



Machine learning assisted improved desalination pilot system design and experimentation for the circular economy

Waqar Muhammad Ashraf^{a,1}, Muhammad Ahmad Jamil^{b,*}, Ghulam Moeen Uddin^c,
Bashar Shboul^d, Kashif Ishfaq^e, Kim Choon Ng^f, Mike Dixon^g, Ben Bin Xu^b,
Muhammad Wakil Shahzad^{b,*}

^a The Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK

^b Mechanical & Construction Engineering Department, Northumbria University, Newcastle Upon Tyne NE18ST, UK

^c Department of Mechanical Engineering, University of Engineering & Technology, Lahore, Punjab 54890, Pakistan

^d Department of Renewable Energy Engineering, Faculty of Engineering, Al Al-Bayt University, Mafraq, Jordan

^e Department of Industrial and Manufacturing Engineering, University of Engineering and Technology Lahore, 548900, Pakistan

^f Water Desalination and Reuse Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

^g Synauta Inc, Canada

ARTICLE INFO

Editor: Guangming Jiang

Keywords:

Water circular economy
Distillate production
Machine learning
Sustainable development goal
Desalination

ABSTRACT

Desalination is among the most feasible solutions to supply sustainable and clean drinking water in water scarcity areas. In this regard, Multi-Effect Desalination (MED) systems are particularly preferred for harsh feeds (high temperature and salinity) because of their robust mode of operation for water production. However, maintaining the efficient operation of the MED systems is challenging because of the large system design and variables' interdependencies that are sensitive to the distillate production. Therefore, this research leverages the power of machine learning and optimization to estimate the optimal operating conditions for the maximum distillate production from the MED system. In the first step, detailed experimentation is conducted for distillate production against hot water temperature (HWT) varying from 38 to 70 °C, and feed water temperature (FWT) is changed from 34 to 42 °C. Whereas, the feed flow rate (FFR) is investigated to be varied nearly from 3.6 to 8.7 LPM in the three stages, i.e., FFR-S1, FFR-S2 and FFR-S3. The compiled dataset is used to make the process models of the MED system by three ML-based algorithms, i.e., Artificial Neural Network (ANN), Support Vector Machine (SVM), and Gaussian Process Regression (GPR) under rigorous hyperparameters optimization. GPR exhibited superior predictive performance than those of ANN and SVM on R^2 value of 0.99 and RMSE of 0.026 LPM. Monte Carlo technique-based variable significance analysis revealed that the HWT has the highest effect on distillate production with a percentage significance of 95.6 %. Then Genetic Algorithm is used to maximize the distillate production with the GPR model embedded in the optimization problem. The GPR-GA driven maximum distillate production is estimated on HWT = 70 ± 0.5 °C, FWT = 40 ± 2.5 °C, FFR-S1 = 6 ± 2.6 LPM, FFR-S2 = 7 ± 1 LPM and FFR-S3 = 7 ± 1. The ML-GA-based system analysis and optimization of the MED system can boost the distillate production that promotes operation excellence and circular economy from the desalination sector.

1. Introduction

Clean drinking water is among the formidable challenges because over 2.2 billion people do not have its availability [1]. The situation is exacerbated with surging demand which is expected to surpass 6000 billion cubic meters by 2030 from 4200 billion cubic meters in 2015 driven by rising population, industrial development, and life quality

standards [2]. To deal with the water scarcity problem, desalination has emerged as the only practical and feasible solution [3]. They also contribute towards the water circular economy. Therefore, different desalination systems are working worldwide, majorly including Multi-stage Flash (MSF), Reverse Osmosis (RO), Multi-effect Desalination (MED), and others (electrodialysis, nanofiltration, ultrafiltration, membrane distillation, etc.) [4]. Meanwhile, a surge in the installed

* Corresponding authors.

E-mail addresses: muhammad2.ahmad@northumbria.ac.uk (M.A. Jamil), muhammad.w.shahzad@northumbria.ac.uk (M.W. Shahzad).

¹ Authors contributed equally.

capacity is observed and is predicted to grow further as shown in Fig. 1 [5,6]. It is also worth mentioning that thermal-based systems have grown significantly over the last decades due to low pretreatment needs, robust design, low maintenance, and versatile feed handling ability [7]. Among thermal-based systems, multi-effect desalination (MED) has shown promising characteristics including better heat transfer, high working efficiency, capability to harness low-grade energy, and hybridization flexibility with other systems [8] resulting in a wide range of installations worldwide as a standalone as well as hybrid systems [9]. However, these systems have high energy consumption, operating costs, and CO₂ emissions compared to membrane-based systems [10]. These perform only within 10 to 13 % of their thermodynamic limit and hence are the subject of extensive research for improved performance [11].

Besides the above-mentioned established technologies, significant research interest has been observed in biomass-based adsorbent materials for water treatment [12,13]. These can be developed from agricultural residues [14], natural fibres [15], wood cellulosic [16], composite material [17] and activated carbons [18]. Hossain et al. [19] proposed a cost-effective adsorbent material for effective adsorption and recovery of rare earth metals Yb (III). In another study [20] they developed an eco-friendly composite adsorbent via an eco-friendly process for successful detection and removing Cd (II). Similarly, they also proposed ligand-immobilized nano-composite adsorbent for efficient cerium(III) detection and recovery [21]. They showed that the proposed adsorbent was also highly selective to Ce (III) with fast kinetic performances Su et al. [22] proposed a microbubble floating-extraction mechanism which has the separation efficiency of molybdenum over 99 %. Awual [23] developed novel nanocomposite materials that were prepared for optical Hg (II) detection/removal and showed that the proposed materials are promising candidates for in situ environmental remediation. Similarly, they developed ligand-based facial conjugate materials for treating copper (II) [24]. They showed that the conjugate materials exhibited high sensitivity and selectivity to Cu(II) ions.

Desalination plays a critical role in the water circular economy to minimize water waste and promote the responsible use and reuse of water resources. Desalination technologies can contribute to this by addressing water scarcity and enhancing water resource management in several ways. The earlier research on MED desalination systems includes experimental and theoretical studies [25]. The experimental studies included the development of the system, hybridization with other systems, and performance investigation [26]. The theoretical developments include numerical modelling, optimization, parametric analysis, and retrofitting of different components for performance improvement [27,28]. These conventional mathematical models require extensive knowledge, complex equation development, and calculation of associated properties [29]. Particularly, under unpredictable demands, and fluctuating energy resources like renewable energy, cogeneration plants,

and hybrid systems, the designing is difficult [30]. Also, the conventional models use assumptions that compromise the accuracy of the predicted characteristics [31]. Furthermore, the development and solution of the mathematical model for performance prediction and optimization is computationally expensive. Owing to the limitations of the conventional methods and the superiority of machine learning models to approximate the complex function space with high accuracy, reduced computational resource requirement, and flexibility to update the trained model, a growing trend of employing machine learning for water desalination systems is observed and plenty of research is conducted as summarized below [32].

1.1. Desalination and machine learning

The application of machine learning (ML) on desalination systems is obviously to benefit from technological advancements and accurate data-driven methods [33]. ML tools can be applied for the system modelling and analysis of the MED system for informed decision-making and operational strategy development for the desalination systems [34]. Plenty of work has been conducted in this regard deploying the ML tools to model different performance indicators of the desalination systems. For instance, Son et al. [35] proposed the utilization of a convolutional neural network in conjunction with a long short-term memory ML structure for pH prediction in water treatment systems. They showed that the proposed framework demonstrated better accuracy with $R^2 \geq 0.998$ than the conventional numerical models with $R^2 < 0.309$. Bonny et al. [36] developed a deep reinforcement learning-based method to optimize the pressure across a trans-membrane reverse osmosis system. They reported improvement in permeate and salt rejection with 99 %. Shahane et al. [37] optimized the evaporator design for higher water production, better heat transfer, and lower scale formation using an artificial intelligence-based non-dominated sorting genetic algorithm. They showed that the proposed model achieved optimal results with an average relative error of <3.24 % and can be used for further improvement. Krzywanski et al. [38] identified the optimal operating conditions to achieve water uptake of 1.65 g/g for a fluidized bed adsorption cooling and desalination system using artificial intelligence. Similarly, He et al. [39] reported a 10 % improvement in freshwater productivity by using AI-based optimization of renewable energy-driven desalination systems.

Karambati et al. [40] optimized a hybrid Stirling engine integrated with a MED system using an analytical hierarchy process that increased energy and water productivity thus reducing the operating costs of the system. They showed that the optimal system can deliver 2.58 kW of power, 19.92 m³/day of water at a rate of 0.29 \$/kW-hr, and 1.6 \$/m³, respectively. Similarly, Pombo et al. [41] designed a small, modular nuclear fission chamber to be used as a hybrid plant; solar energy and wind energy were used to achieve an optimum hybrid plant that improved both safety and energy efficiency. Salem et al. [42] optimized a hybrid solar still and humidification-dehumidification system using a multilayer perceptron. They showed the superiority of the proposed model to predict and enhance system performance in less time and resources. Shakibi et al. [43] developed and optimized a cogeneration plant with a gas turbine cycle combined with a field of heliostats and thermal vapor compression MED system using a multi-objective grey wolf, support vector regression, and grasshopper optimization algorithms. The study reported an exergy efficiency of up to 45.6 %.

The above-mentioned literature review suggests that the use of ML tools in the water sector improved system performance, process control, and effective resource planning in fluctuating situations. However, ML-based process design and performance improvement of the MED systems for maximum distillate production are rarely conducted in the literature potentially because of the fabrication challenges of the MED system and the unavailability of the design and process data. More importantly, a comprehensive analytical framework consisting of a step-by-step approach for carrying out the ML-assisted analysis and optimization

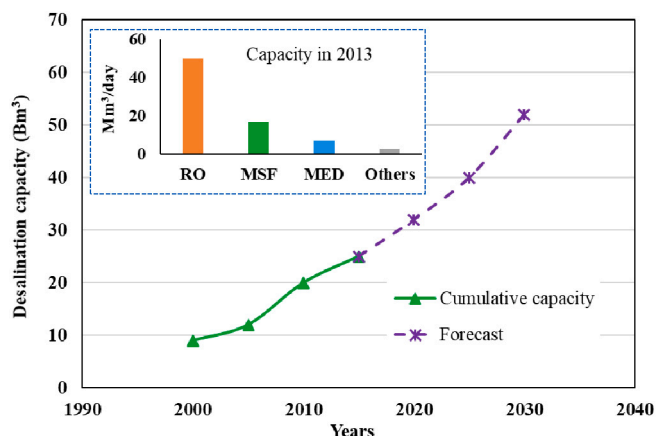


Fig. 1. Global desalination trend and prediction [5,6].

for enhancing MED system performance is missing in the literature which is a major obstacle to the inclusion of ML in the desalination sector. The verification of the ML-based estimated solution may contribute to the knowledge pool of applied ML in the desalination sector demonstrating its effectiveness and usefulness for the performance enhancement of desalination systems.

Thus, the current study provides an ML-based comprehensive analytical framework deploying the advanced modelling algorithm and optimization technique for maximum water production from the MED system. For this task, an extensive experimental investigation of an in-house MED system is conducted first on different operational conditions and the dataset is compiled. ML models like artificial neural network (ANN), support vector machines (SVM), and Gaussian process regression (GPR) based process models for distillate production are constructed under rigorous hyperparameter optimization to train a well-predictive model that can approximate the operating behavior of the designed MED system. A sensitivity analysis employing the Monte Carlo method is conducted to get an insight into the variable's significance on the distillate production that is missing in the literature. The model-based optimization analysis by genetic algorithm is carried out to estimate the optimized operating conditions for the maximum water generation and the results are verified on the experimental set-up thereby demonstrating the major novel aspect of this work. An extensive closed-loop data-driven modelling and optimization assessment for the performance enhancement of the designed MED system and distillate production are carried out by the ML approach which is the major novelty of this work and may help the industrial practitioners and engineers to deploy the proposed framework for the operational excellence of their industrial systems to contribute to water scarcity and sustainability challenges.

2. Materials and methods

The methodology adopted to conduct the current research work is graphically represented in Fig. 2. The first step involves detailed experimentation and data collection at assorted operating conditions of the MED system. Then, the data is visualized and processed to eliminate outliers in the dataset compiled from the experiments performed on the MED system. In the next stage, ML models including ANN, SVM and GPR are developed under extensive hyperparameter tuning. The best predictive model is then used to conduct the significance analysis of the variables and maximize the distillate production under the optimized operating conditions by genetic algorithm. Finally, the investigation of the model-based optimized operating conditions is verified on the experimental system. All these stages are described in detail below.

2.1. Experimentation and data collection

Extensive experimentation is conducted using a Multi-Effect Desalination (MED) test rig as indicated in Fig. 3. The rig contains a steam generator, evaporator stages, a condenser, a cooling tower, a brine cooler, collection tanks, pumps, and a control system. The system operates as the hot water is directed to the steam generator in a closed loop. The steam produced by the generator is then supplied to the first evaporator. The saltwater feed is applied on the outer surface of the evaporators using a pump and nozzles. The feed flow rate is controlled using throttling valves according to the demand. In the first evaporator,

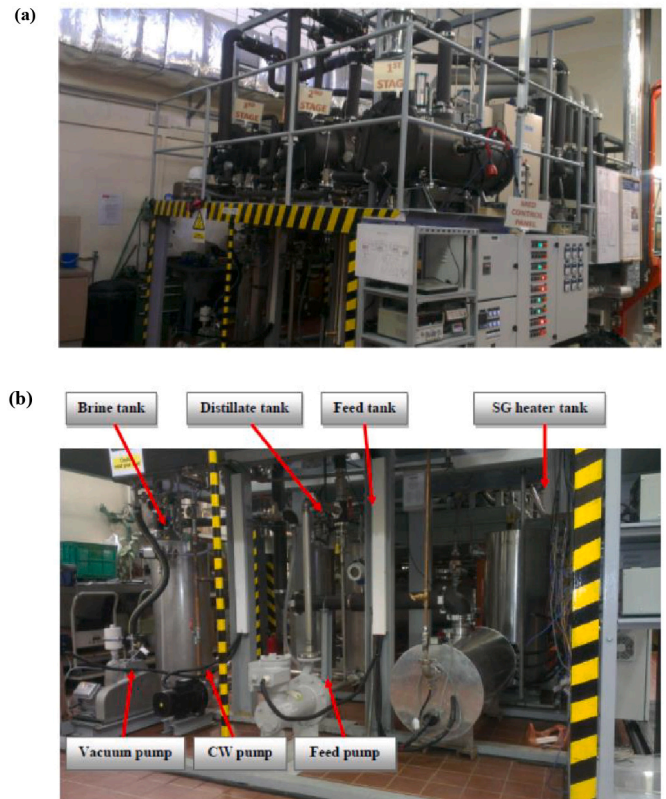


Fig. 3. MED experimental test rig (a) evaporators and (b) pumps and tanks [44].

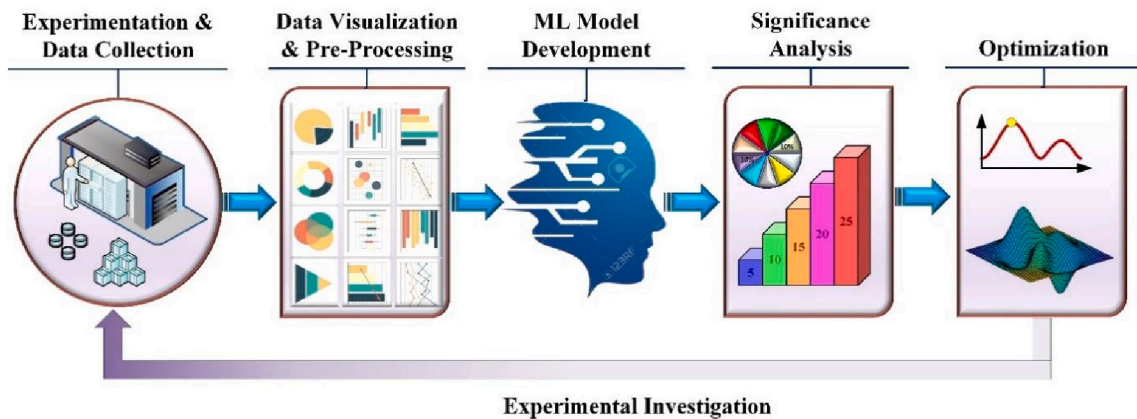


Fig. 2. A stepwise methodology is adopted in this study to maximize the distillate production from the MED system. The experiments are conducted on the experimental set-up, the collected dataset is compiled and visualized and then used to make process models by ML. The significance order of the input variables of the MED system is established by the Monte Carlo technique. The ML model is embedded in the optimization problem to maximize the distillate production and the problem is solved by genetic algorithm. The estimated results for the maximum distillate production are also verified on the experimental setup.

the steam inside the tube transfers heat to the feed water and evaporates a portion of it. The vapours produced in the first evaporator are then directed to the next evaporator tubes and the process goes on. Meanwhile, the steam inside the tubes is condensed and collected as a distillate (freshwater). The brine enters the subsequent effect which produces some additional vapor due to flashing caused by pressure drop between the evaporators. The vapours produced in the final evaporator are sent to the condenser to produce distillate by using a cooling tower. The distillate and brine from all the evaporator stages are collected in the distillate collection tank and brine collection tank respectively using headers. The brine is partially recirculated to mix with feed which improves the system's recovery ratio. Turbine flow meters are utilized to calculate the flow rate inside the system. Meanwhile, the vacuum inside the system is maintained using a vacuum pump to ensure continuous evaporation.

The test rig is completely instrumented to observe and log real-time data. Meanwhile, it is also important to mention that the instruments are vacuum-grade because the system operates under vacuum conditions. Table 1 summarizes the instrumentation details of the system, and an Agilent data logger is used to record the data.

2.2. Data-visualization & pre-processing

The MED system was operated under steady-state conditions for diverse operating scenarios to ensure the stable operation of distillate production in each stage [45]. It was observed that the values of various thermo-physical operating parameters in the second and third stages are effectively influenced by that of the first stage's operating parameters. Therefore, the important system control parameters are considered from the 1st stage for ML-assisted analysis of the aforementioned system due to two reasons. (1) The 1st stage operating parameters serve as the driver of MED (2) ML models work well on the causal and independent nature of the input variables [46].

The operating parameters taken from the 1st stage of MED are hot water temperature (HWT), and feed water temperature (FWT) since these are independently controlled variables. Moreover, these are the most critical parameters in MED system operation. For instance, a higher hot water temperature means higher input energy and higher productivity. Similarly, a higher feed water temperature reduces the preheating requirements and is of significance when waste heat is used. Similarly, the feed flow rate (FFR) corresponding to the three stages is also included which represents the feed of seawater flow to the stages of MED and are represented as FFR-S1, FFR-S2, and FFR-S3 respectively. The feed flow also impacts productivity; however, it needs to be controlled carefully. This is because a very low feed flow rate results in dry patches in the evaporator and reduces system performance. On the other hand, a very high flow rate results in lower evaporation because of higher

Table 1
Instrumentation details for MED system [44].

Instrument	Description	Range	Accuracy
GE: Pressure transducer	2-wire, 6.3 NPT male	Pressure: 0–60 kPa	0.5 % FS
OMEGA: Thermistor probes	Length: 152 mm long, Dia: 3 mm	Temperature: –80 °C to 150 °C	±0.1 °C
OMEGA: Feed flow meter (turbine type)	12.7 mm NPT FM,	Flow: 1.1–11 l/min Temperature: –40 to 85 °C	±1.0 %
OMEGA: Hot and cooling water flow rate (turbine type)	25.4 mm NPT	Flow: 3.8–60 LPM, Temperature: –7 to 107 °C	±1.0 %
Aichi Tokei Denki: Distillate flow meter (turbine type)	ND05-PATAAC	Flow: 0.0–3 LPM, Temperature: 0 to 70 °C	±1.0 %
Sure Instruments: Vapor flow meter (turbine type)	Magnetic pick-up type	Flow: 15–200 m ³ /h, Temperature: –30 to 100 °C	±1.0 %

sensible water heating. Finally, the output of the MED system is the distillate water production (now called distillate production) that is to be modelled on the selected input variables.

The data thus collected is required to be visualized to investigate the data-distribution space. Violin plot is an efficient approach to visualize the data distribution and the data-distribution curve is also plotted thereby showing the density of the data distribution. The visualization of data distribution is an important step before building the ML models since visualizing the data distribution space on the operating ranges of the variables helps ensure data quality. Once the data visualization step is carried out, the collected data is normalized for the development of ML models. Data normalization is essentially important because it ensures the efficient construction of ML models for the given input variables which may have significantly large operating ranges. Thus, scaling all the variables into equal ranges, depending upon the normalization technique, provides fair chances for developing their association with the output variable which could otherwise be biased towards some particular input variable(s). Among the various data-normalization techniques reported in the literature [47] min-max scaler normalization technique has shown promising results in model development [48]. The theoretical description of the minima and maxima normalization method is given as:

$$u' = \frac{u - u_{min}}{u_{max} - u_{min}} \quad (1)$$

Here, u , u_{min} and u_{max} are the actual, minimum, and maximum values of the variable present in the dataset. u' is the normalized value corresponding to u and is mapped in the [0,1] range. Subsequently, the normalized training dataset thus obtained is then used for ML modelling.

2.3. ML modelling algorithms

Three state-of-the-art ML models, i.e., ANN, SVM, and GPR are employed for development of the data-driven models for MED system due to their excellent modelling capabilities and suitability for the system under consideration. ANN is an advanced and efficient data-modelling tool that mimics the functioning of the human brain. Multi-layer perceptron with backpropagation, one of the commonly used network architectures among the scientific community, is proven to be efficient enough for developing engineering solutions for complex and large-scale industrial systems [49]. ANN has fast and effective learning ability in addition to digging the non-linearities present in the large volumes of the data [50]. Moreover, backpropagation is a supervised technique and can develop a generalized model for the given system provided optimized network architecture is constructed [51].

SVM being the supervised learning technique has demonstrated excellent performance in modelling complex systems [52,53]. The normalized training data undergoes space transformation to fit the hyperplane within the decision boundaries and subsequently, the interactions among the input-output variables are captured by Karush-Kuhn-Tucker (KKT) statement [54]. Therefore, SVM has good generalization capability and high prediction accuracy. One of the advanced and competitive features of SVM over ANN is that its computational complexity is independent of input space dimensions as well as it cannot be trapped in the local extreme [55].

GPR is also one of the powerful modelling algorithms of ML and can approximate the complex function underlying the dataset with high accuracy. GPR represents a probabilistic model that relies on non-parametric kernels and is useful for interpolation tasks for the high dimensional input space [56]. A Gaussian process (GP) denotes a group of random variables, and the finite set of random variables has a joint Gaussian distribution. In a GPR model, the response is elucidated by incorporating latent variables, represented as $f(x_i)$ for $i = 1, 2, \dots, N$, derived from a gaussian process, alongside explicit basis functions. The covariance function built on the hidden variables captures the level of smoothness in the output variable while the basis functions project the

input variables “x” into a feature space dimension. The “exploration-exploitation” strategy is implemented, and its trade-off is controlled while modelling the underlying function from the training dataset the trained GPR can then effectively predict the response against the new input observations.

2.3.1. Evaluation criteria

Two commonly used statistical measures, i.e., coefficient of determination (R^2) and root mean square error (RMSE) are selected to measure the performance of the ML models during the model development phase. Their mathematical expressions are defined as:

$$R^2 = 1 - \frac{\sum_i^N (y_i - \hat{y}_i)^2}{\sum_i^N (y_i - \bar{y}_i)^2} \quad (1)$$

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

In this context, y_i represents the observed or actual values of the output variable and \hat{y}_i represents the predicted values of the output variable generated by the model and $i = 1, 2, 3, \dots, N$ equal to the entire count of data points or instances within the dataset. Whereas, \bar{y}_i and $\hat{\bar{y}}_i$ are the mean of actual and model-predicted values. R^2 measures the accuracy of the ML model in predictive analytics and it varies from 0 (poor accuracy) to 1 (excellent accuracy). While RMSE computes the error between the true value and model-simulated response and is to be minimized.

2.4. Variables significance analysis

The ML models are the functional map between the input-output variables shaped by the data associated with the variables. In the next step, it is important to investigate how changing one variable in its operating range impacts the output variables meaning the significance of the input variable towards the output variable can be studied. The importance of the input variable about the output variable in the constructed functional map should be studied as it helps to recognize the sensitivity of the output variable towards the change in the input variable.

The Monte Carlo technique is a global sensitivity analysis method to analyze the impact of the input variable on the output variable. In the Monte Carlo technique, a large number of experiments are constructed and simulated by the developed ML model in a manner that the impact of the input variable on the output variable can be investigated comprehensively. Resultantly, the percentage significance order of the input variables towards the output variables can be established. The details associated with the Monte Carlo technique-based experimental design, simulation from the ML model, and establishment of the variables' significance order can be studied from [55,57].

2.5. Estimating optimized operating conditions for maximizing the distillate production from the MED system

A genetic algorithm is a metaheuristic and nature-inspired global optimization algorithm that can produce quality solutions in a reasonable timeframe and with computational resources [58]. The genetic algorithm is composed of a collection of population sets known as individuals, with each individual defined by a mathematical eq. A random initial population is generated, and its fitness value is evaluated. Subsequently, the individuals undergo modification through mutation and crossover operations, producing offspring. The fitness value of each offspring is recorded, and the most favourable one is selected for the subsequent generation [59]. This process continues until either the

maximum number of generations is reached, during which the input data is modified, or the convergence criteria are met, indicating optimal results for the optimization. Thus, the optimized operating conditions of the input variables can be estimated thereby ensuring the maximum distillate production from the MED system under consideration [60].

2.6. Investigation of the optimized operating conditions

In the last step, the optimized operating conditions for the distillate production from the MED system are investigated on the experimental set-up to check the effectiveness of the model-based analysis for the maximum distillate production. A good agreement between the true distillate production and the model-based estimated distillate production on the determined operating conditions would validate the effectiveness of the ML-based modelling and optimization analysis demonstrating the confidence to desalination community to apply the ML-based analytics for the desalination systems.

3. Result & discussion

3.1. Visualizing the data-distribution space of input-output variables

The experiments are constructed on the wide operating ranges of the input variables taken from the commissioned MED system and are performed on the experimental set-up. The mean, median, and mode values of the distillate production are calculated given the fluctuation in the distillate production corresponding to each experiment. The whole set of experiments are carried out on the commissioned MED system and the distillate production dataset is compiled.

Fig. 4 presents the data-distribution profiles of the input variables hot water temperature (HWT - °C), feed water temperature (FWT - °C), feed flow rate in stage -1 (FFR-S1, LPM), feed flow rate in stage -2 (FFR-S2 - LPM), and feed flow rate in stage -3 (FFR-S3 - LPM) as well as the output variable (distillate production) of the MED system. HWT is varied from 38 °C to 70 °C which is the significantly wide operating range to be investigated for the distillate production from the MED system. HWT provides the thermal energy input to the MED system for the evaporation of the seawater that is condensed to make distillate production. Thus, operating the MED system on the large operating space of the HWT allows us to investigate the performance of the system for the distillation production. Another important input variable of the MED system is the FWT which is the temperature of seawater and is generally maintained around 34 °C to 42 °C with the mean value of 37 °C. The FWT also influences the energy consumption profile of the MED system since the heat duty of the system is dependent on the initial temperature of the seawater being sprayed in the stages of the MED system. The FFR-S1 is varied from 3.6 LPM to 8.7 LPM in the first stage of the designed MED system whereas FFR-S2 and FFR-S3 are maintained around 5.8 LPM to 8.1 LPM. The asymmetric data-distribution profile is observed for the feed flow rates in the stages of the MED system and the mean value is computed to be 7.6 LPM, 6.8 LPM and 7.0 LPM for FFR-S1, FFR-S2 and FFR-S3 respectively. Good data distribution profiles of the experimental data spread over the operating ranges of the variables associated with the MED system is observed. The initial quality of the data collected from the experimental set up of the MED system is ensured since outliers are not present and the data distribution is continuous on the operating ranges of the variables. Thus, the dataset can be used to train the ML based process models to approximate the distillate production behavior from the MED system.

3.2. Development of ANN, SVM, and GPR-based process models for the MED system

Three modelling algorithms of ML including ANN, SVM and GPR are trained to predict the distillate production on the operating conditions of the MED system. The work regarding ML model development and

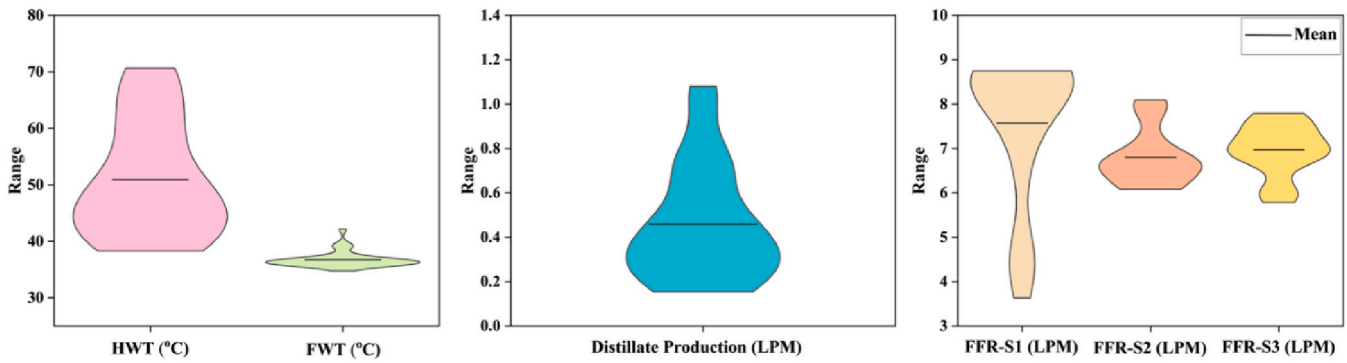


Fig. 4. Data-distribution profiles of the input variables: HWT, FWT, FFR-S1, FFR-S2, FFR-S3, and the output variable: Distillate Production taken from the MED system. The asymmetric data distribution with a continuous distribution of the data observations on the operating ranges of the variables is observed.

simulation is carried out in MATLAB 2021b. The hyperparameters associated with the ML models are to be rigorously and extensively tuned to achieve the good prediction performance of the models. For ANN, the hyperparameters are explored in the following ranges/categories: number of fully connected layers: 1–3, activation function: ReLU, tangent hyperbolic and tangent sigmoidal, regularization strength: 3.03×10^{-7} –3033.3, and layer size: 1–300. Whereas, several hyperparameters are tuned for the SVM model in the operating ranges given as: box constraint: 0.001–1000, epsilon: 0.00028725–28.725 and kernel function: Gaussian, Linear, Quadratic and Cubic. Whereas, for GPR, the following hyperparameters are optimized within their design space which is established as follows: sigma: 0.0001–2.6159, basis function: constant, zero and linear, kernel scale: 0.033639–33.639 and kernel function: Nonisotropic Exponential, Nonisotropic Matern 3/2, Nonisotropic Rotational Quadratic, Nonisotropic Squared Exponential, Isotropic Exponential, Isotropic Rotational Quadratic and Isotropic Squared Exponential.

The cross-validation technique is a rigorous method to reduce the possibility of overfitting and thus promoting the generalization ability of the model. In this work, we have implemented five-fold cross-validation technique for ML models development. Grid-search method in conjunction with the Bayesian optimization technique and expected improvement per second algorithm are deployed for the hyperparameters tuning associated with the ANN, SVM, and GPR for building the well-predictive models. The grid search method systematically explores the effect of the selected combination of the hyperparameters on the model's predictive performance and explores different combination of the hyperparameters unless the ML model achieves good predictive performance. The optimized values of the hyperparameters obtained for the ANN model are as follows: the size of the first hidden layer = 233, the size of the second hidden layer = 113, activation function = tangent hyperbolic, and the regularization strength = 0.00025. The optimized values of hyperparameters for the SVM model are as follows: kernel function = linear, epsilon = 0.014, and box constraint = 779.37. Whereas the optimized values of hyperparameters for the GPR model are as follows: basis function = zero, kernel function = isentropic exponential, kernel scale = 3.8809, and sigma = 0.6157.

Fig. 5 compares the modelling performance of ANN, SVM, and GPR models to predict the distillate production from the MED system. The model-predicted responses are compared with the true values, and the efficacy of the model-predictability is established based on R^2 and RMSE. Referring to Fig. 5(a), a good degree of fit is observed for the ANN model as R^2 of 0.98 and RMSE of 0.037 LPM are computed for the model-based predictions of distillate production. On the other hand, SVM seems to comparatively underperform for the distillate production's modelling task since R^2 of 0.84 and RMSE of 0.103 LPM are computed as shown on Fig. 5(b). Whereas, R^2 of 0.99 and RMSE of 0.026 LPM are computed for the GPR-based predictions for the distillate production. The true data-distribution and model-based data-distribution

profiles are also plotted along the edges of Fig. 5. Both ANN and GPR model-based predictions seem to approximate the true data-distribution profile of distillate production. However, comparing the predictive performance of the three models, it is evident that GPR has developed superior functional mapping to predict distillate production with the relatively lowest predictive error in comparison with those of ANN and SVM. Moreover, the GPR based predictions lie close to the 95 % confidence interval than those of ANN and SVM. Therefore, GPR is selected for performing the further analysis of interest as demonstrated in the following sections.

3.3. Monte Carlo technique-based variables significance analysis

Monte Carlo technique-based variable significance analysis explores the functional ranges of the input variables and constructs the simulated experiments that are simulated from the developed ML model. Usually, a large number of simulated experiments with different possible operating conditions of the input parameters are constructed to investigate the system's response. In this work, we have taken 10 step sizes for the input variable whose significance is to be evaluated and 1000 randomly generated observations within the operating ranges of the other input variables are generated. Thus, 1000 simulated experiments with a constant value of the input variable, whose significance is to be investigated, are constructed and the procedure is repeated unless the complete operating range of the input variable is explored. The simulated experiments are predicted from the developed GPR model and the process is replicated for each of the input variables to investigate the variable's significance.

The parametric effect of the input variables on the distillate production is normalized, and the percentage significance of the input variables towards the distillate production is calculated which is presented in Fig. 6. It is found that hot water temperature (HWT) is the most substantial factor towards distillate production with a significance value of 95.6 %. It is reasoned that as the hot water temperature increases, the energy input to the system increases. This boosts the heat transfer in the evaporator resulting in higher vapor generation and resultantly the distillate production. Feed water temperature (FWT) is the second most significant variable towards distillate production with a significance value of 2 %. This is because at higher feed temperatures, the sensible energy demand for preheating feed in the evaporator is lower. Therefore, hot water energy is utilized for evaporation which increases distillate production. Similarly, the percentage significance value of feed flow rate-S1 (FFR-S1), feed flow rate-S2 (FFR-S2), and feed flow rate-S3 (FFR-S3) is as follows: 2 %, 0.3 %, and 0.1 % respectively. It can be explained considering the domain knowledge of the MED system that a higher feed flow rate requires higher sensible energy input to achieve the evaporation temperature and subsequent vapor production. Meanwhile, it is also important to mention that a lower feed flow rate results in dry patches that diminishes the distillate production. Therefore, the

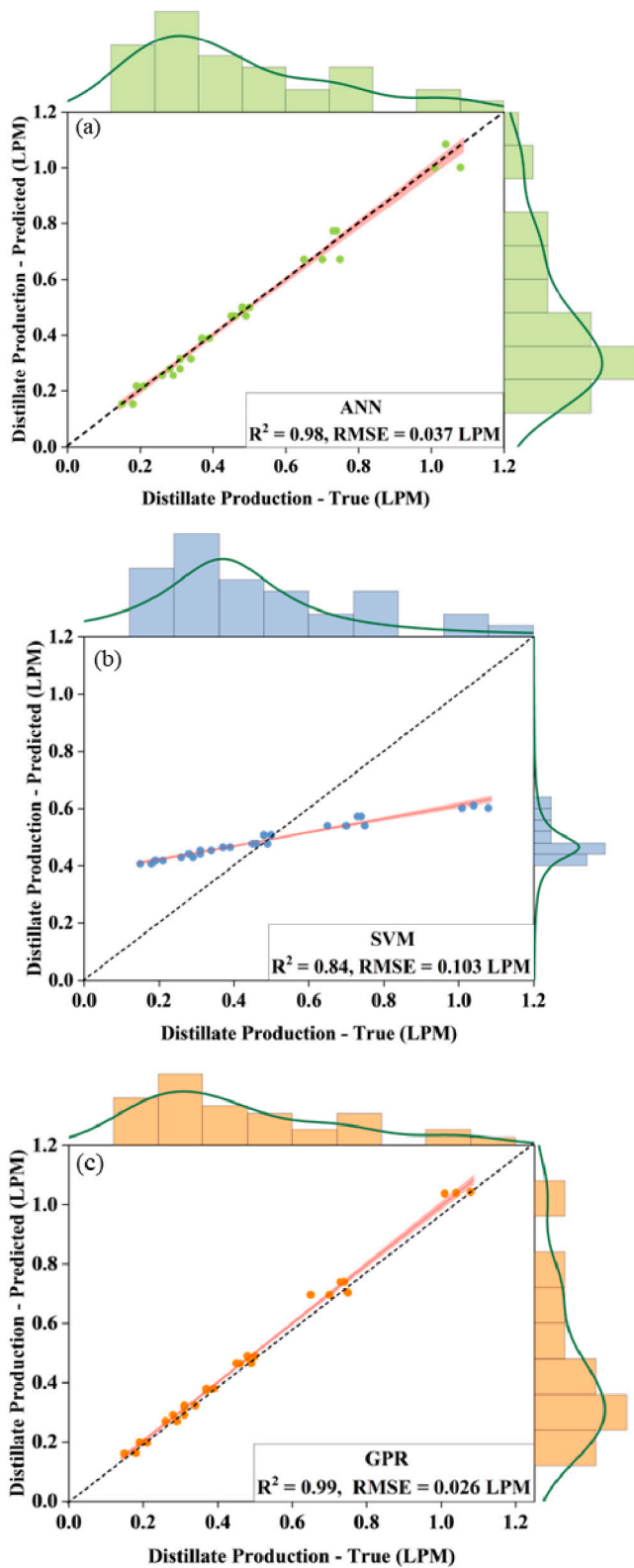


Fig. 5. Development of ML models to predict the distillate production by (a) ANN, (b) SVM, and (c) GPR. The GPR model appears to have comparatively superior modelling performance than those of ANN and SVM. The prediction interval on 95 % confidence interval is also constructed for the model-based predictions.

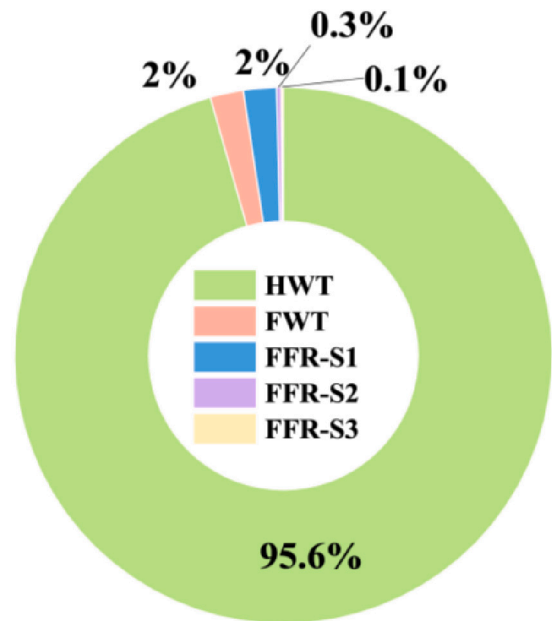


Fig. 6. Significance percentages associated with input variables towards the prediction of distillate production. HWT is the most significant variable towards the distillate production followed by FWT, FFR-S1, FFR-S2, and FFR-S3.

required quantity of feed water flow rate should be maintained.

3.4. Maximizing the distillate production by genetic algorithm

In this work, the GPR model exhibited good performance to model the distillate production from the MED system based on the identified input variables. The MED system is operated in the lab on the design space of the input variables and the distillate production dataset is compiled corresponding to the experimental conditions. It is imperative to investigate the MED system by advanced optimization techniques to estimate the optimized operating conditions of the input variables such that distillate production is maximized. Thus, the optimization problem is formulated that takes the distillate production as the objective function and the design space of the input variables is incorporated to explore it for the determination of the optimized operating conditions to obtain the maximum distillate production. The optimization problem for maximizing the distillate production from the MED system is written as:

$$\max f = \text{Distillate Production}$$

Subject to:

$$38 \text{ C} \leq \text{HWT} \leq 70\text{C}$$

$$34 \text{ C} \leq \text{FWT} \leq 42\text{C}$$

$$3.6 \text{ LPM} \leq \text{FFR} - \text{S1} \leq 8.7 \text{ LPM}$$

$$6.1 \text{ LPM} \leq \text{FFR} - \text{S2} \leq 8.1 \text{ LPM}$$

$$5.8 \text{ LPM} \leq \text{FFR} - \text{S3} \leq 7.8 \text{ LPM}$$

The optimization problem is solved in MATLAB 2021b by genetic algorithm solver and the default values of number of generations, number of population, and other settings are deployed to obtain the solution for the maximum distillate production. Fig. 7 presents the mapping of the optimized values of the hot water temperature and feed water temperature as determined by the GA-based optimization technique on the distillate production profile. The two input variables are deployed for constructing the distillate production profile since they have 97.6 % percentage significance towards it. The mean value of hot

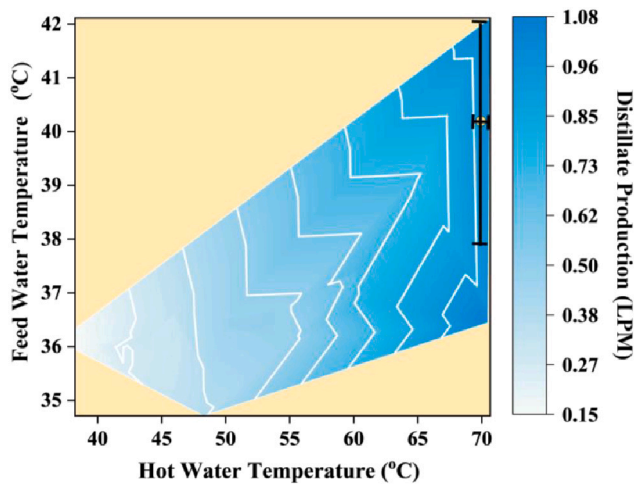


Fig. 7. Mapping the genetic algorithm-driven optimal solution for the maximum distillate production on the response curve of distillate production constructed against the two significant input variables – hot water temperature and feed water temperature.

water temperature is 70 ± 0.5 °C and feed water temperature is 40 ± 2.5 °C corresponding to distillate production of 1.042 LPM – the maximum achievable distillate production from the MED system. Similarly, the optimized operating range for feed flow rate-S1, feed flow rate-S2, and feed flow rate-S3 for maximum distillate production (1.042 LPM) is as follows: 6 ± 2.6 LPM, 7 ± 1 LPM, and 7 ± 1 LPM respectively. Overall, the higher hot water temperature increases the energy input to the system which boosts the heat transfer in the evaporator resulting in higher vapor production and distillate. Meanwhile, higher feed water temperature also increases distillate production because the higher feed temperature results in lower sensible energy demand for preheating feed in the evaporator. Therefore, hot water energy is utilized for evaporation which increases water production. However, the feedwater temperature can only be increased by recuperating energy from the rejected brine, distillate stream, or auxiliary energy source.

3.5. Investigation of the GA-driven optimized solution on the MED set-up

The ML-based studies conducted in the domain of desalination systems focus on modelling its performance indicators and subsequently, the trained models are deployed for predictive analytics. A closed-loop ML-based modelling and optimization framework that can be deployed for estimating the effective solutions for the desalination systems, especially for MED systems is potentially missing in the literature and is an obstacle to the inclusion of ML in the domain of desalination systems. The experimental validation of the model-based optimized solution to obtain the maximum distillate production from the MED system is carried out to evaluate the accuracy of the estimated solution on the experimental setup.

The MED system is operated on the optimized operating range of the input variables and the distillate production dataset is compiled. The important input parameters including HWT, FWT, and FFR-S1, FFR-S2 and FFR-S3 are maintained carefully in the optimized range. This is because the appropriate selection of these parameters increases distillate production. The distribution of the distillate production against the optimized operating ranges of the input variables is presented on Fig. 8. The distillate production is varied from 0.04 LPM to 1.98 LPM during the experimentation while the average distillate water production remained around 0.98 LPM during the experimentation. The fluctuation in the distillate production can be explained by the sensitivity of the sensors and flashing (due to pressure drop) of the distillation production. Therefore, the average value of the distillate production is computed to estimate the distillate production during the experimentation.

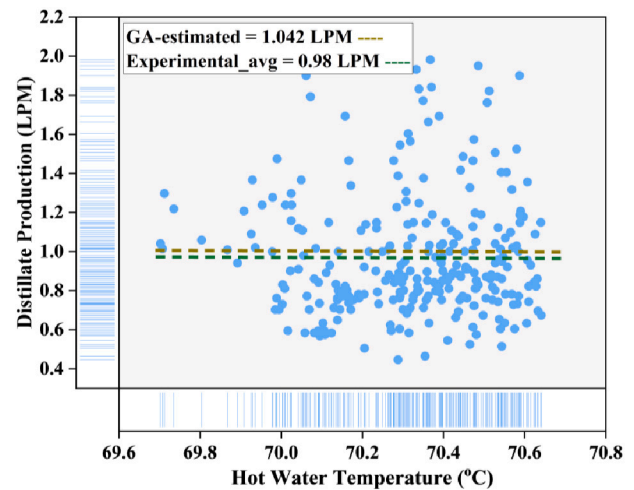


Fig. 8. Investigation of model-based optimized solution for maximum distillation production. A close agreement between the GA-estimated and experimental value of distillate production is observed.

Comparing the true distillate production with the GPR model-based maximum distillate production estimated by GA, i.e., 1.042 LPM, it is apparent that the model-based optimized solution for the maximum distillate production is quite close to the experimental observation. The experimental verification of the ML-based optimized solution demonstrates the effectiveness of the presented ML-based modelling and optimization framework for the maximum distillate production from the MED system. Thus, the data-driven ML approach can be applied for the improved operation of distillate production thereby contributing to operation excellence of the MED system.

4. Conclusion

Desalination technologies hold the promise to meet the water demand for different applications on the face of water scarcity and sustainability issues. The improved process design and operation of the MED system can enhance distillate production boosting the performance of the installed desalination technologies worldwide. In this work, the analytical framework leveraging the power of machine learning and rigor of optimization technique is proposed for the improved design of the MED system to support the circular economy.

- Extensive experimentation is carried out on the in-house built MED system to collect the dataset which is deployed for constructing the ML-based process models. ANN, SVM, and GPR-based process models are trained, under rigorous hyperparameter tuning, to predict the distillate production from the MED system. GPR is turned out to be a superior algorithm having R^2 of 0.99 and RMSE of 0.026 LPM.
- Monte Carlo technique-based variables significance analysis reveals that hot water temperature is the most significant input variable towards the distillate production with a percentage significance value of 95.6 followed by feed water temperature having a percentage significance of 2.
- The trained GPR model is integrated into the optimization problem for the maximum distillate production and the optimization problem is solved by genetic algorithm. The optimized operating conditions for the maximum distillate production are estimated which are as follows: hot water temperature = 70 ± 0.5 °C, feed water temperature = 40 ± 2.5 °C, feed flow rate-S1 = 6 ± 2.6 LPM, feed flow rate-S2 = 7 ± 1 LPM and feed flow rate-S3 = 7 ± 1 .
- The optimized operating values of the input variables are investigated on the experimental MED system and a close agreement

between the true distillate production (0.98 LPM) and estimated distillate production (1.042 LPM) is found.

- The closed-loop investigation of model-based optimization results on the MED system developed by the proposed analytical framework demonstrates the effectiveness of ML for the improved process design, performance enhancement and operation excellence of the MED system that contributes to the circular economy and digitalization of the desalination systems.

Future work

In the future, a comprehensive study on different desalination technologies will be conducted following the ML-based modelling and optimization framework. The multi-objective optimization problem will be investigated for the higher energy efficiency and operation excellence of the desalination systems.

CRedit authorship contribution statement

Waqar Muhammad Ashraf: Software, Formal analysis. **Ghulam Moeen Uddin:** Methodology, Formal analysis. **Bashar Shboul:** Methodology. **Kashif Ishfaq:** Validation, Formal analysis. **Kim Choon Ng:** Writing – review & editing, Conceptualization. **Mike Dixon:** Validation, Software. **Ben Bin Xu:** Writing – review & editing, Investigation. **Muhammad Wakil Shahzad:** Writing – original draft, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The authors would like to thank Northumbria University UK and King Abdullah University of Science and Technology Saudi Arabia for funding this research. Dr. Shahzad also would like to thank the International Desalination Associate (IDA) Fellowship 2022 and RAEng/Leverhulme Trust Research Fellowships Scheme Tranche 19 Reference: LTRF2223-19-103. Waqar Muhammad Ashraf declares to have received the funding from Punjab Education Endowment Fund (PEEF): PEEF/CMMS/21/142 to pursue his PhD at University College London, United Kingdom.

References

- [1] United Nations Children Fund, World health Organisation, Progress on Drinking Water, Sanitation and Hygiene - Joint Monitoring Programme 2017 Update and SDG Baselines, 2017, <https://doi.org/10.1111/tmi.12329>.
- [2] A. Velmurugan, P. Swarnam, T. Subramani, B. Meena, M.J. Kaledhonkar, Water Demand and Salinity, *IntechOpen*, 2020, p. 13, <https://doi.org/10.5772/intechopen.88095>.
- [3] M.A. Jamil, B.A. Qureshi, S.M. Zubair, Exergo-economic analysis of a seawater reverse osmosis desalination plant with various retrofit options, *Desalination* 401 (2016) 88–98.
- [4] M.A. Jamil, M.W. Shahzad, S.M. Zubair, A comprehensive framework for thermo-economic analysis of desalination systems, *Energy Convers. Manag.* 222 (2020), <https://doi.org/10.1016/j.enconman.2020.113188>.
- [5] E. Jones, M. Qadir, M.T.H. van Vliet, V. Smakhtin, S. Mu Kang, The state of desalination and brine production: a global outlook, *Sci. Total Environ.* 657 (2019) 1343–1356, <https://doi.org/10.1016/j.scitotenv.2018.12.076>.
- [6] M.W. Shahzad, M. Burhan, D. Ybyraiymkul, K.C. Ng, Desalination processes' efficiency and future roadmap, *Entropy* 21 (2019) 84, <https://doi.org/10.3390/e21010084>.
- [7] M.W. Shahzad, M. Burhan, L. Ang, K. Choon Ng, Energy-water-environment nexus underpinning future desalination sustainability, *Desalination* 413 (2017) 52–64, <https://doi.org/10.1016/j.desal.2017.03.009>.
- [8] M.A. Jamil, H. Yaqoob, A. Abid, M.U. Farooq, B. Bin Xu, L. Dala, K.C. Ng, M. W. Shahzad, An exergoeconomic and normalized sensitivity based comprehensive investigation of a hybrid power-and-water desalination system, *Sustain Energy Technol Assess* 47 (2021) 101463, <https://doi.org/10.1016/j.seta.2021.101463>.
- [9] M.W. Shahzad, K. Thu, B.B. Saha, K.C. Ng, An emerging hybrid multi-effect adsorption desalination system, *Evergreen* 1 (2014) 30–36.
- [10] N. Ghaffour, T.M. Missimer, G.L. Amy, Technical review and evaluation of the economics of water desalination: current and future challenges for better water supply sustainability, *Desalination* 309 (2013) 197–207.
- [11] K.C. Ng, M. Burhan, Q. Chen, D. Ybyraiymkul, F.H. Akhtar, M. Kumja, R.W. Field, M. W. Shahzad, A thermodynamic platform for evaluating the energy efficiency of combined power generation and desalination plants, *NPJ Clean Water* 4 (2021) 1–10, <https://doi.org/10.1038/s41545-021-00114-5>.
- [12] J. Vievard, A. Alem, A. Pantet, N.D. Ahfir, M.G. Arellano-Sánchez, C. Devouge-Boyer, M. Mignot, Bio-based adsorption as ecofriendly method for wastewater decontamination: a review, *Toxics* Vol. 11 (2023) 404 (11 (2023) 404), <https://doi.org/10.3390/TOXICS11050404>.
- [13] M.R. Awual, M.M. Hasan, Fine-tuning mesoporous adsorbent for simultaneous ultra-trace palladium(II) detection, separation and recovery, *J. Ind. Eng. Chem.* 21 (2015) 507–515, <https://doi.org/10.1016/J.JIEC.2014.03.013>.
- [14] N. Karić, A.S. Maia, A. Teodorović, N. Atanasova, G. Langergraber, G. Crini, A.R. L. Ribeiro, M. Dolić, Bio-waste valorisation: agricultural wastes as biosorbents for removal of (in)organic pollutants in wastewater treatment, *Chem. Eng. J. Adv.* 9 (2022) 100239, <https://doi.org/10.1016/J.CEJA.2021.100239>.
- [15] S. Wu, W. Shi, K. Li, J. Cai, L. Chen, Recent advances on sustainable bio-based materials for water treatment: fabrication, modification and application, *J. Environ. Chem. Eng.* 10 (2022) 108921, <https://doi.org/10.1016/J.JECE.2022.108921>.
- [16] M.N. Hasan, M.A. Shenashen, M.M. Hasan, H. Znad, M.R. Awual, Assessing of cesium removal from wastewater using functionalized wood cellulosic adsorbent, *Chemosphere* 270 (2021) 128668, <https://doi.org/10.1016/J.CHEMOSPHERE.2020.128668>.
- [17] M.R. Awual, M.M. Hasan, A. Islam, M.M. Rahman, A.M. Asiri, M.A. Khaleque, M. C. Sheikh, Offering an innovative composited material for effective lead(II) monitoring and removal from polluted water, *J. Clean. Prod.* 231 (2019) 214–223, <https://doi.org/10.1016/J.JCLEPRO.2019.05.125>.
- [18] M. Raninga, A. Mudgal, V.K. Patel, J. Patel, M. Kumar Sinha, Modification of activated carbon-based adsorbent for removal of industrial dyes and heavy metals: a review, *Mater. Today Proc.* 77 (2023) 286–294, <https://doi.org/10.1016/J.MATPR.2022.11.358>.
- [19] M.S. Hossain, M.A. Shenashen, M.E. Awual, A.I. Rehan, A.I. Rasee, R.M. Waliullah, K.T. Kubra, M.S. Salman, M.C. Sheikh, M.N. Hasan, M.M. Hasan, A. Islam, M. A. Khaleque, H.M. Marwani, K.A. Alzahrani, A.M. Asiri, M.M. Rahman, M. R. Awual, Benign separation, adsorption, and recovery of rare-earth Yb(III) ions with specific ligand-based composite adsorbent, *Process. Saf. Environ. Prot.* 185 (2024) 367–374, <https://doi.org/10.1016/J.PSEPE.2024.03.026>.
- [20] M.C. Sheikh, M.M. Hasan, M.N. Hasan, M.S. Salman, K.T. Kubra, M.E. Awual, R. M. Waliullah, A.I. Rasee, A.I. Rehan, M.S. Hossain, H.M. Marwani, A. Islam, M. A. Khaleque, M.R. Awual, Toxic cadmium(II) monitoring and removal from aqueous solution using ligand-based facial composite adsorbent, *J. Mol. Liq.* 389 (2023) 122854, <https://doi.org/10.1016/J.MOLLIQ.2023.122854>.
- [21] M.R. Awual, M.M. Hasan, A. Shahat, M. Naushad, H. Shiwaku, T. Yaita, Investigation of ligand immobilized nano-composite adsorbent for efficient cerium (III) detection and recovery, *Chem. Eng. J.* 265 (2015) 210–218, <https://doi.org/10.1016/J.CEJ.2014.12.052>.
- [22] S. Su, Y. Huang, S. Yang, B. Liu, H. Sun, G. Han, An innovative strategy for deeply separating macro amounts of molybdenum from tungstate solutions via vulcanization step-by-step and microbubble floating-extraction: theoretical and experimental investigation, *J. Ind. Eng. Chem.* 131 (2024) 635–652, <https://doi.org/10.1016/J.JIEC.2023.10.066>.
- [23] M.R. Awual, Novel nanocomposite materials for efficient and selective mercury ions capturing from wastewater, *Chem. Eng. J.* 307 (2017) 456–465, <https://doi.org/10.1016/J.CEJ.2016.08.108>.
- [24] M.R. Awual, M.M. Hasan, M.A. Khaleque, M.C. Sheikh, Treatment of copper(II) containing wastewater by a newly developed ligand based facial conjugate materials, *Chem. Eng. J.* 288 (2016) 368–376, <https://doi.org/10.1016/J.CEJ.2015.11.108>.
- [25] M.A. Jamil, Z.U. Din, T.S. Goraya, H. Yaqoob, S.M. Zubair, Thermal-hydraulic characteristics of gasketed plate heat exchangers as a preheater for thermal desalination systems, *Energy Convers. Manag.* 205 (2020) 112425, <https://doi.org/10.1016/j.enconman.2019.112425>.
- [26] H.S. Son, M.W. Shahzad, N. Ghaffour, K.C. Ng, Pilot studies on synergetic impacts of energy utilization in hybrid desalination system: multi-effect distillation and adsorption cycle (MED-AD), *Desalination* 477 (2020) 114266, <https://doi.org/10.1016/j.desal.2019.114266>.
- [27] H. Rezvani, P.B. Whittaker, H.T. Chua, New MED based desalination process for low grade waste heat, *Desalination* 395 (2016) 57–71, <https://doi.org/10.1016/j.desal.2016.05.022>.
- [28] M.L. Elsayed, O. Mesalhy, R.H. Mohammed, L.C. Chow, Exergy and thermo-economic analysis for MED-TVC desalination systems, *Desalination* 447 (2018) 29–42.
- [29] S. Rashidi, N. Karimi, W.M. Yan, Applications of machine learning techniques in performance evaluation of solar desalination systems – a concise review, *Eng. Anal. Bound. Elem.* 144 (2022) 399–408, <https://doi.org/10.1016/j.enganabound.2022.08.031>.

- [30] M.W. Shahzad, K.C. Ng, M. Burhan, Q. Chen, M.A. Jamil, N. Imtiaz, B. Bin Xu, Demystifying integrated power and desalination processes evaluation based on standard primary energy approach, *Therm. Sci. Eng. Progr.* 27 (2022) 101153, <https://doi.org/10.1016/j.tsep.2021.101153>.
- [31] Y. Wang, Z. Cao, A. Barati Farimani, Efficient water desalination with graphene nanopores obtained using artificial intelligence, *NPJ 2D Mater. Appl.* 5 (2021), <https://doi.org/10.1038/s41699-021-00246-9>.
- [32] J. Jawad, A.H. Hawari, S. Javaid Zaidi, Artificial neural network modeling of wastewater treatment and desalination using membrane processes: a review, *Chem. Eng. J.* 419 (2021) 129540, <https://doi.org/10.1016/j.cej.2021.129540>.
- [33] S.S. Ray, R.K. Verma, A. Singh, M. Ganesapillai, Y.N. Kwon, A holistic review on how artificial intelligence has redefined water treatment and seawater desalination processes, *Desalination* 546 (2023) 116221, <https://doi.org/10.1016/j.desal.2022.116221>.
- [34] H. Salem, I.M. El-Hasnony, A.E. Kabeel, E.M.S. El-Said, O.M. Elzeki, Deep learning model and classification explainability of renewable energy-driven membrane desalination system using evaporative cooler, *Alex. Eng. J.* 61 (2022) 10007–10024, <https://doi.org/10.1016/j.aej.2022.03.050>.
- [35] M. Son, N. Yoon, K. Jeong, A. Abass, B.E. Logan, K.H. Cho, Deep learning for pH prediction in water desalination using membrane capacitive deionization, *Desalination* 516 (2021) 115233, <https://doi.org/10.1016/j.desal.2021.115233>.
- [36] T. Bonny, M. Kashkash, F. Ahmed, An efficient deep reinforcement machine learning-based control reverse osmosis system for water desalination, *Desalination* 522 (2022) 115443, <https://doi.org/10.1016/j.desal.2021.115443>.
- [37] S. Shahane, H.Q. Jin, S. Wang, K. Nawaz, Numerical modeling based machine learning approach for the optimization of falling-film evaporator in thermal desalination application, *Int. J. Heat Mass Transf.* 196 (2022) 123223, <https://doi.org/10.1016/j.ijheatmasstransfer.2022.123223>.
- [38] J. Krzywanski, D. Skrobek, A. Zylka, K. Grabowska, A. Kulakowska, M. Sosnowski, W. Nowak, A.M. Blanco-Marigorta, Heat and mass transfer prediction in fluidized beds of cooling and desalination systems by AI approach, *Appl. Therm. Eng.* 225 (2023) 120200, <https://doi.org/10.1016/j.applthermaleng.2023.120200>.
- [39] Q. He, H. Zheng, X. Ma, L. Wang, H. Kong, Z. Zhu, Artificial intelligence application in a renewable energy-driven desalination system: a critical review, *Energy AI* 7 (2022) 100123, <https://doi.org/10.1016/j.egyai.2021.100123>.
- [40] B. Mahjoob Karambasi, M. Ghodrati, G. Ghorbani, A. Lalbakhsh, M. Behnia, Design methodology and multi-objective optimization of small-scale power-water production based on integration of Stirling engine and multi-effect evaporation desalination system, *Desalination* 526 (2022) 115542, <https://doi.org/10.1016/j.desal.2021.115542>.
- [41] D.V. Pombo, H.W. Bindner, S.V. Spataru, P.E. Sørensen, M. Rygaard, Machine learning-driven energy management of a hybrid nuclear-wind-solar-desalination plant, *Desalination* 537 (2022), <https://doi.org/10.1016/j.desal.2022.115871>.
- [42] H. Salem, A.E. Kabeel, E.M.S. El-Said, O.M. Elzeki, Predictive modelling for solar power-driven hybrid desalination system using artificial neural network regression with Adam optimization, *Desalination* 522 (2022) 115411, <https://doi.org/10.1016/j.desal.2021.115411>.
- [43] H. Shakibi, A. Shokri, E. Assareh, M. Yari, M. Lee, Using machine learning approaches to model and optimize a combined solar/natural gas-based power and freshwater cogeneration system, *Appl. Energy* 333 (2023) 120607, <https://doi.org/10.1016/j.apenergy.2022.120607>.
- [44] M.W. Shahzad, The Hybrid Multi-effect Desalination (MED) and the Adsorption (AD) Cycle for Desalination | ScholarBank@NUS. <https://scholarbank.nus.edu.sg/handle/10635/49640>, 2013.
- [45] A. Abid, M.A. Jamil, N. Us Sabah, M.U. Farooq, H. Yaqoob, L.A. Khan, M. W. Shahzad, Exergoeconomic optimization of a forward feed multi-effect desalination system with and without energy recovery, *Desalination* 499 (2021), <https://doi.org/10.1016/j.desal.2020.114808>.
- [46] M.A. Jamil, S.M. Zubair, Effect of feed flow arrangement and number of evaporators on the performance of multi-effect mechanical vapor compression desalination systems, *Desalination* 429 (2018) 76–87, <https://doi.org/10.1016/j.desal.2017.12.007>.
- [47] Q. Xie, M. Suvarna, J. Li, X. Zhu, J. Cai, X. Wang, Online prediction of mechanical properties of hot rolled steel plate using machine learning, *Mater. Des.* 197 (2021) 109201, <https://doi.org/10.1016/j.matdes.2020.109201>.
- [48] S. Jain, S. Shukla, R. Wadhvani, Dynamic selection of normalization techniques using data complexity measures, *Expert Syst. Appl.* 106 (2018) 252–262, <https://doi.org/10.1016/j.eswa.2018.04.008>.
- [49] B. Shboul, M.E. Zayed, W.M. Ashraf, M. Usman, D. Roy, K. Irshad, S. Rehman, Energy and economic analysis of building integrated photovoltaic thermal system: seasonal dynamic modeling assisted with machine learning-aided method and multi-objective genetic optimization, *Alex. Eng. J.* 94 (2024) 131–148, <https://doi.org/10.1016/J.AEJ.2024.03.049>.
- [50] W.M. Ashraf, V. Dua, Artificial intelligence driven smart operation of large industrial complexes supporting the net-zero goal: coal power plants, *Digit. Chem. Eng.* 8 (2023) 100119, <https://doi.org/10.1016/J.DCHE.2023.100119>.
- [51] W.M. Ashraf, V. Dua, Data Information integrated Neural Network (DINN) algorithm for modelling and interpretation performance analysis for energy systems, *Energy AI* 16 (2024) 100363, <https://doi.org/10.1016/J.EGYAI.2024.100363>.
- [52] M.E. Zayed, A.E. Kabeel, B. Shboul, W.M. Ashraf, M. Ghazy, K. Irshad, S. Rehman, A.A.A. Zayed, Performance augmentation and machine learning-based modeling of wavy corrugated solar air collector embedded with thermal energy storage: support vector machine combined with Monte Carlo simulation, *J. Energy Storage* 74 (2023) 109533, <https://doi.org/10.1016/J.EST.2023.109533>.
- [53] M.W. Shahzad, V.H. Nguyen, B. Bin Xu, R. Tariq, M. Imran, W.M. Ashraf, K.C. Ng, M.A. Jamil, A. Ijaz, N.A. Sheikh, Machine learning assisted prediction of solar to liquid fuel production: a case study, *Process. Saf. Environ. Prot.* 184 (2024) 1119–1130, <https://doi.org/10.1016/J.PSEP.2024.02.060>.
- [54] W.M. Ashraf, G.M. Uddin, R. Tariq, A. Ahmed, M. Farhan, M.A. Nazeer, R. U. Hassan, A. Naeem, H. Jamil, J. Krzywanski, M. Sosnowski, V. Dua, Artificial intelligence modeling-based optimization of an industrial-scale steam turbine for moving toward net-zero in the energy sector, *ACS Omega* 8 (2023) 21709–21725, https://doi.org/10.1021/ACSOMEGA.3C01227/ASSET/IMAGES/LARGE/AO3C01227_0012.JPEG.
- [55] W.M. Ashraf, G.M. Uddin, S.M. Arafat, J. Krzywanski, W. Xiaonan, Strategic-level performance enhancement of a 660 MWe supercritical power plant and emissions reduction by AI approach, *Energy Convers. Manag.* 250 (2021) 114913, <https://doi.org/10.1016/j.enconman.2021.114913>.
- [56] B. Shboul, M.E. Zayed, R. Tariq, W.M. Ashraf, A.S. Odat, S. Rehman, A. S. Abdelrazik, J. Krzywanski, New hybrid photovoltaic-fuel cell system for green hydrogen and power production: performance optimization assisted with Gaussian process regression method, *Int. J. Hydrog. Energy* 59 (2024) 1214–1229, <https://doi.org/10.1016/J.IJHYDENE.2024.02.087>.
- [57] W.M. Ashraf, G.M. Uddin, S.M. Arafat, S. Afghani, A.H. Kamal, M. Asim, M.H. Khan, M.W. Rafique, U. Naumann, S.G. Niazi, H. Jamil, A. Jamil, N. Hayat, A. Ahmad, S. Changkai, L. Bin Xiang, I.A. Chaudhary, J. Krzywanski, Optimization of a 660 MWe supercritical power plant performance—a case of industry 4.0 in the data-driven operational management Part 1. Thermal efficiency, *Energies* Vol. 13 (2020) 5592 (13 (2020) 5592), <https://doi.org/10.3390/EN13215592>.
- [58] R. Tariq, M. Ali, N.A. Sheikh, M.W. Shahzad, B. Bin Xu, Deep learning artificial intelligence framework for sustainable desiccant air conditioning system: optimization towards reduction in water footprints, *Int. Commun. Heat Mass Transfer* 140 (2023) 106538, <https://doi.org/10.1016/J.ICHEATMASSTRANSFER.2022.106538>.
- [59] R. Tariq, C.E. Torres-Aguilar, J. Xamán, I. Zavala-Guillén, A. Bassam, L.J. Ricalde, O. Carvente, Digital twin models for optimization and global projection of building-integrated solar chimney, *Build. Environ.* 213 (2022) 108807, <https://doi.org/10.1016/J.BUILDENV.2022.108807>.
- [60] W. Muhammad Ashraf, G. Moeen Uddin, H. Afroze Ahmad, M. Ahmad Jamil, R. Tariq, M. Wakil Shahzad, V. Dua, Artificial intelligence enabled efficient power generation and emissions reduction underpinning net-zero goal from the coal-based power plants, *Energy Convers. Manag.* 268 (2022) 116025, <https://doi.org/10.1016/j.enconman.2022.116025>.