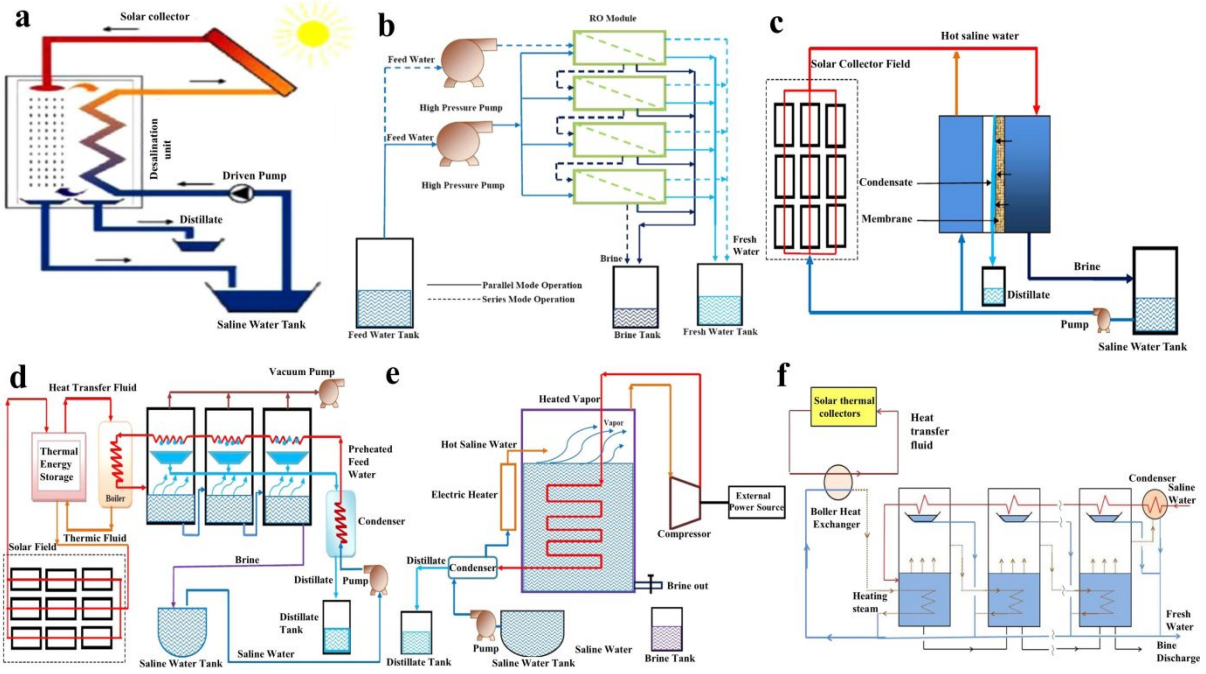


Fig. 2 (a) Humidification and dehumidification desalination unit coupled with a solar collector. (b) Series and parallel arrangement of RO modules. (c) Solar-powered membrane distillation unit. (d) Solar-powered multi-stage flash desalination system. (e) Mechanical vapor compression desalination system. (f) Schematic diagram of solar MED desalination system with feed preheating. (a) Reproduced from ref. 32,33 with permission from Elsevier, copyright 2013. (b) Reproduced from ref. 35,36 with permission from Elsevier, copyright 2015. (c) Reproduced from ref. 35,39 with permission from Elsevier, copyright 2015. (d) Reproduced from ref. 35,42 with permission from Elsevier, copyright 2015. (e) Reproduced from ref. 35,44 with permission from Elsevier, copyright 2015. (f) Reproduced from ref. 45 with permission from Elsevier, copyright 2018.

## 2.2 Direct processes

Direct solar desalination uses a collector system to harness solar radiation and heat seawater, causing it to evaporate and subsequently condense into fresh water. The most representative technology is the SS. Solar energy is one of the simplest forms of solar desalination that does not consume conventional energy and is simple and easy to operate with a yield of about 4-6 l/m$^2$ per day, which is sufficient for households.[47] Although SS are unsuitable for large-scale desalination systems,[48] they are simple and affordable in design and very appropriate in some remote coastal areas with plenty of sunlight but a lack of power and electricity. Based on the need for additional components, direct solar desalination could be subdivided into passive SSs and active SSs.

### 2.2.1 Passive solar stills

A passive SS is a device that uses solar energy to evaporate water and collect pure water. It does not require an external energy supply and relies primarily on solar energy to complete the distillation process. There are various ways to classify passive SSs, such as by the design of the evaporator, different materials[49], thermal storage options, shape and number of basins.[50]

Researchers have investigated many aspects of conventional SSs to improve their performance, including water depth, the angle of inclination of the glass cover, the type of material used, etc. A comprehensive review has been given by Prakash et al.[51] Hansen et al.[52] studied new materials suitable for solar desalination applications and selected unused wick materials for analysis. They determined that water coral fleece material is the best hygroscopic material based on four parameters: porosity, absorbency, capillary rise, and heat transfer coefficient. They also compared the performance of different wick materials under different absorber plate configurations, concluding that water coral fleece with weir mesh-stepped absorber plate yielded the best with distilled water up to 4.28 l/day. Mevada et al.[53] investigated different energy storage materials such as black glass bulb, black granite, and white marble stone were tried in SS to improve the distillate yield. A comparison was also made between conventional solar still (CSS) and solar still with energy storage materials (SSWESM). The results showed that the daily distillate yield was 1.4 kg/m$^2$ and 2.5 kg/m$^2$ for CSS and SSWESM, respectively, and the daily efficiency of SSWESM was 72.6% higher than that of CSS.

In the fabrication of SS, temperature and pressure conditions have a significant impact on certain processing techniques, especially when it comes to the selection of materials, processing methods and equipment assembly. Chen et al.[54] investigated the development of an efficient solar distillation system through the use of thermally expanding materials. By controlling the expansion properties of the materials at elevated temperatures, a porous light-absorbing structure was formed and the rate of water evaporation was increased.

Omara et al.[55] conducted a comparative performance study of an improved and conventional SS by installing reflectors on both vertical sides of the step SS steps. The results showed that the productivity of the modified step SS with internal reflectors was about 75% higher than that of the conventional still, and the daily efficiencies of the modified step still were 56%, 53%, and 34%, respectively. Panchal et al.[56] annually investigated the use of MgO and $TiO_2$ nanofluids at different concentrations to assess the distillate yield of stepped SS, where the nanofluid concentrations investigated ranged from 0.1% to 0.2%. The results showed that the use of MgO nanofluids (0.2% and 0.1% concentrations) and $TiO_2$ nanofluids (0.2% and 0.1% concentrations) increased the step SS distillate yields by 45.8%, 33.33%, 20.4% and 4.1%.

The distillate yield of MgO nanofluid was higher than $TiO_2$ due to lower specific heat capacity and higher thermal conductivity.

Conventional single-basin passive SS generally have low distillation efficiency and productivity. Rajaseenivasan et al.[57] investigated the incorporation of an additional basin in a double-inclined SS and using different materials within the basin. Wick materials such as jute cloth, waste cotton pieces, and black cotton cloth were used to increase the evaporation area. The results showed that double and single basin stills using mild steel sheets had a maximum fire use efficiency of 2.072% and 1.412%, respectively.

Fin is a low cost heat transfer enhancement technique where the fins at the bottom of the solar evaporator improve the performance by increasing the rate of heat transfer from the basin to the water.[58] However, not much work has been done on fins to improve the distillate yield in many research works. Mevada et al.[59] reviewed the effect of fin configuration parameters on SS performance. The results mainly found that fins increased the surface area of water and thus increased the heat transfer rate. Fins can also be used to reduce the heat loss at the bottom of the solar evaporator.

Surfaces used for evaporation and condensation phenomena play essential roles in the performance of basin-type SS. Kabeel et al.[60] devised a concave wick surface that was used for evaporation, while four side pyramidal shapes were still used for condensation, and a jute wick was used to increase the evaporation surface area. The results reveal an average daily distillate productivity of 4.1 $l/m^2$, a maximum instantaneous system efficiency of 45%, and an average daily efficiency of 30%.

The glass's inclination angle has an impact on parameters such as yield and instantaneous efficiency. Tiwari et al.[61] designed single slope passive SSs with three different condensations covering inclinations of 15°, 30° and 45°. The results indicated that a condensation cover inclination angle of 15° yielded the highest annual yield and distillation efficiency. Samee et al.[62] suggested that the optimal glass cover angle for the designed single basin SS was 33.3° in the arid region of South West Pakista n (33.7° N latitude).

2.2.2 Active solar stills

Active SSs typically attach additional components and require an external energy supply system to drive auxiliary equipment such as solar collectors, condensers, pumps, etc. Active SSs typically produce water at a higher rate than passive SSs.

A solar collector is a device that collects and concentrates solar radiation to maximize the conversion of solar radiation into thermal energy to drive water's evaporation and condensation processes. Different heat collection principles can divide solar collectors into flat-plate solar

collectors, concentrating solar collectors, and vacuum tube collectors. Badran et al.[63] designed a single-stage basin SS connected to a conventional flat plate collector. The results showed that the integrated single basin distiller increased its yield by 231% after 24 hours of operation. Shiv et al.[64] designed a single-slope SS integrated with an evacuated tube collector (ETC). The study results demonstrated further improvement, achieving an optimal daily yield of up to 3.9 kg, with energy and fire use efficiencies of 33.8% and 2.6%, respectively. Concentrating collectors include a receiver and a concentrator, which intercept a large area of direct sunlight and focus it into a small absorption area, thereby increasing the radiant flux. Ashraf et al.[65] designed a parabolic SS consisting of a parabolic disk concentrator as shown in Fig. 3a. Their results showed that the average daily efficiency of the distiller is 34.69% better than different types of solar distillers, and the cost is low enough to be used by rural households.

A condenser is a refrigeration system component that converts a gas or vapor into a liquid. Attaching a condenser to an SS can boost productivity by increasing the condensation rate. Condensers can be divided into external condensers and internal condensers depending on where they are attached. Kumar et al.[66] improved the single-slope SS by attaching an external condenser and compared it with the conventional single-slope SS as shown in Fig. 3b. The distillation efficiency of the improved still was found to increase by 39.49% over the conventional still at a lower cost. Kabeel et al.[67] conducted an experiment on SS using an integrated nanofluid and an external condenser. The results showed an increase in distillation yield of about 53.2%. Ahmed[68] designed a single slope distiller with an integrated dual channel condenser against a distiller without a condenser and found that an additional internal condenser improves the performance of the distiller.

Wind turbines have also been used in SS as rotating shafts to increase distillate production. For example, Mohamed et al.[69] designed small wind turbines as rotating shafts installed in the main SS to break the boundary layer at the water surface of the basin (see Fig. 3c). For the same flow rate, the productivity was inversely proportional to the water depth, and the vibration induced by the rotating shaft induced the droplets to flow from the lid into the collection channel.

The integration of solar chimneys allows for the production of both electricity and fresh water. For example, in Fig. 3d, Mostafa et al.[70] attached a solar chimney to a conventional SS where it can absorb heat from solar radiation, generating a chimney effect that induces airflow due to natural convective forces to enhance the desalination process. They established a mathematical model of the airflow inside the solar chimney and put the experimental results of four aspects (i.e., sunny and cloudy weather conditions, the influence of salt concentration conditions, the influence of the water depth in the basin and the unit effect) to simulate and

validate the simulation. The results of solar water stratification experiments of the integrated solar chimney SS under different operating conditions are also compared. It was found that the solar chimney has a higher efficiency than the conventional solar distillers and a 30% higher desalination rate than solar ponds.

Photovoltaic-thermal (PV-T) distillers combine the technology of photovoltaic (PV) panels and solar thermal collectors in a single system that can generate electricity and heat simultaneously. This system can also utilize efficiently solar energy to increase water production. Gajendra et al.[71] conducted an experiment in a double-slope active SS with a solar PV-thermal system, as shown in Fig. 3e. It was found that the productivity of the improved still was 1.4 times higher than that of the one with single-slope PV-thermal technology.

Utilizing solar photovoltaic-operated fan work is reported to be economical and can increase the evaporation rate. For example, Taamneh et al.[72] investigated the effect of forced convection on the performance of a pyramidal solar evaporator, the experimental system as shown in Fig. 3f. The use of a fan and photovoltaic solar panels proved to be economically viable, resulting in a 25 percent increase in the daily production of freshwater.

Solar Pond (SP) is a remarkable development in renewable energy technology which stores solar energy for many solar thermal applications. SP can provide heat for various applications such as solar heating, cooling and refrigeration. Researchers have utilized salinity gradient and SP with SS to increase the yield. Panchal et al.[73] described how SP can be used to increase the yield of SS by providing hot water through its stored thermal energy. The paper states that the optimum salinity value inside the SP is a critical parameter and key to the performance of SP and SS. The paper also reveals the use of shallow and micro SP in combination with SS to improve production.

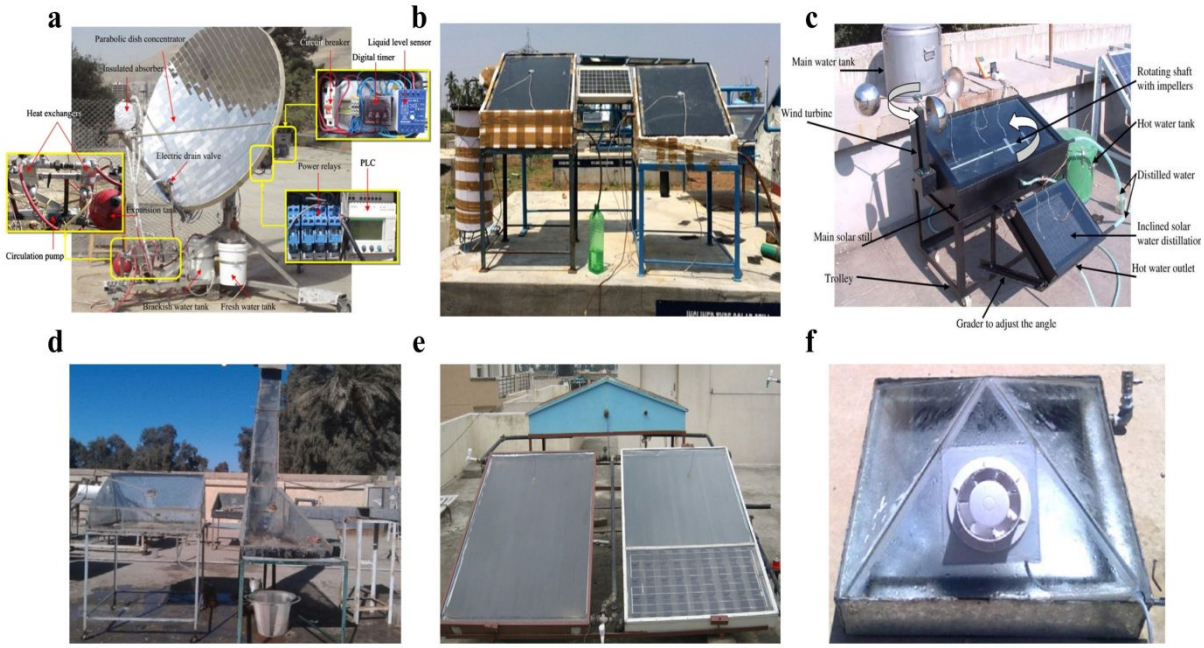


Fig. 3 (a) Photograph of the developed point-focus solar still. (b) Experimental setup of conventional and modified still. (c) Single slope still coupled with wind turbine and inclined solar distiller. (d) Construction of solar basin and solar chimney. (e) Photograph of hybrid photovoltaic thermal (PVT) double slope active solar still. (f) Pyramid solar distiller with fan. (a) Reproduced from ref. 65 with permission from Elsevier, copyright 2014. (b) Reproduced from ref. 66 with permission from Elsevier, copyright 2016. (c) Reproduced from ref. 69 with permission from Elsevier, copyright 2009. (d) Reproduced from ref. 70 with permission from IEEE, copyright 2020. (e) Reproduced from ref. 71 with permission from Elsevier, copyright 2011. (f) Reproduced from ref. 72 with permission from Elsevier, copyright 2012.

## 3 Comparison of different forecasting models

The previous section has outlined the use of design, fabrication, and integration of additional components to modify conventional SS to improve their performance. A lot of research has been carried out on various solar distillation systems. For example, Sharshir et al.[74] reviewed the factors affecting the productivity of SS and techniques for improvement. However, various factors influence the performance of SS, and the expensive and time-consuming nature of experimental work can make performance prediction challenging in certain cases. To be able to predict the performance of SS accurately, different models have been developed, mainly consisting of three methods, as stated in Fig. 4. Accordingly, ML is undoubtedly one of the powerful black-box data analysis tools nowadays, which does not require specific expertise and is able to solve complex situations well. Two other approaches are the numerical solutions of differential equations for heat mass transfer[75,76] and regression modeling.[77,78] In addition, ML has demonstrated its outstanding performance in terms of other domains. For instance, Tao et al.[79] employing ML-assisted nanoparticle synthesis, which focuses on ML algorithms to support nanoparticle synthesis and highlights the key methods for collecting large datasets. Batra et al.[80]

review the emerging materials intelligence ecosystems and discusses the use of emerging ML to address the challenges faced to drive their development. With the improvement of solar distillation technology and the development of artificial intelligence, ML approaches are being progressively employed for solar desalinationas (see Fig. 5). This section then focuses on ML models for predicting SS yield.

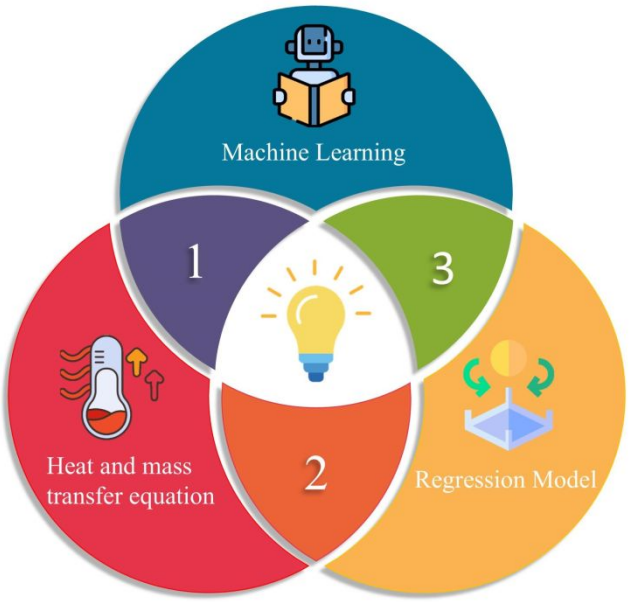


Fig. 4 Three models for predicting the performance of solar stills.

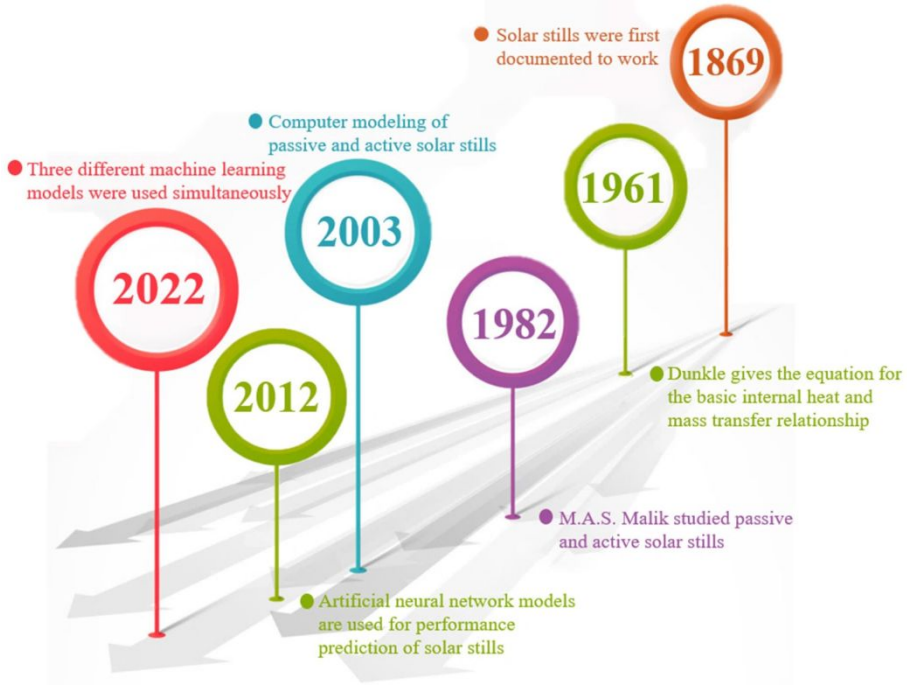


Fig. 5 Evolution of experimental and predictive methods for solar stills.

### 3.1 Use of conventional methods

Previous scholars have used the principle of internal heat-mass transfer to perform mathematical modeling to predict the output of solar desalination systems and Bhatti et al.[81] numerically studied the flow of hybrid nanofluids in porous media. The similar variables were

used to mathematically model the momentum and energy equations, which were solved numerically using the successive linearization method (SLM). Their results show that the new findings are not only consistent but also ensure the accuracy of the present results for mixed nanofluids. In another study, Bhatti et al.[82] discussed diamond (C) and silicon dioxide ($SiO_2$) nanoparticles in a water-based hybrid nanofluid suspended on an exponentially elastic surface to improve the photothermal performance of an energy conversion system, developed nonlinear differential equations, and solved them numerically. The results show that their proposed SLM method is more stable based on numerical comparisons. Shirvan et al.[83] studied the numerical solution of the combined surface radiation and natural convection heat transfer in a solar cavity receiver, which was obtained using the SIMPLE algorithm.

Some scholars have also used regression modeling to predict the production performance of SS. Samuel and Chang[84] designed a prototype SS for use in this area to solve the water problem of the remote islanders, carried out data collection, built a multivariate regression model and used the TMY data of Dongji Islet for quantification of SS. The results showed that the established multivariate linear regression model had R-squared ($R^2$), adjusted R-squared ($R_a^2$), mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) values of 99.5%, 99.4%, 0.144, 0.167, and 9.71%, respectively. This demonstrates that applying multivariate linear regression and optimal subsetting techniques based on TMY data has proven to be a viable approach to modeling the productivity of a prototype SS.

The data-driven approach of Design of Experiments (DOE) has a corresponding application on SS systems. Primarily derived from statistical methods, the DOE approach significantly reduces the cost and time of the data collection process when compared to the traditional one-factor experimental approach. Additionally, the DOE approach considers the interaction of independent variables on system behavior and can be effectively applied for performance prediction and optimization purposes.[85] Response surface methodology (RSM), Taguchi methodology (TM), and analytical factorial design (FD) are three major statistical. SS system involves several operating parameters, such as water flow rate, heat absorption area, inlet temperature, tilt angle, etc. DOE can help optimize these parameters to enhance desalination efficiency. Meanwhile, the operation of SS system depends on environmental conditions, such as sunlight intensity, wind speed, humidity, etc. DOE can help the system to find the best working conditions under various environments and ensure the stability of the system under different climatic conditions. DOE can not only optimize a single objective (e.g., freshwater yield), but also be used for multi-objective optimization. For example, finding the

best balance between freshwater yield, energy consumption and system cost. DOE tools that have been widely used to analyze the performance and optimization of desalination systems. Rejeb[86] developed a polynomial regression mode for predicting the efficiency of solar power generation using numerical equilibrium energy modeling. The author carried out a statistical RSM model to investigate the interactions between the factors under study and their combined effects on daily distilled water productivity. The results showed that the $R^2$ and $R_a^2$ were 0.9906 and 0.9818, respectively. The predictions of the polynomial model were in good agreement with the numerical results of the transient thermal numerical model. Khalifa and Lawal[87] optimized the air-gap membrane distillation (AGMD) desalination system using Taguchi orthogonal design arrays and the RSM. An analysis of variance was then employed to analyze the model and the significant effect of each operating parameter on flux. Their results suggested the maximum fluxes of the Taguchi method and RSM were 76.046 kg/m$^2$ h and 76.998 kg/m$^2$h, respectively, under optimal conditions. Allah et al.[88] used a design of experiments approach to analyze the input factors affecting the performance. The effects of nine factors, such as solar radiation, basin area, and brackish water depth, on the performance of the solar evaporator were investigated, and an accurate theoretical model of the thermal behavior of the solar distiller was developed. The results indicate that the established mathematical model can accurately describe the highly complex behavior of SS.

### 3.2 Use of classical ML models

The ML process applied to predict system performance can be divided into three steps: data processing, model construction, and model validation. The data processing step involves collecting experimental or theoretical results and organizing the data into a training set, a validation set, and a test set. The training set is used for model training and parameter tuning, while the validation set is employed to select the model and fine-tune hyperparameters. Eventually, the test set is used to evaluate the model's performance. Model building is the process of learning and training data to generate mathematical models that can make predictions or decisions on unknown data. The entire model-building process is an iterative one that usually requires several attempts and adjustments to find the best model and parameter combination. Also, as data and tasks change, the model may need to be updated and retrained periodically to maintain its performance and accuracy. Model validation in ML is the process of evaluating the performance of a model on unseen data. The purpose of validating a model is to ensure that the model generalizes well to new data, that the patterns learned by the model during training can be generalized to future data. The most commonly used assessment metrics are the $R^2$, RMSE, MRE, MAPE, overall index (OI), efficiency coefficient (EC), and coefficient of variation

(COV).[89,90] Moreover, ML techniques can be categorized explicitly into classical ML methods and deep learning (DL) methods. Classical ML methods used for the study of desalination systems include Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Regression Models (RM).[85] This subsection focuses on applying classical ML methods in solar distillers.

In SS desalination systems, most studies use ANNs to predict their performance. ANN is an ML model that mimics the structure and function of the human nervous system, consisting of multiple artificial neurons (or nodes) interconnected through connections (or weights). ANNs typically involve an input layer, a hidden layer (optional), and an output layer,[91] where each layer consists of multiple neurons.[92,93] The structure of four of these models of ANN: Multi-Layer Perceptron (MLP), Wavelet Neural Networks (WNN), Radial Basis Function (RBF), and Elman neural network (ENN), as shown in Fig. 6.[30] ANN has the ability to learn and construct nonlinear and complex relational models, which is crucial because many of the relationships between inputs and outputs in desalination systems are nonlinear and complex. Santos et al.[94] used ANN and local weather data to predict yields of two different commercial SS while using ANN to determine the minimum amount of inputs required for accurate SS performance. It was found that 31-78% of the predictions of the ANN model were at 10% of the actual yield depending on the input variables selected. Abdelhafez et al.[95] predicted the thermal efficiency (η) of triple SS utilizing three ANN models, namely Feedforward (FF), Elman, and Nonlinear Autoregressive Exogenous (NARX). As shown in Fig. 7a, the FF model had the best predictive ability with the highest $R^2$ of 0.99838. Abujazar et al.[96] utilized a Cascaded Forward Neural Network (CFNN) model to predict the yield of an inclined stepped SS system. They compared its predictions with those of regression and linear models, using three statistical error terms for evaluation. As a result, the proposed CFNN has minimum values of 22.48%, 18.51%, and -14.51% for RMSE, MAPE, and MBE respectively. Fig. 7b shows that CFNN predicts the productivity of the system more accurately than other models.

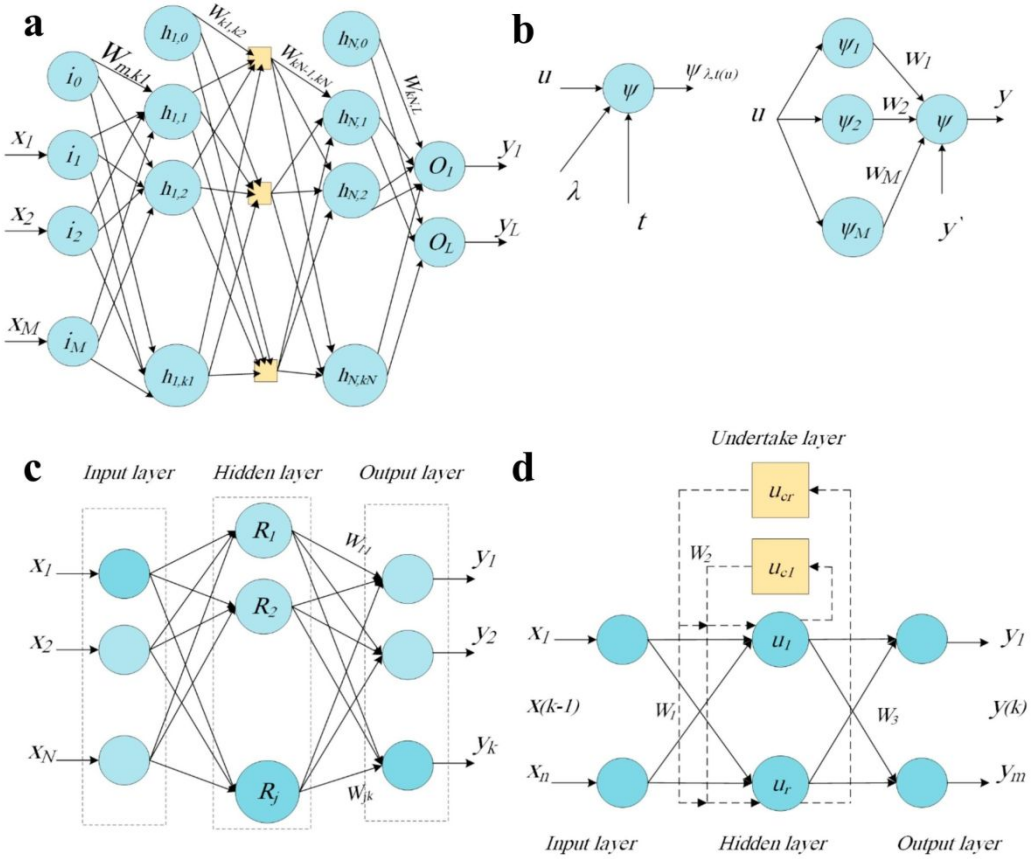


Fig. 6 The structure of (a) MLP, (b) WNN, (c) RBF and (d) ENN. (a-b)Reproduced from ref. 30 with permission from Elsevier, copyright 2019.

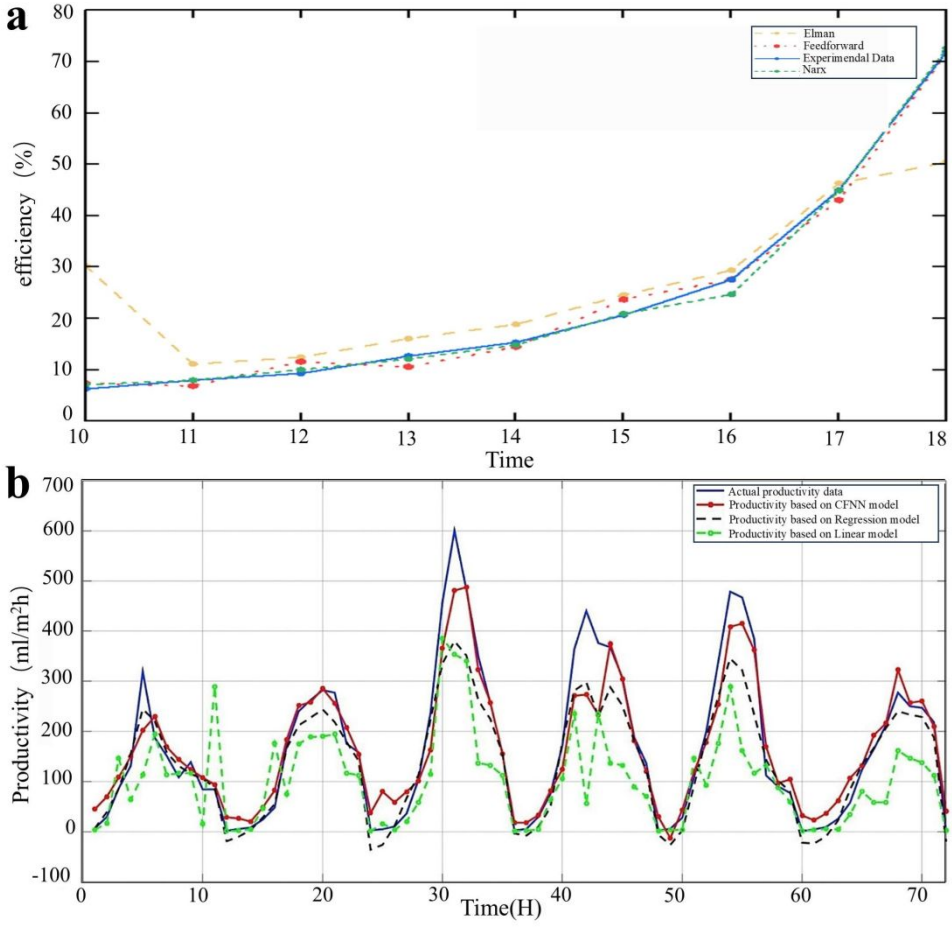

Fig. 7 (a) Comparison between experimental and estimated thermal efficiency. (b) Hourly sample of the predicted the solar still productivity by CFNN, regression and linear models. (a) Reproduced from ref. 95 with permission from JOCET, copyright 2013. (b) Reproduced from ref. 96 with permission from Elsevier, copyright 2018.

Although ANN has a wide range of applications in SS systems, the selection of an appropriate No single ML model is universally best for predicting the performance of a specific SS system. Several studies have used SVM, RF, and ANFIS for modeling and prediction.[97] In particular, SVM can be used to model the relationship between process responses (e.g. freshwater yield, evaporation efficiency, etc.) and process descriptors (e.g. environmental conditions, material properties, operating parameters, etc.). First, data on the input characteristics (descriptors) and output responses of the SS apparatus are collected. To ensure data availability and accuracy, appropriate data cleaning and normalisation is then performed. Then, a suitable kernel function such as RBF kernel is selected to capture the complex interactions between the input features. Finally, model training, tuning and evaluation are then performed to accurately predict the performance of the SS system. This binary classification model is used for classification and regression analysis. Its basic principle involves finding the optimal hyperplane that separates different data classes, maximizing the margin between them to classify or regress the data effectively. The schematic diagram of SVM is presented in Fig. 8a. ANFIS, on the other hand, is an intelligent ML method that combines fuzzy inference with ANN.[98] The structure diagram of ANFIS is depicted in Fig. 8b. The ANFIS model demonstrated a higher ability to deal with real-data applications than other traditional ANN methods.[99,100] It is particularly effective in addressing problems characterized by high uncertainty and ambiguity, owing to its strong adaptive capabilities. Additionally, ANFIS can dynamically adjust system parameters in response to changes in the environment and input data. For example, Elsheikh et al.[101] designed a low-cost heat localization bilayered structure (HLBS) that can efficiently convert the absorbed solar energy into thermal energy to improve conventional SS's performance. They used three machine learning methods (i.e., ANN, SVM, and ANFIS) to predict water production and compared the prediction results between a CSS and an modified solar still (MSS). As shown in Fig. 9a, SVM exhibited a higher accuracy in predicting water production than the other two models. Lastly, RF is an ML classifier that consists of multiple decision trees. The final output category is determined by the majority vote of the individual trees. This method is known for its high accuracy.[102.103] Its main process presented in Fig. 8c could be summarized into three steps: data manipulation,[104] model construction, and model optimization. In a study, Wang et al.[105] designed a tube SS. To accurately predict its performance, they developed an ANN model, an RF model, and a traditional multiple linear

regression model based on experimental data. They optimized and compared these models using Bayesian optimization for hyperparameter tuning. The results showed that the determination coefficients of RF, ANN and multiple linear regression were 0.9745, 0.7098, and 0.9267. As shown in Fig. 9b, the superiority of RF was well demonstrated with minimal errors. Similarly, Kandeal et al.[106] utilized four different ML models (i.e., ANN, RF, Support Vector Regression (SVR), and Linear Support Vector Regression (LVR)) to predict the performance of double slope solar still (DSSS). They optimized these models using the Bayesian optimization algorithm (BOA) and conducted training, testing, and validation for each of the models. As a results, RF was found to be the most accurate ML model with the highest $R^2$ of 0.983 (see Fig. 9c.)

In addition, there are a number of ML models that have been applied to SS system prediction. Random Vector Functional Link (RVFL) is a fast neural network model widely used for modelling and prediction of various nonlinear systems.[107] RVFL is able to efficiently deal with complex nonlinear problems through randomly generated weights and directly connected input feature layers. Previous studies relying on artificial neural networks only used connections between hidden and output layers without considering connections between input and output layers. This can lead to overfitting problems and also reduces the efficiency of the method. RVFL, through its efficient nonlinear fitting capability, is able to quickly predict the output of the system based on the environmental conditions and system parameters input to the SS system. Sharshir et al.[108] proposed a novel developed pyramid solar still (DPSS) integrated with copper plate and graphite nanofluid to predict the hourly freshwater (HF) and instantaneous energy efficiency (IEE) of the DPSS using the FA-RVFL model of firefly algorithm (FA) which simulates the behaviour of fireflies, Fig. 8d illustrates the structure of the developed technique. The prediction results are also compared with RVFL, SVM and conventional ANN. The results show that the proposed FA-RVFL model is characterised by coefficients of determination of 0.981 and 0.999 for the total HF and IEE datasets, respectively, and regression values of 0.996 and 0.999, respectively.It proves that the developed FA-RVFL has an excellent performance in predicting the FA and IEE of DPSS. K-Nearest Neighbors (KNN) is a simple but effective supervised learning algorithm widely used in classification and regression tasks. It finds the K closest neighbours to the target point in the feature space by calculating the distance (usually Euclidean distance) between data points and makes predictions based on the labels of these neighbours. KNN can be used to classify different design options for SS, such as classifying different material combinations, different structural designs, etc. By analysing the performance metrics of different designs, KNN can help researchers to quickly identify superior design

solutions. Alawee et al.[109] came out with an innovative approach to predict the cumulative distillate yield of a double slope solar still using correlation analysis, ReliefF for feature selection, and the KNN algorithm. The analysis was based on a dataset based on six cases, which included variations in distillate yield relative to different operating environmental conditions. Key features that have a significant impact on the overall performance were identified to manage the productivity of the solar still. The results showed that the best model was evaluated based on the $R^2$ , RMSE and CVRMSE of the KNN model and the best model obtained scores of 0.995, 0.0033 and 0.1666, respectively, demonstrating the effectiveness of the proposed machine learning approach in predicting distillate output. Gaussian Process Regression (GPR) is a nonparametric Bayesian regression method widely used to model complex nonlinear relationships. In SS research and applications, GPR can provide accurate performance prediction, optimal design, and uncertainty quantification. Kottala et al.[110] demonstrated the performance of a novel trough collector using a natural recirculation open loop and experimentally evaluated it for several types of solar radiation data. Seven different ML models were used to predict the instantaneous thermal efficiency of the developed system for various solar radiation categories. The results show that the GPR model shows a higher predictive performance than the other developed ML models (i.e., RMSE = 0.0049, $R^2$ = 0.9977).

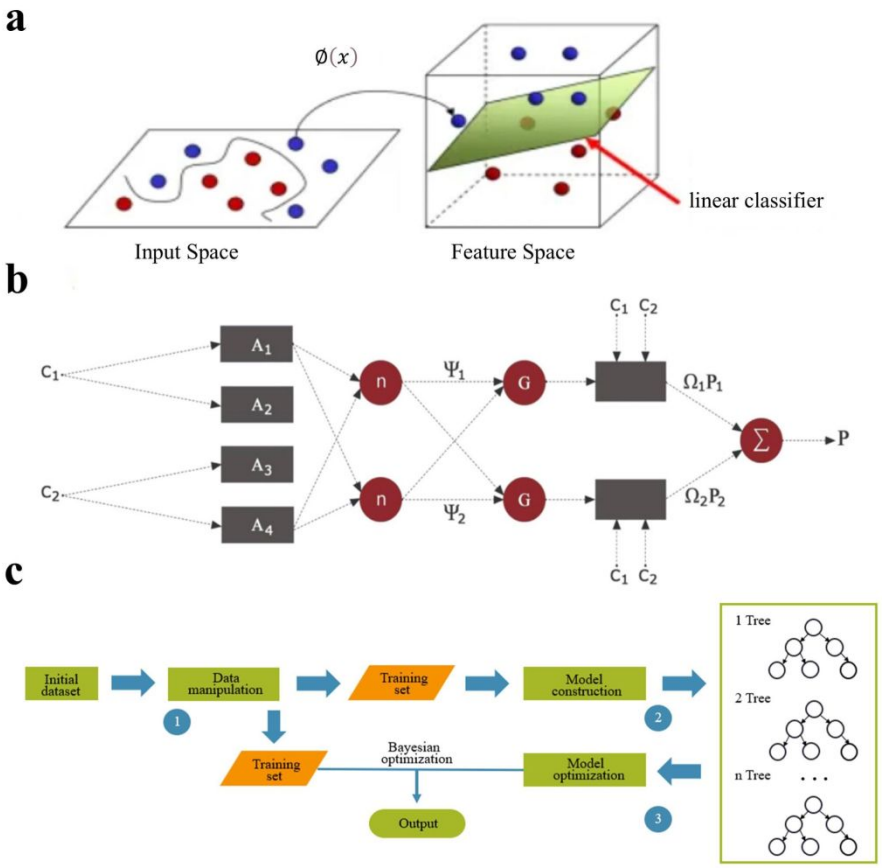

Fig. 8 Structure of (a) SVM and (b) ANFIS (c) Data flow diagram of RF. (b) Reproduced from ref. 111 with permission from Frontiers in Energy Research, copyright 2021. (c) Reproduced from ref. 105 with permission from Elsevier, copyright 2021.

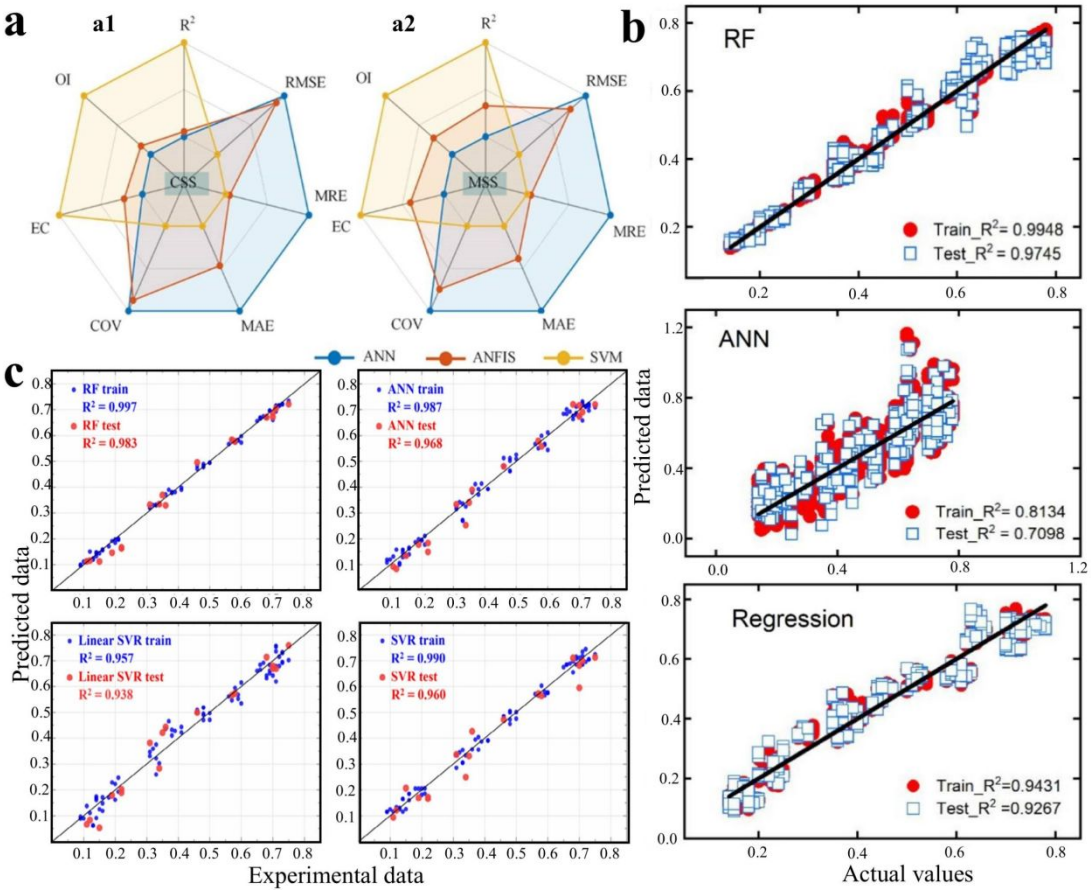


Fig. 9 (a) Spider plot of different statistical measures used to evaluate the performance of the ML models for: a1 CSS; a2 MSS. (b) Prediction results of different models. (c) The linear correlation between the experimental and predicted data for the four ML models. (a) Reproduced from ref. 101 with permission from Elsevier, copyright 2022. (b) Reproduced from ref. 105 with permission from Elsevier, copyright 2021. (c) Reproduced from ref. 106 with permission from Wiley, copyright 2021.

The previous section discussed the integration of solar desalination with other components. In addition, the use of nanomaterials in conjunction with SS can also increase the yield. Micro/nanomaterials hold significant promise for SS due to their high thermal conductivity, extensive surface area, and superior solar energy absorption properties. By leveraging these materials, it is possible to develop highly efficient solar evaporation systems, which are crucial for enhancing solar energy utilization. Additionally, data-driven approaches can be effectively applied to predict the performance of these systems. For example, Bagheri et al.[112] designed an SS incorporating a cylindrical parabolic collector and solar panels to capture additional solar energy. They modeled the system using ANN and compared its performance with mathematical modeling approaches. They found that the ANN model predicts more accurately than the proposed first-principles model and minimum error rate. The best response is obtained for the neural network n = 7 with $R^2$ = 0.999820171, MSE = 1.94E-06, %AAD (average absolute

deviation) = 0.426716116. Bahiraei et al.[113] developed a novel nanofluidic SS equipped with a thermoelectric cooler and applied $Cu_2O$-water nanofluid in a solar distillation cell. They then predicted the yield using a MLP neural network, which was optimized with both the Imperialist Competitive Algorithm (ICA) and the Genetic Algorithm (GA). The results stated that the modified SS had a much higher yield than other SS. The $R^2$ correlation values of the MLP turned out to be 0.9458 and 0.8883 in the training and testing phases, respectively, which were able to predict the yield accurately. In another study, Bahiraei et al.[114] utilized ANFIS and ANN to predict the energy efficiency of a single-slope SS equipped with a thermoelectric module. They also applied $Cu_2O$ nanoparticles to a solar distillation cell with Particle Swarm Optimization (PSO) to augment both ML models. As a result, the training and testing $R^2$ values of ANN-ANFIS were 0.95-0.95 and 0.93-0.99, respectively. It means that the yield was predicted more accurately. Bamasag et al.[115] designed a dish solar distiller (DSD) incorporating phase change material (PCM), specifically paraffin wax blended with CuO nanoparticles. They used three machine learning approaches (i.e., ANN, ANFIS, and SVM) to predict the water productivity of the solar still. The results indicate that the average daily distillation rate of the stepped DSD with phase change material is approximately 178% higher than that of the CSS. Among the models, the SVM demonstrates the highest $R^2$ of 0.99, the smallest RMSE ranging from 2.19 to 3.17, and the smallest normalization error between -0.02 ~ 0.08, as shown in Fig. 10. This suggests that the SVM outperforms both the ANN and ANFIS in accurately predicting the yield. Wang et al.[116] compared RF with pairwise plots and Pearson correlation analysis, to demonstrate that RF is more advanced and accurate in evaluating the importance of factors in materials and systems. This comparison aims to enhance the utilization and design of micro/nanomaterials. They found that RF can obtain reasonable weight values better than traditional methods.

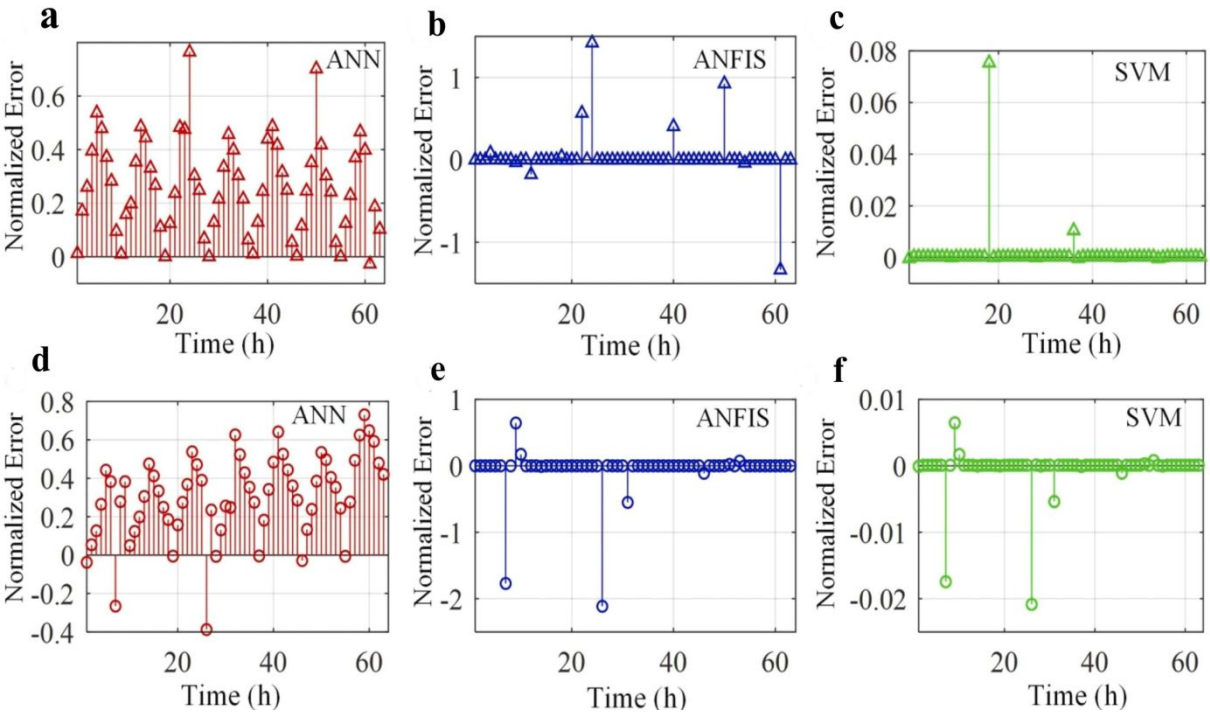


Fig. 10 The normalized error between experimental and predicted data using: (a) ANN for conventional solar distiller; (b) ANFIS for conventional solar distiller; (c) SVM for conventional solar distiller; (d) ANN for modified solar distiller; (e) ANFIS for modified solar distiller; (f) SVM for the modified solar distiller. (a-f) eproduced from ref. 115 with permission from Wiley, copyright 2022.

ML approaches can effectively model the dynamic performance of solar desalination systems. Sohani et al.[117] developed three different ANN models: back-propagation (BP), FF, and RBF. Their goal was to identify a model that accurately predicts the water temperature and yield of the enhanced SS system. The performance of the three ANN models was assessed using MAE and $R^2$. Consequently, FF and RBF were the most accurate models in predicting the water temperature with MAE and $R^2$ of 3.56-2.82 and 0.96-0.98, respectively. Given the varying criteria and performance characteristics in SS desalination technology, it is essential to examine multiple types rather than relying on a single structure to achieve the highest possible prediction accuracy. The production rate of SS can also be predicted from conventional weather information using ML methods. Gao et al.[118] used RF optimized by BOA to predict the yield of two kinds of SSs. They compared it with MLR and traditional prediction models to verify the accuracy of the models. They used conventional weather forecast data as inputs to the models while the actual measured data were used to train the models. As a result, the determination coefficients of the two SSs predicted by RF were 0.935 and 0.929, significantly higher than those of the MLR (0.767) and the conventional model (0.829 and 0.847). This indicates that RF is a reliable method for yield prediction.

Each machine learning model has its unique advantages and disadvantages in SS performance prediction, and the selection of an appropriate model depends on the

characteristics of the data, prediction accuracy requirements, and computational resources. Multi-model integration or hybrid model approaches often lead to better prediction results. Model performance and prediction accuracy can be significantly improved by proper data preprocessing and feature selection. Firstly, linear regression is a simple and effective method suitable for modelling linear relationships, although it is less adaptable to nonlinear data. Secondly, ANN is able to simulate complex nonlinear relationships due to its multilayer structure and is well suited for large-scale datasets, but the training time is long and difficult to interpret. SVM, on the other hand, performs well in small samples and high-dimensional data and is able to deal with complex nonlinear problems, but the tuning of its parameters and the selection of kernel functions are more complicated. In addition, ANFIS combines neural networks and fuzzy logic to deal with uncertainty and ambiguity information, and performs well in performance prediction of complex multivariate systems. ANFIS is able to deal with nonlinear relationships and provide high prediction accuracy, which is particularly suitable for scenarios that require the simultaneous processing of both exact and fuzzy data. DT and RF are also widely used, with the former making decisions through a tree structure and the latter improving prediction accuracy by integrating multiple decision trees, both of which can effectively deal with complex nonlinear relationships and have good noise immunity, but RF may face overfitting problems. RVFL extends the input features by introducing random vectors into the network, which can quickly process large-scale data and give efficient predictions without relying on complex training. In SS, RVFL is suitable for real-time prediction tasks that require fast response by reducing the training time and providing good nonlinear fitting ability. The KNN algorithm is simple to implement and adaptable through distance-based prediction, but has high computational complexity and sensitivity to noise when dealing with high-dimensional data. GPR, a non-parametric Bayesian regression method, is particularly suited to small samples and complex data, and is able to provide quantification of the uncertainty in the prediction, however, its computational complexity is high in the presence of large data volumes.

### 3.3 Use of DL models

DL is a new research direction in the field of ML that has a more complex structure and mainly requires a large amount of data. DL methods applied to desalination systems generally be categorized into three approaches, namely Recurrent Neural Networks (RNN), Conventional Neural Networks (CNN), and Deep Reinforcement Learning (DRL). In SS systems, long-short-term memory neural networks (LSTM) have been used to predict their performance. The LSTM is a recurrent ANN used with DL. The main advantage of this neural network over conventional FF neural networks is its ability to memorize patterns for long periods of time owing to the

advanced structure associated with the feedback connections. Elsheikh et al.[119] designed a stepped SS fitted with copper corrugated absorber plates using the LSTM model for prediction and comparison with CSS. The proposed model was validated using field experimental data for training, and time series of freshwater production were used to train the model. The results revealed that the stepped SS had a production rate that was 128% higher than the CSS. Additionally, the model's coefficients of determination were 0.97 and 0.99, and its RMSE values were 0.0067 and 0.0021, respectively. Consequently, the model accurately predicted the production rate of the SS.

# 4 Optimization

## 4.1 Training models

The used algorithm, the training model, and other factors affect the model's accuracy. No specific general ML model outperforms other models in predicting the desalination system's performance. By optimizing the ML model with an appropriate training model, the accuracy of desalination yield prediction can be further enhanced. Chauhan[120] used a multilayer perceptron neural network based on a supervised learning mechanism to predict the thermophysical properties of moist air inside the chamber of an SS. Specifically, six distinct training algorithms were used to train the model, namely one-step secant, conjugate gradient Powell-Beale restarts, conjugate gradient Fletcher reeves update, resilient backpropagation, scaled conjugate gradient, and Levenberg-Marquardt (LM). The performance of each algorithm was evaluated using six statistical parameters. In Fig. 11a, the correlation coefficients of the LM algorithm for training, testing, and validation are 1, 0.99988, and 0.99995 for all modeling phases, and the slopes and intercept values of the LM algorithm are approximated to be 1 and -0.003 for all the data. Accordingly, the LM algorithm provides more accurate predictions on the thermophysical properties of moist air in SS systems compared to other algorithms. Mashaly and Alazba[121] utilized three algorithms including the conjugate gradient backpropagation with Fletcher-Reeves restarts, the resilient backpropagation, and LM to predict the productivity of SS in hyper-arid environments. The predictions were evaluated by comparing with experimental results using four standard statistical performance metrics, namely RMSE, efficiency coefficient (E), OI, and coefficient of residual mass (CRM). As a result, the model based on the LM algorithm has the highest overall $R^2$ value of 0.99437 , as depicted in Fig. 11b. In addition, the LM algorithm has the minimum average RMSE (0.024), the maximum average E (0.989), the maximum average OI (0.981), and the minimum average CRM (-0.003) for all modeling phases. Hence, LM was found to be the most effective algorithm for predicting solar distillers' productivity. Chauhan[122] also used an ANN model based on the LM algorithm to predict the

fraction yield of CSS and modified solar still integrated with earth distillate (MSSIE) yields. Their results indicated that the ANN model trained by the LM algorithm is accurate in predicting the distillate yield of the distillation unit. In Fig. 11c and 11d, the correlation coefficients of the ANN based on the LM algorithm for the training, testing, and validation of the overall data for CSS and MSSIE were introduced, respectively.

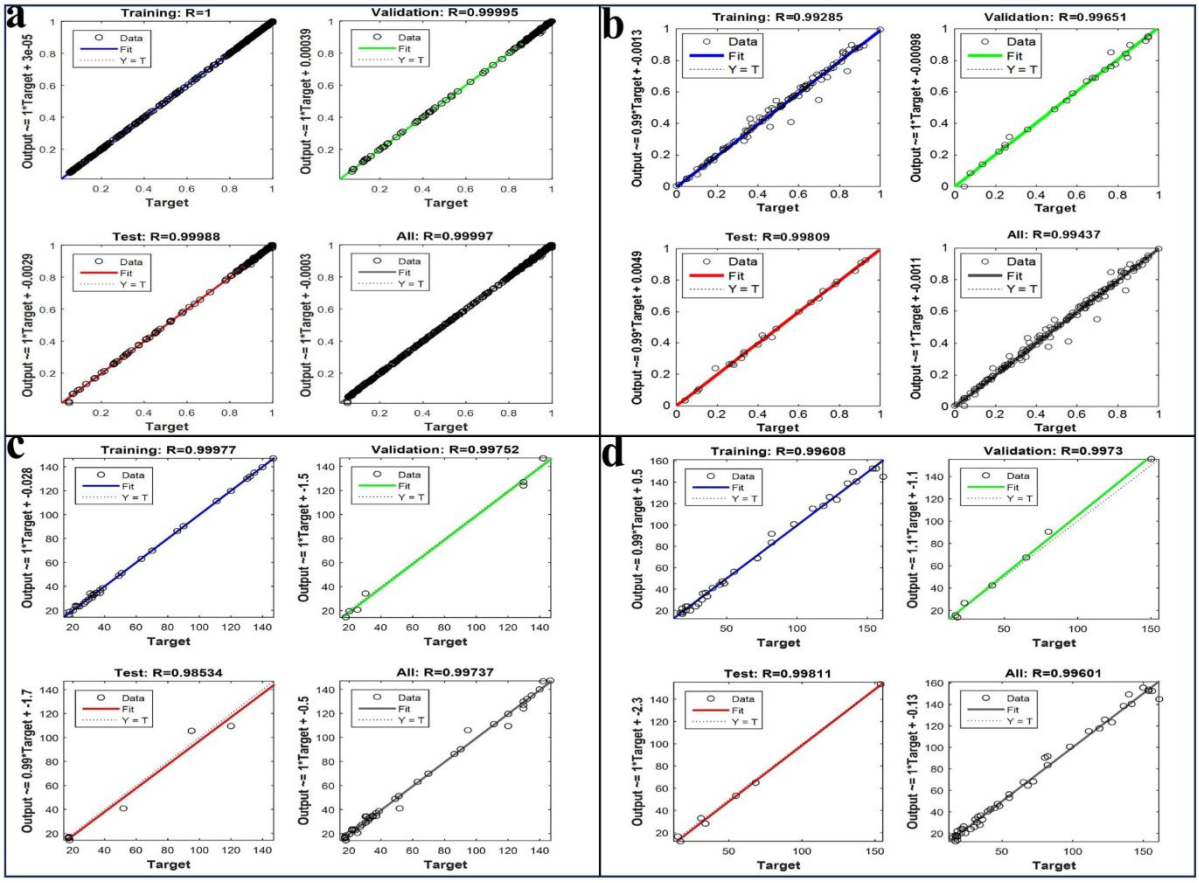


Fig. 11 (a) Regression analysis with training, testing, validation, and all data with the LM algorithm. (b) Regression analysis with training, validation, testing and all data with the LM algorithm. (c) Regression analysis of the LM algorithm in CSS and (d) Regression analysis of the LM algorithm in MSSIE with training, validation, testing, and all data. (a) Reproduced from ref. 120 with permission from Elsevier, copyright 2020. (b) Reproduced from ref. 121 with permission from IWA, copyright 2015. (c,d) Reproduced from ref. 122 with permission from Taylor & Francis, copyright 2022.

### 4.2 Hyper-parameter tuning

In machine learning, tuning the hyperparameters of a model is an important part of improving the performance of an algorithm. Adjusting the different hyperparameters is one of the important factors affecting the accuracy of the data-driven approach in predicting the output of SS. The optimization of hyperparameters can be achieved by different optimization algorithms. Nazari et al.[123] used the ICA algorithm with an optimized ANN model to predict the energy efficiency, exergy efficiency, and water productivity of single-slope SS. Their results indicate that the MAE of the ANN-ICA correspondly reduces the prediction of water productivity,

energy efficiency, and fire efficiency by 54.30%, 40.11%, and 53.35% compared to conventional ANN, whereas the RMSE were approximately 15.77, 1.37, and 0.29, respectively. As shown in Fig. 12a, $R^2$ with MLP-ICA had an increment of 0.0280, 0.0186, and 0.0450 for water productivity, energy efficiency, and fire use efficiency, respectively. Essa et al.[124] used Harris Hawks Optimizer to improve the conventional ANN and predict the yield of three different distillation systems including passive SS, active SS, and active SS with a condenser. Their results were also compared with those obtained from two other models: a support vector machine and a conventional ANN. It was found that the cumulative yield of the active distiller combined with the condenser increased by 53.21%. Notably, Harris Hawks Optimizer's ANN has the highest $R^2$ value of 0.98 and 0.97 for passive and active solar distiller prediction, respectively, and hence the best model of the HHO-ANN type. It should be noted that the choice of hyperparameters is not static. Each optimization algorithm has its own advantages and limitations and needs to be optimized to the specific problem and dataset. Wang et al.[105] built three different models to predict the performance of tubular SSs including classical ANN, RF, and traditional multiple linear regression. All models were optimized using the BOA algorithm. As shown in Fig. 12b, the results show that though the performance of both RF and ANN is improved, the ANN is more sensitive to the response of BOA and the error of ANN is well improved. Bahiraei et al.[113] optimized MLP neural network using ICA[125] and GA[126] was used to predict the water production of nanofluidic solar thermoelectric distillers in 48 samples and 38 records (where 8 were used as training data, and the remaining 10 to evaluate the generalization ability of the developed network). The results show that the RMSE of GA-MLP and ICA-MLP is reduced by 40.49% and 62.01% in the testing phase compared to MLP. So, the application of GA and ICA has a significant effect on the accuracy of MLP, while ICA has a better optimization than GA. Bahiraei et al.[114] utilized the PSO-enhanced ANFIS and MLP neural network to predict the energy efficiency of a solar distiller with a single slope thermoelectric module, respectively. As a results, the $R^2$ values of PSO-ANN and PSO-ANFIS are 0.9225-0.9597 and 0.9471-0.9984, and the RMSE values are 5.7861-3.5158 and 4.0279-1.9956 for PSO-ANN and PSO-ANFIS, respectively, in the training phase compared to the original ones. Therefore, PSO improves the training accuracy of ANFIS more than ANN. It can also be derived from the entire spatial distribution of actual and estimated energy efficiencies, as depicted in the 3D graphs in Fig. 12c. In these graphs, the impact parameters of X8 and X2 are shown on the x- and z-axes, respectively, and the corresponding Y (energy efficiency) is shown on the y-axis.

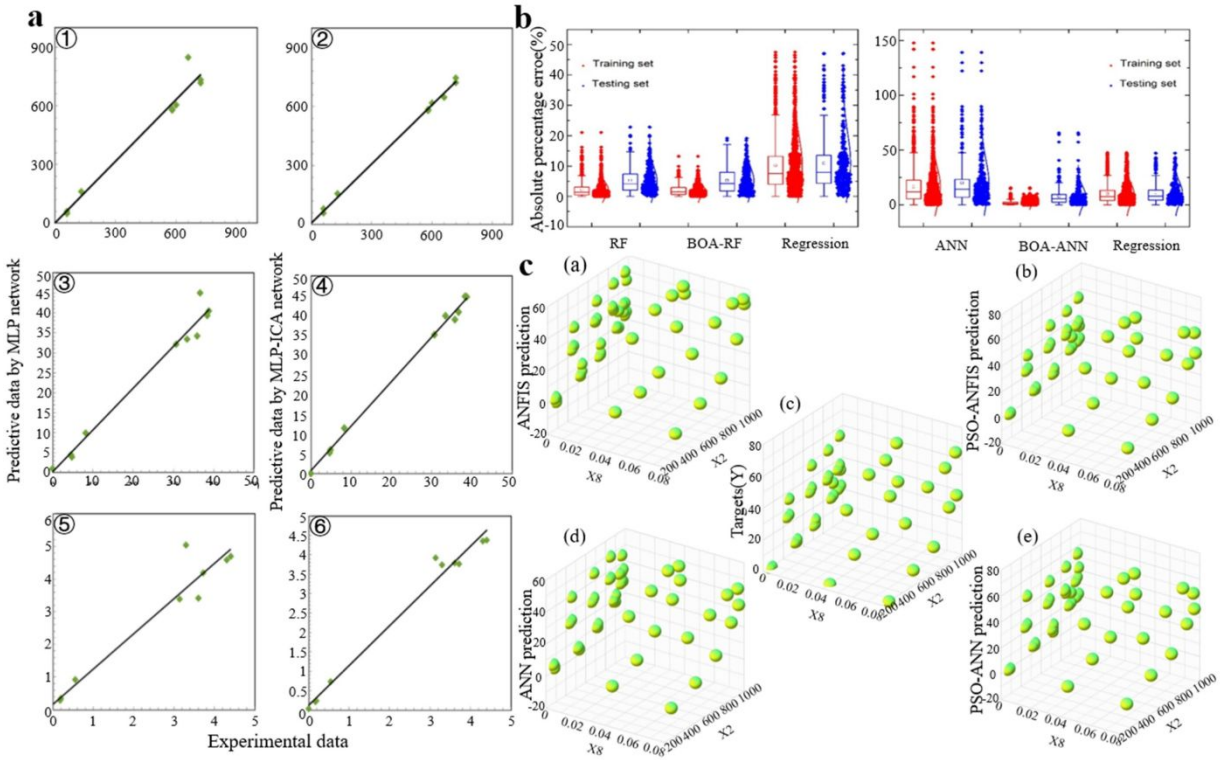


Fig. 12 (a) The correlations of the actual data and predicted data based on the test data for: a1 MLP, water productivity, a2 MLP-ICA, water productivity, a3 MLP, energy efficiency, a4 MLP-ICA, energy efficiency, a5 MLP, exergy efficiency, a6 MLP-ICA, exergy efficiency. (b) The absolute percentage error in different models. (c) The 3D scatter plot of the results of c1 ANFIS, c2 PSO-ANFIS, c3 observed EF, c4 ANN, and c5 PSO-ANN prediction. (a) Reproduced from ref. 123 with permission from Elsevier, copyright 2020. (b) Reproduced from ref. 105 with permission from Elsevier, copyright 2021. (c) Reproduced from ref. 114 with permission from Elsevier, copyright 2021.

### 4.3 Feature selection

Feature selection is the process of reducing the number of input variables in developing a predictive model. The number of input variables is reduced could ensure an optimal computational cost of modeling. Hence, selecting appropriate features is very important. For the SS system, the inputs are chosen mainly from parameters such as solar radiation, wind speed, water depth, water and glass temperatures, and ambient air. Outputs are mainly selected from parameters such as water productivity. The combination of model inputs and the appropriate training and test data lengths is one of the main factors limiting the accuracy in prediction of nonlinear models.[127] There are various methods used to extract the most correlated variables. Among these, RF[128] can be used for feature selection by assessing the importance of individual features. RF calculates the contribution of each feature, allowing for the identification of the most influential factors and providing insights into the relationships between input and output variables. For example, Wang et al.[105] used RF to assess the importance of various performance

parameters, such as solar radiation intensity (IR), wind speed ($V_w$), basin slab temperature ($T_b$), brine temperature ($T_{sw}$), cover temperature ($T_c$), and ambient temperature ($T_\infty$). They investigated how these parameters affect the soil moisture evaporation rate. The results showed that the parameters that had the greatest effect on the rate of water evaporation were $T_{sw}$ and $T_b$, which accounted for 40.87% and 32.43%, respectively(Fig. 13a). Kandeal et al.[106] utilized RF in basin nanofluid ($T_{w,basin}$), vapor ($T_{vapor}$), inner glass ($T_{glass,in}$), outer glass ($T_{glass,out}$) temperature, and r radiation intensity (SR) to screen the factors that have the greatest contribution to freshwater productivity. They stated that $T_{w,basin}$ and $T_{vapor}$ are important parameters, accounting for about 66% and 13%. As shown in Fig. 13b, these parameters mainly affect vapor generation and directly impact freshwater productivity. Gao[118] used the BOA-optimized RF model to predict the performance of SS under the influence of weather variations and to find the weather parameters closely related to SS. Their results suggested that the maximum temperature ($T_{max}$), relative humidity (RH), and minimum temperature ($T_{min}$) are three factors with the highest correlation (i.e., 41%, 20%, and 18%, respectively, see Fig. 13c) with yield. Also, the prediction model is universally applicable so it can be extended to any other type of SS.

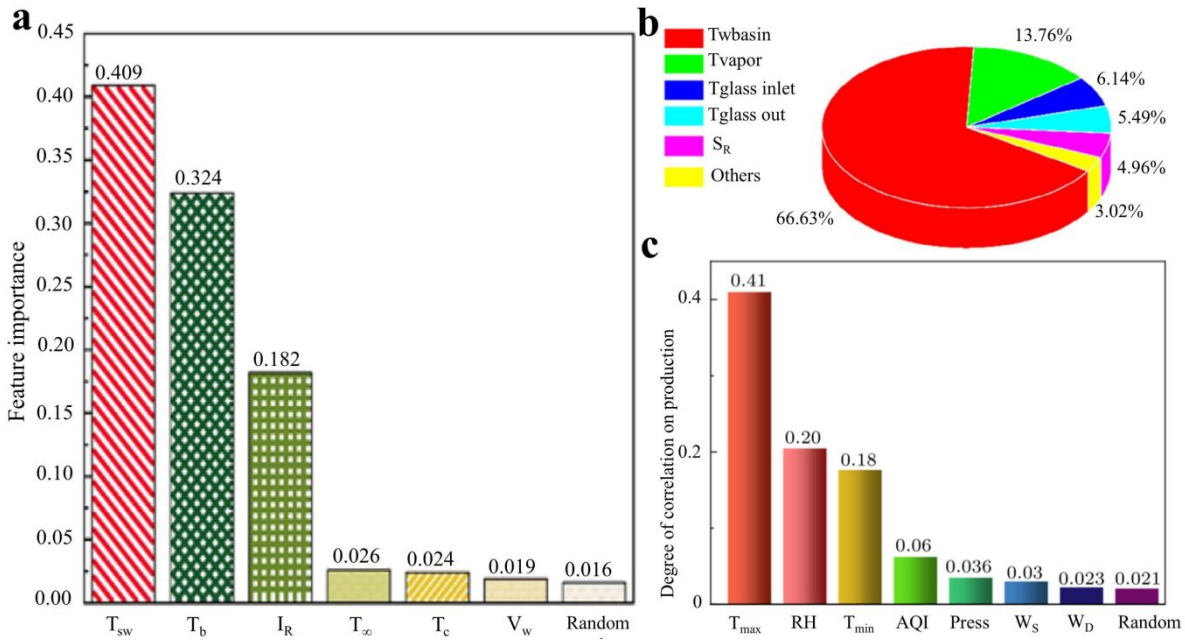


Fig. 13 (a) Results of feature importance. (b) feature importance based on the RF algorithm. (c) Degree of correlation between the production of solar stills and conventional weather forecasting parameters. (a) Reproduced from ref. 105 with permission from Elsevier, copyright 2021. (b) Reproduced from ref. 106 with permission from Wiley, copyright 2021. (c)Reproduced from ref. 118 with permission from Chinese Physics B, copyright 2023.

### 4.4 Datasets

Data sets are an essential part of machine learning and serve as its foundation. The size and quality of the data often determine the performance of machine learning. Therefore, it is more crucial to focus on obtaining high-quality data rather than relying on sophisticated algorithms. However, data collection processes are often time-consuming and expensive. So, the time and cost of data collection procedures should be weighed against the appropriate amount of data needed to develop accurate data-driven models. Fig. 14 presents an overview of the data set sizes obtained from various desalination technologies.[85] In Figure 14a, in the SS desalination system, the size of the dataset is less than 500 regardless of the type of data-driven approach based on. In Fig. 14b, ANN model is the most widely used ML method. Dataset segmentation is also an important task for ML, which divides available data into training, validation, and testing sets to support model development, tuning, and evaluation. The dataset is divided according to these principles: the training set is used for model training and parameter tuning; the validation set is used for model selection and hyperparameter tuning; and the test set is used for the final evaluation of the model's performance. The test set should closely reflect the actual application scenarios to ensure accurate assessment. In SS systems, the training set accounts for 70-80% of the total data volume, whereas the validation and test sets each account for 10-20%. A reasonable dataset division method can improve the generalization ability and performance of the model, which can provide suggestions for data set division in future research.

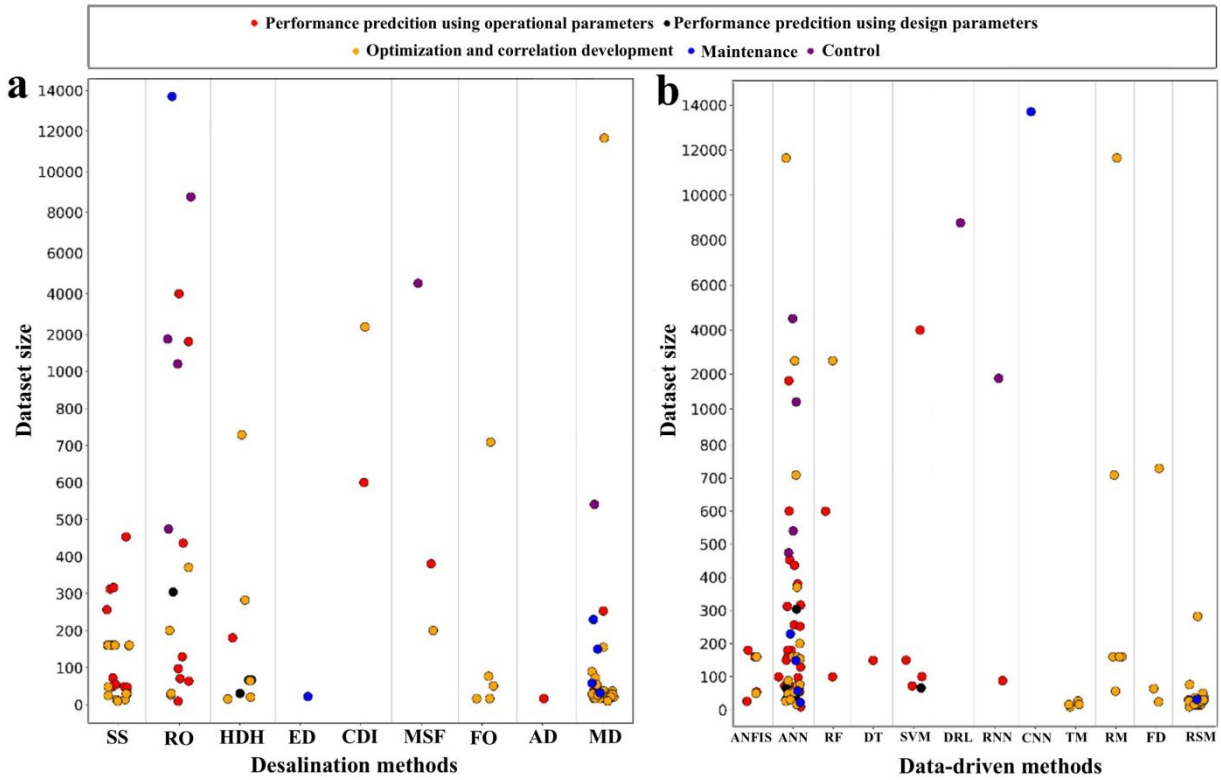


Fig. 14 Size of datasets with respect to (a) desalination systems and (b) data-driven methods for different applications. Reproduced from ref. 85 with permission from Elsevier, copyright 2022.

## 5 Conclusions

This paper reviews the use of solar energy as an abundant renewable energy source to drive the desalination process. We focus on describing the main applications of solar distillers in solar desalination and the use of machine learning in aiding the desalination system of solar distillers, as illustrated in Fig. 15. The main reviews synthesized in this paper are listed below:

- Solar desalination technology does not consume fossil fuels, reduces carbon emissions, and can be operated in remote areas far from the power grid, reducing the dependence on external energy supply and effectively solving the current water and energy shortage problems. Solar desalination systems can also be combined with other renewable energy sources, such as wind power, to form a hybrid energy-driven system, which further improves energy efficiency. Compared with complex mechanical equipment, SS systems are relatively simple in structure and require less maintenance, reducing the cost of repair and replacement during operation.
- Solar desalination technology is divided into direct and indirect processes. The direct process is represented by solar distillers, while the indirect processes are HDH, MED, MSF, VC, and so on. SS offer advantages such as a simple structure, low installation and maintenance costs, among others. Depending on whether additional components are attached or not, they can be categorized into passive SSs and active SSs.
- Machine learning is increasingly recognized as a powerful and emerging approach for modeling and predicting solar desalination, complementing traditional methods based on heat and mass transfer differential equations and regression models. By examining historical data, ML can provide valuable insights into efficiency and system behaviour. Through ML modelling analysis, it is possible to make refined predictions of the cost of SS systems, helping companies to optimise system design and operation, reduce costs and improve the feasibility of commercial applications.
- For SS systems, combining physical models with machine learning models can be considered. This approach can improve prediction accuracy by fusing physical laws (e.g., laws of thermodynamics, water evaporation kinetics, etc.) with data-driven models, especially in data-limited situations. By fusing physics knowledge, ML models are no longer solely dependent on the amount of data, but can be made less data-dependent through the laws of physics, while also ensuring that the model is interpretable.
- Tuning hyperparameters in machine learning and selecting appropriate inputs-outputs play important roles in improving the accuracy of predictive SS systems. For example, grid search or Bayesian optimization methods can be used to tune the hyperparameters of the

model to find the optimal model configuration. Key features such as evaporation efficiency, heat absorption capacity and water temperature gradient can be extracted from the distillation system and environmental conditions.

● The dataset is crucial for modeling machine learning; the inadequacy of the dataset and quality issues are the main current challenges. It is recommended to collect experimental data including distillers' performance parameters, environmental conditions (e.g., temperature, humidity, sunlight intensity), and material properties, as well as to integrate weather station or satellite data covering historical weather data at the operational site.Considering the utilization of larger datasets, this work contributes to the study of solar evaporation and solar stills.

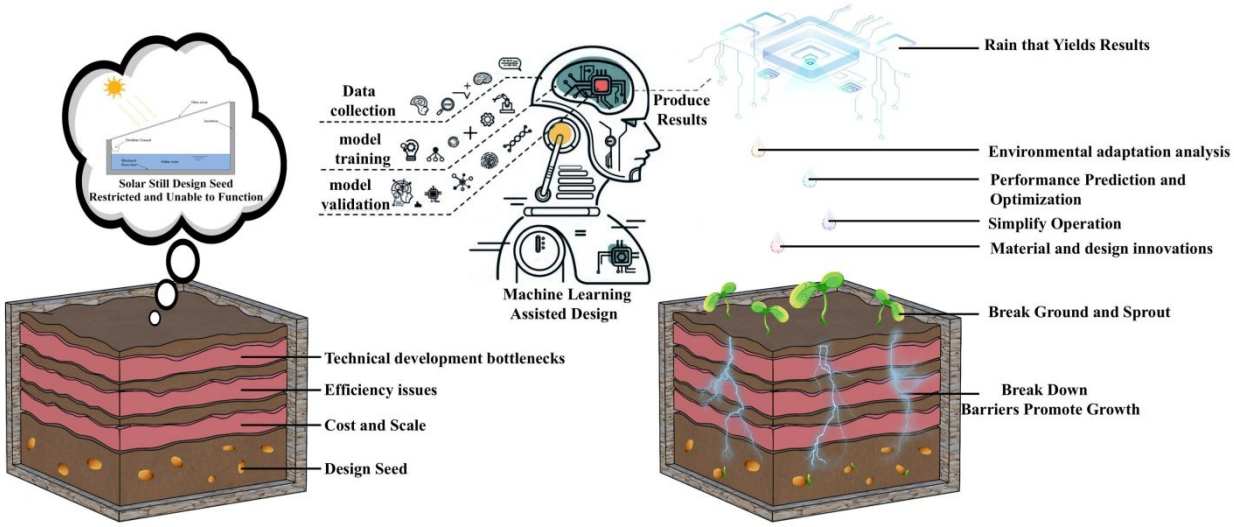


Fig. 15 A look into the future of machine learning- assisted solar stills.

**Acknowledgements**

This work was financially supported by the National Natural Science Foundation of China (No. 22206080) and the Central Plain's Leading Scientific and Technological (244200510007). We also acknowledge the support from the ARC Laureate Fellowship (FL230100095) and the JST-ERATO Yamauchi Materials Space-Tectonics Project (JPMJER2003).

**Conflict of Interest**

The authors declare no conflict of interest.

**References**

1. M. A. Shannon, P. W. Bohn, M. Elimelech, J. G. Georgiadis, B. J. Mariñas and A. M. Mayes, *Nature*, 2008, **452**, 301-310.
2. G. N. Tiwari, H. N. Singh and R. Tripathi, *Sol. Energy*, 2003, **75**, 367-373.
3. F. Macedonio, E. Drioli, A. A. Gusev, A. Bardow, R. Semiat and M. Kurihara, *Chem. Eng. Process*, 2012, **51**, 2-17.
4. M. T. Mito, X. Ma, H. Albuflasa and P. A. Davies, *Renew. Sust. Energ. Rev.*, 2019, **112**, 669-685.
5. P. Droogers, W. W. Immerzeel, W. Terink, J. Hoogeveen, M. F. P. Bierkens, L. P. H. van Beek and B. Debele, *Hydrol. Earth Syst. Sci.*, 2012, **16**, 3101-3114.
6. N. Ghaffour, T. M. Missimer and G. L. Amy, *Desalination*, 2013, **309**, 197-207.
7. R. Borsani and S. Rebagliati, *Desalination*, 2005, **182**, 29-37.
8. C. Sommariva, H. Hogg and K. Callister, *Desalination*, 2003, **158**, 17-21.
9. M. A. Darwish and N. M. Al-Najem, *Desalination*, 1987, **64**, 83-96.
10. A. G. M. Ibrahim and S. E. Elshamarka, *Sol. Energy*, 2015, **118**, 397-409.
11. M. A. Eltawil, Z. Zhengming and L. Yuan, *Renew. Sust. Energ. Rev.*, 2009, **13**, 2245-2262.
12. A. Al-Othman, M. Tawalbeh, M. El Haj Assad, T. Alkayyali and A. Eisa, *Desalination*, 2018, **443**, 237-244.
13. S. Shaaban, *Energ. Convers. Manage.*, 2019, **198**, 111794.
14. J. Y. Choi, T. Lee, A. B. Aleidan, A. Rahardianto, M. Glickfeld, M. E. Kennedy, Y. Chen, P. Haase, C. Chen and Y. Cohen, *J. Environ. Manage.*, 2019, **250**, 109487.
15. J. Kim, J. Kim, J. Lim, S. Lee, C. Lee and S. Hong, *Chem. Eng. J.*, 2019, **374**, 49-58.
16. A. Farsi and I. Dincer, *Desalination*, 2019, **463**, 55-68.
17. S. Sadri, M. Ameri and R. Haghighi Khoshkhoo, *Desalination*, 2017, **402**, 97-108.
18. D. G. Harris Samuel, P. K. Nagarajan, R. Sathyamurthy, S. A. El-Agouz and E. Kannan, *Energ. Convers. Manage.*, 2016, **112**, 125-134.
19. S. Shoeibi, S. A. A. Mirjalily, H. Kargarsharifabad, M. Khiadani and H. Panchal, *Desalination*, 2022, **540**, 115983.
20. S. Shoeibi, H. Kargarsharifabad, N. Rahbar, G. Ahmadi and M. R. Safaei, *Sustain. Energy Techn.*, 2022, **49**, 101728.
21. S. Shoeibi, H. Kargarsharifabad and N. Rahbar, *J. Energy Storage*, 2021, **42**, 103061.
22. L. D. Jathar and S. Ganesan, *Groundwater for Sustainable Development*, 2021, **12**, 100539.
23. L. D. Jathar, S. Ganesan, U. Awasarmol, K. Nikam, K. Shahapurkar, M. E. M. Soudagar, H. Fayaz, A. S. El-Shafay, M. A. Kalam, S. Bouadila, S. Baddadi, V. Tirth, A. S. Nizami, S. S. Lam and M. Rehan, *Environ. Pollut.*, 2023, **326**, 121474.
24. S. K. Singh, C. Sharma and A. Maiti, *J. Environ. Chem. Eng.*, 2021, **9**, 105473.
25. S. Mittal, A. Gupta, S. Srivastava and M. Jain, *Chem. Eng. Process.*, 2021, **164**, 108403.

26. L. D. Jathar, K. Nikam, U. V. Awasarmol, R. Gurav, J. D. Patil, K. Shahapurkar, M. E. M. Soudagar, T. M. Y. Khan, M. A. Kalam, A. Hnydiuk-Stefan, A. E. Gürel, A. T. Hoang and Ü. Ağbulut, *Heliyon*, 2024, **10**, e25407.
27. Q. He, H. Zheng, X. Ma, L. Wang, H. Kong and Z. Zhu, *Energy and AI*, 2022, **7**, 100123.
28. S. W. Sharshir, A. Elhelow, A. Kabeel, A. E. Hassanien, A. E. Kabeel and M. Elhosseini, *Environ. Sci. Pollut. R.*, 2022, **29**, 90632-90655.
29. Z. Xu, J. Shu and D. Meng, *Natl. Sci. Rev.*, 2024, **11**.
30. A. H. Elsheikh, S. W. Sharshir, M. Abd Elaziz, A. E. Kabeel, W. Guilan and Z. Haiou, *Sol. Energy*, 2019, **180**, 622-639.
31. S. Rashidi, N. Karimi and W.-M. Yan, *Eng. Anal. Bound. Elem.*, 2022, **144**, 399-408.
32. M. Shatat, M. Worall and S. Riffat, *Sustain. Cities. Soc.*, 2013, **9**, 67-80.
33. H. Müller-Holst, M. Engelhardt and W. Schölkopf, *Desalination*, 1999, **122**, 255-262.
34. A. Giwa, N. Akther, A. A. Housani, S. Haris and S. W. Hasan, *Renew. Sust. Energ. Rev.*, 2016, **57**, 929-944.
35. H. Sharon and K. S. Reddy, *Renew. Sust. Energ. Rev.*, 2015, **41**, 1080-1118.
36. L. F. Greenlee, D. F. Lawler, B. D. Freeman, B. Marrot and P. Moulin, *Water Res.*, 2009, **43**, 2317-2348.
37. M. A. Darwish and H. K. Abdulrahim, *Desalination*, 2008, **228**, 30-54.
38. S. R. Osipi, A. R. Secchi and C. P. Borges, *Desalination*, 2018, **430**, 107-119.
39. M. R. Qtaishat and F. Banat, *Desalination*, 2013, **308**, 186-197.
40. A. Alkhudhiri, N. Darwish and N. Hilal, *Desalination*, 2012, **287**, 2-18.
41. S. Zhao, C. Jiang, J. Fan, S. Hong, P. Mei, R. Yao, Y. Liu, S. Zhang, H. Li, H. Zhang, C. Sun, Z. Guo, P. Shao, Y. Zhu, J. Zhang, L. Guo, Y. Ma, J. Zhang, X. Feng, F. Wang, H. Wu and B. Wang, *Nat. Mater.*, 2021, **20**, 1551-1558.
42. S. M. A. Moustafa, D. I. Jarrar and H. I. El-Mansy, *Sol. Energy*, 1985, **35**, 333-340.
43. S. Burn, M. Hoang, D. Zarzo, F. Olewniak, E. Campos, B. Bolto and O. Barron, *Desalination*, 2015, **364**, 2-16.
44. A. M. Helal and S. A. Al-Malek, *Desalination*, 2006, **197**, 273-300.
45. Y. Zhang, M. Sivakumar, S. Yang, K. Enever and M. Ramezanianpour, *Desalination*, 2018, **428**, 116-145.
46. M. A. Darwish, F. Al-Juwayhel and H. K. Abdulraheim, *Desalination*, 2006, **194**, 22-39.
47. A. Al-Karaghouli, D. Renne and L. L. Kazmerski, *Renew. Sust. Energ. Rev.*, 2009, **13**, 2397-2407.
48. B. Zala, K. Dodia, H. N. Panchal*, I. J. o. R. R. i. E. S.*, 2013, **2**, 6-11.
49. Y. Su, J. Hu, G. Yuan, G. Zhang, W. Wei, Y. Sun, X. Zhang, Z. Liu, N.-T. Suen, H.-C. Chen and H. Pang, *Adv. Mater.*, 2023, **35**, 2307003.
50. D. Dsilva Winfred Rufuss, S. Iniyan, L. Suganthi and P. A. Davies, *Renew. Sust. Energ. Rev.*, 2016, **63**, 464-496.
51. P. Prakash and V. Velmurugan, *Renew. Sust. Energ. Rev.*, 2015, **49**, 585-609.
52. R. S. Hansen, C. S. Narayanan and K. K. Murugavel, *Desalination*, 2015, **358**, 1-8.
53. D. Mevada, H. Panchal, M. Ahmadein, M. E. Zayed, N. A. Alsaleh, J. Djuansjah, E. B. Moustafa, A. H. Elsheikh and K. K. Sadasivuni, *Case Stud. Therm. Eng.*, 2022, **29**, 101687.
54. C. Chen, Y. Li, J. Song, Z. Yang, Y. Kuang, E. Hitz, C. Jia, A. Gong, F. Jiang, J. Y. Zhu, B. Yang, J. Xie and L. Hu, *Adv. Mater.*, 2017, **29**, 1701756.
55. Z. M. Omara, A. E. Kabeel and M. M. Younes, *Desalination*, 2013, **314**, 67-72.
56. H. Panchal, R. Sathyamurthy, A. E. Kabeel, S. A. El-Agouz, D. Rufus, T. Arunkumar, A. Muthu Manokar, D. P. Winston, A. Sharma, N. Thakar and K. K. Sadasivuni, *J. Therm. Anal. Calorim.*, 2019, **138**, 3175-3182.

57. T. Rajaseenivasan, K. Kalidasa Murugavel and T. Elango, *Desalin. Water Treat.*, 2015, **55**, 1786-1794.
58. A. A. El-Sebaii, M. R. I. Ramadan, S. Aboul-Enein and M. El-Naggar, *Desalination*, 2015, **365**, 15-24.
59. D. Mevada, H. Panchal, K. k. Sadasivuni, M. Israr, M. Suresh, S. Dharaskar and H. Thakkar, *Groundwater for Sustainable Development*, 2020, **10**, 100289.
60. A. E. Kabeel, *Energy*, 2009, **34**, 1504-1509.
61. A. K. Tiwari and G. N. Tiwari, *Int. J. Energ. Res.*, 2007, **31**, 1358-1382.
62. M. A. Samee, U. K. Mirza, T. Majeed and N. Ahmad, *Renew. Sust. Energ. Rev.*, 2007, **11**, 543-549.
63. A. A. Badran, I. A. Al-Hallaq, I. A. E. Salman and M. Z. J. D. Odat, *Desalination*, 2005, **172**, 227-234.
64. S. Kumar, A. Dubey and G. N. Tiwari, *Desalination*, 2014, **347**, 15-24.
65. S. Gorjian, B. Ghobadian, T. Tavakkoli Hashjin and A. Banakar, *Desalination*, 2014, **352**, 1-17.
66. R. A. Kumar, G. Esakkimuthu and K. K. Murugavel, *Desalination*, 2016, **399**, 198-202.
67. A. E. Kabeel, Z. M. Omara and F. A. Essa, *Energ. Convers. Manage.*, 2014, **78**, 493-498.
68. S. J. S. Ahmed and w. technology, *Solar & Wind Technology*, 1988, **5**, 637-643.
69. M. A. Eltawil and Z. Zhengming, *Desalination*, 2009, **249**, 490-497.
70. M. Mostafa, H. M. Abdullah and M. A. Mohamed, *IEEE Access*, 2020, **8**, 219457-219472.
71. G. Singh, S. Kumar and G. N. Tiwari, *Desalination*, 2011, **277**, 399-406.
72. Y. Taamneh and M. M. Taamneh, *Desalination*, 2012, **291**, 65-68.
73. H. Panchal, K. K. Sadasivuni, F. A. Essa, S. Shanmugan and R. Sathyamurthy, *Heat Transf.*, 2021, **50**, 1392-1409.
74. S. W. Sharshir, N. Yang, G. Peng and A. E. Kabeel, *Appl. Therm. Eng.*, 2016, **100**, 267-284.
75. H. B. Bacha, T. Damak, M. Bouzguenda, A. Y. Maalej and H. B. Dhia, *Desalination*, 2003, **156**, 305-313.
76. Y. Gong, X.-l. Wang and Y. Li-xin, *Desalination*, 2005, **172**, 157-172.
77. X. Wang and K. C. Ng, *Appl. Therm. Eng.*, 2005, **25**, 2780-2789.
78. G. Yuan, L. Zhang and H. Zhang, *Desalination*, 2005, **182**, 511-516.
79. H. Tao, T. Wu, M. Aldeghi, T. C. Wu, A. Aspuru-Guzik and E. Kumacheva, *Nat. Rev. Mater.*, 2021, **6**, 701-716.
80. R. Batra, L. Song and R. Ramprasad, *Nat. Rev. Mater.*, 2021, **6**, 655-678.
81. M. M. Bhatti, R. Ellahi and M. Hossein Doranehgard, *J. Mol. Liq.*, 2022, **361**, 119655.
82. M. M. Bhatti, H. F. Öztop, R. Ellahi, I. E. Sarris and M. H. Doranehgard, *J. Mol. Liq.*, 2022, **357**, 119134.
83. K. Milani Shirvan, M. Mamourian, S. Mirzakhanlari, A. B. Rahimi and R. Ellahi, *Int. J. Numer. Method. H.*, 2017, **27**, 2385-2399.
84. A. Samuel and K.-C. Chang, *Water-Sui*, 2022, **14**.
85. P. Behnam, M. Faegh and M. Khiadani, *Desalination*, 2022, **532**, 115744.
86. O. Rejeb, M. S. Yousef, C. Ghenai, H. Hassan and M. Bettayeb, *Case Stud. Therm. Eng.*, 2021, **24**, 100816.
87. A. E. Khalifa and D. U. Lawal, *Desalin. Water Treat.*, 2016, **57**, 28513-28530.
88. M. Al-Abed Allah, M. Abu Abbas and M. Maqableh, *Desalin. Water Treat.*, 2022, **270**, 1-11.
89. K. Elmaadawy, M. A. Elaziz, A. H. Elsheikh, A. Moawad, B. Liu and S. Lu, *J. Environ. Manage.*, 2021, **298**, 113520.
90. A. H. Elsheikh, M. Abd Elaziz and A. Vendan, *Weld. World*, 2022, **66**, 27-44.

91. M. Caner, E. Gedik and A. Keçebaş, *Expert Syst. Appl.*, 2011, **38**, 1668-1674.
92. Y. Kashyap, A. Bansal and A. K. Sao, *Renew. Sust. Energ. Rev.*, 2015, **49**, 825-835.
93. D. B. Jani, M. Mishra and P. K. Sahoo, *Renew. Sust. Energ. Rev.*, 2017, **80**, 352-366.
94. N. I. Santos, A. M. Said, D. E. James and N. H. Venkatesh, *Renew. Energ.*, 2012, **40**, 71-79.
95. E. Abdelhafez and R. A. Haj Khalil, *JOCET*, 2013, **1**, 238-242.
96. M. S. S. Abujazar, S. Fatihah, I. A. Ibrahim, A. E. Kabeel and S. Sharil, *J. Clean. Prod.*, 2018, **170**, 147-159.
97. J. Huang, T. Jin, M. Liang and H. Chen, *Appl. Therm. Eng.*, 2021, **182**, 116053.
98. J. S. R. Jang, *IEEE T. Syst. Man. Cy.*, 1993, **23**, 665-685.
99. A. A. Aldair, A. A. Obed and A. F. Halihal, *Renew. Sust. Energ. Rev.* 2018, **82**, 2202-2217.
100. A. D. Saee, A. Baghban, F. Zarei, Z. Zhang and S. Habibzadeh, *Int. J. Refrig.*, 2018, **96**, 38-49.
101. A. H. Elsheikh, S. Shanmugan, R. Sathyamurthy, A. Kumar Thakur, M. Issa, H. Panchal, T. Muthuramalingam, R. Kumar and M. Sharifpur, *Sustain. Energy Techn.*, 2022, **49**, 101783.
102. V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan and B. P. Feuston, *J. Chem. Inf. Model.*, 2003, **43**, 1947-1958.
103. J. C.-W. Chan and D. Paelinckx, *Remote Sens. Environ.*, 2008, **112**, 2999-3011.
104. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay and G. Louppe, *J. Mach. Learn. Res.*, 2012, **12**.
105. Y. Wang, A. W. Kandeal, A. Swidan, S. W. Sharshir, G. B. Abdelaziz, M. A. Halim, A. E. Kabeel and N. Yang, *Appl. Therm. Eng.*, 2021, **184**, 116233.
106. A. W. Kandeal, M. An, X. Chen, A. M. Algazzar, A. Kumar Thakur, X. Guan, J. Wang, M. R. Elkadeem, W. Ma and S. W. Sharshir, *Energy Technol-Ger*, 2021, **9**, 2100189.
107. C. L. P. Chen and J. Z. Wan, *IEEE T. Syst. Man. Cy. B.*, 1999, **29**, 62-72.
108. S. W. Sharshir, M. Abd Elaziz and M. R. Elkadeem, *Sol. Energy*, 2020, **198**, 399-409.
109. W. H. Alawee, A. A. Jaber, Z. M. Omara, S. A. Mohammed, H. A. Dhahad, Z. H. Khan and L. A. Al-Haddad, *Desalin. Water Treat.*, 2024, **320**, 100683.
110. R. K. Kottala, K. R. Balasubramanian, B. S. Jinshah, S. Divakar and B. K. Chigilipalli, *J. Therm. Anal. Calori.*, 2023, **148**, 7101-7124.
111. M. Alhuyi Nazari, M. Salem, I. Mahariq, K. Younes and B. B. Maqableh, *Front. Energy Res.*, 2021, **9**.
112. A. Bagheri, N. Esfandiari, B. Honarvar and A. Azdarpour, *Math. Comp. Model. Dyn.*, 2020, **26**, 453-480.
113. M. Bahiraei, S. Nazari, H. Moayedi and H. Safarzadeh, *Powder Technol.*, 2020, **366**, 571-586.
114. M. Bahiraei, S. Nazari and H. Safarzadeh, *Powder Technol.*, 2021, **385**, 185-198.
115. A. Bamasag, F. A. Essa, Z. M. Omara, E. Bahgat, A. O. Alsaiari, H. Abulkhair, R. A. Alsulami and A. H. Elsheikh, *Process Safe. Environ.*, 2022, **162**, 112-123.
116. Y. Wang, G. Peng, S. W. Sharshir, A. W. K, e. eal and N. Yang, *ES Materials & Manufacturing*, 2021, **14**, 87-94.
117. A. Sohani, S. Hoseinzadeh, S. Samiezadeh and I. Verhaert, *J. Therm. Anal. Calori.*, 2022, **147**, 3919-3930.
118. W. Gao, L. Shen, S. Sun, G. Peng, Z. Shen, Y. Wang, A. W. Kandeal, Z. Luo, A. E. Kabeel, J. Zhang, H. Bao and N. Yang, *Chinese Phys. B*, 2023, **32**, 048801.
119. A. H. Elsheikh, V. P. Katekar, O. L. Muskens, S. S. Deshmukh, M. A. Elaziz and S. M. Dabour, *Process Safe. Environ.*, 2021, **148**, 273-282.

120. R. Chauhan, S. Sharma, R. Pachauri, P. Dumka and D. R. Mishra, *J. Energy Storage*, 2020, **30**, 101408.
121. A. F. Mashaly and A. A. Alazba, *J. Water Reuse Desal.*, 2015, **5**, 480-493.
122. R. Chauhan, P. Dumka and D. R. Mishra, *Int. J. Ambient Energy*, 2022, **43**, 1389-1396.
123. S. Nazari, M. Bahiraei, H. Moayedi and H. Safarzadeh, *J. Clean. Prod.*, 2020, **277**, 123232.
124. F. A. Essa, M. Abd Elaziz and A. H. Elsheikh, *Appl. Therm. Eng.*, 2020, **170**, 115020.
125. E. Atashpaz-Gargari and C. J. I. C. o. E. C. Lucas, *IEEE CEC*, 2007, 4661-4667.
126. D. T. Pham and D. Karaboga, *P. I. Mech. Eng.*, 1997, **211**, 157-167.
127. B. Choubin and A. Malekian, *Environ. Earth Sci.*, 2017, **76**, 538.
128. R. Genuer, J.-M. Poggi and C. Tuleau-Malot, *Pattern Recogn. Lett.*, 2010, **31**, 2225-2236.

## Data availability statements

Data and figures for this manuscript are available within the main.