



Novel Hybrid Machine Learning Algorithms for Lakes Evaporation and Power Production using Floating Semitransparent Polymer Solar Cells

Ismail Abd-Elaty¹ · N. L. Kushwaha² · Abhishek Patel³

Received: 24 March 2023 / Accepted: 21 July 2023 / Published online: 18 August 2023
© The Author(s) 2023

Abstract

The present study predicts the future evaporation losses by applying novel hybrid Machine Learning Algorithms (MLA). Water resources management is achieved by covering the reservoir water surface with floating semitransparent polymer solar cells. The energy produced by these panels will be used in the irrigation activities. The study is applied for the mass water body of Nasser Lake, Egypt and Sudan. Five MLAs namely additive regression (AR), AR-random subspace (AR-RSS), AR-M5Pruned (AR-M5P), AR-reduced error pruning tree (AR-REPTree), and AR-support vector machine (AR-SVM) were developed and evaluated for predicting future evaporation losses in the years 2030, 2050, and 2070. The study concludes that the hybrid AR-M5P ML model was not only superior to the AR model alone but also outperformed other hybrid models such as AR-RSS and AR-REPTree. The expected total annual water saving are projected to reach 3.47 billion cubic meters (BCM), 3.68 and 3.90 BCM, while the total annual power production is observed to be 1389×10^9 Megawatt (MW), 1535×10^9 MW and 1795×10^9 MW in the years 2030, 2050 and 2070, respectively. These results were achieved by covering the shallow water depths from contour level 0 m to 10 m below the surface water level. Additionally, this study shows the ability of using MLAs in the estimation of reservoir evaporation and addressing the water shortages in high stress regions.

Keywords Machine learning · Evaporation · Solar panels · Water saving and energy

1 Introduction

Reservoir evaporation is natural process for the water cycle and hydrology system in which reflects the exchange between water and air; it is also interesting for the heat transfer and energy production around the world (Katsaros 2001). About 61% of the

Highlights

- Evaporation losses are accelerated by climate changes.
- The expected water saving 3.47, 3.68 and 3.90 billion cubic meter at years 2030, 2050 and 2070, respectively.
- The power production was 1389×10^9 , 1535×10^9 , and 1795×10^9 Megawatt.
- Machine Learning Algorithms are good tools for evaporation prediction.
- Reservoirs covering are better for water saving and production energy to use in agriculture schedules.

Extended author information available on the last page of the article

global precipitation around the world is lost by evaporation (Vishwakarma et al. 2022). The estimation of evaporation volume is essential for the future water resources management and planning, crops modelling, and irrigation scheduling in high water stress regions (Kişi, 2006; Kushwaha et al. 2022). The methods for evaporation estimation include the water budget, evaporation pans, bulk aerodynamic, and Penman's equation. Recently, these methods also include the satellite sensor technology and remote sensing, and the regular hybrid and integrative data-driven models (Kişi 2006; Singh et al. 2021).

The climatic variables are used to estimate evaporation values by numbers of researchers (Duan and Bastiaanssen 2017). These variables are playing important role that affects the evaporation processes including the meteorological parameters of temperature, wind speed, relative humidity, vapor pressure and others (Ghorbani et al. 2021). The metaheuristic algorithms were incorporated for the modeling of hydrological system as a step for the developments in data-driven (Adnan et al. 2020; Kushwaha et al. 2021).

Kişi (2006) showed that the adaptive neuro-fuzzy inference systems (ANFIS) technique is good for the evaporation modeling using the climatic data compare with the the artificial neural networks (ANN) and Stephens–Stewart (SS). Kim and Kim (2008) developed a modeling for the pan evaporation and evapotranspiration using the ANN and genetic algorithm models; the study showed that the two methods are good tools to estimate the hydrological systems.

Tabari et al. (2009) showed that the ANN method is the best compare with the multivariate non-linear regression (MNL) method for evaporation estimation in Iran. Kumar et al. (2012) applied the two techniques of ANN and ANFIS technique for prediction theevaporation in Pantnagar, India. The results showed that the ANFIS technique is good performance than ANN method. Deo et al. (2016) applied three models including the relevance vector machine (RVM), extreme learning machine (ELM), and multivariate adaptive regression splines (MARS) techniques in Australia. The study presented that the RVM method is more accurate than the other methods for forecasting of monthly evaporation. Pammar and Deka (2017) showed that the two methods of discrete wavelet transform (DWT) and support vector regression (SVR) are good for estimation of the pan evaporation in the humid areas (Bajpe region) than semi-arid areas (Bangalore region) in India. Al-Mukhtar (2021) presented that the SVM is a better performance for estimating the evaporation time series in New Delhi and Ludhiana, India than random tree (RT), reduced error pruning tree (REPTree) and random subspace (RSS).

The water evaporation from Aswan high dam reservoir (AHDR), Egypt is about 20% to 30% of the country's water supplies from the Nile River, which translates to 12×10^9 m³ year⁻¹ to 16×10^9 m³ year⁻¹ (Abdel Wahab et al. 2018). Hassan et al. (2007) reduced the evaporation losses in AHDR using a pontoon framework and circular foam sheets, saving, about 1 million cubic meters (MCM) per year by covering area of 0.50 km². Ebaid and Ismail (2010) also reduced the evaporation in AHRD by disconnecting the secondary channels of El khors, saving water about 2.40 billion cubic meters (BCM) per year. The evaporation rates were found to be 2.73 mm day⁻¹ and 9.58 mm day⁻¹ at the middle and the edge the lake respectively.

Elba et al. (2017) increased the water depths in the AHDR by lowering the bed levels of the lake through removing sediments, saving water about 6.5% of the forecasting evaporation losses in year 2100.

The use of floating photovoltaic systems (FPVS) with solar cells for sustainable renewable energy that saves surface water and land resources (Trapani and Millar 2013). FPVS systems have advantages compared with the land solar panels by the lower sunlight

obstacles, efficient power production, and improvement of the aquatic environment and water quality (Sahu, et al. 2016).

The Floating photovoltaic (FPV) cells are better than conventional ground-mounted photovoltaic (GPV) cells, especially in highly populated regions where the open area is limited and competed (Zhang et al. 2020). Current and future prediction indicate that FPV is expected to double. In 2020, the global breakdown of installed FPV in China was the majority at 75.04% while the Japan, Korea, United Kingdom and others countries had 15.98%, 6.01%, 0.99% and 1.98% respectively. The FPV cells cost are higher than GPV with range 4% to 8% (Cazzaniga and Rosa-Clot 2021; Almeida et al. 2022).

Furthermore, the old-fashioned crystalline silicon solar cells are brittle, heavy and hard compare with the semitransparent polymer solar cells (ST-PSCs). FPV is gaining the world attention for green energy revolution and the use of land. ST-PSCs are smart with high absorption, thin layers; provide controlled shading with green electricity and tunable absorption spectra. This system is good for power generation, water evaporation, and algal growth compared to the old-fashioned crystalline silicon cells. The results of the study showed that the ST—PSCs achieved a maximum efficiency of 13% and an average visible transmittance of over 20% (Zhang et al. 2020) (see Fig. 1a). Abd-Elaty et al. (2021) studied the impact of covering irrigation canals for water management in the Nile Delta, Egypt. The study showed that the efficiency of the canals increased by 75.33% and the efficiency of aquifer recharge increased by 20% when using floating PV plus canal lining.

This research aims to investigate and predict the evaporation in Nasser Lake, Egypt using Machine Learning Algorithms (MLA). The objective of this study is also to reduce water losses in the reservoir and increase water supplies by using FPVS to address climate change and increased temperatures. This system will increase the power production for use in agriculture activities to optimize the value of water industry and the unit value of fresh water. This study is important for decision makers in Egypt and around the world for planning and development, increasing water resources and power generation for use in agriculture production.

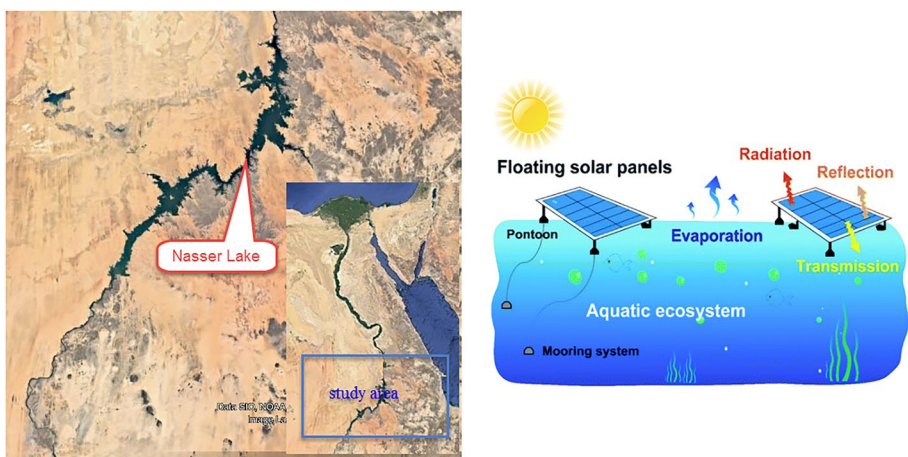


Fig. 1 a A schematic diagram semitransparent polymer solar cells (After Yin et al. 2021) and b Location map of study area

2 Materials and Methods

2.1 Study Area and Climate Data

Reservoirs of dams play an important role saving the human lives and the Communities development by managing the flooding and the fresh water resources (Asmal et al. 2000). The climate change will increase the temperature and the evaporation from water bodies and reservoirs and reduce Egypt's water share from the Nile which estimated to be 20 BCM per year (LNFDC 2008). Figure 1b presents the location of Lake Nasser locates between 23° 58' and 20° 27' N and 30° 70' and 33° 150' E, the total lake length is 500 km in the Egypt border by 350 km and 150 km in the Sudanese border. The High Aswan Dam (HAD) constructed for developing the water resources management in Egypt along the year and protection the infrastructures and human lives from flood; this dam formed the reservoir of HAD (RHAD) in the upstream.

The climate data from 1980 to 2021 were developed for the location of the study area in HAD. Figure 2 is presented these values including the minimum, average and maximum temperature with ranged 1.57° C to 34.85° C, also the humidity is ranged from 34.50% to 49.62%, the minimum, average and maximum wind speed 0.06 m s⁻¹ to 10.41 m s⁻¹ while the saturated and air pressure was ranged from 0.66 Kpa to 2.49 Kpa. The average annual evaporation from the lake of ranges from 4.65 mm day⁻¹ to 7.95 mm day⁻¹.

2.2 Empirical Methods for Evaporation and Water Saving

The evaporation losses were estimated in the current study using the bulk aerodynamic method. This method is used the Harbeck equation based on the climate data for the evaporation estimation from massive lakes and reservoirs. The monthly evaporation was estimated by Rosenberry et al. (2007), the saturated vapor pressure and monthly evaporation volume by Junzeng et al. (2012) while the water saving using the solar cells by Sahu et al. (2016) as the following Eqs. (1), (2), (3) and (4).

The AHDR evaporation was estimated based on the climate data from 1980 to 2021 using Lake Nasser coefficient, saturated vapor pressure at water surface temperature and the actual vapor pressure of the air; the following equation is used to calculate the evaporation (Rosenberry et al. 2007):

$$E = N U_2 (e_s - e_a) \quad (1)$$

The saturated vapor pressure at water surface temperature is estimated using the temperature while the actual vapor pressure of the air is estimated using the relative humidity (Fig. 2) as the following equation (Junzeng et al. 2012);

$$e_s = 0.611 \exp \left(\frac{17.27T}{T + 237.30} \right) \text{ and } e_a = \left(\frac{RH}{100} \right) e_s \quad (2)$$

The evaporation volume in the study area estimated using the evaporation rates, the surface area of lake by 5775 km² and numbers of day in month as the following equation (Junzeng et al. 2012):

$$V_{\text{losses}} = (E * A_{\text{surface}} * n) * 10^3 \quad (3)$$

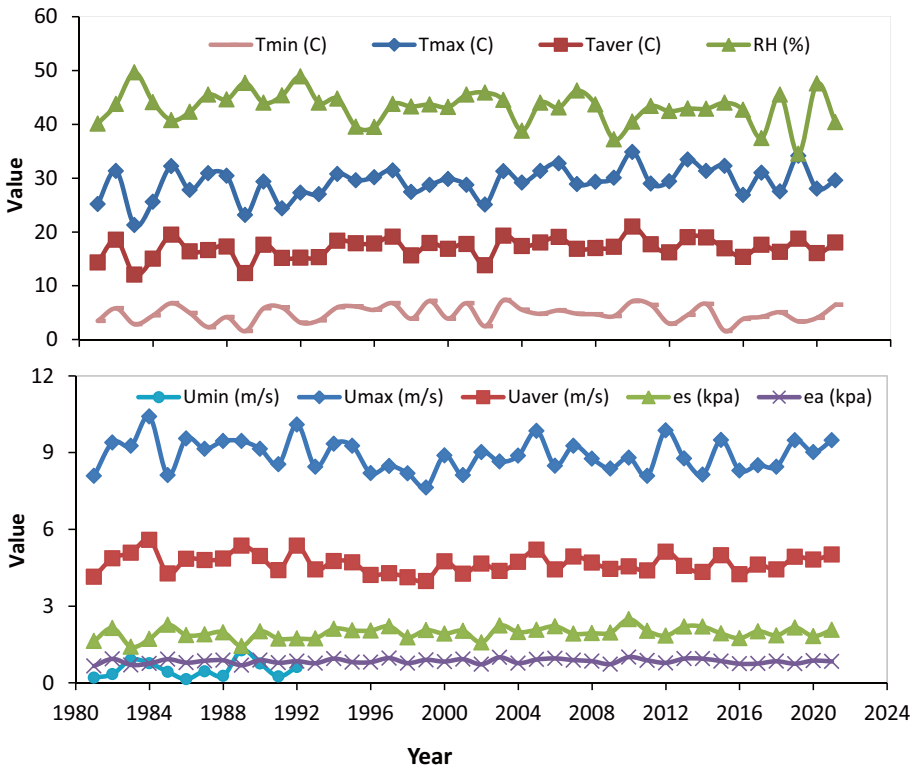


Fig. 2 Climate data for the study area

The volume of water saving for AHDR is estimated using the evaporation rates, the solar cell area and the number of day in month as the following equation (Sahu et al. 2016):

$$V_{\text{saving}} = (E * A_{\text{panel}} * n * 0.70) * 10^3 \tag{4}$$

where: E: evaporation losses (mm day⁻¹); N: Lake Nasser coefficient by 0.0525 (Hassan et al. 2017); U₂: wind speed (m sec⁻¹); es: saturated vapor pressure (kpa) at water surface temperature; ea: actual vapor pressure of the air (kpa); RH: relative humidity; T: temperature (° C); V_{losses}: water losses volume (m³ month⁻¹); E: evaporation rate (mm day⁻¹); A_{surface}: Lake Nasser surface area (km²); n: numbers of day in month (day), V_{saving}: water savings; Apanel: solar cell area.

2.3 Solar Energy Estimation Using Covering the Lake

The saving of water from evaporation losses is carried out in the current study using solar panels. These cells will covering the lake at shallow depths where the evaporation from smaller depths is high compare with deeper depths. The available solar power is ranged between 200 and 250 MW km⁻² for Aswan city., These values of solar power data were obtained from the Solar Atlas of Egypt (Choi 2014). The estimation of energy potential in the study area using this equation (Kosmopoulos et al. 2013):

$$\text{SEP} = A * \text{Ed} * N * H \quad (5)$$

where SEP: is the Solar energy production (MW month^{-1}); A: is the surface area subjected to sun (km^2); Ed: is the energy density (MW Km^{-2}); N: is the number of sun hourly in day (hr), and H is the number of days in month.

2.4 Selection of Best Input Combination for Model Development

The optimal selection and combination of the input parameters for the climate and meteorological is the key for the best performance of the hydrological models Also, the best combination of these parameters has been selected using Relief algorithm (Kira and Rendell 1992). Table 1 presents the ranks of the es, Taver, Uaver, Umax and RH, variables for predicting the evaporation.

2.5 Machine Learning Algorithms

2.5.1 Additive Regression AR

This algorithm is a non-parametric regression technique and developed by Friedman and Stuetzle (1981). It used one smoother function and overcomes the problem of curse dimensionality. The technique using following form:

$$E[y_i | x_{i1}, \dots, x_{ip}] = \beta_o + \sum_{j=1}^p f_i(x_{ij}) \quad (6)$$

where $\sum_{j=1}^p f_i(x_{ij})$: the smooth functions fitted from the data, β_o is regression coefficient.

2.5.2 Random Subspace (RSS)

RSS is ensemble methods of machine learning which it is used for decision trees. It constructs of the decision trees according to the parallel learning algorithm (MA) (Ho 1998). The advantage of this method is being sensibly with the high-dimensional problems in which the features numbers is much larger than the training points numbers (Arabameri et al. 2021).

The most voting for the generated trees is used in the subspace ensemble system for the sample of X, and is estimated by the following form (Skurichina and Duin 2002):

$$\beta(x) = \underset{y \in \{-1, 1\}}{\text{argmax}} \sum_b \delta_{\text{sgn}(C^b(x)), y} \quad (7)$$

where δ is the Kronecker symbol, and $y \in \{-1, 1\}$ is a decision (class label) of the classifier, $C^b(x)$ is the classifier of each subspace b (Skurichina and Duin 2002).

Table 1 Selected attribute and their ranks

Meteorological input parameters	es	Taver	Uaver	Umax	RH
Rank	0.01942	0.01352	0.01056	0.00818	0.0044

2.5.3 M5 Pruned (M5P)

The tree model is developed by Quinlan (1992) and it used for solving the regression problems (Kisi et al. 2017). This technique is developed based on the two steps, the first step, the standard deviation reduction (SDR) is estimated using the following form (Malik et al. 2020; Al-Mukhtar 2021).

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \tag{8}$$

The second step is constructing a multivariate linear regression model for each node at the model subtree (Malik et al. 2020). The optimal model selection according to the minimize value of error for the linear or subtree model (Arabameri et al. 2021). The accuracy forecasting of M5 model is adjusted using this equation:

$$PV(S) = \frac{n_i \times PV(S_i) + k \times M(S)}{n_i + k} \tag{9}$$

where T is a set of cases in the data that reach a node, *sd* is the standard deviation, *T_i* is the subset of cases that have the *ith* outcomes of the potential test, *PV(S_i)* is the predicted value at *S_i*, *n_i* number of training cases, *S_i* is the case follow a branch of subtree *S*, *k* is the smoothing factor, *M(S)* is the value given by the model at *S*.

2.5.4 Reduced Error Pruning Tree (REPTree)

The REPTree is adjusted as simple and computationally speed (Quinlan 1987) and based on the principle of information gain and minimizing the variance error (Chen et al. 2019). It was constructing to compare with other pruning methods (Ganatra and Bhensdadia 2012). Also, the data is split according to the information gain at each node and then the subtrees are pruned by the reduced error (Quinlan 1987). The subtree would be replaced by a leaf when the new induced tree has equal or less error than before. The process is repeated over until any further replacements increases the variance error of the test data.

2.5.5 Support Vector Machine (SVM)

Cortes and Vapnik (1995) established the supervised soft computing kernel-based SVM model that is capable of reducing complexities alongside errors in the estimation. The classifier models of SVMs are applied to problems of data classification under different classes. Another group of the SVM is the SVR, which is used in regression prediction problems. Here, the kernel function (input vector *x*) implicitly transforms the lower-dimensional inputs to a higher-dimensional feature [*β(x)*] such that *w* is the weighting vector and *b* is a bias. These two parameters are estimated using regularized risk function [*R(P)*], as shown in the following equations:

$$f(x) = w\beta(x) + b; \quad R(P) = P \frac{1}{n} \sum_{i=1}^n L(d_i, y_i) + \frac{1}{2} ||w||^2 \tag{10}$$

$$L_{\epsilon}(d, y) = \begin{cases} |d - y| - \epsilon & |d - y| \geq \epsilon \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where; P is a penalty parameter, $\frac{1}{2}||w||^2$ is a regularization term, d_i is the desired value, $P \frac{1}{n} \sum_{i=1}^n L(d_i, y_i)$ is the error term, and ϵ is the tube size of SVM in L_{ϵ} (Table 2).

2.6 Hybridization of Meta-Heuristics Algorithms Using Stacked Generalization

According to purpose of forecasting monthly evaporation, the current experiment used stacking hybrid algorithms. Wolpert (1992) developed the stacking hybrid algorithm proposal. This technique offers a framework for ensemble algorithms, which mix two or more algorithms over the course of training. Studies have shown that stacking hybrid algorithms can improve the predictability of the algorithms (Kushwaha et al. 2022). Moreover, the idea of the stacking hybrid generalization is applying the first-level learners for the train and forecast the training data sets. The new training dataset is creating for the meta learner based on the predicted outcomes from the first-level learners and then it were combined. (Sikora et al. 2015) and (Zhou 2009) presented the complete details on layered hybrid generalization.

2.7 Statistical Performance Assessment of Developed Hybrid Models

The current study was evaluated based on the performance machine learning model For numbers of statistical indices including the mean square error (MSE), root mean square error (RMSE), relative root square error (RRSE), mean absolute error (MAE), relative absolute error (RAE), Nash–Sutcliffe Efficiency (NSE) (Kushwaha et al. 2021), Willmott Index (WI), and determination coefficient of (r^2) (Kushwaha et al. 2022). MSE calculates the proximity of a fitted line to the data points. The root mean square deviation of time series anticipated values from observed values is represented by RMSE statistics. When the error is being reduced in the same dimensions as the quantity being predicted, the relative squared error (RRSE) is measured as its square root. The mean absolute deviation of time series anticipated values from the observed values is represented by MAE statistics, as shown. While the RAE statistic shows the proportion of the measurement's absolute error to the real measurement, it also aids in estimating the absolute error's size relative to the measurement's actual size. The most often employed metric for measuring model performance is Nash–Sutcliffe efficiency. It goes from

Table 2 Machine learning algorithm parameters used for modeling evaporation

Model name	Description of parameters
Random Subspace (RSS)	Random seed = 1, Classifier = REPTree, batch size = 100, numbers of executions slots = 1, number of subspace size = 0.5, and iteration = 10
Additive Regression (AR)	Number of iteration = 30, batch size = 100, Classifier = Bagging, and shrinkage = 1,
M5 Pruned (M5P)	Minimum number of instances = 4 and Batch size = 100
Reduced Error Pruning Tree (REPTree)	Batch size = 100, random seed = 1, Initial count = 0, number of folds = 3, minimum proportion of the variance = 0.001, minimum number = 2, and max depth = 1
Support Vector Machine (SVM)	Kernel = Normalized Poly, Regression Optimizer = SMO Improved, Filter type = Normalize training dat, batch size = 100, C = 1, and cache size = 250000

1 for the ideal fit value of 0 means that the accuracy was the same as the mean value. The index of agreement is another name for the Willmott index (WI). The WI varies from zero to one (0 WI 1); an optimum agreement/fit of about one is possible. In contrast, r^2 is a measurement of the linear relationship between the dependent and independent variables.

For evaporation modelling, the models that have been shown to have higher values of r^2 (closer to 1) and RRSE and lower values of MSE, RMSE, MAE, and RAE are deemed to be comparable better models. O and P stand for observed and predicted or simulated values for the dataset in the following equations; OAvg and PAvg stand for the average or mean magnitude of observed and predicted or simulated values; and N is for the number of observations.

$$MSE = \frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2; \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2} \quad (12)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - O_{Avg})^2}}; \quad MAE = \frac{1}{N} \sum_{i=1}^N |O_i - P_i| \quad (13)$$

$$RAE = \frac{\sum_{i=1}^N |O_i - P_i|}{\sum_{i=1}^N |O_i - O_{Avg}|}; \quad r^2 = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - O_{Avg})^2} \quad (14)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - O_{Avg})^2} \right]; \quad WI = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (|P_i - PAvg| + |O_i - O_{Avg}|)^2} \quad (15)$$

2.8 Study Scenarios

This study aims to estimate evaporation from Nasser Lake using machine learning algorithms based on climate and meteorological data from 1980 to 2021. The method predicts future monthly evaporation values using this data. Using these predictions, we estimated water losses and developed a plan for water saving and energy production by covering shallow water depths in the lake.

3 Application of Machine Learning Models

The goal of this study is to estimate evaporation losses and assess the predictability of hybrid machine learning algorithms/models. In this section, we present the results of modelling evaporation losses at our study site using data-driven machine learning algorithms. We evaluated a total of five models, including AR, AR-M5P, AR-SVM, AR-RSS, and AR-REPTree, for predicting evaporation. These models were trained on a monthly time scale from 1981 to 2009 and tested from 2010 to 2021. We compared the models using various statistical criteria, both numerical and graphical.

3.1 Training of Models (1981–2009)

The training of all ML models was performed with 348 data points (accounting 70% of dataset). The descriptive statistic of observed evaporation and estimated from five ML models (AR, AR-M5P, AR-SVM, AR-RSS, and AR-REPTree) is presented in the form of Violin box plot (Fig. 3a). The Figure depicts the minimum, maximum, median along with 1/4th and 3/4th percentile quartiles, and the distribution of dataset (observed and modeled). The statistical results for testing period related to correlation coefficient, mean absolute error, root mean squared error, relative absolute error (%), and root relative squared error (%), Nash–Sutcliffe efficiency (NSE), and Willmott Index (WI) are collected in the Table 3.

As indicated by the descriptive statistics the estimated evaporation through, AR and AR-M5P model, has the least (0.970) and highest (0.982) correlation with that of observed values respectively. Whereas the RMSE of 0.629 mm (AR-M5P) and 0.789 mm (AR) was observed to be minimum and maximum respectively among all five implemented models. The range of relative absolute error (%) was found to be 16.183–21.518 percent during calibration of the models. Higher NSE and WI statistics for AR-M5P model among all algorithms confirms that it is being better performer during training period. Similarly, statistics depicts that AR algorithm is being least performer based on the lowest NSE (0.941) and WI (0.985). From Table 3 it is clear that the hybrid AR-M5P algorithm confirms optimal values against all descriptive statistics. Moreover, based on statistics, it was observed that during calibration all four hybrid ML algorithms (AR-M5P, AR-SVM, AR-RSS, and AR-REPTree) performs superior than AR model alone. Depicts the monthly time series during training period of observed vs predicted evaporation and respective scatter plots was done. The time series plots provide the capabilities pf models to estimate the evaporation. The scatter plots with liner regression results the highest R-square value of 0.965 for the AR-M5P followed by AR-SVM, AR-RSS, AR-REPTree and AR model with that of 0.962, 0.957, 0.951, and 0.941 respectively. These results indicates that the AR-M5P model provides the best result close to best fit line among all five ML models for E estimation during training period. The interesting fact that was observed from rhis step that the AR model is predicting the overestimated and underestimated values for low and high evaporation values respectively, leading to the model to show weaker prediction power then hybrid models during training period. Further comparative analysis of models was done using the Taylor

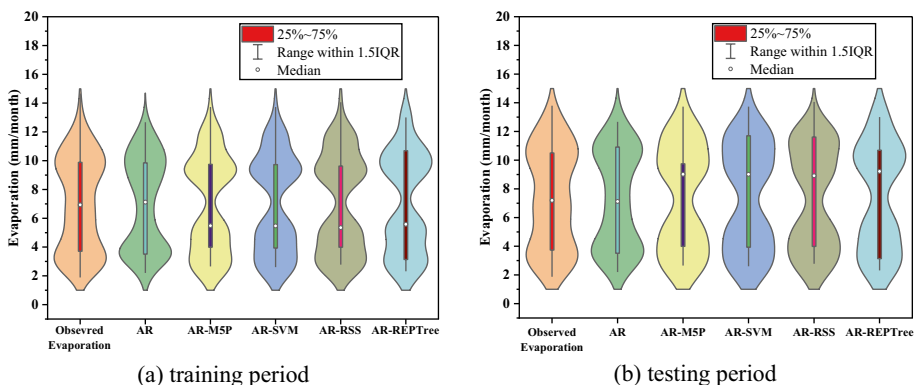


Fig. 3 Violin box plot of predicted and observed monthly evaporation values

Table 3 Statistical results of different ML algorithms for monthly Evaporation during training period (1981–2009) and testing period (2010–2021)

Statistic	ML Models									
	AR		AR-M5P		AR-SVM		AR-RSS		AR-REPTree	
	training	testing	training	testing	training	testing	training	testing	training	testing
R	0.970	0.968	0.982	0.978	0.981	0.977	0.978	0.967	0.975	0.961
MAE	0.625	0.686	0.470	0.569	0.490	0.601	0.532	0.689	0.568	0.736
RMSE	0.789	0.859	0.629	0.738	0.646	0.767	0.699	0.884	0.757	1.005
RAE (%)	21.518	22.509	16.183	18.661	16.858	19.716	18.307	22.627	19.547	24.171
RRSE (%)	24.285	24.904	19.350	21.378	19.878	22.227	21.519	25.618	23.309	29.136
NSE	0.941	0.937	0.963	0.954	0.960	0.950	0.954	0.934	0.946	0.914
WI	0.985	0.984	0.991	0.989	0.990	0.988	0.989	0.983	0.987	0.979
N	348	144	348	144	348	144	348	144	348	144

diagram (Fig. 5a). In Taylor diagram on the basis of correlation and standard deviation hybrid AR-M5P model and AR model are found to be at closest and farthest respectively to the observed/referenced point. This indicated hybrid AR-M5P model has performed best among all five ML models to simulate evaporation during training of the models.

3.2 Testing of Models (2010–2021)

Validation or testing of the trained models is necessary to avoid overfitting of the models. The testing of all ML models was performed with 144 data points (accounting 30% of dataset). For testing period, descriptive statistic of observed evaporation and estimated from five ML models (AR, AR-M5P, AR-SVM, AR-RSS, and AR-REPTree) are presented in the form of Violin box plot (Fig. 3b). The figure depicts the minimum, maximum, median along with 1/4th and 3/4th percentile quartiles, and the distribution of dataset (observed and modeled).

Table 3 presents the results of all ML models tested, along with descriptive statistics for the testing period. The highest correlation between observed and predicted evaporation values was found for the AR-M5P model (0.978), followed by the AR-SVM model (0.977), the AR model (0.968), the AR-RSS model (0.967), and the AR-REPTree model (0.961). The results consistently showed that the AR-M5P ML model had the lowest errors (MAE, RMSE, RAE, and RRSE), along with the highest NSE (0.954) and WI (0.989). In contrast, the AR-REPTree model had the highest errors and the lowest NSE and WI (Table 3).

Scatter plots with regression lines were generated for all models to compare observed and predicted evaporation values. The AR-M5P model had the highest coefficient of determination (R^2) and closely followed the best fit line (1:1). The hybrid AR-REPTree model had the lowest R^2 of 0.923 for the testing period. Time series plots in Fig. 4 depicted the fluctuation of estimated and observed evaporation values for the testing period. The AR-M5P model estimated the variation of evaporation reasonably well, followed by the AR-SVM, AR, AR-RSS, and AR-REPTree models, indicating that the AR-M5P model is superior among the five ML models.

The model testing was further analysed using a Taylor diagram (Fig. 5b) to carry out a comparative evaluation of the models. The hybrid AR-M5P model was found to be the closest to the observed/referenced location based on correlation and standard deviation, while the AR-REPTree model was furthest. This indicates that the hybrid AR-M5P model performed the best among all five ML models for simulating evaporation during testing.

3.3 Estimation of Future Monthly Evaporation

The calibrated model was used to forecast future lake evaporation for the years 2030, 2050, and 2070. The results showed that the maximum evaporation occurred in August and reached 12.33, 13.45, and 14.56 mm/month, with a volume of 2.21, 2.41, and 2.61 BCM/month, respectively. Conversely, the minimum evaporation values occurred in December at 2.58, 2.48, and 2.38 mm/month, with corresponding volumes of 0.46, 0.44, and 0.43 BCM/month (Fig. 6). The results also showed that months with higher temperatures led to an increase in future evaporation, while those with lower temperatures led to a decrease in predicted evaporation because of the impact of climate change. The total annual loss for future lake evaporation for the years 2030, 2050, and 2070 was estimated to be 16.52, 17.54, and 18.56 BCM/year, respectively.

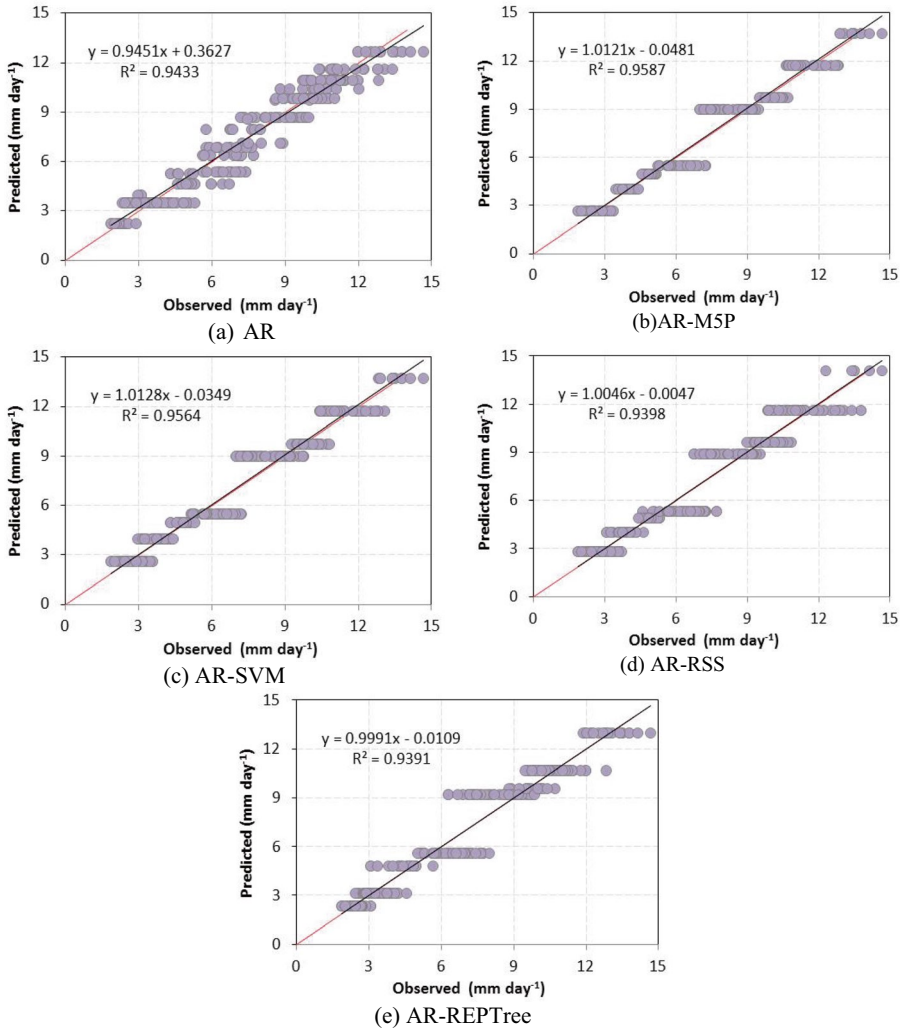


Fig. 4 Scatter plots between observed and predicted monthly evaporation during testing period

3.4 Estimation of Future Water Saving and Energy Production

Figure 7 illustrates the predicted water savings and energy production from the lake by covering shallow water depths from contour level 0 to 10 m from the surface water level in an area of 1732 km², which represents about 30% of the total lake area. The maximum water saving reached 0.46, 0.505, and 0.548 million cubic meters (MCM) per month for the years 2030, 2050, and 2070, respectively, in August. The minimum water savings occurred in December, with values of 0.097, 0.093, and 0.089 BCM per month for the years 2030, 2050, and 2070, respectively. The expected total annual water savings were estimated to be 3.47, 3.684, and 3.898 BCM per year for the years 2030, 2050, and 2070, respectively.

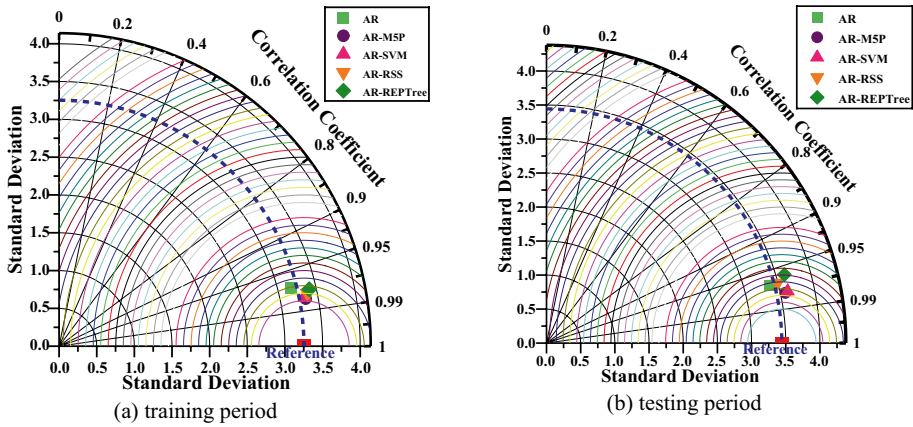


Fig. 5 Taylor diagram representation of model performance

Regarding power production, the maximum predicted power output was estimated to be 0.134×10^9 , 0.151×10^9 , and 0.167×10^9 megawatts (MW) per month for the years 2030, 2050, and 2070, respectively, in July. The minimum power output was estimated

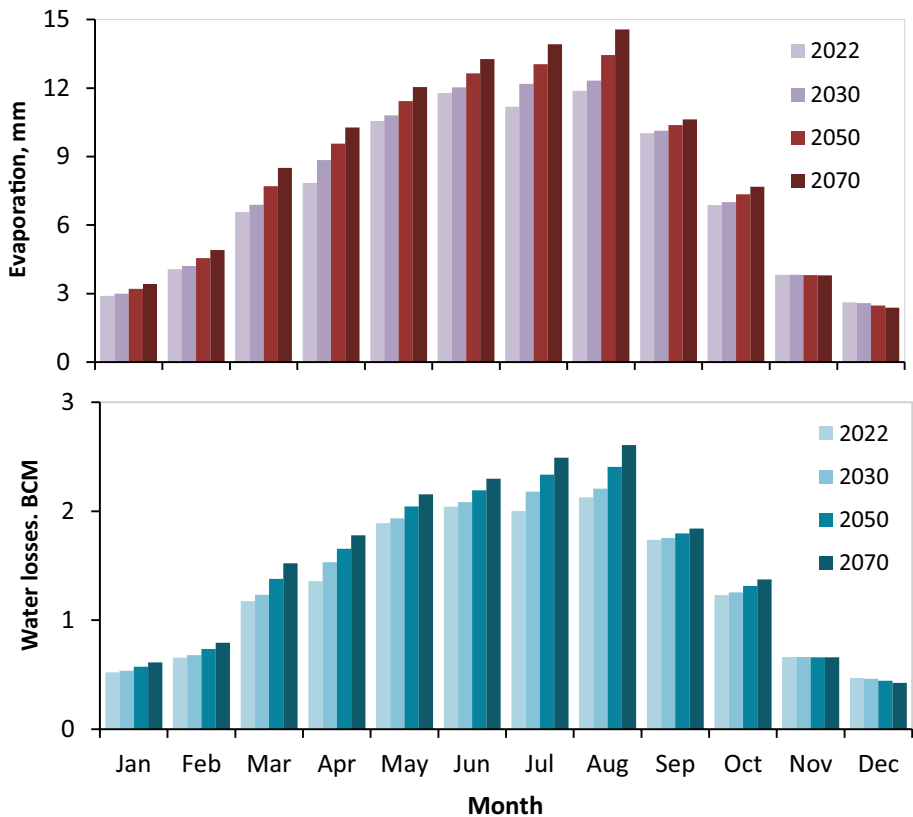


Fig. 6 Predicted Evaporation and water losses from Nasser Lake

to be 0.077×10^9 , 0.087×10^9 , and 0.097×10^9 MW per month for the years 2030, 2050, and 2070, respectively, in February. The total annual power production was estimated to be 1.266×10^9 , 1.425×10^9 , and 1.583×10^9 MW per year for the years 2030, 2050, and 2070, respectively (refer to Table 4).

4 Qualification the Performance of Machine Learning Models for AHDR

In this study hybrid ML models were comparatively evaluated to estimate monthly evaporation. The climate Data from mass water body of Lake Nasser in Egypt and Sudan were employed for the sake of models evaluations using several statistical metrics (MAE, RMSE, NSE, WI and r). Then best combination has been selected using Relief algorithm. As such, the optimal input combinations were es, Taver, Uaver, Umax and RH, indicating that all these variables affect pan evaporation. The results showed that derived evaporation had good correlation with observed evaporation along with low

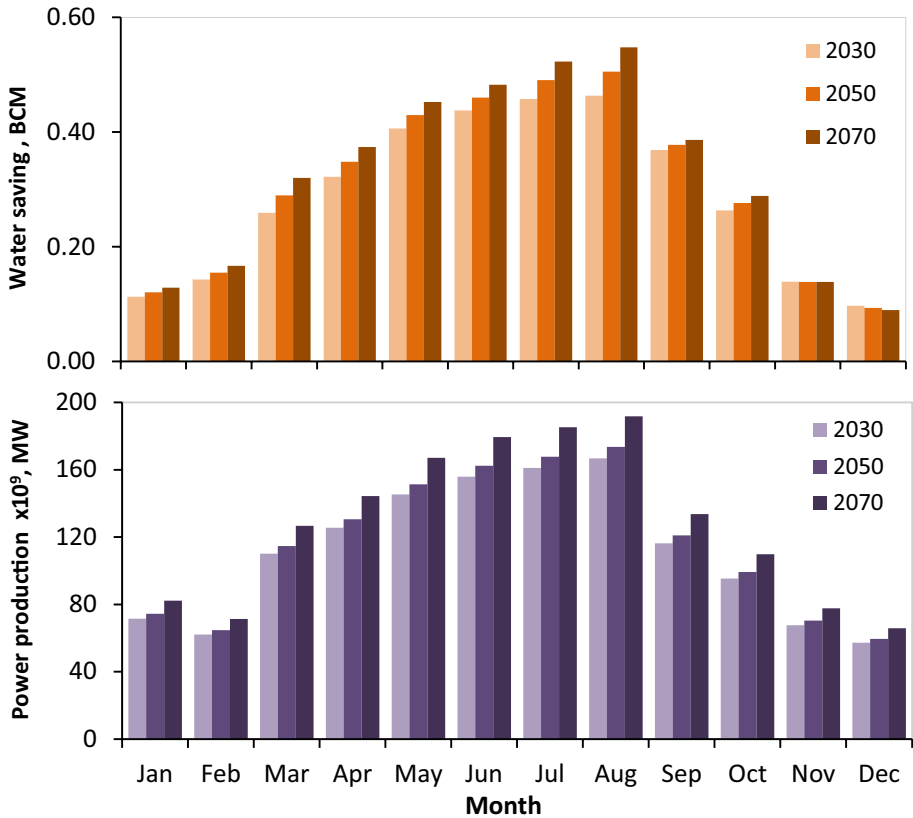


Fig. 7 Predicted water saving and energy production from lake Nasser

Table 4 Nasser Lake for predicted evaporation, water losses and saving and energy production

Month	Scenario	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	annual
day	-	31	28	31	30	31	30	31	31	30	31	30	31	-
Sunlight hours	-	9	8	9.5	10.5	11	12	12.5	11.5	10.5	9.5	8.5	7.5	-
Evaporation (mm)	2030	3.00	4.21	6.89	8.85	10.81	12.03	12.18	12.33	10.13	7.01	3.82	2.58	
	2050	3.21	4.56	7.70	9.56	11.42	12.65	13.05	13.45	10.38	7.34	3.81	2.48	
	2070	3.42	4.91	8.51	10.27	12.04	13.26	13.91	14.56	10.62	7.68	3.81	2.38	
Water losses (BCM)	2030	0.54	0.68	1.23	1.53	1.93	2.08	2.18	2.21	1.75	1.25	0.66	0.46	16.52
	2050	0.57	0.74	1.38	1.66	2.05	2.19	2.34	2.41	1.80	1.31	0.66	0.44	17.54
	2070	0.61	0.79	1.52	1.78	2.16	2.30	2.49	2.61	1.84	1.37	0.66	0.43	18.56
Water saving (BCM)	2030	0.113	0.143	0.259	0.322	0.406	0.438	0.458	0.463	0.369	0.263	0.139	0.097	3.47
	2050	0.121	0.155	0.289	0.348	0.429	0.460	0.490	0.505	0.377	0.276	0.139	0.093	3.684
	2070	0.129	0.167	0.320	0.374	0.453	0.483	0.523	0.548	0.386	0.289	0.138	0.089	3.898
Power (109xMW)	2030	0.097	0.078	0.102	0.109	0.118	0.125	0.134	0.124	0.109	0.102	0.088	0.081	1.266
	2050	0.109	0.087	0.115	0.123	0.133	0.140	0.151	0.139	0.123	0.115	0.099	0.091	1.425
	2070	0.121	0.097	0.128	0.136	0.148	0.156	0.168	0.154	0.136	0.128	0.110	0.101	1.583

values of associated errors which showed that hybrid ML models can predict monthly evaporation reasonably well. Based on RMSE and correlation coefficient M5P model found to be satisfactory performing for prediction of stream flow (Onyari and Ilunga 2013). Moreover, M5P outperformed other hybrid ML models for flow prediction and also for daily water level prediction (Khosravi et al. 2021). Similar to such studies the present study, in particular, highlights the capability of hybrid AR-M5P model to predict monthly evaporation losses. The AR model was found to have the least accuracy during training period; however, it has outperformed two hybrid models (AR-RSS and AR-REPTree) during testing period. The highest R^2 and closeness of linear regression line (fitted between predicted evaporation from AR-M5P model and observed) with 1:1 line (best fit line) reinforces the capability of AR-M5P hybrid ML model to predict evaporation. Figure 5 show the clear comparative picture of the hybrid ML models. The nearer position of AR-M5P model then other ML models to the reference location shows its good performance and best prediction capability among all. Overall, from Fig. 4 it has been observed that the hybrid models are capable to estimate the monthly evaporation with reasonable accuracy (Kushwaha et al. 2021). Moreover, these results, in line with past studies Malik et al. (2020), and Kushwaha et al. (2022) and highlight the ability of ML models to estimate evaporation accurately. Considering the results obtained from this study it is concluded that hybrid AR-M5P provides superior results than all another LM models for both the period (training and testing) as it commits the highest correlation and lowest errors in prediction of evaporation (Table 3).—The study concludes that the hybrid AR-M5P ML model is not only superior to the AR model alone but also outperforms other hybrid models such as AR-RSS and AR-REPTree.

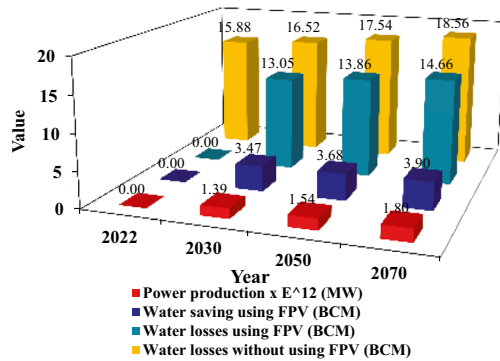
Evaporation is a critical parameter in the hydrological cycle. It can reduce water losses and the water budget for the total inflow. The current results show that the maximum evaporation occurs in August, while the minimum values occur in December. The total annual water losses from Lake Nasser will reach 16.52, 17.54, and 18.56 billion cubic meters (BCM) per year in 2030, 2050, and 2070, respectively. The expected total annual water savings will reach 3.47, 3.684, and 3.898 BCM per year in the same years. The total annual power production will reach 1.266×10^9 , 1.425×10^9 , and 1.583×10^9 megawatt-hours (MWh) per month in 2030, 2050, and 2070, respectively.

This management will be developed by covering the shallow water depths from contour level 0 to 10 m from the surface water level in the lake by a surface area of 1732 square kilometres, which represents about 30% of the total lake area (see Fig. 8). In other words, the management plan will cover about 30% of the surface of Lake Nasser with a floating photovoltaic system (FPVS). This will reduce evaporation from the lake and generate electricity at the same time. The expected water savings and power production are significant, and this management plan could be a valuable tool for conserving water resources in Egypt.

The results of our study agree with those of Abd-Elhamid et al. (2021), who estimated the annual evaporation from Lake Nasser to be 12 billion cubic meters (BCM). They also estimated that annual water savings would reach 2.1, 4.2, 6.3, 7.0, and 8.4 BCM, and annual energy production would reach 2.85×10^9 , 5.67×10^9 , 8.54×10^9 , and 11.38×10^9 megawatt-hours (MWh) by covering 25, 50, 75, and 100% of the total surface area of the lake, respectively. The difference in energy production between the two studies is due to the fact that Abd-Elhamid et al. estimated energy production over 24 h, while our study took into account daily sunshine hours.

Allawi et al. (2019) predicted the reservoir using ML in the Layang Reservoir, Johor River, Malaysia, the study showed that the selection the suitable training period to learn the prediction model plays role to attain good results. Al Sudani and Salem (2022) forecasting the evaporation

Fig. 8 Future lake Nasser water losses, water saving and energy production



rates using advanced ML, the study in two study area for Diyala and Erbil state, Iraq. The study proposed that GMBM model can therefore assist local stakeholders in the management of water resources. Abed et al. (2022) estimated the monthly pan evaporation rates using random forest, convolutional neural network (CNN), and deep neural network (DNN), the study showed that the CNN model was powerful for predict evaporation. Essak and Ghosh (2022) showed that covering global reservoirs with FPV by just 1% would produce 404 GW of power. Antonopoulos and Gianniou (2022) investigated the surface water temperature using energy budget components and Artificial Neural Networks (ANNs models) at the interface of atmosphere and water. The results showed that the ANN models are the most accurate compare net radiation models using the adjusted Slob's equations, and then, by the models with the heat storage changes functions. Agrawal et al. (2022) used 5 ensembled machine learning methods in estimating the daily reference evapotranspiration. The results showed that this approach substantiated by boosting algorithm for Extreme Gradient Boosting (XGBoost) significantly enhance the performance in the estimation method. Faramarzzadeh et al. (2023) developed Machine Learning and Remote Sensing model for investigating the gap-filling daily precipitation data in East Africa, the study showed that the random forest technique is performed the best results among all other methods for sloving the gap data. Beça et al. (2023) optimized the reservoir water management using rule curves and a dynamic assessment of water demands that are dependent on a water transfer system. The study indicated that this method can ensure 100% of the urban water supply, improve the reliability of the irrigation supply from 75% to 86–91%, and provide 92–98% of the river ecological flow. TR et al. (2023) estimated the precise values of the daily reference evapotranspiration using four ensemble techniques in the state of Karnataka, India, from 1979 to 2014. The results showed that these models based on all climatic variables were the most accurate in comparison with other input combinations.

The challenges of our research are the environmental impact of FPV on the ecosystem and water quality. More studies are needed to achieve future zero emissions, water savings, and green solar energy production. Overall, our findings indicate that hybrid models have a stronger predictive value in real-world situations and may be employed more effectively in watersheds with little data. In addition to predicting pan evaporation, these types of models may be used to forecast a wide range of hydrological and water resources phenomena, including ETo. The current study limitations are the cost of using floating semi-transparent polymer solar cells compare with the other solar cells, also the environmental impact of the covering this reservoir on the ecosystem and hydrology cycle in this region.

5 Conclusions

In this study, we applied additive regression and its hybridization with four other machine learning algorithms (AR-M5P, AS-SVM, AS-RSS, and AR-REPTree) to forecast monthly evaporation, water savings, and power production for the years 2030, 2050, and 2070. The algorithms were trained with 29 years of data (1981–2009) and tested against 12 years of data (2010–2021). The results were compared with classic AR to see the accuracy improvement of the new developed methods.

The AR-M5P model was found to be the best performing model among the other evaluated methods, as it showed the least error indices values. Overall, our findings indicate that hybrid models have a stronger predictive value in real-world situations and may be employed more effectively in watersheds with little data. In addition to predicting pan evaporation, these types of models may be used to forecast a wide range of hydrological and water resources phenomena, including ETo.

The descriptive statistics results showed that the highest correlation between observed and predicted evaporation values was found for the AR-M5P model (0.978), followed by the AR-SVM model (0.977), the AR model (0.968), the AR-RSS model (0.967), and the AR-REPTree model (0.961).

The study found that the total annual evaporation losses from Lake Nasser will reach 16.52, 17.54, and 18.56 billion cubic meters (BCM) per year, the also the expected total annual water saving reached 3.47, 3.684 and 3.898 BCM year¹ while the total annual power production reached 1.266×10^9 , 1.425×10^9 and 1.583×10^9 MW month⁻¹ in 2030, 2050, and 2070, respectively. This is a significant amount of water, and it is therefore important to find ways to reduce evaporation losses. The study recommends extending water management studies for shallow water depths to minimize evaporation losses and save freshwater. This could be done by covering shallow water depths with floating solar panels, which would reduce evaporation and generate electricity at the same time. The study also recommends that policymakers take steps to address the expected rising in temperature, climate changes, rising in sea levels and drought. This could include investing in water conservation measures, such as water recycling and desalination, and developing drought-resistant crops. By taking these steps, policymakers can help to ensure that Egypt has a sustainable water supply in the future.

Acknowledgements The authors thank the Department of Water and Water Structure Engineering, Faculty of Engineering, Zagazig University, Zagazig 44519, Egypt, for constant support during the study.

Authors' Contributions Ismail Abdelaty, N. L. Kushwaha and Abhishek Patel: Conceptualisation, Methodology, Investigation, Formal analysis, Data curation. Ismail Abdelaty, N. L. Kushwaha and Abhishek Patel: Visualisation, writing—original draft, Writing—review & editing, Resources. Ismail Abdelaty, N. L. Kushwaha and Abhishek Patel: Supervision, Writing—review & editing.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Availability of Data and Material Upon request.

Code Availability Upon request.

Declarations

Ethics Approval Not applicable.

Consent to Participate Yes.

Consent for Publication Yes.

Conflicts of Interest/Competing Interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdel Wahab M, Essa Y, Khalil A, Elfadli K, Giulia P (2018) Water loss in Egypt based on the lake Nasser evaporation and agricultural evapotranspiration. *Environ Asia* 11:192–204
- Abd-Elaty I, Pugliese L, Bali KM, Grismer ME, Eltarabily MG (2021) Modelling the impact of lining and covering irrigation canals on underlying groundwater stores in the Nile Delta, Egypt. *Hydrol Process* 36. <https://doi.org/10.1002/hyp.14466>
- Abd-Elhamid HF, Ahmed A, Zelenáková M, Vranayová Z, Fathy I (2021) Reservoir management by reducing evaporation using floating photovoltaic system: A case study of Lake Nasser, Egypt. *Water* 13:769. <https://doi.org/10.3390/w13060769>
- Abed M, Imteaz MA, Ahmed AN et al (2022) Modelling monthly pan evaporation utilising Random Forest and deep learning algorithms. *Sci Rep* 12:13132. <https://doi.org/10.1038/s41598-022-17263-3>
- Adnan R, Heddad S, Yaseen Z, Shahid S, Kisi O, Li B (2020) Prediction of potential evapotranspiration using temperature-based heuristic approaches. *Sustainability* 13:297. <https://doi.org/10.3390/su13010297>
- Agrawal Y, Kumar M, Ananthakrishnan S et al (2022) Evapotranspiration modeling using different tree based ensemble machine learning algorithm. *Water Resour Manag* 36:1025–1042. <https://doi.org/10.1007/s11269-022-03067-7>
- Allawi MF, Binti Othman F, Afan HA, Ahmed AN, Hossain MS, Fai CM, El-Shafie A (2019) Reservoir evaporation prediction modeling based on artificial intelligence methods. *Water* 11(6):1226. <https://doi.org/10.3390/w11061226>
- Almeida RM, Schmitt R, Grodsky SM, Flecker AS, Gomes CP, Zhao L, Liu H, Barros N, Kelman R, McIntyre PB (2022) Floating solar power could help fight climate change—Let us get it right. *Nature* 606:246–249. <https://doi.org/10.1038/d41586-022-01525-1>
- Al-Mukhtar M (2021) Modeling of pan evaporation based on the development of machine learning methods. *Theor Appl Climatol* 1–19. <https://doi.org/10.1007/s00704-021-03760-4>
- Al Sudani ZA, Salem GS (2022) Evaporation rate prediction using advanced machine learning models: a comparative study. *Adv Meteorol* 2022(1433835):13. <https://doi.org/10.1155/2022/1433835>
- Antonopoulos VZ, Gianniou SK (2022) Analysis and modelling of temperature at the water – atmosphere interface of a lake by energy budget and ANNs models. *Environ Process* 9:15. <https://doi.org/10.1007/s40710-022-00572-0>
- Arabameri A, Pal SC, Rezaie F, Nalivan OA, Chowdhuri I, Saha A, Lee S, Moayedi H (2021) Modeling groundwater potential using novel GIS-based machine-learning ensemble techniques. *J Hydrol Reg Stud* 36:100848. <https://doi.org/10.1016/j.ejrh.2021.100848>
- Asmal K, Lakshmi CJ, Henderson J, Lindahl G, Scudder T, Carino J, Blackmore D et al (2000) Dams and development, a new framework for decision-making: The report of the world commission on dams. Earthscan, London
- Beça P, Rodrigues AC, Nunes JP et al (2023) Optimizing reservoir water management in a changing climate. *Water Resour Manag* 37:3423–3437. <https://doi.org/10.1007/s11269-023-03508-x>
- Cazzaniga R, Rosa-Clot M (2021) The booming of floating PV. *Sol Energy* 219:3–10. <https://doi.org/10.1016/j.solener.2020.09.057>
- Chen J-L, Yang H, Lv M-Q, et al (2019) Estimation of monthly pan evaporation using support vector machine in Three Gorges Reservoir Area, China. *Theor Appl Climatol* 138:1095–1107. <https://doi.org/10.1007/s00704-019-02871-3>
- Choi YK (2014) A study on power generation analysis of floating PV system considering environmental impact. *Int J Softw Eng Appl* 8:75–84. <https://doi.org/10.14257/ijseia.2014.8.1.07>

- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1007/BF00994018>
- Deo RC, Samui P, Kim D (2016) Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. *Stoch Environ Res Risk Assess* 30:1769–1784. <https://doi.org/10.1007/s00477-015-1153-y>
- Duan Z, Bastiaanssen W (2017) Evaluation of three energy balance-based evaporation models for estimating monthly evaporation for five lakes using derived heat storage changes from a hysteresis model. *Environ Res Lett* 12:024005. <https://doi.org/10.1088/1748-9326/aa568e>
- Elba E, Urban B, Ettmer B, Farghaly D (2017) Mitigating the impact of climate change by reducing evaporation losses: Sediment removal from the High Aswan Dam reservoir. *Am J Clim Chang* 6:230–246. <https://doi.org/10.4236/ajcc.2017.62012>
- Ebaid HMI, Ismail SS (2010) Lake Nasser Evaporation Reduction Study. *J Adv Res* 1:315–322. <https://doi.org/10.1016/j.jare.2010.09.002>
- Essak L, Ghosh A (2022) Floating photovoltaics: A review. *Clean Technol* 4(3):752–769. <https://doi.org/10.3390/cleantechnol4030046>
- Faramarzadeh M, Ehsani MR, Akbari M et al (2023) Application of machine learning and remote sensing for gap-filling daily precipitation data of a sparsely gauged basin in East Africa. *Environ Process* 10:8. <https://doi.org/10.1007/s40710-023-00625-y>
- Friedman JH, Stuetzle W (1981) Projection pursuit regression. *J Am Stat Assoc* 76:817–823. <https://doi.org/10.1080/01621459.1981.10477729>
- Ganatra A, Bhensdadia CK (2012) Improved decision tree induction algorithm with feature selection, cross validation, model complexity and reduced error pruning data center networking view project big data view project. *J Comput Sci Inf Technol* 3:3427–3431
- Ghorbani MA, Jabejdar MA, Yaseen ZM, Inyurt S (2021) Solving the pan evaporation process complexity using the development of multiple mode of neurocomputing models. *Theor Appl Clim* 145:1521–1539. <https://doi.org/10.1007/s00704-021-03724-8>
- Hassan RMA, Hekal NTH, Mansor NMS (2007) Evaporation reduction from lake nasser using new environmentally safe techniques. *Int Water Technol Conf Sharm El-Sheikh* 179–194
- Hassan A, Ismail S, Elmoustafa A, Khalaf S (2017) Evaluating evaporation rates using numerical model (Delft3D). *Curr Sci Int* 6:402–411
- Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20:832–844. <https://doi.org/10.1109/34.709601>
- Junzeng XU, Qi W, Shizhang P, Yanmei YU (2012) Error of saturation vapor pressure calculated by different formulas and its effect on calculation of reference evapotranspiration in high latitude cold region. *Procedia Eng Int Conf Mod Hydraul Eng* 2:43–48. <https://doi.org/10.1016/j.proeng.2012.01.680>
- Katsaros K (2001) Evaporation and humidity. In *Encyclopedia of Ocean Sciences* Elsevier. <https://doi.org/10.1006/rwos.2001.0068>
- Khosravi K, Golkarian A, Boojij MJ, Barzegar R, Sun W, Yaseen ZM, Mosavi A (2021) Improving daily stochastic streamflow prediction: Comparison of novel hybrid data-mining algorithms. *Hydrol Sci J* 66(9):1457–1474. <https://doi.org/10.1080/02626667.2021.1928673>
- Kim S, Kim HS (2008) Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modelling. *J Hydrol* 351:299–317. <https://doi.org/10.1016/j.jhydrol.2007.12.014>
- Kira K, Rendell LA (1992) A practical approach to feature selection. In: Sleeman, D., Edwards, P. (Eds.), *Mach Learn Proc*. Morgan Kaufmann, San Francisco (CA), pp. 249–256. <https://doi.org/10.1016/B978-1-55860-247-2.50037-1>
- Kişî Ö (2006) Daily pan evaporation modelling using a neuro-fuzzy computing technique. *J Hydrol* 329:636–646. <https://doi.org/10.1016/j.jhydrol.2006.03.015>
- Kisi O, Shiri J, Demir V (2017) Hydrological time series forecasting using three different heuristic regression techniques, 1st edn. Elsevier Inc, Handbook of Neural Computation
- Kosmopoulos P, Kazadzis S, El-Askary H (2013) The solar atlas of Egypt. Available online: <http://www.nrea.gov/Content/files/SOLAR%20ATLAS%202018%20digital1.pdf>. Accessed 11 Mar 2021
- Kumar P, Kumar Jaipaul D, Tiwari AK (2012) Evaporation estimation using artificial neural networks and adaptive Neuro-Fuzzy inference system techniques. *Pak J Meteorol* 8(16):81–88. <https://doi.org/10.7763/IJCTE.2012.V4.424>
- Kushwaha NL, Rajput J, Elbeltagi A, Elnaggar AY, Sena DR, Vishwakarma DK, Mani I, Hussein EE (2021) Data intelligence model and meta-heuristic algorithms-based pan evaporation modelling in two different agro-climatic zones: A case study from Northern India. *Atmosphere* 12:1654. <https://doi.org/10.3390/atmos12121654>

- Kushwaha NL, Rajput J, Sena DR, Elbeltagi A, Singh DK, Mani I (2022) Evaluation of data-driven hybrid machine learning algorithms for modelling daily reference evapotranspiration. *Atmos Ocean* 62:1–22. <https://doi.org/10.1080/07055900.2022.2087589>
- LNFD (2008) Climate change and its effects on water resources management in Egypt lake nasser flood and drought control (LNFD) PROJECT REpORTs. Ministry of Water Resources and Irrigation—Planning Sector, Giza
- Malik A, Kumar A, Kim S, Kashani MH, Karimi V, Sharafati A, Ghorbani MA, Al-Ansari N, Salih SQ, Yaseen ZM, Chau KW (2020) Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model. *Eng Appl Comput Fluid Mech* 14:323–338. <https://doi.org/10.1080/19942060.2020.1715845>
- Onyari EK, Ilunga FM (2013) Application of MLP neural network and M5P model tree in predicting streamflow: A case study of Luvuvhu catchment, South Africa. *Int J Innov Manag Technol* 4(1):11. <https://doi.org/10.7763/IJIMT.2013.V4.347>
- Pammar L, Deka PC (2017) Daily pan evaporation modeling in climatically contrasting zones with hybridization of wavelet transform and support vector machines. *Paddy Water Environ* 15:711–722. <https://doi.org/10.1007/s10333-016-0571-x>
- Quinlan JR (1987) Simplifying decision trees. *Int J Man-Mach Stud* 27:221–234
- Quinlan JR (1992) Learning with continuous classes. *Aust Jt Conf Artif Intell* 92:343–348
- Rosenberry DO, Winter TC, Buso DC, Likens GE (2007) Comparison of 15 evaporation methods applied to a small mountain lake in the northeastern USA. *J Hydrol* 340:149–166. <https://doi.org/10.1016/j.jhydrol.2007.03.018>
- Sahu A, Yadav N, Sudhakar K (2016) Floating photovoltaic power plant: A review. *Renew Sustain Energy Rev* 66:815–824
- Sikora R, Al-Laymoun O, Sikora R, Al-Laymoun O (2015) A modified stacking ensemble machine learning algorithm using genetic algorithms through big data analytics. In: *Handbook of Research on Organizational Transformations through Big Data Analytics*. IGI Global 43–53. <https://doi.org/10.58729/1941-6679.1061>
- Singh A, Singh R, Kumar AS, Kumar A, Hanwat S, Tripathi V (2021) Evaluation of soft computing and regression-based techniques for the estimation of evaporation. *J Water Clim Chang* 12:32–43. <https://doi.org/10.2166/wcc.2019.101>
- Skurichina M, Duin R (2002) Bagging, boosting and the random subspace method for linear classifier. *Pattern Anal Appl* 5:121–135. <https://doi.org/10.4028/www.scientific.net/msf.440-441.77>
- Tabari H, Marofi S, Sabziparvar AA (2009) Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. *Irrigation Science* 28(5):399–406. <https://doi.org/10.1007/s00271-009-0201-0>
- TR J, Reddy NS, Acharya UD (2023) Modeling daily reference evapotranspiration from climate variables: Assessment of bagging and boosting regression approaches. *Water Resour Manag* 37:1013–1032. <https://doi.org/10.1007/s11269-022-03399-4>
- Trapani K, Millar DL (2013) Proposing offshore photo voltaic (PV) technology to the energy mix of the Maltese islands. *Energy Convers Manag* 67:18–26. <https://doi.org/10.1016/j.enconman.2012.10.022>
- Vishwakarma DK, Pandey K, Kaur A, Kushwaha NL, Kumar R, Ali R, Elbeltagi A, Kuriqi A (2022) Methods to estimate evapotranspiration in humid and subtropical climate conditions. *Agric Water Manag* 261:107378. <https://doi.org/10.1016/j.agwat.2021.107378>
- Wolpert DH (1992) Stacked generalization. *Neural Netw* 5:241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- Yin L, Zhou Y, Jiang T, Xu Y, Liu T, Li N, Zhou K, Yu L, Guo C, Murto P, Xu X (2021) Semitransparent polymer solar cells floating on water: selected transmission windows and active control of algal growth†. *J Mater Chem c* 9:13132–13143. <https://doi.org/10.1039/D1TC03110D>
- Zhang N, Jiang T, Guo C, Qiao L, Ji Q, Yin L, Yu L, Murto P, Xu X (2020) High-performance semitransparent polymer solar cells floating on water: Rational analysis of power generation, water evaporation and algal growth. *Nano Energy* 77. <https://doi.org/10.1016/j.nanoen.2020.105111>
- Zhou Z-H (2009) Ensemble learning. In: Li, S.Z., Jain, A. (Eds.), *Encycl Biomet Springer US, Boston, MA*, pp. 270–273. https://doi.org/10.1007/978-0-387-73003-5_293

Authors and Affiliations

Ismail Abd-Elaty¹ · N. L. Kushwaha²  · Abhishek Patel³

✉ Ismail Abd-Elaty
Eng_abdelaty2006@yahoo.com

N. L. Kushwaha
nand.kushwaha@icar.gov.in

Abhishek Patel
abhi.patel.ape121@gmail.com

¹ Department of Water and Water Structures Engineering, Faculty of Engineering, Zagazig University, Zagazig 44519, Egypt

² Division of Agricultural Engineering, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India

³ ICAR-Central Arid Zone Research Institute, Regional Research Station, Kukma, Bhuj, Gujrat 370105, India