Environmental Sciences Europe

RESEARCH

Open Access

New strategy based on Hammerstein–Wiener and supervised machine learning for identification of treated wastewater salinization in Al-Hassa region, Saudi Arabia

Syed Muzzamil Hussain Shah[1], Sani I. Abba[1*], Mohamed A. Yassin[1,2], Dahiru U. Lawal[1], Farouq Aliyu[3], Ebrahim Hamid Hussein Al-Qadami[4], Haris U. Qureshi[5], Isam H. Aljundi[1,6], Hamza A. Asmaly[7], Saad Sh. Sammen[8*] and Miklas Scholz[9,10,11,12,13*]

Abstract

The agricultural sector faces challenges in managing water resources efficiently, particularly in arid regions dealing with water scarcity. To overcome water stress, treated wastewater (TWW) is increasingly utilized for irrigation purpose to conserve available freshwater resources. There are several critical aspects affecting the suitability of TWW for irrigation including salinity which can have detrimental effects on crop yield and soil health. Therefore, this study aimed to develop a novel approach for TWW salinity prediction using artificial intelligent (AI) ensembled machine learning approach. In this regard, several water quality parameters of the TWW samples were collected through field investigation from the irrigation zones in Al-Hassa, Saudi Arabia, which were later assessed in the lab. The assessment involved measuring Temperature (T), pH, Oxidation Reduction Potential (ORP), Electrical Conductivity (EC), Total Dissolved Solids (TDS), and Salinity, through an Internet of Things (IoT) based system integrated with a real-time monitoring and a multiprobe device. Based on the descriptive statistics of the data and correlation obtained through the Pearson matrix, the models were formed for predicting salinity by using the Hammerstein-Wiener Model (HWM) and Support Vector Regression (SVR). The models' performance was evaluated using several statistical indices including correlation coefficient (R), coefficient of determination (R^2), mean square error (MSE), and root mean square error (RMSE). The results revealed that the HWM-M3 model with its superior predictive capabilities achieved the best performance, with R^2 values of 82% and 77% in both training and testing stages. This study demonstrates the effectiveness of AI-ensembled machine learning approach for accurate TWW salinity prediction, promoting the safe and efficient utilization of TWW for irrigation in water-stressed regions. The findings contribute to a growing body of research exploring AI applications for sustainable water management.

Keywords Salinization, Hammerstein–Wiener model, Support vector regression, Artificial intelligence, Machine learning, SDG6

*Correspondence:
Sani I. Abba
sani.abba@kfupm.edu.sa
Saad Sh. Sammen
Saad123engineer@yahoo.com
Miklas Scholz
Miklas.Scholz@bso-oberursel.de; mscholz@uj.ac.za; miklas.scholz@kunststoff-technik-adams.de; miklas@nexusbysweden.com
Full list of author information is available at the end of the article

Introduction

Water scarcity in arid and semi-arid regions is often exacerbated by a combination of natural and human-induced factors [1]. To address these challenges, utilizing TWW for irrigation is gaining prominence as a practical and cost-effective alternative, alleviating pressure on natural water resources [2]. Globally, the utilization of TWW for irrigation is on the rise, with notable examples such as Israel, Spain, and California, where 85%, 40%-70%, and 30% of TWW, respectively, are reutilizing it for irrigation. In Saudi Arabia, the irrigation sector is suggested to account for about 71% of the annual freshwater consumption. To ensure the growing gap between water demand and supply, treated wastewater has been recognized as a potential solution. Thus, the initial regulatory step called "Treated Sanitary Wastewater and Its Reuse Regulations" was taken in 2000 which required either secondary or tertiary treatment before its use. Later, in 2006, the Ministry of Environment, Water, and Agriculture [MEWA] released two booklets focusing on the use of treated wastewater in agriculture [3]. Beyond water conservation, TWW carries the added advantage of being nutrient-rich, reducing the necessity for fertilization. The United Nations (UN) agencies, like the Food and Agriculture Organization (FAO), acknowledge that repurposing TWW for irrigation has immense potential and is a vital aspect of resolving serious ecological problems worldwide [4, 5]. Salinity levels found in treated wastewater tend to be higher than in the source water. About 30 million tons of salt (NaCl) is consumed annually in the European Union alone, resulting in a significant global consumption. This extensive salt consumption contributes to heightened salinity in urban effluents and, consequently, in the resulting treated wastewater [6].

The World Health Organization (WHO) published a number of guidelines in 1973, 1989, and 2006 which specified secure practices for the use of treated wastewater in irrigation as a way to address the risks associated to public health [7–10]. The primary objective of these guidelines was to reinforce government regulations regarding wastewater treatment, focusing on the thresholds for TWW quality standards [8, 11, 12]. However, the FAO published two guidelines on the use of TWW for irrigation. The first recommendation separates irrigation water into three categories based on characteristics like toxicity, salinity, sodicity, and other risks [13]. This categorization exposes the possible crop production issues associated with conventional water sources. In the second guideline, the FAO segmented the application of water reuse in irrigation into three groups, taking into account the type of irrigated crops [14]. Besides, the Environmental Protection Agency (EPA) issued water reuse guidelines in 1980, 1992, 2004, and 2012. The most recent version is viewed as a refinement of the 2004 guideline, with the goal of promoting wastewater reuse by serving as trustworthy references, drawing from a compilation of global experiences [15, 16]. In general, the latest guideline places greater emphasis on environmental and health preservation compared to its predecessor [17]. The guidelines set forth by WHO, EPA, and FAO serve as foundational principles for the formulation of regulations in various countries globally. Therefore, in the absence of national guidelines in any country, it is recommended to turn to the guidelines provided by WHO, EPA, and FAO as a viable solution.

Al-Hassa has been reported as one of the largest irrigation zones in the Kingdom of Saudi Arabia. In AL-Hassa, the primary source of water used for irrigation includes the water from the groundwater wells mixed with treated wastewater and partially agricultural drainage. The sewage treatment plants (STPs) namely Hofuf STP, Umran STP, Oyun STP, and Aramco STP derive almost 200,000 m^3/day of treated wastewater used for irrigation [18]. One of the studies reported elevated levels of salinity and nitrate in the groundwater wells in Al-Ahsa oasis while investigating the groundwater quality. The groundwater utilized for irrigation in Al-Ahsa Oasis is noted for its high salinity and falls within the categories of low to medium sodicity. This salinity and sodicity profile could potentially be attributed to the extensive extraction from groundwater wells [19].

It is vital to note that various plants and crops may tolerate various levels of salinity since it is a crucial factor that indicates the water's usefulness and possible environmental impact after its release. It is therefore crucial to comprehend and anticipate salinity levels [20, 21]. Lately, the application of Machine Learning (ML) approaches to foresee wastewater quality has drawn more attention. With the help of the given data set, ML models can possibly be trained to identify the correlations between different variables [22]. For instance, Hejabi et al. [23] assessed AI models' efficacy in simulating effluent quality parameters. The metrics related to influent quality were classified as independent, whilst the parameters related to effluent quality were classified as dependent [23]. Similarly, Mustafa et al. [24] employed ML and ensemble techniques to model TDS concentrations in wastewater treated with Salvinia molesta plants. The study demonstrated enhanced prediction accuracy for TDS concentrations by ensemble learning using artificial neural network (ANN), support vector machine (SVM), adaptive neuro-fuzzy inference system (ANFIS), and multi-linear regression (MLR) [24]. Moreover, Banadkooki et al. [25] adopted different

ML techniques including ANFIS, SVM and ANN to predict the quantity of TDS. These ML techniques were optimized using moth flam optimization (MFO), cat swarm optimization (CSO), particle swarm optimization (PSO), shark algorithm (SA), grey wolf optimization (GWO), and gravitational search algorithm (GSA). The ANFIS-MFO and ANFIS-CSO models showed superior performance over the other models [25]. One of another studies by Abba et al. [26] utilized various models, including general regression neural network (GRNN), Hammerstein–Wiener (HW), non-linear autoregressive exogenous model (NARX), and least square support vector machine (LSSVM), to develop a multi-parametric model for a water treatment plant. The NARX model demonstrated great predictive capabilities for pH, while the HW model demonstrated exceptional simulation abilities for hardness, turbidity, and suspended particles [26]. Mokhtar et al., [27] for instance, assessed three machine learning algorithms, namely SVR, extreme gradient boosting (XGB), and random forest (RF) and four multiple regressions, i.e., stepwise regression (SW), principal components regression (PCR), partial least squares regression (PLS), and ordinary least squares regression (OLS) to predict six IQWI parameters. The study suggested SW as the optimal regression model for IWQI prediction, and SVR as the best AI model, providing insightful information to improve irrigation water quality [27]. Moreover, Hamada et al., [28] employed gaussian process regression (GPR), RF, XGB, and light gradient boosting machine (LightGBM) to predict total suspended solids (TSS), chemical oxygen demand (COD), and biochemical oxygen demand (BOD) concentrations in Gaza wastewater treatment plant effluent a day in advance. The GPR demonstrated the highest accuracy compared to RF, XGB, and LightGBM, with pH and temperature identified as crucial parameters in wastewater quality prediction, emphasizing GPR's suitability for optimal wastewater treatment selection based on original characteristics and standards [28].

Building upon the foundation laid by [23–27] and [28], herein, this study aims at identifying the prediction capabilities of the SVR and HWM models to assess the salinity dynamics in the TWW believing that the study outcomes would help improve the efficacy of the wastewater treatment plant. Furthermore, regardless of the significance of SVR and the HWM in various fields, their representation in Scopus appears relatively scarce. The search for articles about SVR and the HWM within the Scopus database yields limited results as shown in Fig. 1, which indicates an underexplored area within the domain of academic research. Therefore, this highlights the potential for further exploration and in-depth investigation into these predictive modelling techniques and system identification methods across various disciplines. Moreover, the choice of SVR and HWM in this study was mainly due to the versatile capabilities of these two models in dealing with complex systems, particularly in the area of environmental and water resource studies. SVR is known for its

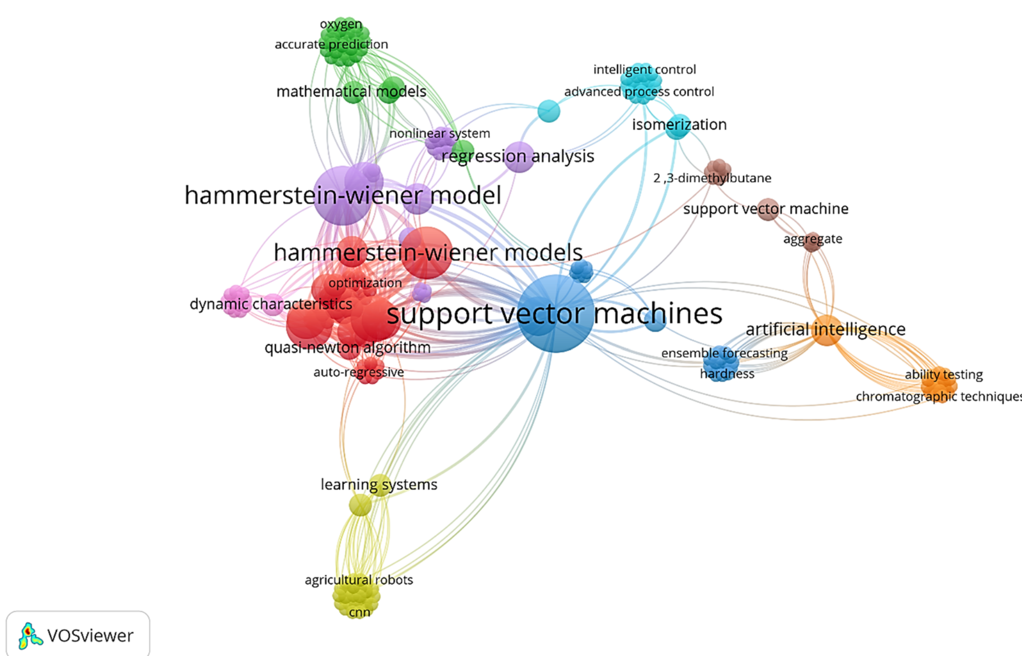

Fig. 1 The major keywords used over the literature on Hammerstein–Wiener and Support Vector Machine Database from Scopus

ability to handle high-dimensional data and non-linearity, whereas the HWM, with its capacity to capture both linear and nonlinear dynamics, offers a unique advantage in accurately representing the relationships.

Methodology

Study area

The Al-Hassa Agricultural zone is renowned for its cultivation which acts a vital component of the region's economy [29]. This region has adopted contemporary farming methods in recent years, utilizing cutting-edge technologies to boost output while maintaining the fundamentals of conventional farming. Groundwater wells, treated wastewater, and partially agricultural drainage water are currently Al-Ahsa Oasis's main sources of irrigation water. The treated wastewater used for agriculture primarily comes from municipal wastewater treatment plants [19]. These plants receive wastewater from residential, commercial, and industrial sources within the city. After undergoing treatment processes to remove contaminants and pathogens, the treated wastewater, also known as reclaimed water, is repurposed for agricultural irrigation. Through the use of treated wastewater that would otherwise be released into the environment, this technique aids in the conservation of freshwater resources [30]. To assure that the treated wastewater satisfies particular quality requirements and is suitable for agricultural use, it passes through a number of treatment processes, including physical, biological, and occasionally advanced treatment techniques. These treatments aim to remove solids, organic matter, and harmful substances to make the water suitable for irrigation without posing risks to crops, soil, or human health.

Using treated wastewater for agriculture aligns with efforts to conserve freshwater resources and supports sustainable agricultural practices. However, it's crucial to ensure that the quality of treated wastewater meet standards to prevent potential risks associated with irrigation. Herein, the study encompassed gathering water samples from diverse irrigation zones in Al-Hassa, as depicted in Fig. 2. These samples were subjected to thorough laboratory analysis, evaluating a range of water quality parameters. These included salinity, temperature, pH, ORP, EC, resistivity, turbidity, and TDS, featured with real-time monitoring facilitated by an Internet of Things (IoT)-based system as well as a multiprobe device. The analyzed samples were then used to predict the key parameters accordingly. The detailed study approach can be referred in Fig. 3.

AI-infused process flow

The implementation of SVR and HWM Model served as the core framework for addressing the research objectives. This paper illuminates a systematic approach that begins with data collection and preprocessing, ensuring the dataset's suitability for both models. The SVR implementation involved data partitioning, model training, and hyperparameter optimization to attain an optimal classification framework. Simultaneously, the HWM was applied, emphasizing data formatting, system identification, or regression techniques tailored for the specific dataset. The integration phase, a key part of this study, carefully combined the results of these models. It does this by using techniques that blend them or by using the predictions of one model as rules for the other. Throughout the paper, the unique strengths of each model demonstrate how their collaboration leads to a more comprehensive and robust solution for the research problem. This approach is briefly depicted in the AI-Infused workflow as shown in Fig. 4, visually capturing the sequential steps and integration points between SVR and the HWM, explaining the comprehensive nature of the methodology. For instance, the first phase involves collecting TWW samples from the field which undergoes

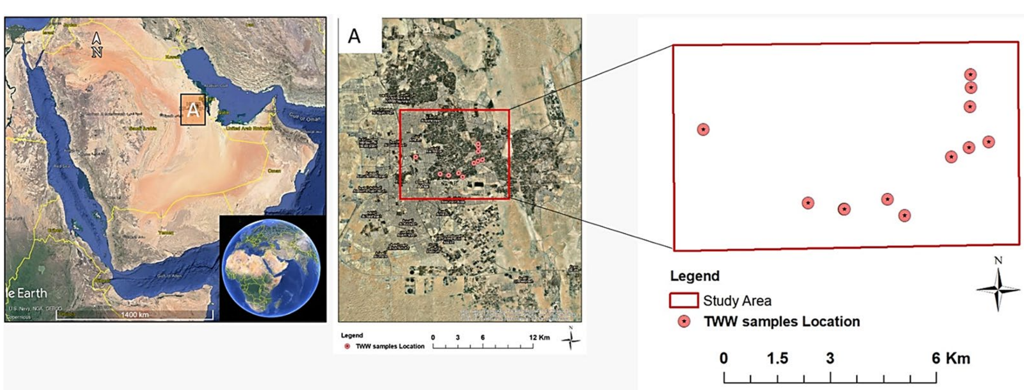

Fig. 2 Study Area—Al-Hassa Farms, Eastern Province, Saudi Arabia

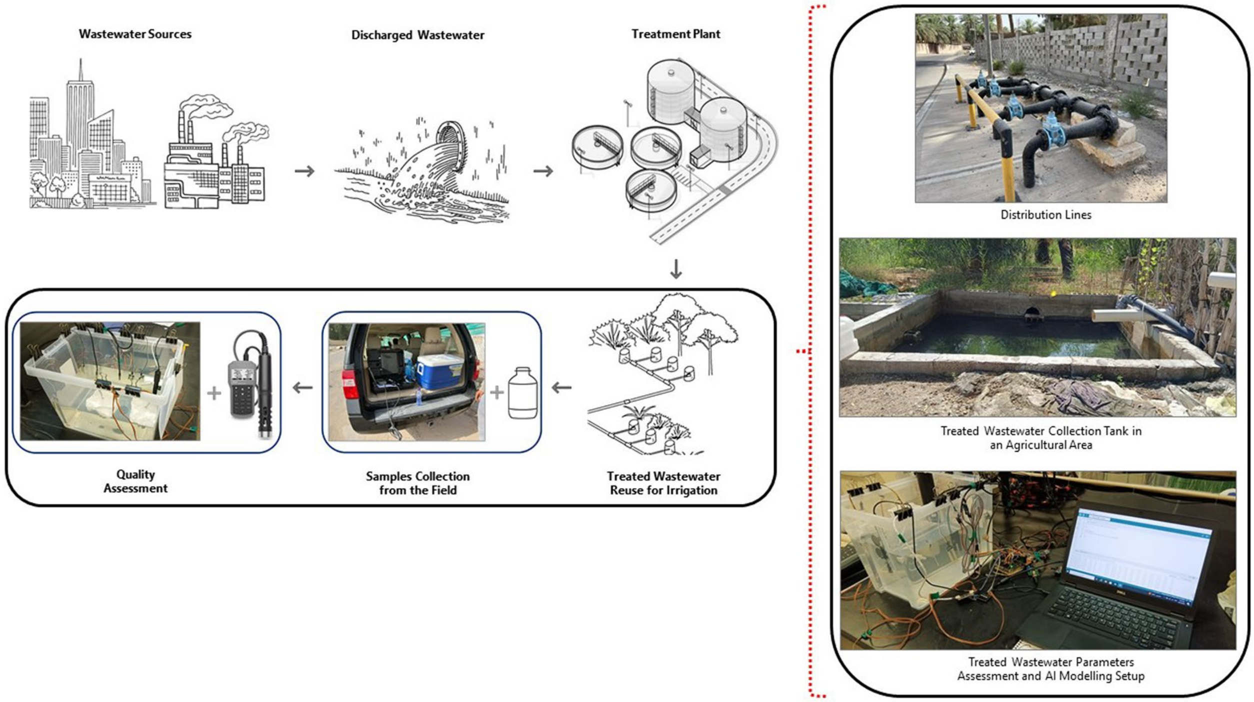

Fig. 3 The setup overview highlighting **a** the treated waste water distribution lines, **b** the collection tanks in the farms and **c** the experimental setup integrated with Arduino and Multiprobe

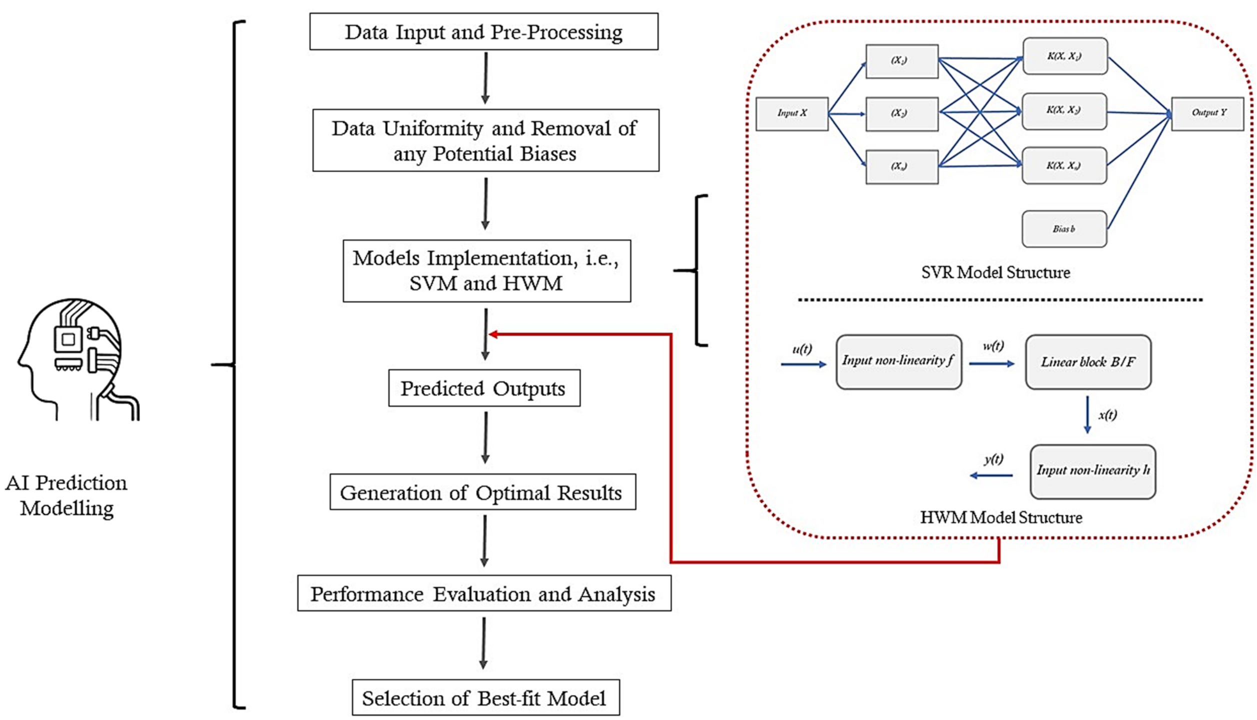

Fig. 4 AI-infused process flow

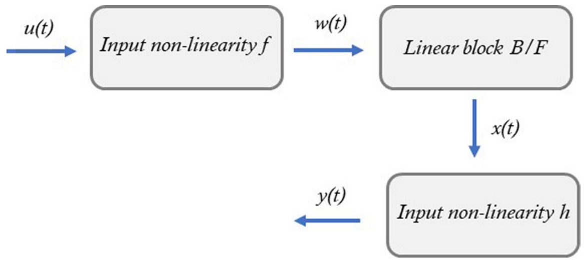

Fig. 5 Schematic diagram of HWM [adapted from [33]]

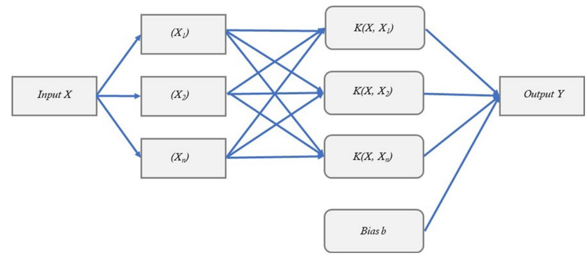

Fig. 6 Schematic diagram of SVR model [adapted from 34]

pre-processing to ensure its quality and consistency for analysis. Later, it involves ensuring all the data uses the same units followed by utilization of two machine learning models. The processed data then predicted the salinity levels in the TWW samples which were then compared to the actual measured values to evaluate the accuracy of the models, followed by the selection of the best fit model.

Hammerstein–Wiener Model (HWM)

The HWM is composed of both nonlinear and linear block systems. This model can be effectively employed as a block-box model, offering versatility in handling various variables and parameters [31]. Its performance surpasses that of linear and nonlinear systems like MLR and ANN, as the HWM considers both linearity and nonlinearity within a dataset. The architectural components of HWM consist of a linear dynamic block, a static input nonlinear block, and static output nonlinear blocks [32] and [26] as shown in Fig. 5. The HWM employs a mechanism that involves transforming nonlinear functions into linear input blocks, which are then converted back into a nonlinear state as output. The key equations associated with the HWM model are as follow:

$$U(t) = f(x(t)) \tag{1}$$

$$y_l(\mathrm{t}) = G_l * u(t) \tag{2}$$

$$y(t) = h\left(y_l(t)\right) \tag{3}$$

$$y(t) = h\left(G_l * f(x(t))\right) \tag{4}$$

where, Eq. 1 describes the nonlinear input block, contains a pair of elements: *u (t)* indicating the input of the system and $f(\cdot)$ denoting a function that is nonlinear which transforms the input *x (t)*. Equation 2 defines the linear dynamic block, where the convolution operation is represented by $*$, the block's output is $y_l(t)$, and G_l is the transfer function of the block. On the other hand, the linear dynamic block's output y_l (t) is obtained by applying the nonlinear function $h(\cdot)$ to the static output nonlinear block, which is represented by Eq. 3. Finally, the overall output of the model is represented by Eq. 4.

Support vector regression (SVR)

The SVR in artificial intelligence are models used for classification tasks. It is a machine learning approach primarily designed to address classification challenges involving small sample sizes, nonlinearities, and high-dimensional data [34–36] as seen in Fig. 6. It involves identifying the most effective hyperplane within a multi-dimensional space to distinctly separate various classes within a dataset. The fundamental equation:

$$y = w \cdot x + b \tag{5}$$

The above equation represents the hyperplane, where y is the predicted target variable, w is the weight vector, x is the input vector, and b is the bias term. The objective is to maximize the margin between this hyperplane and the closest data points, known as support vectors, from distinct classes. SVR aim to minimize misclassifications while maximizing the distance between classes, creating a robust decision boundary [24].

Data and evaluation criteria

A rigorous pre-processing and post-processing approach were used in the extensive process of forecasting salinity in treated wastewater samples taken from the Al-Hassa region in order to assure the accuracy and dependability of the SVR and HWM predictions. Initially, the collected data underwent a normalization procedure, a crucial step aimed at scaling the variables to a standard range, ensuring that each parameter contributes proportionately to the models [37, 38]. To further perform the in-depth analysis, the dataset underwent descriptive analysis. During this stage, the key aspects of the data were outlined providing an in-depth understanding of the fundamental trends. Subsequently, to further investigate the correlations between the obtained water quality metrics, the Pearson correlation matrix was employed for assessing the possible connections and selecting the relevant input variables

[39, 40]. Based on that, the model combinations were decided with a robust framework to estimate salinity. The performance metrics are commonly used to evaluate the overall performance of a system [41]. The comprehensive dataset encompassing almost 7700 values was compiled. Following rigorous model selection procedures, the dataset was divided, allocating 70% of the data for the training phase and reserving the remaining 30% for the testing phase to ensure robust model evaluation. Through this approach, key performance metrics were calculated to gauge the predictive accuracy of the chosen models. This division and subsequent analysis aimed to validate the effectiveness of the models in accurately representing the complex interdependencies among the variables within the treated wastewater samples. Herein, the computed and the measured data was monitored through the four statistical measures, namely R^2, R, MSE, and RMSE as discussed below:

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{N} (Sal._o - Sal._P)^2}{\sum_{i=1}^{N} \left(Sal._o - \overline{Sal._O} \right)^2} \right] \tag{6}$$

$$R = \frac{\sum_{i=1}^{N} \left(Sal._o - \overline{Sal._o} \right) \left(Sal._p - \overline{Sal._p} \right)}{\sqrt{\sum_{i=1}^{N} \left(Sal._o - \overline{Sal._o} \right)^2 \sum_{i=1}^{N} \left(Sal._p - \overline{Sal._p} \right)^2}} \tag{7}$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (Sal._o - Sal._P)^2 \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Sal._o - Sal._P)^2} \tag{9}$$

where, $Sal._P$ and $Sal._O$ represents the predicted and observed salinity, whereas $\underline{Sal._O}$ and $\underline{Sal._P}$ represents the observed and predicted salinity with its corresponding averages for N data points, respectively.

Result and discussions

Performance measure

The descriptive statistical analysis as shown in Table 1 represents the descriptive statistics ensuring that the data was properly aligned, well understood, finely processed and fulfilled the requirements of the chosen AI models. This leads to more effective model development and reliable results [37]. Furthermore, based on the correlation matrix, an appropriate combination of inputs was identified (refer to

Table 1 Descriptive statistics of the data

Parameters	Mean	SD	Kurtosis	Skewness	Min	Max
Temp. [°C]	23.280	0.028	−0.783	0.619	23.24	23.34
pH	8.063	0.0186	24.013	−4.466	7.90	8.07
ORP [mV]	169.584	2.031	−0.866	0.340	166.60	174.50
EC [µS/cm]	2841.975	1.973	−1.071	−0.284	2837.00	2846.00
RES [Ohm-cm]	351.924	0.264	8.353	−3.217	351.00	352.00
TDS [ppm]	1421.230	1.021	−0.955	−0.212	1419.00	1423.00
Turb. [FNU]	0.333	0.071	471.996	13.690	0.20	3.10

Table 2 Pearson correlation matrix between the inputs and output

	Temp. [°C]	pH	ORP [mV]	EC [µS/cm]	RES [Ohm-cm]	TDS [ppm]	Turbidity [FNU]	Sal. [PSU]
Temp.[°C]	1							
pH	0.116	1						
ORP [mV]	−0.810	−0.595	1					
EC [µS/cm]	0.813	0.301	−0.887	1				
RES [Ohm−cm]	−0.492	−0.061	0.370	−0.442	1			
TDS [ppm]	0.797	0.287	−0.864	0.969	−0.494	1		
Turb. [FNU]	−0.011	−0.350	0.196	−0.048	0.114	−0.052	1	
Sal. [PSU]	0.633	0.265	−0.760	0.856	−0.199	0.842	0.000	1

Table 2). The study utilized the holdout validation method, a variation of k-fold cross-validation. This validation process offered diverse approaches like k-fold cross-validation, holdout, and leave-one out, among others. The holdout method simplifies k-fold by randomly dividing data into two sets: training and testing phases [26, 32].

The choice of the models for predicting salinity was based on the given outputs. A total of three candidate models were selected as follows:

$$\text{Model I}: \ \text{EC} + \text{TDS} \tag{10}$$

$$\text{Model II}: \ \text{EC} + \text{TDS} + \text{ORP} + \text{Temp.} \tag{11}$$

$$\text{Model III}: \ \text{EC} + \text{TDS} + \text{ORP} + \text{Temp.} + \text{pH} \tag{12}$$

Predictive analysis

In predicting salinity, the choice of input variables is crucial for accurate modeling. The utilization of different combinations of variables across three distinct models, namely Model I, Model II, and Model III, using SVR and HWM to explore varying levels of complexity and information integration as shown in Figs. 7 and 8, respectively. Model I employed a straightforward combination of two variables, i.e., EC and TDS. This minimalist approach seeks to establish a foundational understanding of salinity prediction by relying on the fundamental factors known to influence salinity levels. For Model II, additional two variables, namely ORP and Temp. were introduced alongside the variables used for Model I. Incorporating additional parameters aimed to capture a wider range of factors that could further enhance the predictive capabilities. Model III further broadened the range of input variables by incorporating pH, thus incorporating an even more extensive collection of factors that could influence salinity. Analyzing the patterns ranging from simpler to complex combinations enabled us to investigate how adding more parameters impacts the predictive abilities of both SVR and HWM. Moreover, the selected variables also play a crucial role in environmental and chemical contexts, contributing effectively to salinity variations.

As depicted in above figures, the SVR models, despite their inherent strength in handling complex relationships through the use of kernel functions, might have limitations in capturing intricate nonlinearities presented in the dataset. The different SVR configurations may not have been as successful in capturing the complex and varied correlations between the salinity and the input

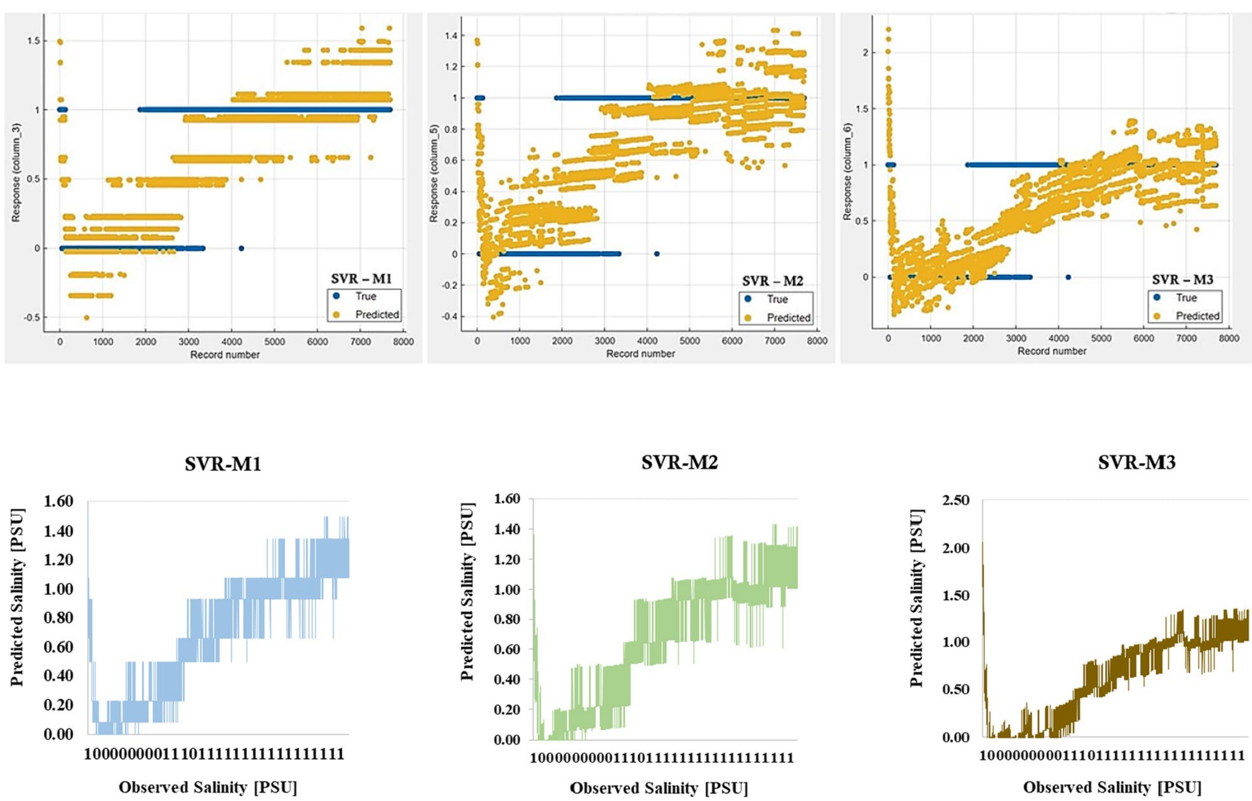

Fig. 7 SVR Models response plot between true and predicted values

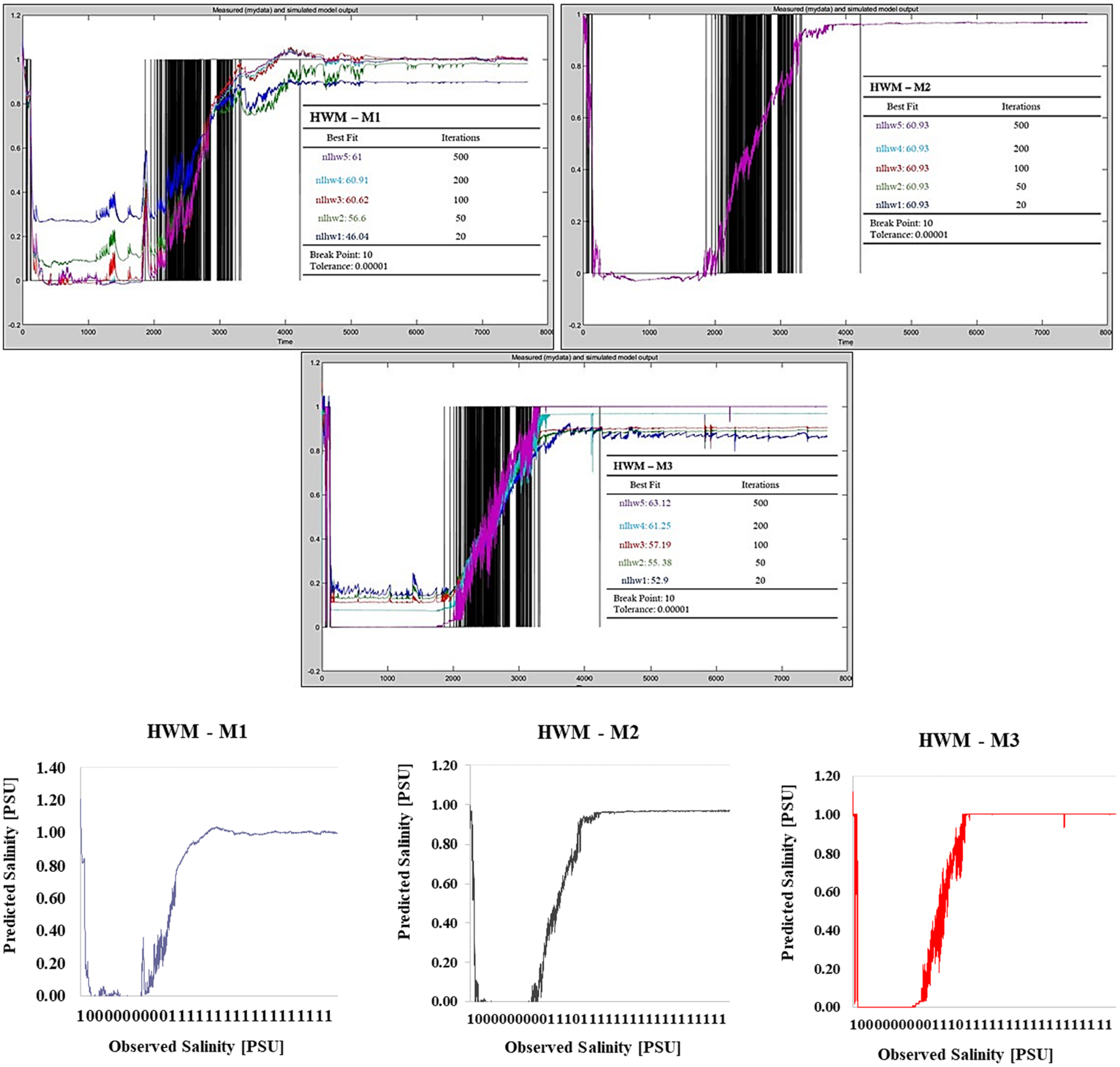

Fig. 8 HVM Structure and best fit models

Table 3 Results of SVR for Modelling Salinity

Models	Training				Testing			
	R^2	MSE	R	RMSE	R^2	MSE	R	RMSE
SVR-M1	0.7300	0.0672	0.8544	0.2592	0.6800	0.0351392	0.8246	0.1875
SVR-M2	0.7240	0.0687	0.8509	0.2621	0.6740	0.0261140	0.8210	0.1616
SVR-M3	0.6776	0.0802	0.8231	0.2833	0.6200	0.0206101	0.7874	0.1436

variables as the more sophisticated structure of HWM. Specifically, the superiority of the HWM-M3 over other models could be justified based on several assumptions. Firstly, the HWM is capable to incorporate the static and the nonlinear elements that could help capture the strong relation between the input variable and the output. Secondly, it uses advanced preprocessing methods which is flexible to handle nonlinearities. The extensive training and testing further assured resilience across he various datasets. The performance metrics for SVM and HWM models can be referred in Tables 3 and 4, respectively.

From Fig. 9, the R^2 values for both the models can be depicted in the form of radar graph. The radar graph tells the effectiveness of the dataset in a particular situation via thorough assessment of the performance. Upon comparision, an exceptional performance of HWM model over SVR was witnessed in both testing and training phases. There seems a potential need for regularization to assure model refinement for the SVR

Table 4 Results of HWM for Modelling Salinity

Models	Training				Testing			
	R^2	MSE	R	RMSE	R^2	MSE	R	RMSE
HWM-M1	0.8076	0.0479	0.8986	0.2188	0.7576	0.0000236	0.8704	0.0049
HWM-M2	0.8086	0.0476	0.8992	0.2182	0.7586	0.0011274	0.8710	0.0336
HWM-M3	0.8279	0.0428	0.9099	0.2070	0.7779	0.0000042	0.8820	0.0021

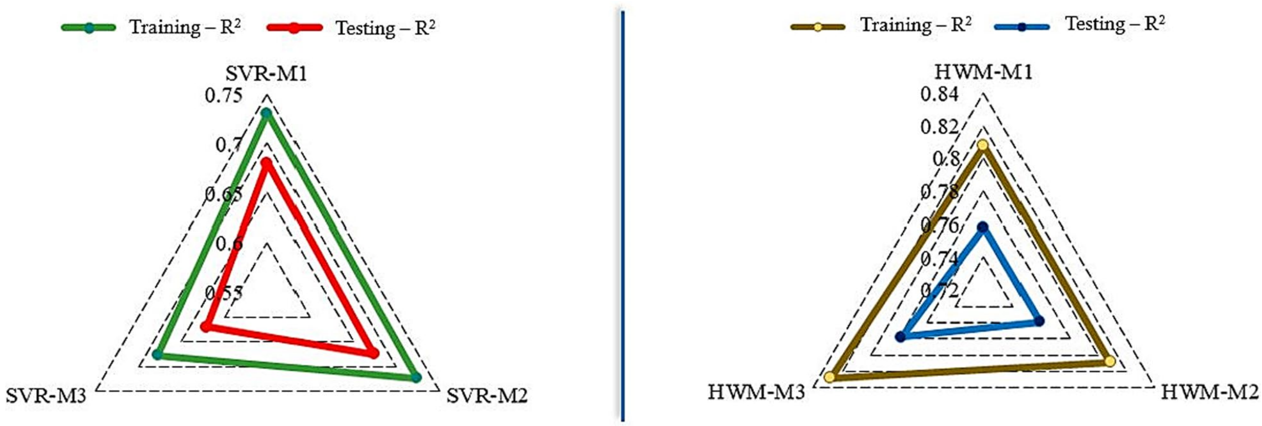

Fig. 9 Radar graph depicting R-Squared values for SVR and HWM models

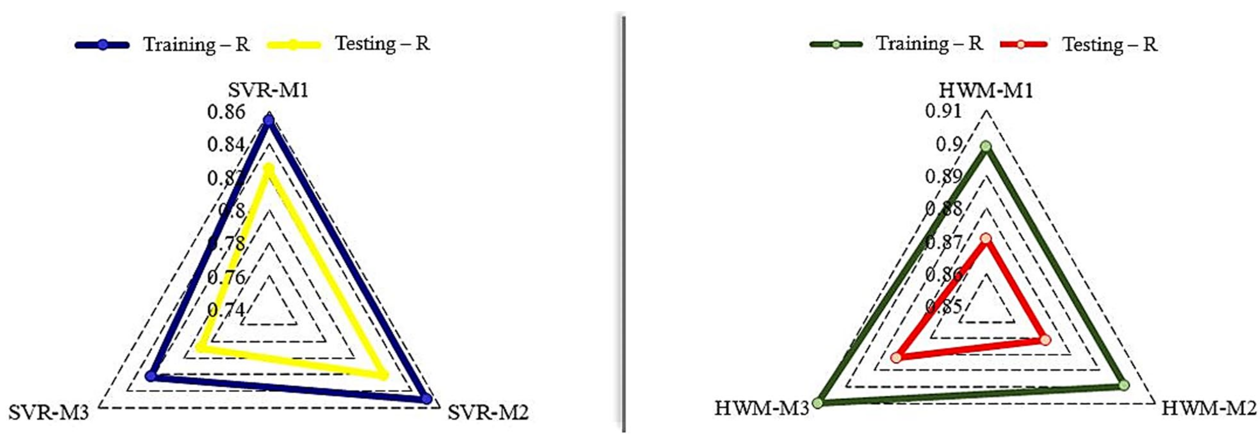

Fig. 10 Radar graph depicting R values for SVR and HWM models

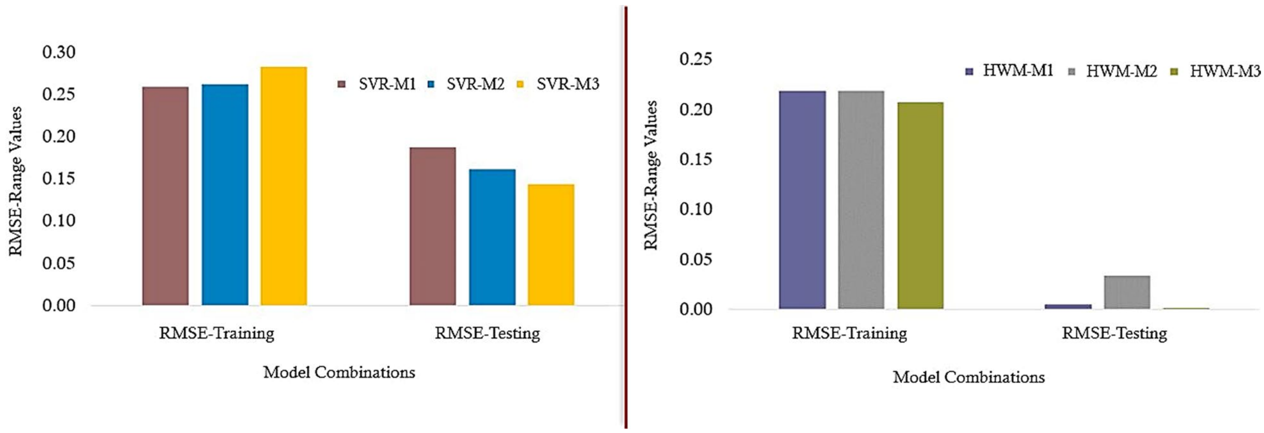

Fig. 11 RMSE values obtained from the training and testing phases of SVR and HWM models

models due to the reduced R^2 values. Conversely, the higher R^2 values for the HWM models indicates it capability to capture underlying patterns.

From Fig. 10 a slight decline in the R values can be seen from the training to the testing stages which indicates that the models were functioning in a balanced manner. The recorded R values indicated that both the models can effectively capture and predict the underlying patterns in the dataset. The HWM models witnessed to perform better somewhat, thus it can be further explored and implemented for practical applications.

Figure 11 demonstrates the quantitative assessment of the models' performance in both training and testing phase. The variations in RMSE values helped understand and assess the model's dependability. In generalization, the SVR-M3 performed well, whereas the HWM-M1 performed better all around. The degree to which SVR-M3 and HWM-M1 generalize to new data during testing is a positive indication of their resilience. It suggests that these models may have succeeded in striking a balance between fitting training data and adapting to observations.

As illustrated in Fig. 12, the performance of the six AI models, namely SVR-M1, SVR-M2, SVR-M3, HWM-M1, HWM-M2, and HWM-M3 was assessed using MSE values that are acquired during both the training and testing stages. For the SVR models, the training MSE values exhibit a gradual increase from SVR-M1 (0.0672) to SVR-M3 (0.0802). In contrast, during the testing phase, the MSE values decrease from SVR-M1 (0.035) to SVR-M3 (0.020), indicating improved generalization performance. Similarly, for the HWM models, the training MSE values show a slight decrease from HWM-M1 (0.0479) to HWM-M3 (0.0428). Notably, during testing, the MSE values for HWM-M1 (0.0000236), HWM-M2 (0.0011274), and HWM-M3 (0.0000042) are remarkably low, highlighting the predictive capabilities of the

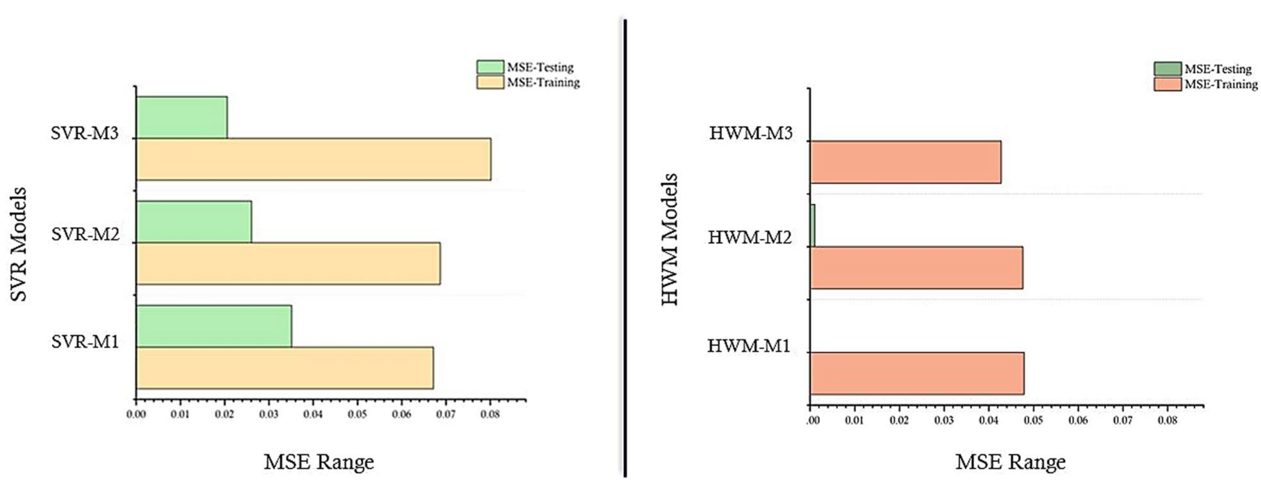

Fig. 12 MSE values obtained from the training and testing phases of SVR and HWM models

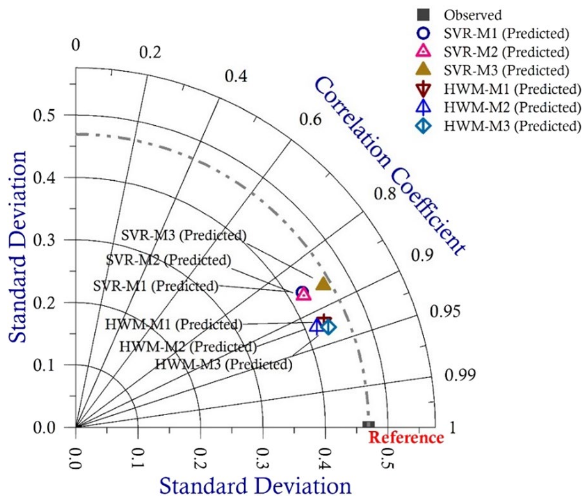

Fig. 13 Taylor diagram with the SD ranges obtained for the observed and predicted salinity

models. These results suggest that the HWM models outperform the SVR models, particularly in the testing phase, showcasing their efficacy in accurately predicting outcomes. Moreover, it is crucial to understand the numerical results using a two-dimensional plot known as Taylor diagram (Fig. 13). Taylor diagram has been in science and engineering to depict the extent of SD, RMSE and R values in one plot. From the plot it can be justified that HWM-3 proved more reliable than the other models with RMSE = 0.0021 in testing phase. The predictive skills of HWM is not surprising owing to the fact that, it is nonlinear system identification approach and we are working with pilot plant system to identify the patterns and influence of TWW used in agricultural sector.

It is imperative to compare our outcomes with the existing state-of-the-art approaches, for instance, Poursaeid et al., [42] employed a hybrid metaheuristic AI model known as the wavelet self-adaptive extreme learning machine (WSAELM) to simulate groundwater parameters in the Mighan Plain, Iran, spanning from 2002 to 2017. The WSAELM model demonstrated an impressive accuracy in predicting salinity, achieving a value of 0.991. However, it is noteworthy that the proposed hybrid model outperformed our results having an R^2 value ranging between 77 and 82%. The superior performance of WSAELM is attributed to the integration of hybrid models than relying on a single model, which has been widely acknowledged in the literature to outperform single models in various contexts. Likewise, in the study conducted by Mosavi et al. [43] ML models were employed to forecast groundwater salinity, with significant factors identified through simulated annealing feature selection. After testing six different models, it was discovered that the SVM model performed the best. The main variables that were shown to have the biggest effects on groundwater salinity prediction were soil type, precipitation, land use, elevation, and groundwater extraction. The SVM model performed better than the others, despite the fact that all models showed excellent accuracy levels (above 0.82). A consistent agreement with our predicted output is evident when comparing these findings with our results. Furthermore, Tran et al., [44] investigated cutting-edge machine learning methods to forecast groundwater salinity in the coastal aquifers of Vietnam's Mekong River Delta, with a testing phase goodness-of-fit of 84%, demonstrating the efficacy of the approach. Additionally, the results showed a modest increase over the current results. In a recent study, Trabels et al. [45] focused on evaluating the capabilities of the ML models to predict groundwater quality for irrigation in Tunisia's downstream Medjerda river basin. Among the models assessed, the AdaBoost stood out for its ability to produce more accurate and concise predictions with the least amount of input parameters (R = 0.89). The outcomes demonstrated a noteworthy concordance with our study results which were obtained from HWM-M3 (R = 0.88). Looking ahead, it seems that the prediction capabilities could be further enhanced by introducing cutting-edge machine learning methods. By doing so, the accuracy could improve even more.

Conclusions

The thorough analysis of water quality parameters integrated with real-time monitoring under supervised machine learning fairly contributed towards the better understanding of salinity dynamics. Based on the performance metrics, the SVR-M1 exhibits higher R^2 values for both training (0.7300) and testing (0.6800) among all SVR models indicating a reasonable predictive accuracy. However, the HWM models consistently outperformed the SVR counterparts, with HWM-M3 highlighting the highest R^2 values in both training (0.8279) and testing (0.7779) phases. It further demonstrated the lowest MSE and RMSE in both training and testing, highlighting its superior predictive performance compared to other models. The potential of HWM-M3 in particular, showed superior predictive capabilities which we believe could be utilized to assure sustainable irrigation practices. Moreover, the comparison of these models further pondered on the need for regularization for SVR models. It is believed that the emphasis should be placed on refining AI models, including exploring novel hybrid models and incorporating additional relevant features for improved predictive accuracy. Future studies may also explore the impact of climate change conditions on

TWW salinity dynamics, considering potential shifts in agricultural practices. Additionally, investigating the economic feasibility and social acceptance of adopting these predictive models in diverse agricultural contexts will contribute to their successful implementation. Comparative analyses across various arid regions would identify unique challenges and opportunities, guiding the development of tailored irrigation strategies. This notable study aims to inform policy, optimize water resource management, and support the development of smart, automated irrigation systems, ultimately contributing to agricultural sustainability and food security in arid landscapes.

Acknowledgements
The authors would like to acknowledge all support from the Interdisciplinary Research Center for Membranes and Water Security, King Fahd University of Petroleum and Minerals.

Author contributions
Author Contributions: Conceptualization, SMHS, SIA, and MAY; methodology, SIA, SMHS, DUL and FA; software, SMHS, SIA, and DUL; validation, IHA, FA, and HAA; formal analysis, SIA, EHHA and MAY; investigation, SMHS and SIA; resources, MAY; data curation, HUQ and DUL; writing—original draft preparation, SIA, MAY and SMHS; writing—review and editing, IHA, HAA, MS, and SSS; visualization, HAA, SSS and IHA; supervision, IHA, and SSS; project administration, MAY; funding acquisition, MAY, MS, and SSS. All authors have read and agreed to the published version of the manuscript.

Funding
This research was funded by the Deanship of Research Oversight and Coordination (DROC) at King Fahd University of Petroleum & Minerals (KFUPM) under the Interdisciplinary Research Center for Membranes and Water Security [Grant Number: INMW2314].

Data availability
No datasets were generated or analysed during the current study.

Declarations

Competing interests
The authors declare no competing interests.

Author details
[1]Interdisciplinary Research Center for Membranes and Water Security, King Fahd University of Petroleum and Minerals, 31261 Dhahran, Saudi Arabia. [2]Department of Geosciences, College of Petroleum Engineering & Geosciences, King Fahd University of Petroleum & Minerals, 31261 Dhahran, Saudi Arabia. [3]Center of Excellence in Development of Non-Profit Organizations, King Fahd University of Petroleum and Minerals, 31261 Dhahran, Saudi Arabia. [4]Eco Hydrology Technology Research Centre (Eco-Hytech), Faculty of Civil Engineering & Built Environment, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia. [5]Department of Civil Engineering, Sir Syed University of Engineering and Technology, Karachi 75300, Pakistan. [6]Department of Chemical Engineering, King Fahd University of Petroleum and Minerals, 31261 Dhahran, Saudi Arabia. [7]Interdisciplinary Research Centers of Hydrogen Technologies and Carbon Management, King Fahd University of Petroleum and Minerals, 31261 Dhahran, Saudi Arabia. [8]Department of Civil Engineering, College of Engineering, University of Diyala, Diyala Governorate, Baqubah 32001, Iraq. [9]Department of Urban Drainage, Bau & Service Oberursel (BSO), Postfach 1280, 61402 Oberursel (Taunus), Germany. [10]Department of Civil Engineering Science, School of Civil Engineering, and the Built Environment, Faculty of Engineering and the Built Environment, University of Johannesburg, Kingsway Campus, PO Box 524, Aukland Park, Johannesburg 2006, South Africa. [11]Kunststoff-Technik Adams, Specialist Company According to Water Law, Schulstraße 7, 26931 Elsfleth, Germany. [12]Nexus by Sweden, Skepparbacken 5, 722 11 Västerås, Sweden. [13]Department of Town Planning, Engineering Networks and Systems, South Ural State University (National Research University), 76, Lenin prospekt, Chelyabinsk 454080, Russian Federation.

Received: 8 February 2024 Accepted: 17 April 2024
Published online: 12 June 2024

References
1. El Bilali A, Taleb A (2020) Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. J Saudi Soc Agric Sci 19(7):439–451. https://doi.org/10.1016/j.jssas.2020.08.001
2. Kalavrouziotis IK et al (2015) Current status in wastewater treatment, reuse and research in some mediterranean countries. Desalin Water Treat 53(8):2015–2030. https://doi.org/10.1080/19443994.2013.860632
3. Alzahrani F, Elsebaei M, Tawfik R (2023) Public acceptance of treated wastewater reuse in the agricultural sector in Saudi Arabia. Sustainability 15(21):15434. https://doi.org/10.3390/su152115434
4. UN-Water, "Wastewater: The Untapped Resource," 2017.
5. Shtull-Trauring E, Cohen A, Ben-Hur M, Tanny J, Bernstein N (2020) Reducing salinity of treated waste water with large scale desalination. Water Res. 186:116322. https://doi.org/10.1016/j.watres.2020.116322
6. Nirit B (2009) Contamination of soils with microbial pathogens originating from effluent water used for irrigation. Contam Soils Environ Impact, Dispos Treat 1:473–486
7. WHO, "Reuse of Effluents-Methods of Wastewater Treatment and Health Safeguards, Report of a WHO Meeting of Experts," Geneva, Switzerland, 1973.
8. WHO, "WHO Guidelines for the Safe Use of Wastewater Excreta and Greywater," Geneva, Switzerland, 2006.
9. WHO, "HealthGuidelines for the Use of Wastewater in Agriculture and Aquaculture," Geneva, Switzerland, 1989.
10. Carr R (2005) WHO guidelines for safe wastewater use—more than just numbers. Irrig Drain 54(S1):S103–S111. https://doi.org/10.1002/ird.190
11. Mara D, Kramer A (2008) The 2006 WHO guidelines for wastewater and greywater use in agriculture: a practical interpretation. Efficient Management of Wastewater, Berlin, Heidelberg: Springer, Berlin Heidelberg 1:1–17. https://doi.org/10.1007/978-3-540-74492-4_1
12. Mara DD, Sleigh A, Blumenthal UJ, Carr RM (2007) Health risks in wastewater irrigation: Comparing estimates from quantitative microbial risk analyses and epidemiological studies. J Water Health 5(1):39–50. https://doi.org/10.2166/wh.2006.055
13. FAO, *Water Quality for Agriculture*. Rome, Italy, 1985
14. FAO, "Waste-Water Treatment and Use in Agriculture," Rome, Italy, 1992
15. USEPA, "Guidelines for Water Reuse," Washington, D.C - USA, 2004
16. USEPA, "Guidelines for Water Reuse," Washington, D.C - USA, 2012
17. Jaramillo M, Restrepo I (2017) Wastewater reuse in agriculture: a review about its limitations and benefits. Sustainability 9(10):1734. https://doi.org/10.3390/su9101734
18. Al-Saikhan MS, Badr ESA, Babeker MY (2020) Study of sewage sludge use for the cultivation of plants and its effects on soil properties in Al Ahsa. Sci J King Faisal Univ 21(2):1–7
19. Badr E-SA, Tawfik RT, Alomran MS (2023) An assessment of irrigation water quality with respect to the reuse of treated wastewater in Al-Ahsa Oasis, Saudi Arabia. Water 15(13):2488. https://doi.org/10.3390/w15132488
20. Silva JA (2023) Wastewater treatment and reuse for sustainable water resources management: a systematic literature review. Sustainability 15(14):10940. https://doi.org/10.3390/su151410940
21. Al-Aizari H, Fegrouche R, Al-Aizari A, Darwsh N, Al-Kadasi F, Chaouch A (2020) Assessment of treated wastewater quality and impact of using it on the soil in Wadi Al-Mawaheb-Dhamar, Republic of Yemen. Egypt J Aquat Biol Fish 24(2):535–547. https://doi.org/10.21608/ejabf.2020.87820
22. Shams MY, Elshewey AM, El-kenawy E-SM, Ibrahim A, Talaat FM, Tarek Z (2023) Water quality prediction using machine learning models based on grid search method. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-16737-4
23. Hejabi N, Saghebian SM, Aalami MT, Nourani V (2021) Evaluation of the effluent quality parameters of wastewater treatment plant based on

uncertainty analysis and post-processing approaches (case study). Water Sci Technol 83(7):1633–1648. https://doi.org/10.2166/wst.2021.067
24. Mustafa H, Hayder G, Abba S, Algarni A, Mnzool M, Nour A (2023) Performance evaluation of hydroponic wastewater treatment plant integrated with ensemble learning techniques: a feature selection approach. Processes 11(2):478. https://doi.org/10.3390/pr11020478
25. Banadkooki FB, Ehteram M, Panahi F, Sammen SS, Othman FB, El-Shafie A (2020) Estimation of total dissolved solids (TDS) using new hybrid machine learning models. J Hydrol 587:124989
26. Abba SI, Nourani V, Elkiran G (2019) Multi-parametric modeling of water treatment plant using AI-based non-linear ensemble. J Water Supply Res Technol 68(7):547–561. https://doi.org/10.2166/aqua.2019.078
27. Mokhtar A, Elbeltagi A, Gyasi-Agyei Y, Al-Ansari N, Abdel-Fattah MK (2022) Prediction of irrigation water quality indices based on machine learning and regression models. Appl Water Sci 12(4):76. https://doi.org/10.1007/s13201-022-01590-x
28. Hamada MS, Zaqoot HA, Sethar WA (2024) Using a supervised machine learning approach to predict water quality at the Gaza wastewater treatment plant. Environ Sci Adv 3(1):132–144. https://doi.org/10.1039/D3VA00170A
29. M. Yassin, S. Isah Abba, A. Garba Usman, and I. Aljundi, "Spatiotemporal and hydrogeological assessment of groundwater supported by soft computing modeling of heavy metal in Al-Hassa, Eastern Province, Saudi Arabia," 2023. https://www.researchgate.net/publication/369031006
30. Mishra S, Kumar R, Kumar M (2023) Use of treated sewage or wastewater as an irrigation water for agricultural purposes- Environmental, health, and economic impacts. Total Environ Res Themes 6:100051. https://doi.org/10.1016/j.totert.2023.100051
31. Gaya MS et al (2017) Estimation of turbidity in water treatment plant using Hammerstein–Wiener and neural network technique. Indones J Electr Eng Comput Sci 5(3):666. https://doi.org/10.11591/ijeecs.v5.i3.pp666-672
32. Nourani V, Elkiran G, Abba SI (2018) Wastewater treatment plant performance analysis using artificial intelligence—an ensemble approach. Water Sci Technol 78(10):2064–2076. https://doi.org/10.2166/wst.2018.477
33. Pham QB et al (2019) Potential of hybrid data-intelligence algorithms for multi-station modelling of rainfall. Water Resour Manag 33(15):5067–5087. https://doi.org/10.1007/s11269-019-02408-3
34. Wang X et al (2020) A hybrid model for prediction in asphalt pavement performance based on support vector machine and grey relation analysis. J Adv Transp 1–14:2020. https://doi.org/10.1155/2020/7534970
35. Abdi MJ, Giveki D (2013) Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules. Eng Appl Artif Intell 26(1):603–608. https://doi.org/10.1016/j.engappai.2012.01.017
36. Liu Z, Cao H, Chen X, He Z, Shen Z (2013) Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. Neurocomputing 99:399–410. https://doi.org/10.1016/j.neucom.2012.07.019
37. Abba SI et al (2022) Integrating feature extraction approaches with hybrid emotional neural networks for water quality index modeling. Appl Soft Comput 114:108036. https://doi.org/10.1016/j.asoc.2021.108036
38. Maharana K, Mondal S, Nemade B (2022) A review: data pre-processing and data augmentation techniques. Glob Transitions Proc 3(1):91–99. https://doi.org/10.1016/j.gltp.2022.04.020
39. Wu CL, Chau KW, Fan C (2010) Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. J Hydrol 389(1–2):146–167. https://doi.org/10.1016/j.jhydrol.2010.05.040
40. Mehdizadeh S, Ahmadi F, Kouzehkalani A (2023) Development of wavelet-based hybrid models to enhance daily soil temperature modeling: application of entropy and τ-Kendall pre-processing techniques. Stoch Environ Res Risk Assess. 37(2):507–526. https://doi.org/10.1007/s00477-022-02268-1
41. Ghali UM et al (1871) Advanced chromatographic technique for performance simulation of anti-Alzheimer agent: an ensemble machine learning approach. SN Appl Sci 2(11):2020. https://doi.org/10.1007/s42452-020-03690-2
42. Poursaeid M, Mastouri R, Shabanlou S, Najarchi M (2020) Estimation of total dissolved solids, electrical conductivity, salinity and groundwater levels using novel learning machines. Environ Earth Sci 79(19):453. https://doi.org/10.1007/s12665-020-09190-1
43. Mosavi A et al (2021) Susceptibility mapping of groundwater salinity using machine learning models. Environ Sci Pollut Res 28(9):10804–10817. https://doi.org/10.1007/s11356-020-11319-5
44. Tran DA et al (2021) Evaluating the predictive power of different machine learning algorithms for groundwater salinity prediction of multi-layer coastal aquifers in the Mekong Delta, Vietnam. Ecol Indic 127:107790. https://doi.org/10.1016/j.ecolind.2021.107790
45. Trabelsi F, Bel-Hadj-Ali S (2022) Exploring machine learning models in predicting irrigation groundwater quality indices for effective decision making in Medjerda River Basin, Tunisia. Sustainability 14(4):2341. https://doi.org/10.3390/su14042341

Publisher's Note