Water Science & Technology

IWA PUBLISHING

Water Science & Technology Vol 87 No 11, 2793 doi: 10.2166/wst.2023.161

Optimization of water reuse and modelling by saline composition with nanoparticles based on machine learning architectures

Koppula Srinivas Rao[a], Vineet Tirth[b,c], Hamad Almujibah[d], Abdullah H. Alshahri[d], V. Hariprasad[e] and N. Senthilkumar [f,*]

[a] Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad, Telangana, India
[b] Mechanical Engineering Department, College of Engineering, King Khalid University, Abha, Asir 61421, Saudi Arabia
[c] Research Center for Advanced Materials Science (RCAMS), King Khalid University, Guraiger, P.O. Box 9004, Abha, Asir 61413, Saudi Arabia
[d] Department of Civil Engineering, College of Engineering, Taif University, P.O. Box 11099, Taif City 21974, Saudi Arabia
[e] Department of Aerospace Engineering, Jain (Deemed-to-be) University, Jain Global Campus, Jakkasandra Post, Kanakapura 562112, India
[f] Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Thandalam, Chennai 602105, India
*Corresponding author. E-mail: nskmfg@gmail.com

NS, 0000-0002-2441-1061

ABSTRACT

Water is a necessary resource that enables the existence of all life forms, including humans. Freshwater usage has become increasingly necessary in recent years. Facilities for treating seawater are less dependable and effective. Deep learning methods have the ability to improve salt particle analysis in saltwater's accuracy and efficiency, which will enhance the performance of water treatment plants. This research proposes a novel technique in optimization of water reuse with nanoparticle analysis based on machine learning architecture. Here, the optimization of water reuse is carried out based on nanoparticle solar cell for saline water treatment and the saline composition has been analyzed using a gradient discriminant random field. Experimental analysis is carried out in terms of specificity, computational cost, kappa coefficient, training accuracy, and mean average precision for various tunnelling electron microscope (TEM) image datasets. The bright-field TEM (BF-TEM) dataset attained a specificity of 75%, kappa coefficient of 44%, training accuracy of 81%, and mean average precision of 61%, whereas the annular dark-field scanning TEM (ADF-STEM) dataset produced specificity of 79%, kappa coefficient of 49%, training accuracy of 85%, and mean average precision of 66% as compared with the existing artificial neural network (ANN) approach.

Key words: machine learning, nanoparticle analysis, optimization, solar cell, water reuse, water treatment

HIGHLIGHTS

- Deep learning methods have the ability to improve salt particle analysis in saltwater's accuracy and efficiency, which will enhance the performance of water treatment plants.
- This research proposes a novel technique in the optimization of water reuse with nanoparticle analysis based on machine learning architecture. Here, the optimization of water reuse is carried out based on nanoparticle solar cells for water treatment.

1. INTRODUCTION

Disposal of solid wastes by landfill is a more widely utilized disposal technique compared to others due to its economic advantages as well as simplicity (Rajput *et al.* 2022). Leachate develops in landfill areas as a result of water's impact on waste composition, precipitation, and physical, chemical, and biological processes. Wastewater with a high level of contaminants, including monovalent and multivalent ions, bacteria, and heavy metals, is referred to as leachate. Chemical composition of leachate is influenced by the type of trash in the landfill, waste content, landfill age, and environmental conditions (Zhong *et al.* 2022). Chemical oxygen demand (COD), total phosphorus (TP), and organic nitrogen concentrations in normal leachate vary between 140 and 152,000 mg/L, 0.1 and 30 mg/L, and 14 and 2,500 mg/L. In landfill regions, liners with low hydraulic conductivity ($k \leq 10 - 7$ cm/s) are employed to stop leachate from getting into groundwater. Numerous variables, including void ratio, medium porosity, pore size, soil type, liquid type, and contaminant concentration, all affect hydraulic conductivity (Fu *et al.* 2021). The structure and characteristics of the clay are altered throughout the leachate's

passage through the clayey soil, altering its hydraulic conductivity. The leachate's increased ion concentration, ionic valence, and organic matter concentration all help to boost hydraulic conductivity, while its increased microbial population and concentration of suspended solids help to lower it. Some of the factors that contribute to improvement in hydraulic conductivity as a result of leachate-liner interaction include thinned diffuse double layer (DDL) thickness of the clay, agglomeration, flocculation of clay particles, formation of cracks in soil structure, and a rise in void ratio. Growing global population increases water usage and daily pollution production (Alshehri *et al.* 2021). Approximately 2.4 billion people worldwide do not have access to reliable sanitation services, both in urban and rural regions. Within 20 years, 2 billion more people are expected to reside in towns and cities, mostly in developing nations, increasing the demand for sanitation. Sewage is released untreated and damaging lakes, rivers, and coastal areas in more than 90% of developing nations. Traditional sanitation practises, which rely on water-wasting toilets, are not an appropriate answer for industrialized and developing nations. The sanitation systems were created and planned under the assumption that human excreta was a type of waste that could be disposed of in the environment and be assimilated by it. EcoSan is a cycle of a sustainable closed system used in sanitation systems that can bridge the gap between agriculture and sanitation. The EcoSan strategy emphasizes resource management and offers a comprehensive approach to environmentally and financially sound sanitation. The major goal is to promote sustainable development by reducing nutrient and water cycles while using less resources and energy. The EcoSan system is a different method for avoiding the drawbacks of traditional wastewater systems (Melesse *et al.* 2020).

Smart technology and newly developed artificial intelligence (AI) and machine learning (ML) are filling a void in water applications that were previously left unfilled by conventional methods and ways of thinking. Smart technology and newly developed AI and ML are filling a void in water applications that were previously left unfilled by conventional methods as well as ways of thinking. According to some projections, the water industry will spend up to 10% of over $90 billion in investments that are anticipated to mature by 2030 on AI. Because of its universality, resilience, and relative ease of design, AI, ML, and smart methods are projected to methods as well as solve complex and challenging problems in water applications, resulting in cost savings and process optimization. Water as well as wastewater treatment, monitoring of natural methods, and precision/water-based agriculture are a few water-related applications that have seen substantial ML usage. These industry studies rely on a variety of ML methods, with the most popular ones being artificial neural network (ANN), recurrent neural network (RNN), radio frequency (RF), support vector machine (SVM), and adaptive neuro-fuzzy inference system (ANFIS), with sporadic use of AI techniques like fuzzy inference systems (FISs). ANN-RF and SVM-RF are two examples of hybrid algorithms that combine two ML systems in applications. Studies have shown success using both AI and ML to optimize modelling processes in water-based settings (Chang *et al.* 2022).

Contribution of this research is as follows:

1. To propose a novel technique in the optimization of water reuse with nanoparticle analysis based on ML architecture.
2. Optimization of water reuse is carried out based on nanoparticle solar cells for water treatment.
3. Saline composition has been analyzed using a gradient discriminant random field based on saline water treatment.

2. RELATED WORKS

Intelligent scissors, distance regularized level set evolution, and graph cuts are some recently developed advanced segmentation algorithms that have enhanced performance, which reduces high-throughput advantage of computerized data analysis. The concentration of dissolved oxygen (DO) in wastewater can be significantly reduced by a high temperature, which causes aquatic species to perish (Aghilesh *et al.* 2021). A drop in DO will cause nitrogen to be transformed into nitrate molecules, which will also produce unpleasant odours. The wastewater's odour and health may be harmed if the hue gradually turns from grey to blackish brown (Wang *et al.* 2021). Wastewater must be treated before being released freely into a body of water in order to protect public health (Jiang *et al.* 2021). Water is necessary for human living, but wastewater needs our attention because it is the primary source of most diseases. Additionally, the main source of sickness is water (Wei *et al.* 2022). 80% of disease outbreaks in 2002 were attributed to water contamination by microbes, according to WHO data (Wang *et al.* 2020). Because losses are a result of the uncertainty introduced by risks, risk management is essential for lowering losses (Li *et al.* 2021). Effective management is required to reduce these risks. As part of this management, procedures must be carried out effectively in order to accomplish an objective. Pre-analysis, prediction, and control are all effective ways to lower possible risks (Nam *et al.* 2022). Therefore, risk management is required to lower the likelihood of failure and offer convenience for handling prospective risks. The methodical process of identifying environmental threats, reviewing probable

outcomes, and controlling the amount of environmental risk is one specific illustration of environmental risk management (Astray *et al.* 2021). The four steps of risk management for the environment are issue formulation, risk analysis, risk characterization, and risk management (Zhao *et al.* 2020). Large datasets can be efficiently classified using ML techniques in a variety of domains, including biology and text mining (Alqahtani *et al.* 2022). Particle shape categorization is now accomplished using deep learning techniques like CNN algorithms and ANN (Bagheri *et al.* 2019). This method produces good results for particular particle forms, but in order to produce training datasets, it requires prior knowledge about the latter. An unsupervised ML technique, which does not need prior input or data training, is preferable to accomplish generalization in data analysis (Hoa *et al.* 2019). Unsupervised ML techniques may effectively evaluate the size as well as shape information of nanoparticles, according to recently published studies; however, they are only applicable to nanoparticles with distinct shapes, significant visual contrast, and good dispersions, such as gold nanorods (isolated particles). Due to a variety of nanoparticle compositions, dispersions, and imaging settings, it is actually challenging to ensure a sharp contrast as well as a homogeneous background in energy management images (Zeng *et al.* 2018; Chen *et al.* 2021).

Wastewater processing plants can benefit from adaptive ML models by minimizing downtime and increasing profits through more efficient management. The use of an ensemble technique increased prediction accuracy by 5% when compared to using only the individual base models (Zaghloul & Achari 2022). Controlling chlorination using AI approaches has been shown to be successful while predicting disinfection by-products (DBP) concentrations and key parameters for adsorption and membrane-filtration processes using ML models has shown similar promise. By establishing the connection between input and output among system variables, we may improve AI and ML application efficiency that suffers from differences in the quality of data used for testing and training (Lowe *et al.* 2022). Successful real-time surveillance, optimization, uncertainty prediction, and fault identification for intricate environmental systems are all possible due to ML algorithms, which proved useful in predicting the uncertain behaviours of treatment procedures. Hyperparameter tweaking is used during the machine learning algorithm (MLA) selection process to discover the optimal solution in the least amount of time while using the fewest available computing resources (Sundui *et al.* 2021).

From the related study findings, it was found that there is a gap in the application of nanoparticles in solar panels to improve water reuse and to determine the saline composition. Also, the application of ML techniques in water reuse is very limited. Incorporating these approaches in the present study makes it novel and distinct with the adoption of datasets from bright-field tunnelling electron microscope (BF-TEM) and annular dark-field scanning TEM (ADF-STEM).

3. SYSTEM MODEL

This section discusses a novel method in the optimization of water reuse with nanoparticle analysis based on ML architecture. Here, the optimization of water reuse is carried out based on nanoparticle solar cells for water treatment. For the aforementioned reasons of easiness, acceptable accuracy, speed, and high performance without the requirement to be skilled professionals in physical issues, ML models are increasingly being utilized to handle most of the complicated practical, real-world problems faced in numerous domains of life (Sarker 2021). The nanoparticles have a porous structure that allows them to absorb large amounts of water while retaining their ability to reject dissolved salts and other contaminants. Nanoparticles, due to their low bulk density, active surface state, and distinctive photoluminescent, and biocompatible features, nanoparticles can take in infrared light and generate visible light, which solar cells can then convert into extra electricity. A considerable increase in photocurrent is achieved when light that hits on the solar panel strikes the nanoparticles, scatters, and may hit additional nanoparticles and bounce several times within the system (Olaimat *et al.* 2021). The saline composition has been analyzed using a gradient discriminant random field based on saline water treatment. The proposed architecture is shown in Figure 1.

Pre-processing begins with separating the image's background from the particles. Setting a global brightness threshold to remove the backdrop from the image is one approach that is frequently used to do this. This method takes advantage of contrast in brightness between particles as well as background in TEM image. The inhomogeneity of the background presents a problem non-depending on this method to create accurate background-removed photos. This inhomogeneity might be caused by differences in the sample backdrop matrix's thickness, surface unevenness, or by uneven electron microscope illumination. When a global filtering threshold is used, inhomogeneous background causes regional background to be both over- as well as under-filtered in the removal method. Because of this, the technique either fails to effectively isolate particles or introduces

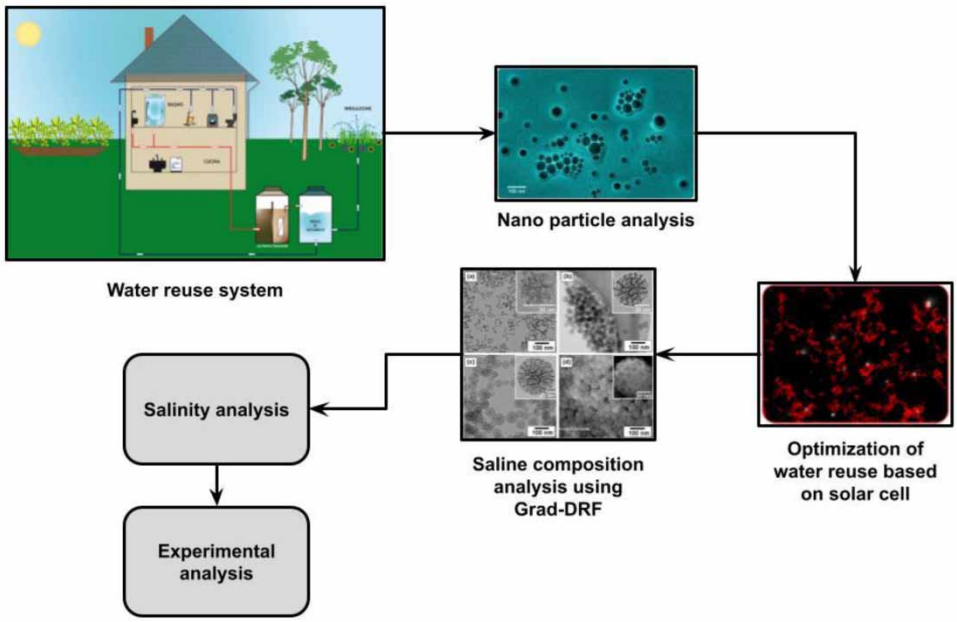

Figure 1 | Proposed architecture.

artefacts, which distort shapes. Such insufficient background removal can compromise the precision of particle shape data (Wen *et al.* 2021).

Solar energy can be absorbed by PV cells. Due to the fact that they are semiconductor devices, they may transform the absorbed energy into useful electrical power. P-N junction diodes are used in the construction of the PV module structure. These diodes can transform incident light that reaches their surface into electrical energy (Pastuszak & Węgierek 2022). Figure 2 illustrates the fundamental design, wiring, and operation of a PV module:

In constructing the PV, there are two layers of silicon present: a negative N layer and a positive P layer that is doped with boron. The tempered glass-coated PV module absorbs solar energy when it is exposed to the sun. After a certain amount of time, the energy absorbed exceeds the band gap energy level, which causes electrons to travel from conduction band to valence band via that band. Due to this, electrons in the conduction band can flow freely and form electron–hole pairs.

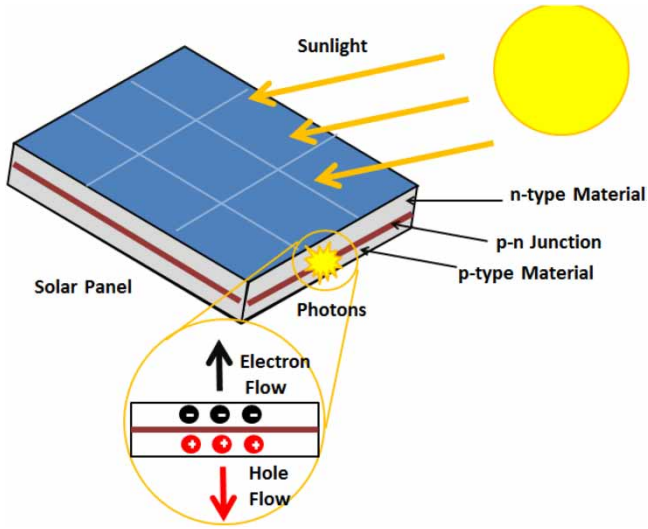

Figure 2 | Working of photovoltaic.

Since the flow of electric current is caused by the movement of electrons, the electricity produced during the process is used to power the load. In order to generate adequate electricity, an array configuration is insufficient because these systems experience numerous losses (Al-Ezzi & Ansari 2022). The ideal method for maximizing each string's effectiveness is MPPT (maximum power point tracking). The most power possible from PV modules can be obtained by using such control approaches.

The energy management system (EMS) keeps state of charge (SOC) between SOC_{max} and SOC_{min} (20–90%) within a predetermined range. The EMS operating techniques implemented state that generating power from solar panels and wind turbines takes precedence over other sources of energy in order to meet load needs. The excess power will be used to charge the battery with a reference value if the amount of power supplied by the PV and wind is greater than the amount needed to power the load, or $P_{PV} + P_{Wind} > P_{Load}$ by Equation (1):

$$P_{B-ref} = -P_{Load} + (P_{PV} + P_{Wind}) \tag{1}$$

where P_{PV} and P_{Wind} are the power generated by the PV and wind turbines, respectively, and $P_{B\text{-}ref}$ is the reference battery power that will be used to charge the battery. Power demand is P_{Load}. The battery keeps charging until its state of charge, or SOC, reaches its maximum point, or SOC_{max}, which is 90%. The electrolyzer (ELZ) will then use the remaining power to produce hydrogen for the solid oxide fuel cell (SOFC) until the pressure reaches its maximum level. The extra energy produced will then be exported to the grid. The energy storage system's overall size and weight can be decreased by using a DC/DC converter with an incorporated magnetic structure. Additionally, a magnetic structure that is integrated helps lessen the output current ripple (Kumar *et al.* 2020).

Referencing overall DC-link voltage and energy time-variations should be done to emphasize the workings of the proposed hybrid energy storage electric converter (HESEC). These can be accomplished using Equations (2) and (3):

$$C_{DC}\frac{dV_{DC}}{dt} = i_B - i_{DC}\ ,\ \ i_{DC} = c_H i_H + c_L i_L \tag{2}$$

$$\frac{dE_{DC}}{dt} = V_{DC}(i_B - i_p)\ ,\ i_p = v_H i_H + v_L i_L \tag{3}$$

where the equivalent DC-link and power currents, i_{DC} and i_p, are proportional to energy exchanged by electrical drive and DC-link. In reference to (5), c_H and c_L are per-unit capacitances in Equation (4), and C_{DC} stands for the equivalent DC-link series capacitance:

$$C_{DC} = \frac{C_H C_L}{C_H + C_L},\ \ c_H = \frac{C_H}{C_H + C_L},\ \ c_L = \frac{C_L}{C_H + C_L} \tag{4}$$

The per-unit capacitor voltages are represented by v_H and v_L in Equation (5):

$$v_H = \frac{V_H}{V_H + V_L}\ ,\ \ v_L = \frac{V_L}{V_H + V_L} \tag{5}$$

Still referring to both (4) and (5), it can be said that if suitable combinations of (c_H,c_L) and (v_H,v_L) exist, V_{DC} and E_{DC} can be changed separately. DC-link energy base value might be provided as Equation (6) to support the preceding claim:

$$\hat{E}_{DC} = \frac{l}{2} C_L V_{DC}^2 \tag{6}$$

The per-unit energy content of the DC-link can thus be simply obtained as Equation (7) by multiplying (4) by (7):

$$e_{\mathrm{DC}} = \zeta v_H^2 + (I - v_u)^2 \tag{7}$$

in which ζ denotes unbalanced capacitance factor:

$$\zeta = \frac{C_H}{C_L} \tag{8}$$

Therefore, by altering the continuous operating (Uc) voltage within a specific range, one can usefully introduce DC-link voltage unbalances while maintaining the total DC-link voltage. In order to regulate the energy flows among each HESEC component optimally, proper DC-link energy and voltage rules are required. Solar panels are used to gather this electricity from the ambient sunlight. Light energy is converted directly into DC electrical energy by the solar panel. To charge the battery, the DC–DC converter adjusts the DC voltage. The wireless sensor networks (WSN) node is powered by a rechargeable battery. The sensor measurement unit is used by the WSN node to measure the desired physical quantity. This data is processed by a microcontroller in the computing unit (Akgül *et al.* 2018). Figure 3 represents the solar energy harvesting methodology.

Gradient discriminant random field-based saline water composition analysis:

Variations in batch-wise training come from the gradient variance. The use of a random sample has the advantage of requiring much fewer computations per iteration while the disadvantage is noisy gradient. Note that iterations are used to calculate the convergence rate in this section. We must define the Lyapunov process by Equation (9) in order to study the training dynamics per iteration (Hou & Yi 2020):

$$h_t = \left\| \mathbf{w}^t - \mathbf{w}^* \right\|_2^2 \tag{9}$$

The variable h_t is a random number, and the equation evaluates the separation between present solution, $\mathbf{w}^t$, and ideal solution, $\mathbf{w}^*$. Consequently, using Equations (10) and (11), it is possible to determine the stochastic gradient descent's (SGD)

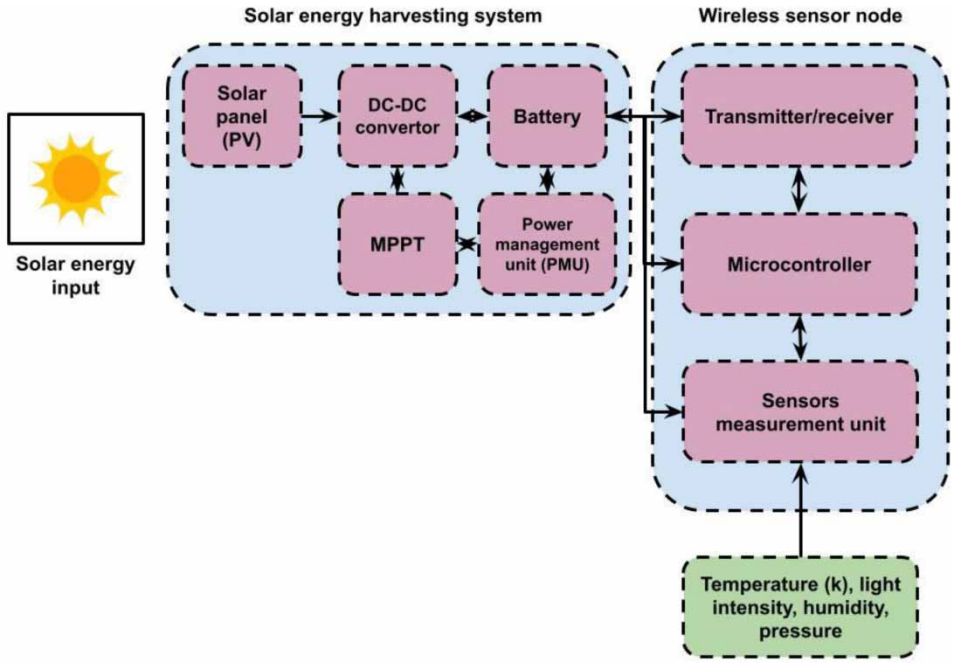

Figure 3 | Block diagram of solar energy harvesting.

convergence rate (Hu *et al.* 2021):

$$h_{t+1} - h_t = \left\|\mathbf{w}^{t+1} - \mathbf{w}^*\right\|_2^2 - \left\|\mathbf{w}^t - \mathbf{w}^*\right\|_2^2 = (\mathbf{w}^{t+1} + \mathbf{w}^t - 2\mathbf{w}^*)(\mathbf{w}^{t+1} - \mathbf{w}^t) = (2\mathbf{w}^t - 2\mathbf{w}^* - \eta_t \nabla\psi_{\mathbf{w}}(\mathbf{d}_t))(-\eta_t \nabla\psi_{\mathbf{w}}(\mathbf{d}_t)) \tag{10}$$

$$= -2\eta_t(\mathbf{w}^t - \mathbf{w}^*)\nabla\psi_{\mathbf{w}}(\mathbf{d}_t) + \eta_t^2(\nabla\psi_{\mathbf{w}}(\mathbf{d}_t))^2 \tag{11}$$

It implies how much progress can be made in one iteration. The convergence rate is improved by decreasing VAR$\{\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)\}$. Average convergence rate at an iteration's precision is produced by the expectation of Equations (12) and (13):

$$\mathbf{E}\{h_{t+1} - h_t\} = -2\eta_t(\mathbf{w}^t - \mathbf{w}^*)\mathbf{E}\{\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)\} + \eta_t^2\mathbf{E}\{(\nabla\psi_{\mathbf{w}}(\mathbf{d}_t))^2\} \tag{12}$$

$$= -2\eta_t(\mathbf{w}^t - \mathbf{w}^*)\mathbf{E}|\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)\} + \eta_t^2(\mathbf{E}\{\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)\})^2 + \mathbf{VAR}\{\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)\} \tag{13}$$

Let's assume that $\psi_{\mathbf{w}}(\mathbf{d}_t)$ is convex so as to make the analysis of Equation (14) simpler.

$$h_{t+1} - h_t < 0 - (\mathbf{w}^t - \mathbf{w}^z)\mathbf{E}\{\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)\} < 0 \tag{14}$$

The objective estimation of $\mathbf{E}\{\nabla\psi_{\mathbf{w}}(\mathbf{d})\}$ is $\mathbf{E}\{\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)\}$. Therefore, increasing an iteration's contribution is equivalent to minimizing VAR$\{\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)\}$. This angle has been adequately covered. With respect to $\mathbf{d}_t$, an iteration's contribution, h_{t+1} h_t, changes. Equations (15) and (16) show the following as variance of h_{t+1} h_t:

$$\mathrm{VAR}\{h_{t+1} - h_t\} = 4\eta_t^2(\mathbf{w}' - \mathbf{w}^*)^2\mathrm{VAR}\{\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)\} + \eta_t^4\mathbf{VAR}\{(\nabla\psi_{\mathbf{w}}(\mathbf{d}_t))^2\} \tag{15}$$

$$- 2\eta_t^3(\mathbf{w}_t - \mathbf{w}_z)\mathbf{COV}(\nabla\psi_{\mathbf{w}}(\mathbf{d}_t), \nabla\psi_{\mathbf{w}}(\mathbf{d}_t)^2) \tag{16}$$

The equation shows that VAR$\{h_{t+1} - h_t\}$, 0, indicating that gradient updates do not contribute uniformly. It's intriguing to note that $\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)$ 2 and $\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)$, the deciding variables in this equation, depend on $\mathbf{d}_t$, indicating a relationship between h_{t+1} h_t and $\mathbf{d}_t$.

This original idea drives us to research the variables in $\mathbf{d}_t$ that influence convergence rate h_{t+1} h_t and how to solve load balance issues in training. There has been numerous research on the variance reduction on $\nabla\psi_{\mathbf{w}}(\mathbf{d}_t)$, but few have gone in this way. Every iteration k, the update shown in Equation (17) is carried out:

$$z(k+1) = z(k) - \alpha_k\bar{g}(k) \tag{17}$$

where stepsizes satisfy $\alpha_k = 1/\mu k$, and $\bar{g}(k) = (1/n)\sum_{i=1}^n g_i(z(k), \xi_i(k))$, i.e., $\bar{g}(k)$ is average of n noisy gradients evaluated at $z(k)$. As a result, using more than one gradient improves the accuracy of the gradient estimation. In fact, by Equation (18) is what we have from assumption:

$$\mathbb{E}\,\|\,\bar{g}(k) - \nabla f(z(k))\|^2] = \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\left[\|g_i(z(k), \xi_i(k)) - \nabla f_i(z(k))\|^2\right] \le \frac{\sigma^2}{n} \tag{18}$$

We will focus on two error terms, which are different from analysis for SGD. Expected squared distance between $z(k)$ and z^* is defined by the first term, called expected optimization error, and expected consensus error, or expected $\mathbb{E}\left[\|\bar{z}(k) - z^*\|^2\right]$ error, gauges disparities of individual estimates among all agents. Equation (19) specifies the average squared distance between each iterate $z_i(k)$ and ideal z:

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\|z_i(k) - z^*\|^2\right] = \mathbb{E}\left[\|\bar{z}(k) - z^*\|^2\right] + \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\|z_i(k) - \bar{z}(k)\|^2\right] \tag{19}$$

Therefore, examining the two terms will give us information about how distributed stochastic gradient descent (DSGD) is performing. Denote $U(k) = \mathbb{E}\left[\|\bar{z}(k) - z^*\|^2\right], V(k) = \sum_{i=1}^{n} \mathbb{E}\left[\|z_i(k) - \bar{z}(k)\|^2\right], \forall k$ to simplify the notation., $\forall k$ We start by looking for an inequality that limits $U(k)$, which is comparable to $\mathbb{E}\left[\|z(k) - z^*\|^2\right]$ in SGD and was inspired by the study for SGD. Equation turns out to be one of these relationships as in Equation (20):

$$U(k+1) \leq \left(1 - \frac{1}{k}\right)^2 U(k) + \frac{2L}{\sqrt{n}\mu}\frac{\sqrt{U(k)V(k)}}{k} + \frac{L^2}{n\mu^2}\frac{V(k)}{k^2} + \frac{\sigma^2}{n\mu^2}k^2 \tag{20}$$

The anticipated consensus error $V(k)$ representing the extra disturbances brought on by the variations in solutions. Two extra terms are probably insignificant in the long run, though, assuming $V(k)$ decays quickly enough relative to $U(k)$, and we would infer that convergence rate of $U(k)$ is similar to $R(k)$ for SGD. $V(k) \leq \mathcal{O}(n/(1-\lambda)^2)(1/k^2)$ for $k \geq \mathcal{O}(1/(1-\lambda))$. Plugging this into Equation (21) leads to inequality $U(k) \leq \theta^2\sigma^2/((1.5\theta - 1)n\mu^2 k) + \mathcal{O}(1/(1-\lambda)^2)(1/k^2)$. Hence, when $k \geq \mathcal{O}(n/(1-\lambda)^2)$, we have that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\|z_i(k) - z^*\|^2\right] \leq \frac{\sigma^2}{n\mu^2 k}\mathcal{O}(1) \tag{21}$$

With the same amount of processing power, DSGD performs similarly to a centralized stochastic gradient descent algorithm after a transient. Using a multilevel logistic prior, a type of Markov random field (MRF), the geographical information is included. In the Bayesian technique, the posterior distribution is often maximized to provide classification results. Maximal a posterior (MAP) $y_{\text{map}} = \arg\max y\ p\ y|X$ predicts class result by maximizing $p\ y\ |X$ while taking into account the input X. MAP estimate can be expressed as an Equation (22) according to Bayes' theorem:

$$y_{\text{map}} = \underset{y}{\arg m}\ p(y|X) = \underset{y}{\arg m}\ p(X|y)p(y) \tag{22}$$

where $p(X \mid y)$ is a likelihood function, which signifies the likelihood of input data given the labels, and $p(y)$ denotes the prior over the labels in y in Equation (23) and in Equation (24):

$$p(X|y) = \prod_{i=1}^{i=m} p(x_i|y_i) \tag{23}$$

$p(X \mid y)$ can be written as

$$p(y|X) = \frac{1}{p(X)}p(X|y)p(y) = \prod_{i=1}^{i=m}\frac{p(x_i)}{p(X)}\prod_{i=1}^{i=m}\frac{p(y_i|x_i)}{p(y_i)}p(y)\ = \alpha(x)\prod_{i=1}^{i=m}\frac{p(y_i|x_i)}{p(y_i)}p(y) \tag{24}$$

where $\alpha(x) = \prod_{i=1}^{i=\text{m}} p(x_i)/p(X)$ is a coefficient independent of y. Finally, the MAP categorization is used as Equation (25):

$$y_{map} = arg\ \underset{y}{\overset{m}{max}}\left\{\sum_{i=1}^{m} logp(y_i|x_i) + logp(y)\right\} \tag{25}$$

The final classification outcome for the suggested method will be the MAP classification. In (25), the MRF-based multilevel logistic (MLL) prior will be utilized as the prior $p(y)$ to encode the spatial information, while the density $p\ y_i \mid x_i$ given by the pixel classifier modelled by CNN with spectral characteristics. This issue is resolved using the MRF-based local binary patterns (LBP) technique, which can also discover the association between spectral and spatial data. The Gibbs distribution, according to the Hammersley–Clifford theorem, expresses the likelihood that an MRF will be relevant. As a

result, Equation (26) provides the MLL prior

$$p(y) = \frac{1}{Z} exp\left(\mu \sum_{(i,j)\in C} \delta(y_i - y_j)\right) \tag{26}$$

where Z stands for the normalizing constant and is the smoothness parameter, C stands for a pair of adjacent pixel set and is a unit impulse function, where $\delta(y = 0) = 1$ and $\delta(y \neq 0) = 0$. The MLL prior favours piecewise smooth classification since it increases the likelihood that adjacent pixels will have the same label. Finally, the equation provides the MAP estimation when (25) and (26) are combined (27):

$$y_{map} = \arg\max_{y} \left\{ \sum_{i=1}^{m} logp(y_i|x_i) - \mu \sum_{(i,j)\in C} \delta(y_i - y_j) \right\} \tag{27}$$

A challenging combinatorial optimization issue is referred to as the minimization of the equation in (19).

4. PERFORMANCE ANALYSIS

The machine used for the experiment has the following hardware components: an Intel Core i5 7200U processor, 8 GB of RAM, a 1 TB hard drive, and NVIDIA GTX 760MX graphics. Additionally, Python 3.5 environments were utilized to simulate how the suggested strategy might be put into practise. In order to establish results of offered technique, we carried out a statistical analysis by evaluating expected performance.

Nanoparticle description: Drop-casting nanoparticle dispersions onto Cu TEM grids with the holey carbon film coating was used to create the TEM samples of nanoparticles. ADF-STEM or BF-TEM modes of the JEOL F200 (Akishima, Tokyo, Japan) were used to photograph the nanoparticle TEM samples.

Table 1 gives analysis for TEM dataset. Here, the dataset analyzed are BF-TEM and ADF-STEM in terms of specificity, computational cost, kappa coefficient, training accuracy, and mean average precision.

Figure 4(a)–4(e) gives analysis for BF-TEM dataset. Proposed technique attained specificity of 75%, computational cost of 51%, kappa coefficient of 44%, training accuracy of 81%, and mean average precision of 61%; existing ANN attained specificity of 71%, computational cost of 45%, kappa coefficient of 41%, training accuracy of 77%, and mean average precision of 55%; CNN attained specificity of 73%, computational cost of 48%, kappa coefficient of 43%, training accuracy of 79%, and mean average precision of 59%. A glass cuvette filled with water was used to create the sample solution that was utilized to create the scattering pattern. Sodium chloride was added in intervals of 10 ppt. Salt solution's maximum concentration was 100 ppt, while its minimum concentration was 0 ppt. Using the ready samples, two classes were produced, which are tabulated in Table 1. When tested with our dataset, the suggested ML model outperformed these models in terms of saline water particle identification accuracy, but the other techniques' Kappas remained the same (Arabameri *et al.* 2021). Figures 4 and 5 depict other performance measurement metrics utilized for evaluation, which are detailed below.

The following parameters must be understood in order to perceive and comprehend the categorization model that has been put into use: True Positives (Tp) – If the actual class labels and expected class labels match, this calculates to true. True

Table 1 | Analysis based on various TEM image dataset

Dataset	Techniques	Specificity	Computational cost	Kappa coefficient	Training accuracy	Mean average precision
BF-TEM	ANN	71	45	41	77	55
	CNN	73	48	43	79	59
	OWR_SC_MLA	75	51	44	81	61
ADF-STEM	ANN	72	51	45	79	60
	CNN	76	53	48	83	63
	OWR_SC_MLA	79	55	49	85	66

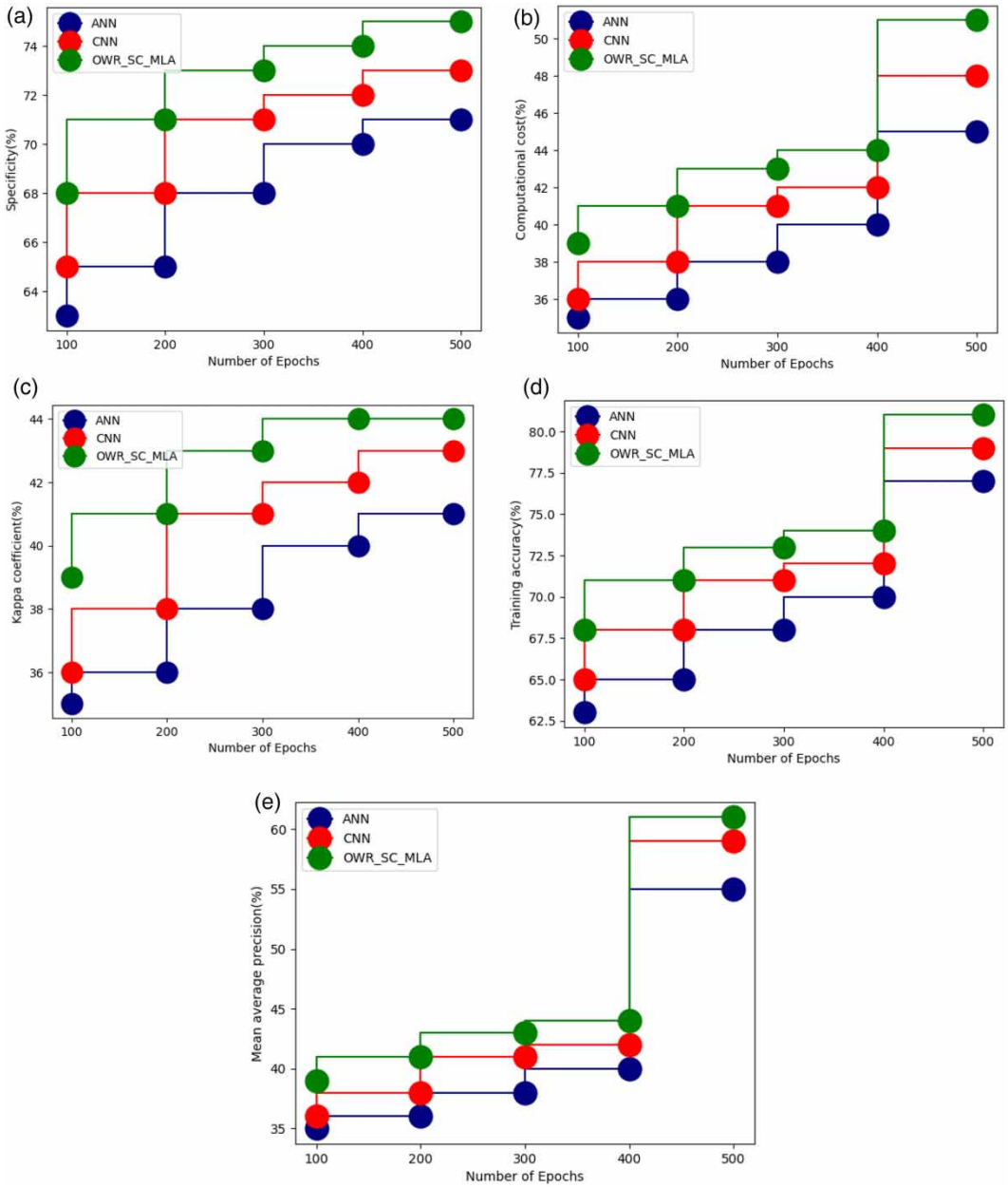

Figure 4 | Analysis for BF-TEM in terms of (a) specificity, (b) computational cost, (c) kappa coefficient, (d) training accuracy, and (e) mean average precision.

Negatives (Tn) – If actual class labels match projected class labels, this evaluates to false. False Positives (Fp) and False Negatives (Fn) specifications indicate whether actual class labels correspond to predicted class by the model or the opposite (Patel *et al.* 2022; Gerevini *et al.* 2023).

Figure 5(a)–5(e) gives analysis for ADF-STEM dataset. Proposed technique attained specificity of 79%, computational cost of 55%, kappa coefficient of 49%, training accuracy of 85%, and mean average precision of 66%; existing ANN attained specificity of 72%, computational cost of 51%, kappa coefficient of 45%, training accuracy of 79%, and mean average precision of 60%; CNN attained specificity of 76%, computational cost of 53%, kappa coefficient of 48%, training accuracy of 83%, and mean average precision of 63%. These metrics scores show that by categorizing salt particles, our deep learning classification model makes precise classifications and predictions. The implemented method was compared with comparable salinity-based prediction methods in order to demonstrate that our classification approach is reliable and produces cutting-edge results

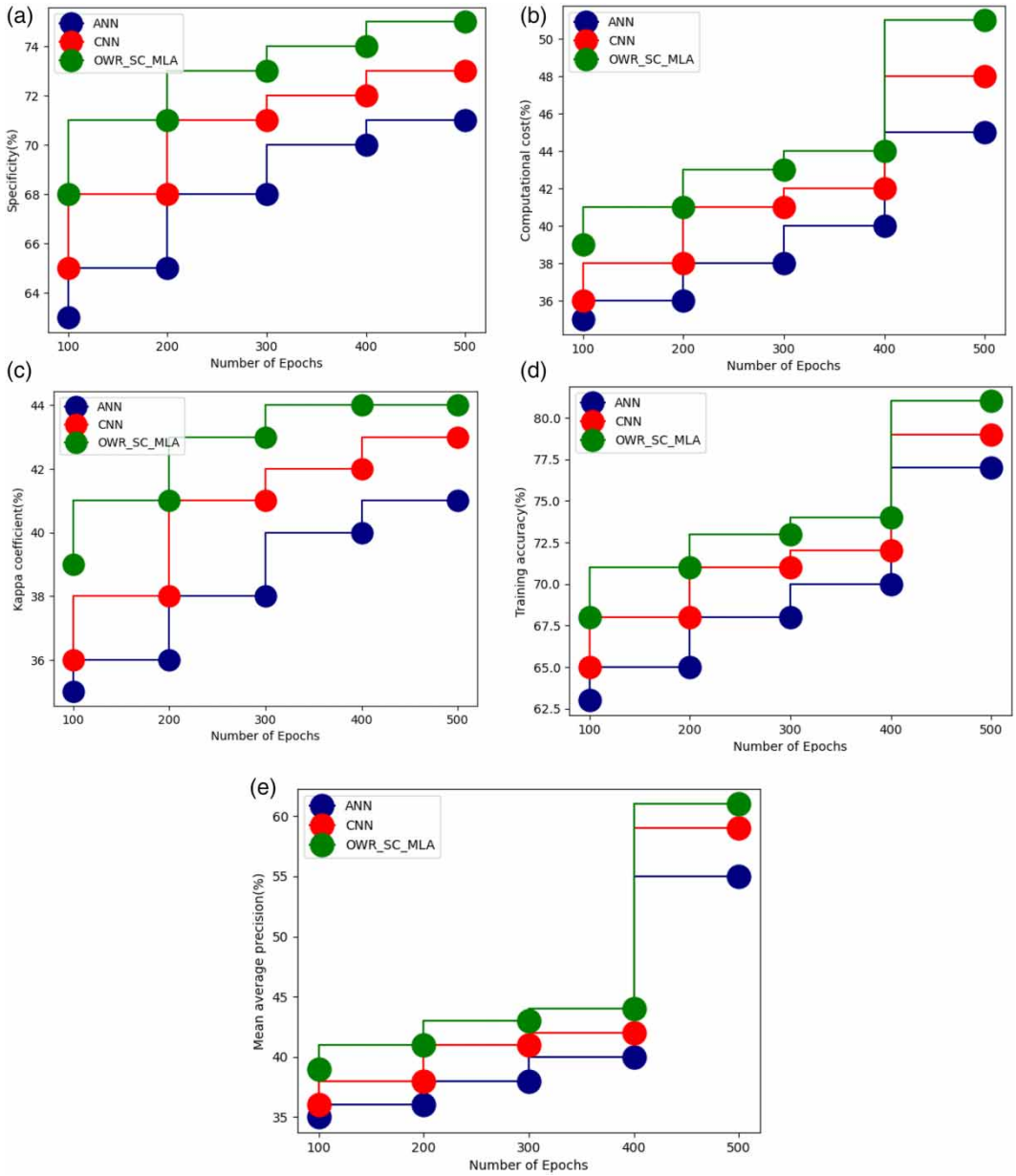

Figure 5 | Analysis for ADF-STEM in terms of (a) specificity, (b) computational cost, (c) kappa coefficient, (d) training accuracy, and (e) mean average precision.

(Khan *et al.* 2022). We sought to compare it with best ML algorithms used to evaluate salinity because there has not been much research done on identification of seawater salinity concentration, making it impossible to compare it with other methods used in this field of research.

5. CONCLUSION

This research proposes a novel technique in the optimization of water reuse with nanoparticle analysis based on ML architecture. The saline composition has been analyzed using gradient discriminant random field based on saline water treatment. The produced solution functions as seawater and utilized to identify concentration of salt particles in seawater because concentration of salt particles in sea or seawater ranges to about 35 ppt. This could be helpful for desalinated water treatment plants. It was noted that our adopted method had shown to be effective and precise in terms of precision. Additionally, a

real-world dataset was used to target salinity of seawater utilizing a combination of ML methods, allowing performance as well as resilience of proposed method to be evaluated. Proposed technique attained specificity of 79%, computational cost of 55%, kappa coefficient of 49%, training accuracy of 85%, and mean average precision of 66%.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University Abha 61421, Asir, Kingdom of Saudi Arabia for funding this work through the Large Groups Project under grant number RGP.2/142/44.

FUNDING

Deanship of Scientific Research at King Khalid University Abha 61421, Asir, Kingdom of Saudi Arabia through the Large Groups Project under grant number RGP.2/142/44.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Aghilesh, K., Mungray, A., Agarwal, S., Ali, J. & Garg, M. C. 2021 Performance optimisation of forward-osmosis membrane system using machine learning for the treatment of textile industry wastewater. *Journal of Cleaner Production* **289**, 125690.

Akgül, B., Hasoğlu, M. & Haznedar, B. 2018 Investigation and implementation ultra-low power PIC-based sensor node network with renewable energy source and decision-making unit. *Wireless Sensor Network* **10**, 41–58.

Al-Ezzi, A. S. & Ansari, M. N. M. 2022 Photovoltaic solar cells: a review. *Applied System Innovation* **5** (4), 67.

Alqahtani, A., Shah, M. I., Aldrees, A. & Javed, M. F. 2022 Comparative assessment of individual and ensemble machine learning models for efficient analysis of river water quality. *Sustainability* **14** (3), 1183.

Alshehri, M., Kumar, M., Bhardwaj, A., Mishra, S. & Gyani, J. 2021 Deep learning based approach to classify saline particles in sea water. *Water* **13** (9), 1251.

Arabameri, A., Pal, S. C., Rezaie, F., Nalivan, O. S., Chowdhuri, I., Saha, A., Lee, S. & Moayedi, H. 2021 Modeling groundwater potential using novel GIS-based machine-learning ensemble techniques. *Journal of Hydrology: Regional Studies* **36**, 100848.

Astray, G., Soto, B., Barreiro, E., Gálvez, J. F. & Mejuto, J. C. 2021 Machine learning applied to the oxygen-18 isotopic composition, salinity and temperature/potential temperature in the Mediterranean sea. *Mathematics* **9** (19), 2523.

Bagheri, M., Akbari, A. & Mirbagheri, S. A. 2019 Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: a critical review. *Process Safety and Environmental Protection* **123**, 229–252.

Chang, H. M., Xu, Y., Chen, S. S. & He, Z. 2022 Enhanced understanding of osmotic membrane bioreactors through machine learning modeling of water flux and salinity. *Science of The Total Environment* **838** (1), 156009.

Chen, F., Du, Y., Qiu, T., Xu, Z., Zhou, L., Xu, J., Sun, M., Li, Y. & Sun, J. 2021 Design of an intelligent variable-flow recirculating aquaculture system based on machine learning methods. *Applied Sciences* **11** (14), 6546.

Fu, C., Tian, A., Zhu, D., Zhao, J. & Xiong, H. 2021 Estimation of salinity content in different saline-alkali zones based on machine learning model using FOD pretreatment method. *Remote Sensing* **13** (24), 5140.

Gerevini, L., Cerro, G., Bria, A., Marrocco, C., Ferrigno, L., Vitelli, M., Ria, A. & Molinara, M. 2023 An end-to-end real-time pollutants spilling recognition in wastewater based on the IoT-ready SENSIPLUS platform. *Journal of King Saud University – Computer and Information Sciences* **35** (1), 499–513.

Hoa, P. V., Giang, N. V., Binh, N. A., Hai, L. V. H., Pham, T. D., Hasanlou, M. & Tien Bui, D. 2019 Soil salinity mapping using SAR sentinel-1 data and advanced machine learning algorithms: a case study at Ben Tre Province of the Mekong River Delta (Vietnam). *Remote Sensing* **11** (2), 128.

Hou, S. C. & Yi, X. X. 2020 Quantum Lyapunov control with machine learning. *Quantum Information Processing* **19**, 8.

Hu, B., Seiler, P. & Lessard, L. 2021 Analysis of biased stochastic gradient descent using sequential semidefinite programs. *Mathematical Programming* **187**, 383–408.

Jiang, W., Pokharel, B., Lin, L., Cao, H., Carroll, K. C., Zhang, Y., Galdeano, C., Musale, D. A., Ghurye, G. L. & Xu, P. 2021 Analysis and prediction of produced water quantity and quality in the Permian Basin using machine learning techniques. *Science of The Total Environment* **801**, 149693.

Khan, M. S. I., Islam, N., Uddin, J., Islam, S. & Nasir, M. K. 2022 Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. *Journal of King Saud University – Computer and Information Sciences* **34** (8A), 4773–4781.

Kumar, D., Nema, R. K. & Gupta, S. 2020 A comparative review on power conversion topologies and energy storage system for electric vehicles. *International Journal of Energy Research* **44**, 7863–7885.

Li, L., Rong, S., Wang, R. & Yu, S. 2021 Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: a review. *Chemical Engineering Journal* **405**, 126673.

Lowe, M., Qin, R. & Mao, X. 2022 A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. *Water* **14** (9), 1384.

Melesse, A. M., Khosravi, K., Tiefenbacher, J. P., Heddam, S., Kim, S., Mosavi, A. & Pham, B. T. 2020 River water salinity prediction using hybrid machine learning models. *Water* **12** (10), 2951.

Nam, S. N., Kim, S., Her, N., Choong, C. E., Jang, M., Park, C. M., Heo, J. & Yoon, Y. 2022 Performance assessment and optimization of forward osmosis–low pressure ultrafiltration hybrid system using machine learning for rhodamine B removal. *Desalination* **543**, 116102.

Olaimat, M. M., Yousefi, L. & Ramahi, O. M. 2021 Using plasmonics and nanoparticles to enhance the efficiency of solar cells: review of latest technologies. *Journal of the Optical Society of America B* **38**, 638–651.

Pastuszak, J. & Węgierek, P. 2022 Photovoltaic cell generations and current research directions for their development. *Materials* **15** (16), 5542.

Patel, J., Amipara, C., Ahanger, T. A., Ladhva, K., Gupta, R. K., Alsaab, H. O., Althobaiti, Y. S. & Ratna, A. 2022 A machine learning-based water potability prediction model by using synthetic minority oversampling technique and explainable AI. *Computational Intelligence and Neuroscience* **2022**, 9283293.

Rajput, S. P., Webber, J. L., Bostani, A., Mehbodniya, A., Arumugam, M., Nanjundan, P. & Wendimagegen, A. 2022 Using machine learning architecture to optimize and model the treatment process for saline water level analysis. *Water Reuse* **13** (1), 51–67.

Sarker, I. H. 2021 Machine learning: algorithms, real-world applications and research directions. *SN Computer Science* **2**, 160.

Sundui, B., Calderon, O. A. R., Abdeldayem, O. M., Lázaro-Gil, J., Rene, E. R. & Sambuu, U. 2021 Applications of machine learning algorithms for biological wastewater treatment: updates and perspectives. *Clean Technologies and Environmental Policy* **23**, 127–143.

Wang, N., Xue, J., Peng, J., Biswas, A., He, Y. & Shi, Z. 2020 Integrating remote sensing and landscape characteristics to estimate soil salinity using machine learning methods: a case study from Southern Xinjiang, China. *Remote Sensing* **12** (24), 4118.

Wang, J., Peng, J., Li, H., Yin, C., Liu, W., Wang, T. & Zhang, H. 2021 Soil salinity mapping using machine learning algorithms with the Sentinel-2 MSI in Arid Areas, China. *Remote Sensing* **13** (2), 305.

Wei, Q., Nurmemet, I., Gao, M. & Xie, B. 2022 Inversion of soil salinity using multisource remote sensing data and particle swarm machine learning models in Keriya Oasis, Northwestern China. *Remote Sensing* **14** (3), 512.

Wen, H., Luna-Romera, J. M., Riquelme, J. C., Dwyer, C. & Chang, S. L. Y. 2021 Statistically representative metrology of nanoparticles via unsupervised machine learning of TEM images. *Nanomaterials* **11** (10), 2706.

Zaghloul, M. S. & Achari, G. 2022 Application of machine learning techniques to model a full-scale wastewater treatment plant with biological nutrient removal. *Journal of Environmental Chemical Engineering* **10** (3), 107430.

Zeng, W., Zhang, D., Fang, Y., Wu, J. & Huang, J. 2018 Comparison of partial least square regression, support vector machine, and deep-learning techniques for estimating soil salinity from hyperspectral data. *Journal of Applied Remote Sensing* **12** (2), 022204.

Zhao, L., Dai, T., Qiao, Z., Sun, P., Hao, J. & Yang, Y. 2020 Application of artificial intelligence to wastewater treatment: a bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse. *Process Safety and Environmental Protection* **133**, 169–182.

Zhong, H., Yuan, Y., Luo, L., Ye, J., Chen, M. & Zhong, C. 2022 Water quality prediction of MBR based on machine learning: a novel dataset contribution analysis method. *Journal of Water Process Engineering* **50**, 103296.

First received 29 January 2023; accepted in revised form 11 May 2023. Available online 24 May 2023