Wasit Journal of Engineering Sciences

Journal homepage: https://ejuow.uowasit.edu.iq

# A hybrid model to improve reference evapotranspiration prediction: Integrating ANN and PSO

Hadeel Essa Khairan [1, 2], and Salah L. Zubaidi [2]

**Affiliations**
[1] Ministry Of Water Resources, State department for the operation of irrigation and drainage projects, Water Resources Directorate in Wasit;
[2] Department of Civil Engineering, Wasit University, Wasit, 52001, Iraq;
**Correspondence**
Hadeel Essa Khairan, Department of Civil Engineering, Wasit University, Wasit,Iraq.

Email:

h.khairan627@uowasit.edu.iq

**Received**
3-June-2023
**Revised**
22-June-2023
**Accepted**
14-August-2023
Doi: 10.31185/ejuow.Vol11.Iss3.450

**Abstract**
Reference evapotranspiration (ETo), one of the key elements of the hydrological cycle, is crucial for managing irrigation and drainage systems. In order to estimate monthly ETo, this study tested the prediction abilities of a unique hybrid methodology that coupled data pre-processing with a hybrid model composed of an artificial neural network (ANN) and particle swarm optimisation (PSO). In order to train and evaluate the model, monthly meteorological data were collected in Al-Kut City, Iraq, from 1990 to 2020. A range of statistical indicators were used to assess the model, including RMSE, NSE, and $R^2$. The outcomes showed that the model, with a coefficient of determination of 0.93, is effective and has good simulation levels.

**Keywords:** Reference evapotranspiration, ANN, PSO, MI, Al Kut.

**الخلاصة:** يعتبر التبخر المرجعي (ETo) أحد العناصر الرئيسية للدورة الهيدرولوجية ، وهو أمر بالغ الأهمية لإدارة أنظمة الري والصرف. من أجل تقدير ETo شهريًا ، اختبرت هذه الدراسة قدرات التنبؤ لمنهجية هجينة فريدة تجمع المعالجة المسبقة للبيانات بنموذج هجين يتكون من شبكة عصبية اصطناعية (ANN) وخوارزمية تحسين سرب الجسيمات (PSO). من أجل تدريب النموذج وتقييمه ، تم جمع بيانات الأرصاد الجوية الشهرية في مدينة الكوت ، العراق ، من 1990 إلى 2020. تم استخدام مجموعة من المؤشرات الإحصائية لتقييم النموذج ، بما في ذلك RMSE و NSE و $R^2$. أظهرت النتائج أن النموذج ، مع معامل تحديد 0.93 ، فعال ولديه مستويات محاكاة جيدة.

## 1. INTRODUCTION

In the hydrologic cycle, evapotranspiration (ETo) is one of the crucial processes. The ETo dynamics have a significant impact on a number of environmental factors, including the sustainable management of water supplies [1]. Direct measurement of this crucial parameter with lysimeters is nearly impossible, time-consuming, and expensive [2]. Consequently, it is essential to estimate ETo accurately in order to increase the effectiveness of water application [3]. The Penman Monteith (PM) model is still recognised as the standard reference for computing ETo as of this writing, after the United Nations Food and Agriculture Organisation's approval [4, 5]. However, it requires a variety of meteorological characteristics over a long period of time as input variables [6]. Also, many weather stations in developing countries lack comprehensive long-term meteorological data due to their level of development and lack of funding [7].

This is what prompted the researchers to use machine learning (ML) to estimate ETo. ML has a high degree of accuracy and precision in predicting variables, and it is quick and inexpensive to implement [8]. For example, Ashrafzadeh, Kişi [9] predicted monthly ETo up to 24 months in advance using support vector machines (SVM), the seasonal autoregressive integrated moving average (SARIMA), and the group method of data handling

(GMDH). The authors examined meteorological data from four weather stations located in Iran's Guilan Plain. Based on the error metrics, it was determined that all of the suggested models accurately predicted monthly ETo. Lu, Fan [10] looked at the performance of the hybrid extreme gradient boosting (XGBoost) model in combination with the Grey Wolf Optimizer (GWO) algorithm for forecasting multi-step-ahead ETo. Also, three other conventional machine learning models were used, namely standalone XGBoost, multi-layer perceptrons (MLP), and M5 model trees (M5), in the subtropical zone of China. And all of them achieved good results.

A literature study also reveals that researchers have successfully used a number of hybrid models that combine traditional neural networks with meta-heuristic techniques to estimate ETo [11]. So, the purpose of this study was to add to the body of knowledge currently available about the ability to precisely model monthly ETo using an artificial neural network (ANN) in conjunction with a bio-inspired optimisation algorithm, in this case, particle swarm optimisation (PSO).

Recently, Khairan, Zubaidi [12] examined simulating the models for predicting ETo and recommended combining ML models with optimisation algorithms that draw inspiration from nature to help improve the precision of standalone models. This study also recommended that:

1. performing data pre-processing In order to prevent data noise and focus on finding the best predictor combination.
2. Using the meta-heuristic algorithms to select and fine-tune One of the most important variables affecting the results and effectiveness of the ANN model is the learning rate coefficient.

The main goals of the current study are to:

1. selecting the ideal model input scenario by utilising average mutual information (AMI) and using single-spectrum analysis (SSA) to denoise the data.
2. Integrate particle swarm optimisation (PSO) with the ANN model to find the ideal ANN parameters.
3. Assess how effectively the PSO-ANN technique simulates ETo.

## 2. Area of Study and Data Set

Al Kut City is located in southeast Iraq on the Tigris River [13]. It is situated between two latitudes (32o 21" and 32o 34") north and two longitudes (45o 54" and 45o 45") east, with an average elevation of roughly 20 metres [14]. The selected area experiences chilly winters and dry, hot summers, with 300–150 mm of annual rainfall on average [15]. The monthly meteorological parameters required for this study include maximum temperature (Tmax), minimum temperature (Tmin) (C), dew forest (DF) (C), wind speed at a height of 2 metres (U2) (m/s), solar radiation (Rs) (MJ/m2/day) and relative humidity (RH) (present). were compiled over 30 years, from January 1, 1990, to December 31, 2022.

## 3. Methodology

### 3.1. FAO-56 Penman–Monteith Model

The ETo values were computed using the meteorological data inputted into the FAO-56 PM model. This strategy is widely acknowledged by scientific societies and has become a popular solution in instances where obtaining ETo values experimentally is challenging [16].

The FAO-56 PM model is represented as follows:

$$\boldsymbol{ETo} = \frac{\mathbf{0.408}\Delta(\boldsymbol{Rn}-\boldsymbol{G})+\boldsymbol{\gamma}\dfrac{\mathbf{900}}{\boldsymbol{Tave}+\mathbf{273}}\boldsymbol{U2}(\boldsymbol{es}-\boldsymbol{ea})}{\Delta+\boldsymbol{\gamma}(\mathbf{1}+\mathbf{0.34}\boldsymbol{U2})} \qquad (1)$$

Where ETo is reference evapotranspiration (mm/day), Rn is net radiation ($MJm^{-2}day^{-1}$), G is soil heat flux ($MJm^{-2}day^{-1}$), T is the monthly mean air temperature (°C), U2 is the wind speed at 2 m height (m/s), es is mean saturation vapour pressure (kPa), ea is actual vapour pressure (kPa), Δ is the slope of vapour pressure function (kPa/°C), and γ is psychometric constant (kPa/°C). The formula and the ETo principle are clearly explained in FAO Irrigation and Drainage Study No. 56 by Allen, Pereira [17].

## 3.2. Pre-processing Data

In this study, more than one pre-processing method was used for the data. First, the data was normalised by using the natural logarithm. Normalisation makes the data have a normal distribution or one that is close to it [18]. The statistical programme SPSS 24 is used in this study. After that, the SSA approach was used to denoise the normalised data. SSA decomposes the time series into a group of signals; in this case, the decomposition was done into twelve signals, and the first signal, which is the trend, was used as it bears 98 per cent of the signal strength. Finally, MI was used to determine the optimal lag time, as shown in the figure 1.

## 3.3.Particle Swarm Optimisation (PSO)

This meta-heuristic optimisation technique is inspired by the swarm intelligence paradigm, which mimics the cooperative behaviours of fish and birds [19]. Each potential solution in the PSO algorithm is treated as a point or particle. All potential solutions group together as a particle swarm. Each particle has two properties: velocity and position. Velocity stands for the particle's movement speed, and position stands for the particle's movement direction. Each particle conducts a separate search for the ideal solution in the search space and logs it as the current individual extremum. Subsequently, the particle swarm as a whole shares the individual extremum. The present global optimal solution for the entire particle swarm is the best individual extremum [4, 20]. There are numerous technical and scientific areas where it is used successfully. For example, agricultural modelling [21], simulation of rainfall-runoff [22]. Additional details about PSO can be found in [19].

## 3.4.Artificial Neural Network (ANN)

ANN is an effective computational method for handling non-linear systems [23]. The multilayer perceptron network (MLP) is a typical neural network architecture [24]. ANN architectures typically have three layers: input, hidden, and output [25]. The input layer accepts and distributes the input variables (computed ETo using FAO-56 PM), the hidden layer contains two sigmoid tangent activation functions to process the data, and the output layer contains a variable that simulates the inputs (forecasted ETo). However, a time-consuming trial-and-error approach is not always the best solution. As a result, meta-heuristic methods were integrated with ANN to determine the best learning rate and number of neurons for hidden layers [12].

## 3.5.Model Performance Evaluation

The performance of the proposed model was evaluated using the root mean squared error (RMSE), Nash-Sutcliffe model efficiency (NSE), and coefficient of determination ($R^2$). Equations (2) to (4) are used to define these indicators.

$$RMSE = \sqrt{\sum_{i}^{N} \frac{(Ri - Pi)^2}{N}} \tag{2}$$

$$NSE = 1 - \left\{ \frac{\sum_{i=1}^{N} (Ri - Pi)^2}{\sum_{i=1}^{N} \left(Ri - \overline{Ri}\right)^2} \right\} \tag{3}$$

$$R^2 = \left( \frac{\sum_{i=1}^{N} (Ri - \overline{Ri})(Pi - \overline{Pi})}{\sqrt{\sum (Ri - \overline{Ri})^2 \sum (Pi - \overline{Pi})^2}} \right)^{\wedge 2} \tag{4}$$

Where:

$R_i$: measure ETo,

$p_i$: predicted ETo,

$\overline{Ri}$: average of predicted ETo,

$\overline{Pi}$: average of measure ETo, and

$N$: sample size.

Also, a graphical test was used to check the PSO-ANN predictive performance.

## 4. Results and Discussion

### 4.1.Development Model Input

The natural logarithm was used to normalise the data. Then, the SSA method was used to acquire noise-free ETo time series data, and this was achieved by analysing the normalised time series into twelve components. Figure 1 displays the normalised time series and the first four elements of the ETo parameter. Data pre-processing enhances the correlation coefficients of the monthly ETo between the goal and predictors (Lags); for example, the correlation coefficient of Lag 1's raw data rose dramatically from 0.83 to 0.99. The correlation coefficients for the first four lags of the denoise time series were 0.99, 0.96, 0.91, and 0.85, respectively.

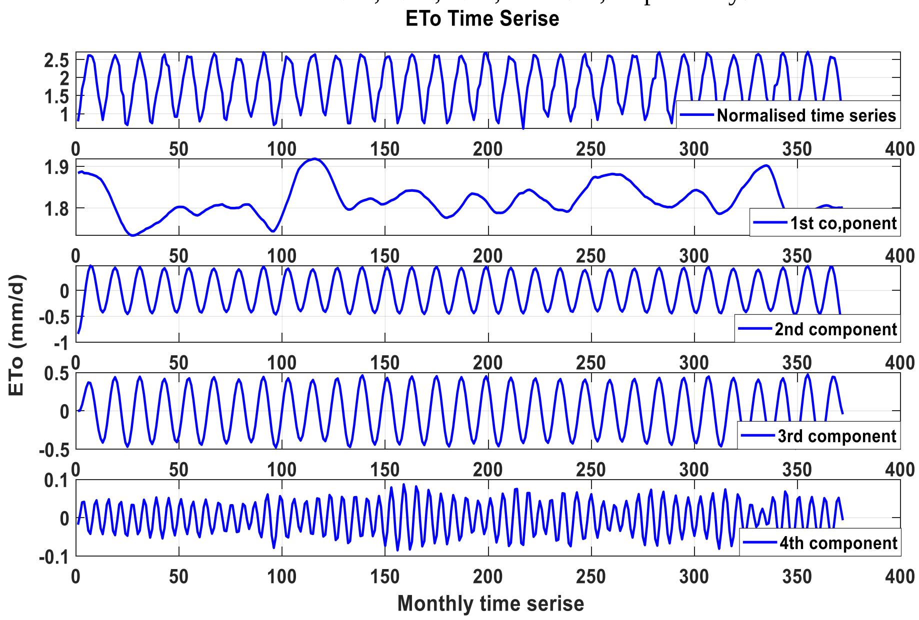


Figure 1: Normalised time series and the first four components obtained by the SSA approach.

The MI technique, which was utilised to select the best scenario for model input, is shown in Figure 2. The time lag is selected as the initial minimum of average mutual information(AMI), according to [26]. Four lags (Lag1 through Lag4) of the monthly ETo value were utilised to simulate future monthly ETo based on the AMI figure.

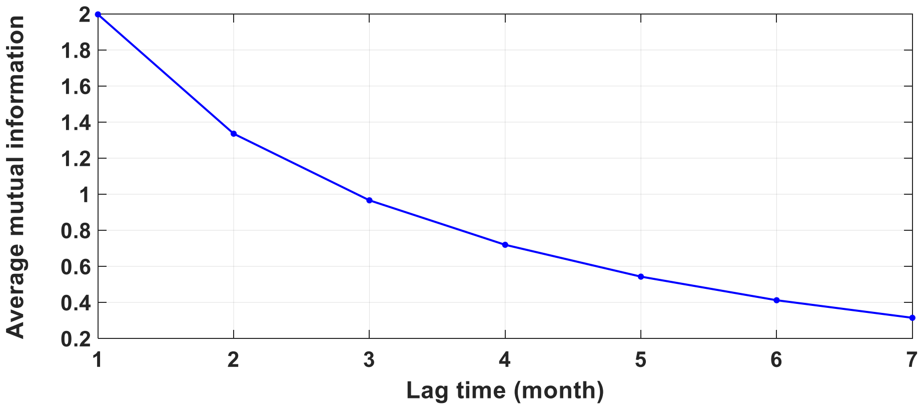


Figure 2. Lag time obtained by AMI

After that the time series were split into three categories: training (70%), testing (15%), and validation (15%).

### 4.2.Analysis of the PSO Algorithm

To establish the ANN model's ideal hyper-parameters (N1, N2, and Lr), the PSO algorithm was integrated with the ANN model. The MATLAB toolbox was used to run the PSO-ANN algorithms. To obtain a minimal fitness function (RMSE) in this work, swarm sizes of 10, 20, 30, 40, and 50 were used five times for each swarm, each with 200 iterations. Figure 3 illustrates the PSO-ANN algorithm by displaying the best fitness function for each swarm.

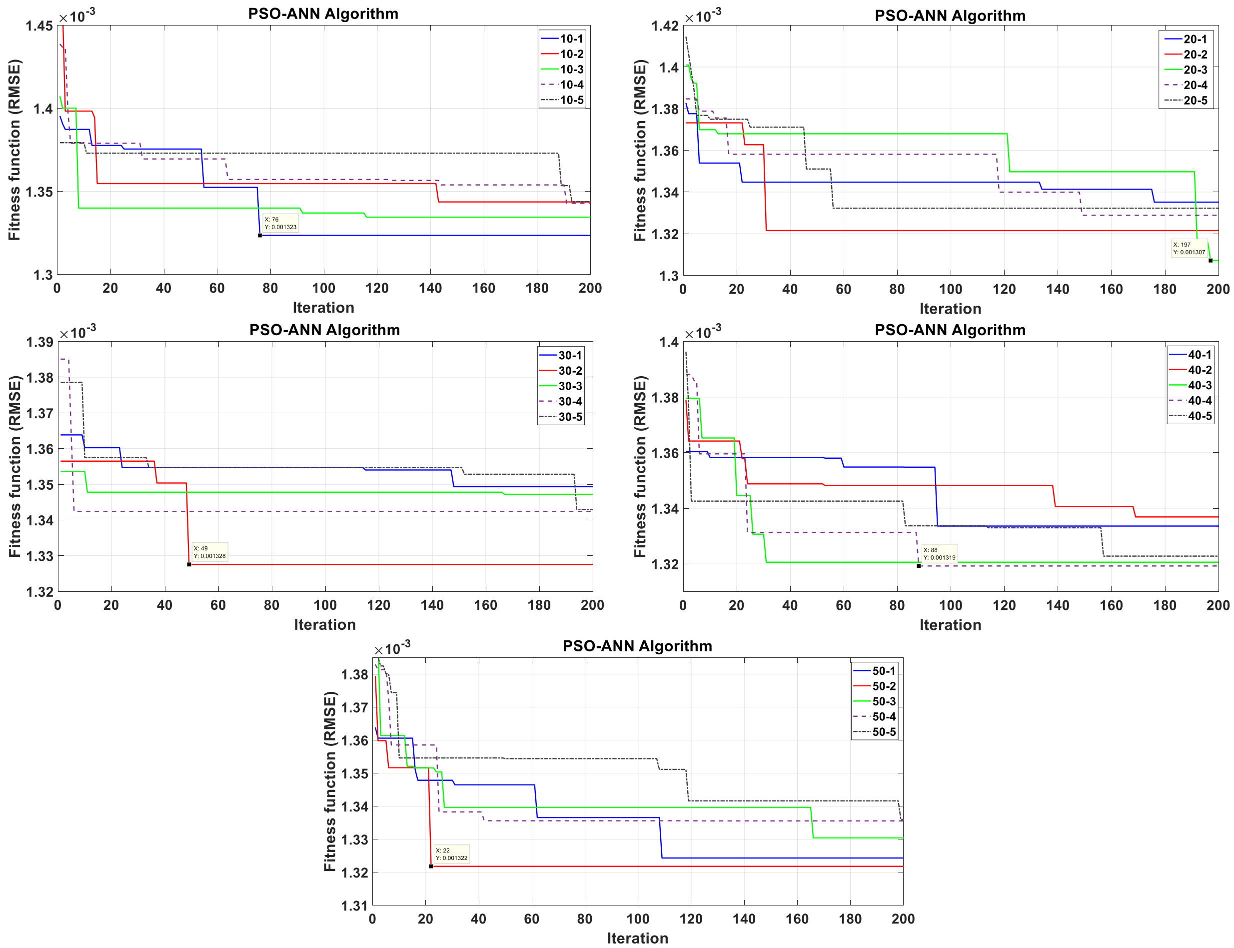


Figure 3. Performance of the PSO algorithm.

The optimal swarm size (20_3) for the PSO-ANN technique is shown in Figure 4 (RMSE = 0.0013, after 197 iterations). The ANN model's best swarm size produced the following hyperparameter values: Lr = 0.5291; N1 = 5; N2 = 9.

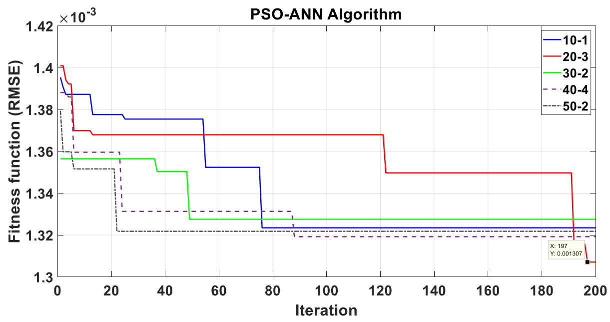


Figure 4. Fitness function of PSO-ANN algorithmm.

## 4.3.Performance Assessment

A variety of statistical indicators were utilised to assess the effectiveness of the technique. The RMSE, NSE, and $R^2$ values are shown in Table 1. According to Dawson, Abrahart [27], the PSO -ANN model performed good in terms of ETo prediction.

Table 1. PSO performance evaluation standards for the validation stage.

| Model | RMSE | NSE | $R^2$ |
|---|---|---|---|
| PSO-ANN | 0.071 | 0.92 | 0.93 |

Figure 6 shows the comparison of the PSO-ANN model's observed and predicted data for the validation stage. The scale of the figure shows that the simulated data closely match the observed data pattern (trend + periodicity) along the time series.

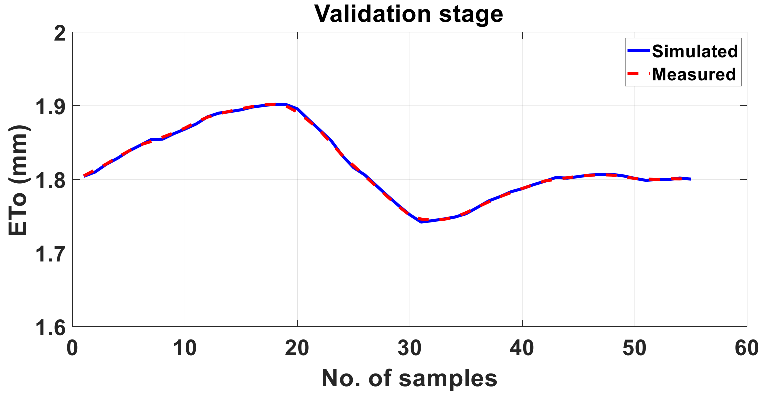


Figure 5. The observed and simulated data at the validation stage

Overall, the data presented above demonstrate that:

1. These findings demonstrate the potential effectiveness of SSA to denoise data and AMI approaches to choose scenarios for explanatory factors.
2. PSO, which integrates the ANN method, is a reliable method for monthly ETo prediction.
3. Statistical analyses showed that the suggested method correctly predicted monthly ETo.

## 5. Conclusions

This study offered a novel hybrid methodology to simulate ETo using 30 years' worth of meteorological data. An ANN model, the PSO algorithm, and data preparation techniques are all used in the process. Monthly data on Tmax, Tmin, U2, Rs, DF, and RH were used to construct ETo using the FAO-56 PM equation. The results of this study showed how crucial data pre-processing methods, such as SSA and MI, are for improving the quality of the raw data and selecting the best lagged scenario. In terms of several statistical indicators, such as RMSE, NSE, and $R^2$, and graphical approaches, the forecasting model performed well. With $R^2 = 0.93$, RMSE = 0.071, and NSE = 0.92, the model demonstrates that the provided methodology is a reliable method for forecasting monthly ETo.

## References


1. Adnan, R.M., et al., Estimating reference evapotranspiration using hybrid adaptive fuzzy inferencing coupled with heuristic algorithms. Computers and Electronics in Agriculture, 2021. 191.
2. Ahmadi, F., et al., Application of an artificial intelligence technique enhanced with intelligent water drops for monthly reference evapotranspiration estimation. Agricultural Water Management, 2021. 244.
3. Alizamir, M., et al., Modelling reference evapotranspiration by combining neuro-fuzzy and evolutionary strategies. Acta Geophysica, 2020. 68(4): p. 1113-1126.
4. Chia, M.Y., Y.F. Huang, and C.H. Koo, Swarm-based optimization as stochastic training strategy for estimation of reference evapotranspiration using extreme learning machine. Agricultural Water Management, 2021. 243.
5. Chia, M.Y., Y.F. Huang, and C.H. Koo, Improving reference evapotranspiration estimation using novel inter-model ensemble approaches. Computers and Electronics in Agriculture, 2021. 187.
6. Jiao, P. and S.-J. Hu, Optimal Alternative for Quantifying Reference Evapotranspiration in Northern Xinjiang. Water, 2021. 14(1).
7. Dong, J., et al., Comparison of four bio-inspired algorithms to optimize KNEA for predicting monthly reference evapotranspiration in different climate zones of China. Computers and Electronics in Agriculture, 2021. 186.

8. El-Kenawy, E.M., et al., Improved weighted ensemble learning for predicting the daily reference evapotranspiration under the semi-arid climate conditions. Environ Sci Pollut Res Int, 2022.
9. Ashrafzadeh, A., et al., Comparative Study of Time Series Models, Support Vector Machines, and GMDH in Forecasting Long-Term Evapotranspiration Rates in Northern Iran. Journal of Irrigation and Drainage Engineering, 2020. 146(6).
10. Lu, X., et al., Forecasting Multi-Step Ahead Monthly Reference Evapotranspiration Using Hybrid Extreme Gradient Boosting with GreyWolf Optimization Algorithm. Computer Modeling in Engineering & Sciences, 2020. 125(2): p. 699-723.
11. Majhi, B. and D. Naidu, Differential evolution based radial basis function neural network model for reference evapotranspiration estimation. SN Applied Sciences, 2021. 3(1).
12. Khairan, H.E., et al., Parameter Optimisation Based Hybrid Reference Evapotranspiration Prediction Models A Systematic Review of Current Implementations and Future Research Directions. Atmosphere, 2022. 14(1): p. 77.
13. Eiben, A.E. and C.A. Schippers, On evolutionary exploration and exploitation. Fundamenta Informaticae, 1998. 35(1-4): p. 35-50.
14. F. Balket, S. and N. M. Asmael, Study the Characteristics of Public Bus Routes in Al Kut City. Journal of Engineering and Sustainable Development, 2021. 25(Special): p. 3-186-3-194.
15. Al-Abadi, A. and A.H.D. Al-Aboodi, Optimum rain-gauges network design of some cities in Iraq. Journal of Babylon University/Engineering Sciences, 2014. 22(4): p. 946-958.
16. Roy, D.K., et al., Optimization algorithms as training approaches for prediction of reference evapotranspiration using adaptive neuro fuzzy inference system. Agricultural Water Management, 2021. 255.
17. Allen, R.G., et al., Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome, 1998. 300(9): p. D05109.
18. Alawsi, M.A., et al., Tuning ANN Hyperparameters by CPSOCGSA, MPA, and SMA for Short-Term SPI Drought Forecasting. Atmosphere, 2022. 13(9): p. 1436.
19. Kennedy, J. and R. Eberhart. Particle swarm optimization. in Proceedings of ICNN'95-international conference on neural networks. 1995. IEEE.
20. Yu, J., et al., A PSO-XGBoost Model for Estimating Daily Reference Evapotranspiration in the Solar Greenhouse. Intelligent Automation & Soft Computing, 2020. 26(5): p. 989-1003.
21. Nieto, P.G., et al., A new predictive model for the filtered volume and outlet parameters in micro-irrigation sand filters fed with effluents using the hybrid PSO–SVM-based approach. 2016. 125: p. 74-80.
22. Xu, Y., et al., Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. 2022. 608: p. 127553.
23. Tikhamarine, Y., et al., Estimation of monthly reference evapotranspiration using novel hybrid machine learning approaches. Hydrological Sciences Journal, 2019. 64(15): p. 1824-1842.
24. Gocić, M. and M. Arab Amiri, Reference Evapotranspiration Prediction Using Neural Networks and Optimum Time Lags. Water Resources Management, 2021. 35(6): p. 1913-1926.
25. Maroufpoor, S., O. Bozorg-Haddad, and E. Maroufpoor, Reference evapotranspiration estimating based on optimal input combination and hybrid artificial intelligent model: Hybridization of artificial neural network with grey wolf optimizer algorithm. Journal of Hydrology, 2020. 588.
26. Stergiou, N., Nonlinear analysis for human movement variability. 2018: CRC press.
27. Dawson, C.W., R.J. Abrahart, and L.M. See, HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environmental Modelling Software, 2007. 22(7): p. 1034-1052.