Exploring machine learning models to predict atmospheric water harvesting with an ion deposition membrane

Journal of Water Process Engineering 72 (2025) 107476

Contents lists available at ScienceDirect

Journal of Water Process Engineering

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

# Exploring machine learning models to predict atmospheric water harvesting with an ion deposition membrane

Giulio Barletta [a,b], Shashwata Moitra [a], Sybil Derrible [c], Alex Mathew [d], Anoop Muraleedharan Nair [d], Constantine M. Megaridis [a,*]

[a] Department of Mechanical and Industrial Engineering, University of Illinois Chicago, 842 W Taylor St, Chicago IL 60607, United States
[b] Department of Energy, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino 10129, Italy
[c] Department of Civil, Materials, and Environmental Engineering, University of Illinois Chicago, 842 W Taylor St, Chicago IL 60607, United States
[d] Biomass Energy Systems, 1175 Central Avenue, University Park IL 60484, United States

ARTICLE INFO

Editor: Soroush Abolfathi

*Keywords:*
Machine learning
Atmospheric water harvesting
Membrane

ABSTRACT

This work investigates the performance of a novel membrane-based atmospheric water harvesting (AWH) unit under various operating conditions of ambient temperature, relative humidity (RH), and carrier fluid flow rate. Ion deposition membranes (IDMs) were selected for their ability to enhance water uptake by lowering the water vapor saturation pressure at the gas-membrane interface. This effect, achieved through metal ion implantation into PTFE-based membranes, improves water harvesting rates – especially under low RH conditions – by up to a factor of four compared to untreated membranes. The benchmark design was tested over all possible combinations of four distinct carrier fluid flow rates, three temperatures, and six RH values. The yield with a lab-scale prototype was as high as 354 ml/day of water, with an average of 155 ml/day, corresponding to water harvesting rates of 22.13 kg/m$^2$/day and 9.69 kg/m$^2$/day, respectively. The experimental dataset obtained was used to build three machine learning (ML) regression models to predict the amount of water harvested under specific operating conditions. The ML techniques are: Support Vector Regression, Gradient Boosting Regression, and Multilayer Perceptron. These methods achieved accuracy scores as high as 89 %, proving suitable for implementation in the regulation of AWH plants featuring this technology. The best-performing model (Multilayer Perceptron) was used to predict the water harvesting potential on a typical spring day in Jeddah, Saudi Arabia, a region facing severe water scarcity.

## 1. Introduction

Water is a fundamental life resource on Earth. While it covers 71 % of the Earth's surface, most of the available water is seawater. Water scarcity has been and remains a major issue. Modest shortages affect two-thirds of the global population for at least one month per year, and half a billion people face severe water scarcity issues throughout the year [1,2]. According to a study published by the United Nations (UN) in 2018, 2.7 to 3.2 billion people are projected to live in regions of severe water scarcity by 2050 [3]. The causes behind this world-challenging problem are several, and the resulting negative impacts on both the environment and humans are numerous. Climate change has had a major role, leading to the impoverishment of water reservoirs, the reduction of fluvial flows, and the retreat of glaciers. Other factors that play a significant role are the pollution of the water sources and their unsuitability or inaccessibility [4]. Existing freshwater sources are under high pressure due to their uneven distribution, as 90 % of the global population resides in developing countries. Water scarcity issues also represent a public health hazard in geographical regions that lack access to clean water: around 2 million deaths per year are caused by unsafe water and poor hygiene [5–7].

Ultimately, developing feasible technologies for freshwater generation is a pressing task as it can help alleviate water scarcity issues. Among these technologies, Atmospheric Water Harvesting (AWH) systems represent an attractive prospect, as they are accessible everywhere and can be driven by renewable energy sources [2]. A plentiful reservoir of freshwater resides as humidity in the air: it is present everywhere, it is naturally and continuously replenished, and it amounts to almost 13,000 trillion liters [5,8]. Available AWH technologies can be classified into two main categories: a) condensation-based, such as vapor

* Corresponding author.
*E-mail address:* CMM@uic.edu (C.M. Megaridis).

https://doi.org/10.1016/j.jwpe.2025.107476
Received 27 December 2024; Received in revised form 7 March 2025; Accepted 11 March 2025
Available online 23 March 2025

compression [9,10], thermoelectric cooling [11,12], or absorption/adsorption refrigeration [13,14]; b) sorption-based, which rely on desiccants [15–18]. A broad landscape of other innovative technologies is continuously expanding, with the development of new devices that cannot be classified as belonging to either of the two groups. These include technologies such as wind-turbine-powered vapor compression cycles developed by Solís-Chaves et al. [19], solar chimneys proposed by Kashiwa et al. [20] and Ming et al. [21], solar-assisted desiccant wheel dehumidifiers proposed by Milani et al. [22], membrane-assisted dehumidifiers proposed by Bergmair et al. [23], and hybrid systems using multi-stage desiccant wheels and vapor compression cycles proposed by Tu et al. [24].

The development and availability of accurate predictive tools for these water harvesting technologies can be beneficial; however, physics-based models are often lacking or difficult to implement. Machine Learning (ML) methods offer a reliable solution to overcome the challenges of physics-based modeling, as demonstrated by numerous studies predicting sorption properties and capacities of various systems. Zhang et al. [25] investigated the aqueous adsorption of organic compounds onto granular activated carbons, carbon nanotubes, biochars, and resins, combining ML techniques and the linear free energy relationship (LFER) approach. After finding that Neural Networks (NN) achieved better prediction accuracy over Support Vector Machine (SVM) and Bagging, they successfully built four NN-LFER models displaying outstanding predictability. Panapitiya et al. [26] proposed an ML model aided by an *ab-initio* Density Functional Theory (DFT) approach to predict the adsorption energy in thiolated, Ag-alloyed Au nanoclusters. After having established that tree ensemble methods performed better than NN and SVM, the authors created more complicated features combining the most relevant ones from the original data set, achieving high prediction accuracy. Zhu et al. [27] adopted ML methods to determine the latent relationships of $CO_2$ adsorption capacity on porous carbon materials (PCMs), properties and adsorption conditions. The authors built highly accurate random forest (RF) predictive models using six sets of $CO_2$ adsorption data at different temperatures and three pressure ranges. Similarly, Trezza et al. [28] exploited ML regressors trained to predict the adsorption capacities of Metal-Organic Frameworks (MOFs) to establish a minimal set of important crystallographic features, and investigated the role of such "genetic code" when using Sequential Learning (SL) algorithms. Guo et al. [29] used two deep learning algorithms — multilayer perceptron (MLP) and long short-term memory (LSTM) network — to predict the gas adsorption capacity of MOFs with methane and carbon dioxide at different pressures. The algorithms were used to build models for predicting adsorption capacities of methane at pressures of 1, 5.8, 35, and 65 bar. Zhang et al. [30] proposed a new method to predict the hydrogen adsorption capacity of MOFs at cryogenic temperatures and high pressures by training extrapolative ML models on data at very low pressures and room temperature. Li et al. [31] used Monte Carlo methods and three ML algorithms to study the performance of AWH on MOFs, implementing RF, Gradient Boosting Regression (GBR), and Neighbor Component Analysis. Fanourgakis et al. [32] proposed a new set of descriptors and investigated the performance of the RF algorithm to predict methane adsorption in nanoporous materials. Xie et al. [33] applied ML to predict the $CO_2$ adsorption capacity of porous carbon based on pressure, textural properties, and elemental compositions. The authors built four models — RF, Linear Regression (LR), Support Vector Regression (SVR), and MLP — and achieved the best accuracy using RF. Furthermore, they utilized SHAP to uncover the relationship between feature values and the $CO_2$ uptake. Brião et al. [34] compared statistical and ML approaches to predict the percentage of dysprosium removal from aqueous solution through the use of expanded vermiculite. They achieved high accuracy with both Artificial NNs trained with the Levenberg-Marquardt back-propagation algorithm and Adaptive Neuro-Fuzzy Inference Systems, demonstrating a better performance of ML over statistical methods. Yettou et al. [35] used Artificial NNs, Least Squared SVR, and regular SVR to model a ternary adsorption system of different heavy metal ions on several adsorbents. They also found that SVR performed slightly better, being more accurate, stable, and rapid than the other two algorithms.

Despite the numerous efforts of ML applied to predict the sorption properties of materials, no work has so far investigated the use of ML in combination with developed, fully operational devices. Here, we describe a novel technology consisting of a modular and compact unit with a large capacity for water sorption and low energy usage. The harvesting effect is obtained through the interplay of an advanced Ion Deposition Membrane (IDM) with a suitable hygroscopic carrier liquid, flowing within channels underneath the membrane. We designed multiple supports for the membrane, each having particular channel shape and dimensions, and tested their water harvesting capacity under various operating conditions of temperature, relative humidity, and carrier fluid flow rate. Furthermore, we extend the application of ML to AWH by investigating the performance of three ML techniques in modeling the harvesting capacity of the unit, with the aim of developing a proven tool to predict water collection rates by the fabricated prototype under prescribed operating conditions.

## 2. Materials and methods

### 2.1. Ion deposition membrane technology

The device analyzed in this study consists of a modular and compact AWH unit (also called "leaf") using a proprietary advanced IDM supplied by Biomass Energy Systems Inc. This advanced membrane technology enhances water adsorption efficiency by modifying the membrane surface through metal-ion implantation. The IDM is manufactured by rapid 3D expansion of Polytetrafluoroethylene (PTFE) microtubes, followed by ion-beam treatment at 50–100 keV, depositing metal ions at a depth of a few nanometers. This process lowers the water vapor saturation pressure at the gas–membrane interface, significantly improving water uptake — particularly under low relative humidity (RH) conditions — by up to a factor of four compared to untreated membranes. The key features associated with this technology are its energy and capital efficiency, as well as its small ecological footprint, compactness, and affordability. These peculiarities make this technology a competitive alternative to the much easier (yet highly energy demanding) condensation processes, and are the main reasons behind its development and implementation. The required power can be supplied using waste heat or traditional small electric generators, and also using renewable sources, such as solar or gravitational energy. This technology is designed to work under any ambient condition, thus being most suited for remote and arid environments. The idea behind the IDM technology is the implementation of a high-surface-area membrane treated with ionic deposition in combination with a hygroscopic carrier liquid for the capture of water from the atmosphere. A conceptual flow diagram of the process is shown in Fig. 1. Water laden atmospheric air (A) passes over the membrane (B) surface. The membrane together with the carrier (C) form a strongly hydrophilic adsorption system that transfers the atmospheric water molecules through the membrane pores onto the carrier, which is then transferred to the desorber system (D), where the water molecules (E) are removed and the carrier is returned to the sorption system.

The most crucial part of this process is clearly the membrane-carrier boundary, whose properties have the strongest influence on the sorption capacity of the system. The IDM material is inert to water, robust, and flexible enough to meet the requirements of chemical stability, wettability, porosity, and structural integrity. The pore size is below 10 μm as seen in the microscopic structure of the membrane depicted in Fig. S1 in the Supplementary information.

The membrane pore size plays a crucial role in the water adsorption process. As water molecules are much smaller than most contaminants carried in the air, a micropore-sized membrane ensures selective transfer of water molecules. By design, this is further aided by the ion deposition

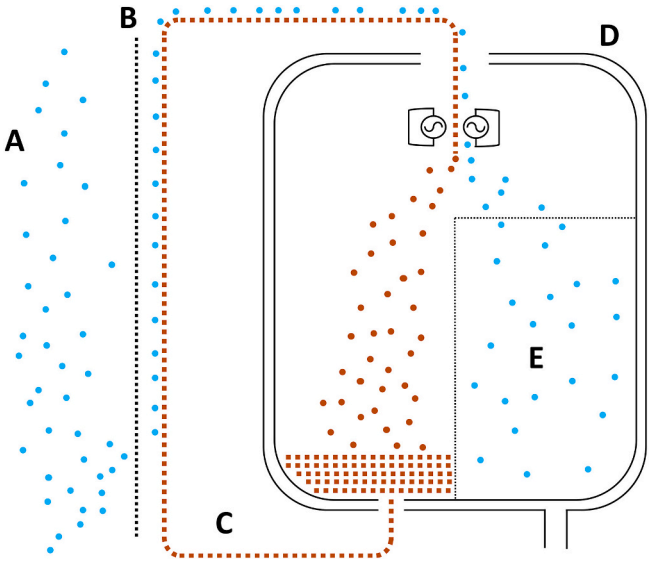


**Fig. 1.** Flow Diagram of the IDM Process. Atmospheric air with natural humidity (A) passes over the membrane surface (B). Atmospheric water is transferred through the pores of the membrane to the carrier fluid (C), which is then entering the desorber system (D), where the water molecules (E) are removed, and the carrier fluid is returned to the sorption system, thus completing the harvest cycle.

treatment on the membrane. Upon passing through the pores, the water remains clean and sterile, while dust and particulates stay on the side of the membrane exposed to the vapor-containing ambient atmosphere. For sustained adsorption to occur, the membrane is subjected to a continuous airflow to facilitate the supply of sufficient moisture content in its vicinity.

The carrier liquid properties are equally critical in facilitating water adsorption and transport. In IDM technology, the carrier liquid does not form strong chemical bonds with water molecules but instead functions as a transport medium. The membrane's inner surface, which comes into contact with the carrier liquid, is pre-treated with ion deposition, enhancing its interaction with the carrier. When the carrier liquid contacts the deposited metal ions, localized modifications to its crystalline structure occur, temporarily reducing intermolecular forces and allowing the liquid to adsorb water molecules via an exothermic process. This mechanism enables efficient water transfer from the membrane into the carrier, optimizing overall system performance.

### 2.2. Experimental setup and data preparation

The only chemical used in this work was phosphoric acid (85 % Solution in water, Lab Grade) sourced from Lab Alley. The liquid was used as a carrier fluid to adsorb the water molecules from the environment at the membrane-carrier interface, and transport them towards the collection vessel. A peristaltic pump (DIP Lab Peristaltic Pump by Kamoer) drew the carrier from the reservoir and moved it through the channels of the lab-scale prototype (LSP) leaf composed of a proprietary IDM, hot pressed on a 3D printed base, all manufactured by Dewcon Instruments Inc. The outlet of the system drained into a beaker, where the mixture of carrier and water was collected. During the experiments, the leaf was kept inside an environmental chamber (Model HCP50, Memmert GmbH + Co. KG) that maintained the temperature and RH at fixed set values (with an accuracy of ±0.1 °C and ± 0.1 % RH). A flow diagram for the LSP layout is shown in Fig. 2.

A typical test campaign, consisting of four consecutive experiments, started with the priming of the system. 50 ml of carrier fluid was pumped through the leaf to fill the channels and the tubes connecting it to the pump (upstream) and the collection vessel (downstream). At the

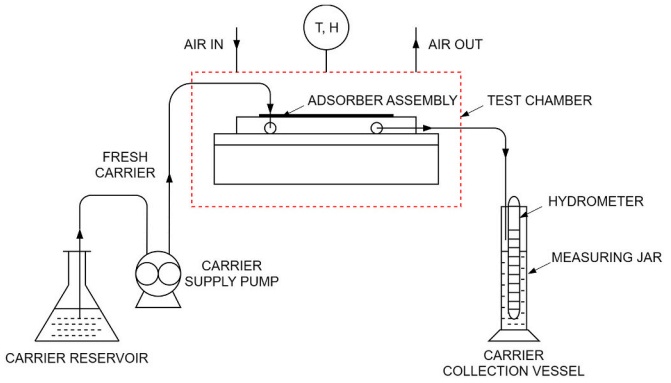


**Fig. 2.** Flow diagram for the lab-scale prototype used in this work. The main components are shown along with the flow path of the carrier fluid, which takes up and then releases the collected water.

same time, the temperature was set in the chamber, but the RH control remained off. Once the temperature reached the set point, the RH was set to the desired level. After the chamber reached the desired atmospheric conditions, the first experiment was started.

The carrier density values at the inlet and the outlet of the adsorber offer an indication of the water mass taken in. Since the densities of water and the carrier fluid are 0.997 g $ml^{-1}$ and 1.684 g $ml^{-1}$, respectively, water adsorption would result in a decrease of the density of the carrier at the outlet. The experiment started by calculating the density of fresh carrier fluid. The weight of a defined volume of fresh carrier was calculated by taking the difference in weight between an empty graduated cylinder and the same graduated cylinder filled with 100 ml carrier (accuracy ±1 ml). This measurement produced the density of the carrier fluid. For cross-verification, the same density was also measured directly using a hydrometer with an accuracy ±0.01 g $ml^{-1}$. In a typical test, a peristaltic pump was used to circulate 100 ml of carrier from a carrier reservoir to the entire system of channels and tubes, which was then collected on the other end in a carrier collection vessel. Once the pump was stopped, the weight and volume of the carrier-water mixture inside the collection vessel were determined, providing the density of the collected mixture and thus, the amount of water in the mixture.

Each experiment was run under different conditions, prescribed by three relevant parameters, namely temperature, RH, and carrier fluid flow rate. The values chosen for these quantities are listed in Tables S1 and S2 in the Supplementary information. Air flow around the unit was another operating parameter whose effect was assessed. In order to do so, an electric fan was mounted inside the chamber, which kept a constant air flow rate. At the end, the variations in the amount of water collected under different air flow rate conditions were too small to discern, thus the air flow rate was not considered as a variable for the experiments.

### 2.3. Membrane leaf designs

Different leaf designs (7 overall) were tested to assess the influence of the channel geometry on harvesting performance. The cross-sectional drawings for each design are reported in the Supplementary information. For Designs 1–5, the cross-sectional shape of the channels was trapezoidal. These designs differ from the benchmark in terms of either depth or width, to assess the influence of one dimension when the other was fixed or when the channel cross-sectional area remained the same. For the remaining two designs (6 and 7), different shapes were chosen: triangular (isosceles) for Design 6, and trapezoidal with a semicircle at the bottom for Design 7. The geometric characteristics of the designs are listed in Table 1. The tolerance for the channel geometric dimensions was ±0.3 mm. Although different leaf channel designs were tested, all of them featured the same surface roughness, and the effect of this

**Table 1**
Geometric dimensions of the leaf channels.

| Design | Top width (W) [mm] | Bottom width (w) [mm] | Depth (d) [mm] | Radius [mm] | Area [$mm^2$] |
|---|---|---|---|---|---|
| 1 | 2 | 1.7 | 1 | – | 1.85 |
| 2 | 1 | 0.85 | 2 | – | 1.85 |
| 3 | 2 | 1.7 | 1.5 | – | 2.775 |
| 4 | 2 | 1.7 | 2 | – | 3.7 |
| 5 | 1.5 | 1 | 1 | – | 1.25 |
| 6 | 2 | 0 | 1.85 | – | 1.85 |
| 7 | 2 | – | 1.25 | 0.85 | 3.45 |

characteristic on the harvesting capacity was not studied. However, it is reasonable to assume that the impact of said feature is negligible overall, given the small dimensions of the channels and the fact that the fluid inside each channel is always in creeping flow.

The benchmark design (Design 1) was tested under all the relevant possible combinations of temperature, RH and carrier flow rate, while the other designs were tested only under some limited conditions to investigate the effect of the channel geometry on the water harvesting process. In particular, we compared the performance of Designs 1 and 2 to assess how the volume of collected water varied when changing both the width and the depth of the channels, keeping the same cross-sectional area. Designs 3 and 4 were compared to understand the impact of channel depth over the sorption process, with both top and bottom widths being fixed. Similarly, Design 5 provided information about the effect of channel width, keeping the depth constant as Design 1. Experiments with Designs 6 and 7 were conducted to understand how the cross-sectional shape of the channels influenced the amount of water collected from the atmosphere. For Design 6, the cross-sectional area was kept constant as in Design 1; this was not possible for Design 7, as that would require a larger opening on the top, thus causing sagging of the membrane.

### 2.4. Machine learning methodology

Many studies have established how ML methods can help in predicting the sorption properties and capacities of various systems for which physics-based models are lacking or difficult to implement [25–30,32–35].

This study compares the stability and accuracy of the models obtained by implementing three ML regression techniques: SVR [36,37], GBR [38,39], and MLP [40,41]. All algorithms were chosen based on their recognized ability to describe nonlinear relationships, even when the dataset has few variables and few data points. SVR is robust to outliers, easy to implement, and can be generalized with high prediction accuracy. GBR is flexible and can be optimized on different loss functions. MLP provides quick predictions after training and excels at capturing non-linear relationships. The scope of the models is to predict the amount of water harvested based on the known parameters, namely operating temperature, RH, and carrier fluid flow rate, and the geometric characteristics of the channels (width at the top, at the bottom, and depth). The full list of variables used, along with their variability, is shown in Table 2. Furthermore, the complete experimental dataset used for training and testing the models is available in Zenodo [42].

To validate the trained models, we adopt the following process. The convention is to split a dataset once into two mutually exclusive sets, one for *training* and one for *testing*. Here, we chose to do it 50 times, each with a fixed random seed. In other words, we split the dataset 50 times, each time with 80 % of the data (122 data points) being used for training and 20 % for testing (30 data points), so the data points included in the training and testing sets are always different. This approach prevents any single train-test split from disproportionately influencing the results and ensures that model performance is not an artifact of a particular data partitioning (which can be a problem for small datasets, as is our case since the testing set only has 30 data points). Next, for each of these 50 splits, we performed independent hyperparameter (HP) optimization using a GridSearch (GS) process. Specifically, one GS was conducted per split, tuning the HPs of each algorithm separately. Rather than selecting the best HPs from a single GS, we treated each of the 50 GS results as a member of a committee, where the most frequently suggested optimal HP combination across all splits was chosen as the final set of HPs for each model. Once the optimal HPs were determined, we trained a separate model for each of the 50 dataset splits, ensuring that a total of 50 models were developed and evaluated independently. Each model was trained using k-fold cross-validation (CV) within the training set and subsequently tested on the corresponding test set. This process was repeated across all 50 splits, and the final model performance was assessed by computing the average and standard deviation of key evaluation metrics. Among the 50 trained models for each algorithm, the one achieving the best performance was saved and used to predict the water harvesting potential during a typical spring day in Jeddah, Saudi Arabia, demonstrating a practical application of the approach. All of the computing work was performed using the *Scikit-learn* package [43].

**Table 2**
Summary of experimental variables used.

| Variable type | Variable | Number of samples | Range | Mean |
|---|---|---|---|---|
| Independent variable ($x$) | Temperature [°C] | 152 | 25.7–36.6 | 31.9 |
| Independent variable ($x$) | Relative Humidity [%] | 152 | 29.2–81.2 | 55.2 |
| Independent variable ($x$) | Carrier flow rate [ml/min] | 152 | 1.45–3.13 | 2.3 |
| Independent variable ($x$) | Top width [mm] | 152 | 1.0–2.0 | 1.8 |
| Independent variable ($x$) | Bottom width [mm] | 152 | 0.0–1.7 | 1.4 |
| Independent variable ($x$) | Depth [mm] | 152 | 1.0–2.0 | 1.4 |
| Dependent variable ($y$) | Water harvested [ml] | 152 | 1.0–11.0 | 4.8 |

Because ML algorithms can be negatively affected by scale changes when dealing with experimental data — different value ranges may result in a biased assessment [44] — both input ($x$) and output ($y$) variables were scaled according to the common min-max scaling method

$$z_j = \frac{x_j - min(x)}{max(x) - min(x)} \tag{1}$$

where $z_j$ is the scaled value of the $j$-th data point, $x_j$ is the original value of the $j$-th data point, and min($x$) and max($x$) are the minimum and maximum value in each feature, respectively. The scaling was performed after splitting the dataset in training and testing sets, using parameters based on the former set. In this way, it was possible to avoid revealing testing set information to the training set prior to the training of the model and to prevent data snooping issues [45].

The models were evaluated on the basis of three common metrics: goodness-of-fit ($R^2$), root-mean-square error (RMSE), and mean absolute error (MAE). The variables in the following equations are the observed value of the $i$-th output ($y_i$), the predicted value for the $i$-th output ($\hat{y}_i$), and the average of the observed $y$ values ($\bar{y}$). The goodness-of-fit parameter

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{2}$$

quantifies the accuracy of the prediction by representing the proportion of variance of the dependent variable that has been explained by the independent variables in the model. The two other measures

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) \tag{3}$$

and

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{4}$$

are used to measure the deviation between real and predicted data, while having different approaches to outliers in the data [46].

## 3. Results and discussion

### 3.1. Experimental results

The tests were performed under controlled environmental and flow operating conditions, with set values for the temperature, RH and carrier flow rate. See Tables S1 and S2 in the Supplementary information for the actual values. The geometry of the channels for each design is reported in Table 1. During each test, the environmental chamber maintained the set-point conditions of temperature and RH, while 100 ml of carrier liquid were pumped from the carrier reservoir through the leaf, after which, the water laden carrier leaving the leaf was collected in a vessel. The volume collected at the outlet was measured and the variation from the initial value determined the volume of water harvested from the atmosphere. This final value was included in the dataset used to train the ML models.

Design 1 underwent a comprehensive experimental campaign, being tested under all the possible combinations of operating conditions: three values of temperature (25 °C, 30 °C, and 35 °C), six values of RH (30 % to 80 %, with increments of 10 %), and four values of carrier flow rate (1.5 ml/min, 2.0 ml/min, 2.5 ml/min, and 3.0 ml/min). This granted a general understanding of the water collection performance as a function of the three working parameters. The results are shown in Fig. 3.

The amount of water collected increased visibly with RH, and less sharply with temperature. This was expected, due to the larger amount of moisture present in the air in more humid conditions. Nonetheless, the variation of the harvested water is not linear, as seen in Fig. 3. This intrinsic non-linearity of the harvesting behavior is the very reason why ML techniques should be of value to this problem, in the absence of an accurate physics-based model.

An intriguing result was the trend vs. carrier flow rate, which showed that the water collection increased with increasing flow rate to a maximum value, and started to decline thereafter. This behavior was most probably related to a trade-off between the effect of residence time and a phenomenon analogous to concentration polarization [47–49]. This phenomenon is usually observed in membrane filtration processes, due to the accumulation of solutes, as well as any organic and inorganic impurity, in the proximity of the membrane surface. This creates a concentration gradient between the boundary layer and the bulk of the solution. The concentrated layer at the membrane surface increases the membrane resistance to permeate flow, consequently reducing the permeate flux through the membrane. In the particular case analyzed in this study, a similar — yet opposite — phenomenon appeared to be in play: the water adsorbed from the atmosphere accumulated on the carrier liquid at the interface with the membrane, forming a more dilute acid layer with weaker hygroscopic properties, thus making this layer less capable of attracting more water molecules from the air. Simultaneously, mass transfer mechanisms acted to remove the accumulated water from the vicinity of the membrane surface. Mass transfer of water through the carrier liquid occurred via diffusion (independent of the velocity) and convection (enhanced at higher velocities), since there was no turbulence in the flow, as can be seen from the values of the Reynolds number with carrier flow rate, see Table S3 in the Supplementary information. It is possible to calculate the diffusion coefficient $D$ of water in the carrier fluid according to the analytical relation proposed by Wilke and Chang [50] for unassociated liquids

$$D = 7.4 \times 10^{-8}\frac{TM^{0.5}}{\mu V^{0.6}} \tag{5}$$

which can then be used to evaluate the Sherwood number [51] as a function of the carrier flow rate. As seen from the Sh values in Table S3 in the Supplementary information, the convective mass transfer prevailed over diffusive transport. The competition between the water accumulation under the membrane and the mass transfer away from this layer resulted in the consistent trend of absorbed water vs. carrier flow rate.

Leaves with different channel cross-sections were tested in order to compare the results with the benchmark (Design 1), and correlate the geometric characteristics of the channels to the water collection capacity of the leaf. The geometry of each design is described by the parameters listed in Table 1. As it was not practical to perform tests for all ambient conditions with each leaf design, only two temperatures (30 °C and 35 °C) and two RH (40 % and 70 %) were tested. Hence, a total of 16 combinations for each design were examined, and the results were included in the same dataset with Design 1. The empirical results obtained from the tests on Design 7 were not incorporated in the dataset used to train the ML models, since the different cross-section for this design could not be included as a feature in a regression problem. Empirical results obtained from the tests on Design 6 were included in the training dataset because, although the design had a different cross-section, it could be described by the chosen geometric features (w = 0). Graphs that compare the performance of Designs 2 through 7 against Design 1 for one set of ambient conditions (30 °C and 70 % RH) are shown in Fig. 4.

The scope of testing Design 2 was to infer a qualitative relationship between the top width of the channels — i.e., the one exposed to the membrane — and the water harvesting capacity of the leaf. In order to do so, the cross-section area was kept fixed at 1.85mm$^2$, the top and

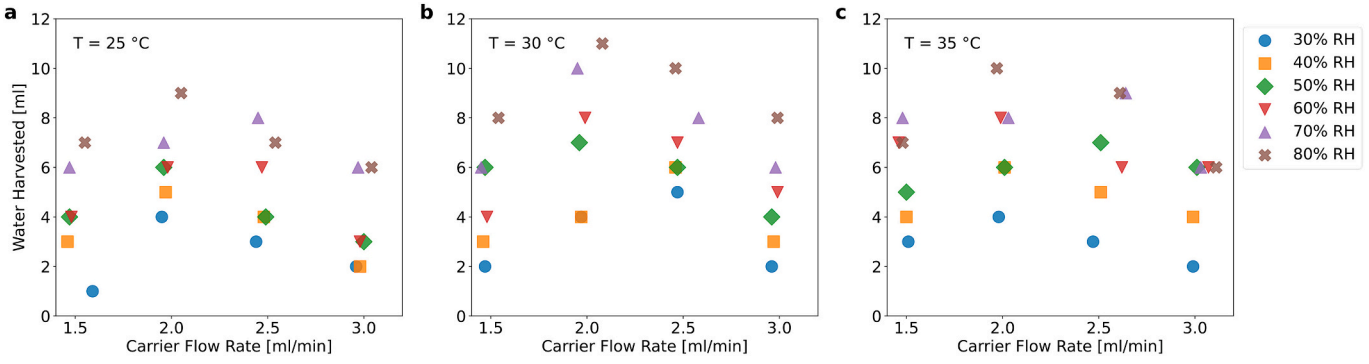


**Fig. 3.** Volume of water [ml] collected by Design 1 leaves when circulating 100 ml of carrier fluid at various ambient temperatures, relative humidities, and carrier flow rates. Ambient (a) 25 °C; (b) 30 °C; (c) 35 °C.

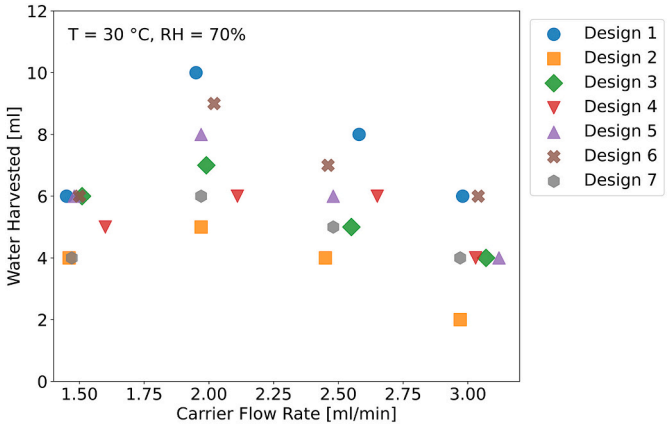


**Fig. 4.** Cross comparison of Designs 2–7 (against Design 1) by circulating 100 ml of carrier fluid under the same ambient conditions (30 °C and 70 % RH) and varying flow rates [ml/min]. The plot displays the amounts of water harvested [ml] for each case to infer the effect of channel geometry characteristics. Design 2 was considered to infer the effect of the width/depth ratio. Designs 3 and 4 were tested to infer the effect of channel depth. Design 5 was explored to infer the effect of channel width. Designs 6 and 7 were studied to examine the effect of channel cross-section shape.

bottom width were decreased to 1 and 0.85 mm respectively, and the depth was increased to 2 mm. This change implied that the same velocity corresponded to the same flow rate in both Designs 1 and 2, so that there was no effect of carrier fluid velocity variations. It is immediately observed that a reduction in the top width of the channel — and thus in exposed surface area of carrier fluid — caused a ∼ 50% drop in the amount of water collected.

Designs 3 and 4 were tested in order to assess the influence of the channel depth over the water harvesting capacity of the leaf. As already reported in Table 1, Design 3 featured a depth of 1.5 mm, while for Design 4 the depth was 2 mm, vs. the 1 mm depth of Design 1.

The behavior for various operating parameters (temperature, RH, and flow rate) can be explained by the same logic used for Designs 1 and 2. The main effect of the increased depth (and thus of the increased cross-section area) was a decreased flow velocity when the flow rate remained constant (again 1.5 ml/min, 2.0 ml/min, 2.5 ml/min, and 3.0 ml/min). The reduction in water collection from Design 1 to Designs 3 and 4 could be explained by the dilution of carrier fluid due to water adsorption in the upper layer. As already stated, decreasing the flow velocity implied a less effective removal of this diluted layer, which in turn hindered the attraction of more molecules from the atmosphere.

The tests on Design 5 were particularly interesting, as they allowed to compare the effects of various changes on the leaf harvesting capacity. Design 5 differed from Design 1 both in the top and bottom width, as well as the cross-sectional area: the top and bottom widths were reduced to 1.5 mm and 1 mm, respectively, while the depth was kept constant at 1 mm. These changes had three primary effects on the performance of the leaf: first, the exposed surface area of the carrier fluid — and thus the active area of the membrane — was reduced, while the higher fluid velocity entailed both a lower residence time and an increased convective mass transfer, thus reducing the negative impact of concentration polarization. It can be seen that the harvesting capacity of Design 5 was significantly lower than the benchmark design.

The objective with Designs 6 and 7 was to understand whether, and possibly how, a different shape for the cross-section of the channels would affect the harvesting capacity of the device. In particular, the shape chosen for Design 6 was an isosceles triangle with the base coincident with the channel aperture, 2 mm wide and 1.85 mm deep. In this way, both the active surface area and the cross-sectional area were kept constant (as in Design 1), so there were no effects related to changes in the velocity of the carrier fluid or channel width under the membrane.

It can be seen that the results from Design 6 are comparable with those obtained with Design 1. This could be expected, as all relevant geometrical features were the same, and only the shape of the cross-section varied. In theory, the different shapes could affect the performance of the leaf by changing the pressure drop due to friction inside the channels, as the cross-section in Design 6 had a larger perimeter (6.21 mm) than Design 1 (5.72 mm), which should result in higher frictional losses. However, the fluid was in creeping flow, and the impact of such a geometric change was trivial. The situation with Design 7 is different: as the cross section is larger than in Design 1, similarly to what happened with Designs 3 and 4, a larger area implied a reduction in the flow velocity for the same flow rate and thus, weaker removal of the diluted top layer. Therefore, the harvesting capacity of Design 7 was lower than Design 1.

### 3.2. Machine learning results

The experimental results were assembled in a dataset and used to train and test the ML models. These models were obtained using algorithms that can find highly nonlinear correlations between the input parameters (temperature, RH, carrier fluid flow rate, channel width at top and bottom, and channel depth), and the output value, namely the volume of water collected.

The dataset was split into 50 different and mutually exclusive couples of training and testing sets, and a model was developed for each couple. This was done to assess the stability of each technique, that is, the degree of independence between predictive performance and the way the dataset is split into subsets.

The first step in the development of the models was the HP tuning through grid-search (GS). The HPs searched upon for each algorithm are reported in Table 3, along with the values that form the best combinations, i.e., higher CV scores. Once these optimal combinations were determined, one model for each algorithm was trained over the training set and validated through *k*-fold CV. After this step, the model was used to predict the outputs of the testing set. This procedure was repeated 50 times, each time using a different pair of training and testing set. During each iteration, the CV score and the accuracy score $R^2$, as well as the MAE and RMSE, were recorded in arrays, so that the scores for each iteration could be accessed. The mean and standard deviation values of the CV score, $R^2$, MAE, and RMSE were calculated for each algorithm. These values were compared to determine which algorithm achieved the best performance, taking into account as performance metrics the accuracy in making predictions, the stability in CV, and the independence from the characteristics of the training set.

Figure 5(a–c) shows the best performance of the three ML techniques in predicting the amount of water harvested by the system. The plots represent the one instance out of fifty in which each algorithm achieved the best combination of CV score and test $R^2$ (meaning the highest values simultaneously, although clearly there might be instances in which one of the two measures was higher). The axes of the plots are the actual

**Table 3**
Best combination of hyper-parameters (HPs) for each model, tuned through a grid-search optimization.

| Algorithm | Hyper-parameter | Value |
|---|---|---|
| SVR | Kernel | *rbf* |
| | $\varepsilon$ | 0.007 |
| | C | 0.5 |
| | $\gamma$ | *scale* |
| GBR | Criterion | *friedman mse* |
| | Learning rate | 0.085 |
| | Estimators | 200 |
| | Max depth | 2 |
| MLP | Activation function | *logistic* |
| | Neurons in hidden layer | 6 |
| | Solver | *lbfgs* |

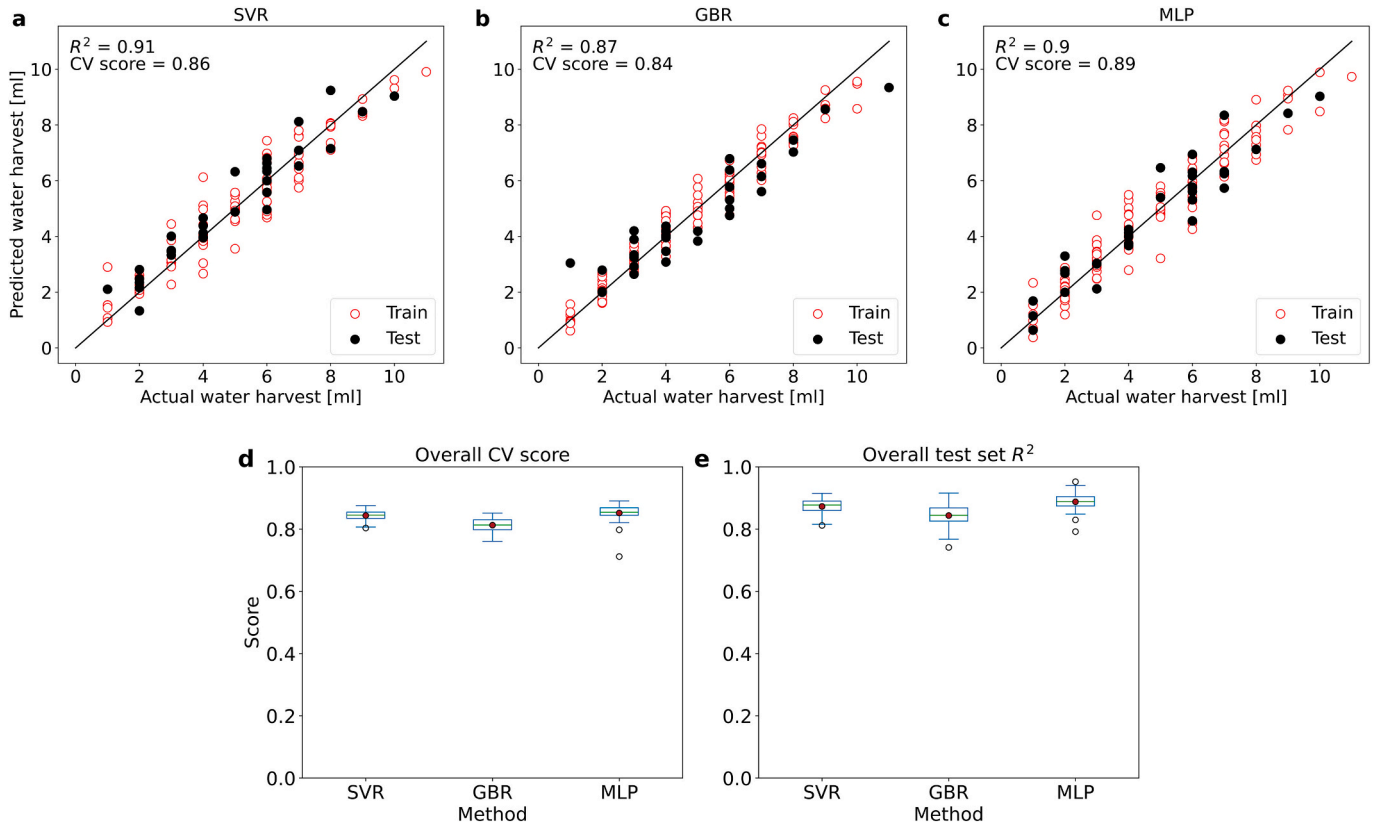


**Fig. 5.** (a–c) Prediction of harvested water volume [ml] at the best instances (highest CV score and test $R^2$ values) for (a) SVR, (b) GBR, and (c) MLP. The black diagonal (slope equal to 1) serves as a reference for the perfect prediction. (d, e) Overall performance evaluation of SVR, GBR and MLP. Red circles mark the mean values, while green lines mark the median. Whiskers extend to 1.5× the inter-quartile range.

amount of water harvested and the amount predicted by the individual models, in volume [ml], and each point in the plane is a sample from the dataset. The predictions for both the training and testing sets are reported. As seen from these graphs, all the algorithms produced models that performed well. Looking at the predictions on the training set samples, the scattering around the reference line for the perfect prediction suggests that all models were immune to over-fitting. Similarly, the performance of the test set was not optimal and all models made errors.

We further analyzed both predictive accuracy and stability across different dataset splits. This provides insights into how each algorithm responds to variations in the training and testing sets, which is crucial for ensuring robustness. The exact values for the average and standard deviation of CV score and $R^2$ for each method, along with the respective MAE and RMSE values, are reported in Table 4. CV and $R^2$ scores are also shown in Fig. 5(d–e). Each data point in the box plot is the score of a model on one distinct split of train/test set, and data points outside 1.5× the interquartile range are shown as outliers. Noticeably, all three methods performed fairly well, meaning that all three are suitable for the data being studied.

Despite their overall strong performance, some differences between the models emerged. The average CV scores were similar, ranging from 0.81 to 0.85, with SVR and MLP performing better than GBR. Among the three, MLP achieved the highest average CV score, indicating stronger generalization capabilities. When evaluating testing accuracy, all three methods exhibited remarkable scores – higher than 0.84 and reaching up to ∼ 0.89 – with MLP again demonstrating the best overall accuracy. However, SVR appeared to be the most stable method, exhibiting lower deviations in both CV and testing accuracy. Additionally, GBR displayed a slightly higher variance in test performance compared to SVR and MLP, suggesting that GBR might be more sensitive to the specific characteristics of the training set. This variability could impact reliability when applied to new data. In contrast, SVR maintained more consistent accuracy across different splits, making it a robust choice despite not achieving the absolute highest accuracy in some instances. Overall, while MLP demonstrated the highest predictive accuracy, SVR stood out as the most stable across dataset variations, and GBR, despite its slight variability, remained a viable option. This comparison underscores the importance of balancing accuracy and stability when selecting ML models for this application.

**Table 4**
Metrics for evaluation of tested ML methods.

| Algorithm | CV score | $\sigma_{CV}$ | $R^2$ | $\sigma_{R^2}$ | MAE [ml] | RMSE [ml] |
|---|---|---|---|---|---|---|
| SVR | 0.844 | 0.016 | 0.873 | 0.025 | 0.651 | 0.815 |
| GBR | 0.813 | 0.021 | 0.844 | 0.036 | 0.723 | 0.902 |
| MLP | 0.852 | 0.027 | 0.888 | 0.026 | 0.615 | 0.764 |

### 3.3. Simulation of real world application

Given the high accuracy achieved by the three ML models used here, one could think of applying such methods first to design, and then to regulate the operation of a real commercial scale plant, composed of multiple units continuously running and harvesting water from the atmosphere. Having access to reliable models, which can predict the amount of water collected under certain weather conditions, makes it easier to evaluate the feasibility of installing a plant, rather than using more accurate but also computationally expensive physics-based models. A demonstration of how such ML models can be utilized in this context is given below. The best performing model (MLP) was used to predict the water harvesting potential during a typical spring day in Jeddah, Saudi Arabia, a representative area facing severe water scarcity.

Considering a plant made of Design 1 units, the carrier flow rate was set to 2 ml/min, as this value was experimentally found to correspond to maximum water collection from the atmosphere. The temperature and RH values were taken from an online weather repository [52]. Figure 6 shows the temporal evolution of the predicted water volume [ml] collected throughout the day for a single lab-scale unit, against the time-varying temperature and RH. Figure 6 reveals the strong dependence of the Water Harvesting Rate (WHR) on RH. The rate showed three clear trends throughout the day. In the early hours, the RH was higher and so was the WHR. During the morning and the following hours, the RH dropped and the volume of water collected increased only by a small amount; eventually the rate increased again as the RH began rising in the late afternoon. Although this prediction seems reasonable, it is important to point out that in some cases, the operating conditions (temperature and RH) were outside the range of the data in the training set. This caused the prediction to be not completely backed up by data, as the model had no precise knowledge about these regions of the predictor space. To address this problem, tests should be run in the lower and higher ranges of temperature and RH values.

As seen in Fig. 6, the yield was approximately 135 ml/day (WHR = 8.44 kg/m$^2$/day). However, these results do not take into account the desorption process needed to separate the freshwater from the carrier fluid. Naturally, this is of paramount importance in determining the overall performance of the technology; in order for the approach to be competitive, the desorption process should have both high yield and low energy usage. Nonetheless, the harvesting capacity itself is still remarkable.

## 4. Conclusions and final remarks

This work investigated the water yield of a membrane-based water harvesting device tested over a wide range of operating conditions maintained in a controlled ambient environment. The experimental dataset thus obtained was used to build and evaluate regression models for predicting the amount of water harvested under these conditions. The benchmark prototype design was tested over all possible combinations of four distinct hygroscopic fluid flow rates, three ambient temperatures and six relative humidity values. The prototype achieved WHR as high as 22.13 kg/m$^2$/day for atmospheric temperature 30 °C, RH 80 %, and carrier flow rate 2.5 ml/min. Different structural designs of the membrane unit were tested under a reduced subset of conditions. The results revealed a nonlinear relationship between the amount of water collected and the carrier fluid flow rate, which held true for all the designs and all temperature and RH conditions. This behavior was attributed to a trade-off between the residence time of the hygroscopic fluid inside the unit and the adverse effect of concentration polarization.

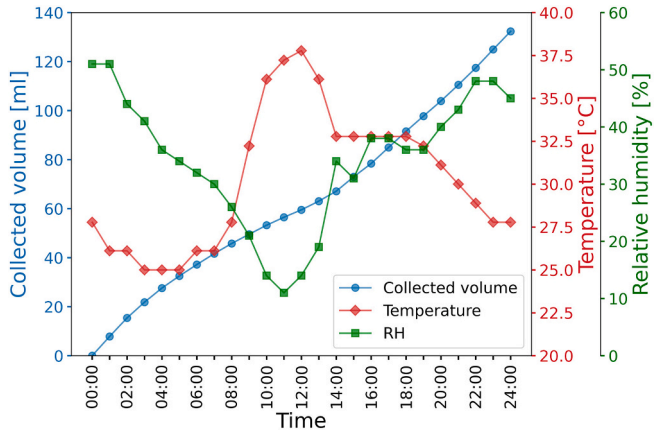


**Fig. 6.** ML-based prediction of the collected water volume using the present membrane approach during a typical spring day in Jeddah, Saudi Arabia [52]. Both ambient temperature and RH are plotted through the day.

Benchmarking our AWH system against existing technologies reveals competitive performance. The current system achieves an average water collection rate of 9.69 kg/m$^2$/day, with a peak performance of 22.13 kg/m$^2$/day. When compared to sorption-based AWH technologies, Ji et al. [53] and Wang et al. [54] achieved > 1.2 kg/m$^2$/day and 2.25 kg/m$^2$/day, respectively. These values suggest that our system operates within the upper range of sorption-based methods. Similarly, conventional membrane-based technologies have reported lower collection rates, with Bergmair et al. [23] achieving 1.8 kg/m$^2$/day, and Zhao et al. [55] reporting a WHR of 1.27 kg/m$^2$/day. Taking into account an estimated combined total loss of 15 % of harvested water due to carrier tubing and the desorption process, these results indicate that our system exceeds the performance of existing AWH technologies in terms of water collection efficiency.

The scalability and cost-effectiveness of the proposed system depend on both the manufacturing process of the IDM leaves and the efficiency of their layout within the system. Based on the average water harvesting rate of 9.69 kg/m$^2$/day, and accounting for an estimated 15 % loss, the required adsorption area for a water production capacity of 100 l/day should be 12.2 m$^2$. However, the required adsorption area may vary depending on the relative humidity conditions at the deployment site. To enhance cost-effectiveness, mass production of IDM leaves through injection molding is a viable approach, particularly for large-capacity systems (50–100 l/day). The current leaf design can be readily adapted for injection molding, allowing for efficient manufacturing, while maintaining a suitable size for transportation and handling. For a fixed leaf size, the required adsorption area can be attained by using a number of leaves, and double-sided leaves (with flow channels and membrane bonding on both sides) could be utilized to further reduce the system footprint. Additionally, an optimized leaf layout, following established industry standards, can minimize space requirements and improve overall cost efficiency.

The experimental data consisted of six input parameters, namely, ambient temperature and RH, carrier fluid flow rate, channel top width, bottom width, and depth, and one output value, the volume of water harvested from each experiment. To predict the performance of the harvesting device, three ML techniques were employed, SVR, GBR and MLP. The performance of each model was evaluated by comparing the average cross-validation score and the prediction accuracy over 50 different and mutually exclusive couples of training and testing sets. All three methods displayed high CV (0.81–0.85) and prediction accuracy $R^2$ (0.84–0.89) scores, and showed stable performance with low variation in the accuracy metrics over the multiple iterations. MLP had the highest average CV score (0.852) and prediction accuracy (0.888), followed by SVR (0.844, 0.873) and GBR (0.813, 0.844). These results show the benefit of employing ML techniques to reduce the burden of time and resource-intensive testing.

In summary, this study identified reliable ML models based on experimental data to predict the volume of water that could be harvested from an AWH unit under specific operating conditions. This knowledge and the ability to accurately predict the water yield make ML techniques a useful resource in both the design and the operation of commercial plants, providing a reliable tool to forecast the amount of water the plant could harvest at any time throughout the year from available weather data.

## CRediT authorship contribution statement

**Giulio Barletta:** Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Shashwata Moitra:** Writing – review & editing, Methodology, Data curation. **Sybil Derrible:** Writing – review & editing, Software. **Alex Mathew:** Methodology. **Anoop Muraleedharan Nair:** Methodology. **Constantine M. Megaridis:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Code availability

The codes used to obtain the models and the results of this study are publicly available on github at https://github.com/giuliobarl/MLWaterHarvesting.

## 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.

## Acknowledgements

G.B. thanks Professor Eliodoro Chiavazzo and Dr. Giovanni Trezza of the Politecnico di Torino multi-Scale ModeLing Laboratory, and Peter Nikurs and Ivan Lee of the UIC Micro/Nanoscale Fluid Transport Laboratory for their assistance during the study. Biomass Energy Systems Inc. provided funding for this work. Biomass Energy Systems Inc. and Dewcon Instruments Inc. provided the IDM units.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jwpe.2025.107476.

## Data availability

Processed datasets and trained models of this study are publicly available in Zenodo at DOI: https://doi.org/10.5281/zenodo.10533012 [42].

## References

[1] M.M. Mekonnen, A.Y. Hoekstra, Four billion people facing severe water scarcity, Sci. Adv. 2 (2016) e1500323, https://doi.org/10.1126/sciadv.1500323.
[2] Y. Tu, R. Wang, Y. Zhang, J. Wang, Progress and expectation of atmospheric water harvesting, Joule 2 (2018) 1452–1475, https://doi.org/10.1016/J.JOULE.2018.07.015.
[3] UNESCO World Water Assessment Programme, World Water Development Report 2018: Nature-based Solutions for Water. Tech. Rep, UNESCO, Paris, 2018.
[4] M. Ejeian, R.Z. Wang, Adsorption-based atmospheric water harvesting, Joule 5 (2021) 1678–1703, https://doi.org/10.1016/J.JOULE.2021.04.005.
[5] E.H. Oelkers, J.G. Hering, C. Zhu, Water: is there a global crisis? Elements 7 (2011) 157–162, https://doi.org/10.2113/GSELEMENTS.7.3.157.
[6] L. Wang, P. D'Odorico, J.P. Evans, D.J. Eldridge, M.F. McCabe, K.K. Caylor, et al., Dryland ecohydrology and climate change: critical issues and technical advances, Hydrol. Earth Syst. Sci. 16 (2012) 2585–2603, https://doi.org/10.5194/HESS-16-2585-2012.
[7] F. Macedonio, E. Drioli, A.A. Gusev, A. Bardow, R. Semiat, M. Kurihara, Efficient technologies for worldwide clean water supply, Chem. Eng. Process. Process Intensif. 51 (2012) 2–17, https://doi.org/10.1016/J.CEP.2011.09.011.
[8] K. Yang, T. Pan, Q. Lei, X. Dong, Q. Cheng, Y. Han, A roadmap to sorption-based atmospheric water harvesting: from molecular sorption mechanism to sorbent design and system optimization, Environ. Sci. Technol. 55 (2021) 6542–6560, https://doi.org/10.1021/ACS.EST.1C00257.
[9] B.A. Habeebullah, Potential use of evaporator coils for water extraction in hot and humid areas, Desalination 237 (2009) 330–345, https://doi.org/10.1016/j.desal.2008.01.025.
[10] A. Magrini, L. Cattani, M. Cartesegna, L. Magnani, Production of water from the air: the environmental sustainability of air-conditioning systems through a more intelligent use of resources. the advantages of an integrated system, Energy Procedia 78 (2015) 1153–1158, https://doi.org/10.1016/j.egypro.2015.11.081.
[11] M. Jradi, N. Ghaddar, K. Ghali, Experimental and theoretical study of an integrated thermoelectric–photovoltaic system for air dehumidification and fresh water production, Int. J. Energy Res. 36 (2012) 963–974, https://doi.org/10.1002/er.1848.
[12] M. Eslami, F. Tajeddini, N. Etaati, Thermal analysis and optimization of a system for water harvesting from humid air using thermoelectric coolers, Energy Convers. Manag. 174 (2018) 417–429, https://doi.org/10.1016/J.ENCONMAN.2018.08.045.
[13] A. Scrivani, U. Bardi, A study of the use of solar concentrating plants for the atmospheric water vapour extraction from ambient air in the Middle East and Northern Africa region, Desalination 220 (2008) 592–599, https://doi.org/10.1016/J.DESAL.2007.04.060.
[14] N.I. Ibrahim, F.A. Al-Sulaiman, R. Saidur, Performance assessment of water production from solar cooling system in humid climate, Energy Convers. Manag. 127 (2016) 647–655, https://doi.org/10.1016/J.ENCONMAN.2016.09.056.
[15] H. Kim, S. Yang, S.R. Rao, S. Narayanan, E.A. Kapustin, H. Furukawa, et al., Water harvesting from air with metal-organic frameworks powered by natural sunlight, Science 356 (2017) 430–434, https://doi.org/10.1126/SCIENCE.AAM8743.
[16] J.Y. Wang, J.Y. Liu, R.Z. Wang, L.W. Wang, Experimental research of composite solid sorbents for fresh water production driven by solar energy, Appl. Therm. Eng. 121 (2017) 941–950, https://doi.org/10.1016/J.APPLTHERMALENG.2017.04.161.
[17] S. Guo, Y. Zhang, S.C. Tan, Device design and optimization of sorption-based atmospheric water harvesters, Device 1 (2023), https://doi.org/10.1016/j.device.2023.100099.
[18] D.K. Nandakumar, Y. Zhang, S.K. Ravi, N. Guo, C. Zhang, S.C. Tan, Solar energy triggered clean water harvesting from humid air existing above sea surface enabled by a hydrogel with ultrahigh hygroscopicity, Adv. Mater. 31 (2019) 1806730, https://doi.org/10.1002/adma.201806730.
[19] J.S. Solís-Chaves, C.M. Rocha-Osorio, A.L. Murari, V.M. Lira, A.J.S. Filho, Extracting potable water from humid air plus electric wind generation: a possible application for a brazilian prototype, Renew. Energy 121 (2018) 102–115, https://doi.org/10.1016/J.RENENE.2017.12.039.
[20] B.A. Kashiwa, C.B. Kashiwa, The solar cyclone: a solar chimney for harvesting atmospheric water, Energy 33 (2008) 331–339, https://doi.org/10.1016/J.ENERGY.2007.06.003.
[21] T. Ming, T. Gong, R.K. de Richter, Y. Wu, W. Liu, A moist air condensing device for sustainable energy production and water generation, Energy Convers. Manag. 138 (2017) 638–650, https://doi.org/10.1016/J.ENCONMAN.2017.02.012.
[22] D. Milani, A. Qadir, A. Vassallo, M. Chiesa, A. Abbas, Experimentally validated model for atmospheric water generation using a solar assisted desiccant dehumidification system, Energy Build. 77 (2014) 236–246, https://doi.org/10.1016/J.ENBUILD.2014.03.041.
[23] D. Bergmair, S.J. Metz, H.C.D. Lange, A.A. van Steenhoven, System analysis of membrane facilitated water generation from air humidity, Desalination 339 (2014) 26–33, https://doi.org/10.1016/J.DESAL.2014.02.007.
[24] R. Tu, Y. Hwang, Performance analyses of a new system for water harvesting from moist air that combines multi-stage desiccant wheels and vapor compression cycles, Energy Convers. Manag. 198 (2019) 111811, https://doi.org/10.1016/j.enconman.2019.111811.
[25] K. Zhang, S. Zhong, H. Zhang, Predicting aqueous adsorption of organic compounds onto biochars, carbon nanotubes, granular activated carbons, and resins with machine learning, Environ. Sci. Technol. 2020 (2020) 7008–7018, https://doi.org/10.1021/acs.est.0c02526.
[26] G. Panapitiya, G. Avendano-Franco, P. Ren, X. Wen, Y. Li, J.P. Lewis, Machine-learning prediction of co adsorption in thiolated, Ag-alloyed Au nanoclusters, J. Am. Chem. Soc. 140 (2018) 17508–17514, https://doi.org/10.1021/JACS.8B08800.
[27] X. Zhu, D.C. Tsang, L. Wang, Z. Su, D. Hou, L. Li, et al., Machine learning exploration of the critical factors for $CO_2$ adsorption capacity on porous carbon materials at different pressures, J. Clean. Prod. 273 (2020) 122915, https://doi.org/10.1016/J.JCLEPRO.2020.122915.
[28] G. Trezza, L. Bergamasco, M. Fasano, E. Chiavazzo, Minimal crystallographic descriptors of sorption properties in hypothetical mofs and role in sequential learning optimization, npj Comput. Mater. 8 (2022) 123, https://doi.org/10.1038/s41524-022-00806-7.
[29] W. Guo, J. Liu, F. Dong, R. Chen, J. Das, W. Ge, et al., Deep learning models for predicting gas adsorption capacity of nanomaterials, Nanomaterials 12 (2022) 3376, https://doi.org/10.3390/NANO12193376/S1.
[30] X. Zhang, Q.R. Zheng, H.Z. He, Machine-learning-based prediction of hydrogen adsorption capacity at varied temperatures and pressures for MOFs adsorbents, J. Taiwan Inst. Chem. Eng. 138 (2022) 104479, https://doi.org/10.1016/J.JTICE.2022.104479.
[31] L. Li, Z. Shi, H. Liang, J. Liu, Z. Qiao, Machine learning-assisted computational screening of metal-organic frameworks for atmospheric water harvesting, Nanomaterials 12 (2022) 159, https://doi.org/10.3390/NANO12010159/S1.
[32] G.S. Fanourgakis, K. Gkagkas, E. Tylianakis, E. Klontzas, G. Froudakis, A robust machine learning algorithm for the prediction of methane adsorption in nanoporous materials, J. Phys. Chem. A 123 (2019) 6080–6087, https://doi.org/10.1021/ACS.JPCA.9B03290.
[33] C. Xie, Y. Xie, C. Zhang, H. Dong, L. Zhang, Explainable machine learning for carbon dioxide adsorption on porous carbon, J. Environ. Chem. Eng. 11 (2023) 109053, https://doi.org/10.1016/J.JECE.2022.109053.
[34] A. Yettou, M. Laidi, A. El Bey, S. Hanini, M. Hentabli, O. Khaldi, et al., Ternary multicomponent adsorption modelling using ANN, LS-SVR, and SVR approach – case study, Kem. Ind. (2021), https://doi.org/10.15255/KUI.2020.071.
[35] H. Drucker, C.J. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines, Adv. Neural Inf. Proces. Syst. 9 (1996).
[36] F. Zhang, L.J. O'Donnell, Support vector regression, in: Machine Learning, Elsevier, 2020, pp. 123–140, https://doi.org/10.1016/B978-0-12-815739-8.00007-9.
[37] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (2001) 1189–1232, https://doi.org/10.1214/aos/1013203451.
[38] P. Nie, M. Roccotelli, M.P. Fanti, Z. Ming, Z. Li, Prediction of home energy consumption based on gradient boosting regression tree, Energy Rep. 7 (2021) 1246–1255, https://doi.org/10.1016/J.EGYR.2021.02.006.
[39] F. Murtagh, Multilayer perceptrons for classification and regression, Neurocomputing 2 (1991) 183–197, https://doi.org/10.1016/0925-2312(91)90023-5.

[40] M.-C. Popescu, V.E. Balas, L. Perescu-Popescu, N. Mastorakis, Multilayer perceptron and neural networks, WSEAS Trans. Circuits Syst. 8 (2009) 579–588.
[41] G. Barletta, Models and Datasets for "Exploring Machine Learning Models to Predict Atmospheric Water Harvesting With an Ion Deposition Membrane", 2024, https://doi.org/10.5281/zenodo.10533012.
[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[43] M. Jafari Gukeh, S. Moitra, A.N. Ibrahim, S. Derrible, C.M. Megaridis, Machine learning prediction of $TiO_2$-coating wettability tuned via UV exposure, ACS Appl. Mater. Interfaces 13 (2021) 46171–46179, https://doi.org/10.1021/acsami.1c13262.
[44] H. White, A reality check for data snooping, Econometrica 68 (2000) 1097–1126, https://doi.org/10.1111/1468-0262.00152.
[45] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination r-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, Peer J. Comput. Sci. 7 (2021) 1–24, https://doi.org/10.7717/PEERJ-CS.623/SUPP-1.
[46] E. Matthiasson, B. Sivik, Concentration polarization and fouling, Desalination 35 (1980) 59–103, https://doi.org/10.1016/S0011-9164(00)88604-X.
[47] L. Song, M. Elimelech, Theory of concentration polarization in crossflow filtration, J. Chem. Soc. Faraday Trans. 91 (1995) 3389–3398, https://doi.org/10.1039/FT9959103389.
[48] S. Sablani, M. Goosen, R. Al-Belushi, M. Wilf, Concentration polarization in ultrafiltration and reverse osmosis: a critical review, Desalination 141 (2001) 269–289, https://doi.org/10.1016/S0011-9164(01)85005-0.
[49] C. Wilke, P. Chang, Correlation of diffusion coefficients in dilute solutions, AIChE J. 1 (1955) 264–270, https://doi.org/10.1002/aic.690010222.
[50] N. Frossling, The Evaporation of Falling Drops, Atomic Energy Research Establishment, Harwell, England, 1963.
[51] Weather Underground, Saudi Arabia Weather History, Jeddah, Mecca, 2022. https://www.wunderground.com/history/daily/sa/jeddah/OEJN/date/2022-4-10. (Accessed 20 April 2023).
[52] J. Ji, R. Wang, L. Li, New composite adsorbent for solar-driven fresh water production from the atmosphere, Desalination 212 (2007) 176–182, https://doi.org/10.1016/j.desal.2006.10.008.
[53] J. Wang, J. Liu, R. Wang, L. Wang, Experimental investigation on two solar-driven sorption based devices to extract fresh water from atmosphere, Appl. Therm. Eng. 127 (2017) 1608–1616, https://doi.org/10.1016/j.applthermaleng.2017.09.063.
[54] B. Zhao, L.-Y. Wang, T.-S. Chung, Enhanced membrane systems to harvest water and provide comfortable air via dehumidification & moisture condensation, Sep. Purif. Technol. 220 (2019) 136–144, https://doi.org/10.1016/j.seppur.2019.03.034.