sustainability

MDPI

Article

Exploring the Applicability of Regression Models and Artificial Neural Networks for Calculating Reference Evapotranspiration in Arid Regions

Mohamed K. Abdel-Fattah [1,*], Sameh Kotb Abd-Elmabod [2,3,4], Zhenhua Zhang [5,6,*] and Abdel-Rhman M. A. Merwad [1]

1 Soil Science Department, Faculty of Agriculture, Zagazig University, Zagazig 44519, Egypt; amerwad@yahoo.com
2 Soils & Water Use Department, Agricultural and Biological Research Division, National Research Centre, Cairo 12622, Egypt; sk.abd-elmabod@nrc.sci.eg
3 Agriculture and Food Research Council, Academy of Scientific Research and Technology (ASRT), Cairo 11562, Egypt
4 State Key Laboratory of Desert and Oasis Ecology, Key Laboratory of Ecological Safety and Sustainable Development in Arid Lands, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China
5 Jiangsu Key Laboratory for Bioresources of Saline Soils, School of Wetlands, Yancheng Teachers University, Yancheng 224007, China
6 School of Agriculture and Environment, The University of Western Australia, Crawley, WA 6009, Australia
* Correspondence: mkabdelfattah@zu.edu.eg (M.K.A.-F.); zhenhuaz70@hotmail.com (Z.Z.)

Abstract: Reference evapotranspiration (ET_0) is critical in agriculture and irrigation water management, particularly in arid and semi-arid regions. Our study aimed to develop an accurate and efficient model for estimating ET_0 using various climatic variables as predictors. This research evaluated two model techniques, i.e., stepwise regression and artificial neural networks (ANNs), to identify the most effective model for calculating ET_0. The two models were developed and tested based on climate data obtained from the whole climatic station of Egypt. The CLIMWAT 2.0 program was used to acquire the climate data for Egypt from a total of 32 stations. This software is a dedicated meteorological database created specifically to work with the CROPWAT computer program. The models were developed using average climate data spanning 29 years, from 1991 to 2020. The obtained data were utilized to compute reference evapotranspiration using CROPWAT 8, based on the Penman–Monteith equation. The results showed that the ANN model demonstrated superior performance in ET_0 calculations compared to other methods, achieving a coefficient of determination (R^2) of 0.99 and a mean absolute percentage error (MAPE) of 2.7%. In contrast, the stepwise model regression yielded an R^2 of 0.95 and an MAPE of 8.06. On the other hand, the most influential climatic variables were maximum temperature, humidity, solar radiation, and wind speed. The findings of this study could be applied in various fields, such as agriculture, irrigation, and crop water requirements, to optimize crop growth under limited water resources and global environmental changes. Furthermore, our study identifies the limitations and challenges of applying these models in arid regions, such as data availability constraints and model complexity. We discuss the need for more extensive and reliable datasets and suggest future research directions, including ensemble modeling, remote sensing data integration, and evaluating climate change's impact on ET_0 estimation. Overall, this study contributes to the understanding of ET_0 estimation in arid regions and provides valuable insights into the applicability of regression models and ANNs. The superior performance of ANNs offers potential advancements in water resource management and agricultural planning, enabling more accurate and informed decision-making processes.

Keywords: regression; stepwise; models; artificial neural networks; reference evapotranspiration; water requirements

Citation: Abdel-Fattah, M.K.; Kotb Abd-Elmabod, S.; Zhang, Z.; Merwad, A.-R.M.A. Exploring the Applicability of Regression Models and Artificial Neural Networks for Calculating Reference Evapotranspiration in Arid Regions. *Sustainability* **2023**, *15*, 15494. https://doi.org/10.3390/su152115494

Academic Editors: Yen-Ming Chiang, Wen-Ping Tsai and Yu-Jia Chiu

Received: 10 September 2023
Revised: 18 October 2023
Accepted: 26 October 2023
Published: 31 October 2023

1. Introduction

Reference evapotranspiration (ET_0) refers to the amount of water lost through evaporation and transpiration from a hypothetical reference crop, assuming standardized crop height, surface resistance, and albedo under specific weather conditions [1–3]. ET_0 is commonly used to estimate the water requirements for crops and vegetation, and the typical method for measuring ET_0 is the Penman–Monteith equation that considers several weather parameters like temperature, humidity, wind speed, and solar radiation [1]. Forecasting ET_0 is vital in irrigation management, water resource planning, and climate change impact assessment. Accurate forecasts of ET_0 can assist farmers in optimizing irrigation scheduling and improving crop yields by minimizing water use and at the same time reducing the risk of water stress. ET_0 forecasts can help water managers and decision-makers allocate water availability and distributions in water resource planning, especially in regions with limited water resources.

Additionally, ET_0 estimates can be used to assess the potential impacts of climate change on water availability and vegetation, allowing for better climate adaptation and mitigation strategies [2–4]. Therefore, accurate ET_0 predictions are crucial for sustainable water management and agricultural production. Various models, including regression models, artificial neural networks, and machine learning algorithms, have been developed to forecast ET_0 using weather parameters data, soil moisture, and satellite remote sensing data [2–4]. These models have improved our ability to predict ET_0 and enable decision-makers to make better sustainable water resource management decisions.

Stepwise multiple regression and artificial neural networks (ANNs) are two popular methods in statistics and machine learning for predictive modelling and variable selection. Stepwise multiple regression is a statistical technique that identifies the most relevant subset of predictor variables based on their statistical significance and contribution to the model's predictive power. ANN, inspired by the human brain's structure, consists of interconnected nodes that process and transform input data to produce output predictions. It can learn complex patterns and relationships through training, adjusting the strength of its connections to minimize prediction errors. ANN does not require a priori assumptions about variable relationships, allowing it to capture nonlinear and interactive effects. However, it often requires more data and can be computationally intensive. Both artificial neural networks (ANNs) and regression models use weather parameters such as temperature, humidity, and solar radiation to predict ET_0. Also, many studies [3,5] have used these models to predict ET_0 using these weather parameters. In a study by Zhang et al. [3], an ANN model was developed using temperature, humidity, and radiation as input variables to predict daily ET_0 in the arid region of northwestern China. The study found that the ANN model outperformed a multiple linear regression model in predicting ET_0. Another study by Ismail et al. [5] used a regression model to predict ET_0 in Malaysia using temperature, humidity, and solar radiation as input variables. The study found that the regression model could accurately predict ET_0 for the region, demonstrating the model's usefulness in water resource management. ET_0 can be used as a predictor variable in regression models for estimating crop water requirements. For example, a study by Jia et al. [6] used ET_0 as a predictor variable in a multiple linear regression model to estimate the irrigation water requirement of maize crops in northern China. The research revealed that including ET_0 in the model enhanced precision in estimating irrigation water requirements. ET_0 can also be used as a predictor variable in artificial neural networks (ANN) and regression models to estimate crop water requirements. For example, a study by Alizadeh et al. [4] used ET_0 as an input variable in an ANN model to predict the water requirement of grapevines. The study found that the ANN model with ET_0 as an input variable outperformed a regression model in predicting the water requirement of grapevines. Temperature is one of the most important weather parameters affecting ET_0, as it affects the water evaporation rate from the soil and plants. Higher temperatures result in higher evaporation and transpiration rates, leading to higher ET_0 values. Humidity is another important weather parameter affecting ET_0. Higher humidity levels reduce the rate of evaporation and transpiration, leading to lower ET_0 values. Humidity can also affect the efficiency of water use by plants, as high humidity can reduce the amount of water that can be extracted from the soil by roots. Solar radiation is also important for predicting ET_0, as it drives evaporation and transpiration by providing the energy needed to convert water into water vapor. Higher solar radiation levels generally lead to higher ET_0 values, as more energy is available to convert water into water vapor [1,7,8]. Therefore, the temperature, humidity, and solar radiation are used in predicting ET_0, as well as in the applications of ET_0 in water resource management and agriculture. Zhang et al. [3] developed an artificial neural network (ANN) model to predict daily reference evapotranspiration (ET_0) in the arid region of northwestern China using temperature, humidity, and radiation as input variables. The study found that the ANN model outperformed a multiple linear regression model in predicting ET_0. Alizadeh et al. [4] used an ANN model to predict the water requirement of grapevines using ET_0, temperature, and solar radiation as input variables. The study found that the ANN model with ET_0 as an input variable outperformed a regression model in predicting the water requirement of grapevines. Both studies demonstrate the usefulness of ANN models in predicting ET_0 and crop water requirements and highlight the importance of using weather parameters such as temperature, humidity, and solar radiation for accurate ET_0 forecasting. Dimitriadou and Nikolakopoulos [9] used multiple linear regression models to predict reference evapotranspiration in Greece's Peloponnese region. The MLR5 model performed best, with five input variables. Other models with two to six inputs also showed satisfactory predictions. The study demonstrated MLR's potential for accurate predictive models. The study by Traore et al. [10] compared four alternative methods for estimating reference evapotranspiration (ET_0) in Burkina Faso, focusing on wind. The generalized regression neural network (GRNN) outperformed the reference model for Burkina Faso (RMBF), Hargreaves (HRG), and Blaney–Criddle (BCR), indicating its reliability in semi-arid zones of Africa. Wind was found to be a sensitive parameter in ET_0 estimation, highlighting the potential of ANN in such areas. The study by Abdullahi et al. [11] used a three-layered feed forward neural network (FFNN) to forecast monthly evapotranspiration (ET_0) in Northern Cyprus from 2017 to 2050. The FFNN models were more accurate than the multi-linear regression (MLR) models, and wind speed (U2) was found to be the most important input factor for accurate ET_0 estimation.

The preceding assessments of the literature show that the neural network technique has outperformed stepwise regression in many areas worldwide where it has been employed to estimate reference evapotranspiration. The main goal of this research was to develop a model that could effectively and accurately forecast reference evapotranspiration (ET_0) based on meteorological data. It also sought to evaluate several modeling approaches, including stepwise regression and artificial neural networks, to determine the best method for modeling ET_0. Additionally, the research sought to establish the importance of various climatic parameters in the model and which climatic factors have the greatest effects on ET_0. The study also sought to develop a realistic and trustworthy ET_0 model that could be applied to various fields, including irrigation, agriculture, and water management, to enhance crop development and preserve water supplies.

2. Materials and Methods

2.1. Study Area

Egypt is geographically situated in the northern region of Africa, with its boundaries defined by the Mediterranean Sea to the north, Libya to the west, Sudan to the south, and the Red Sea to the east (Figure 1). Egypt has an average elevation of 321 m above sea level. This study utilized the monthly average climatic data from 32 meteorological stations across Egypt, as depicted in Figure 1. Climatic data of Egypt were obtained using the Climwat 2.0 software, a climatic database specifically designed to work in tandem with the CROPWAT 8.0 computer program.

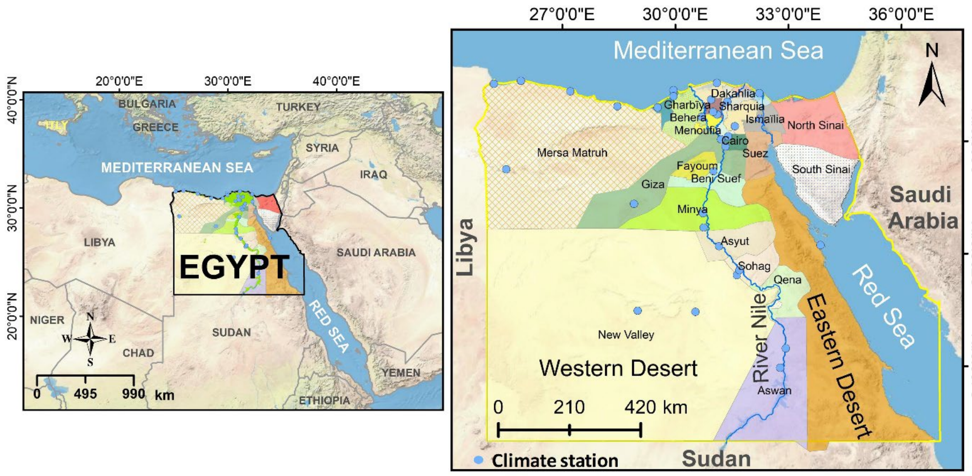

Figure 1. The location of the study area. In the left location of Egypt, and Egyptian provinces and climatic station on the right.

This software pairing facilitates the computation of crop water requirements, irrigation supply, and irrigation scheduling for various crops under diverse climatic circumstances and locations globally. With over 5000 stations worldwide, CLIMWAT 2.0 provides observed agroclimatic data.

2.2. Computed Reference Evapotranspiration (ET_0)

The obtained data were utilized to compute reference evapotranspiration using CROPWAT 8, based on the Penman–Monteith equation. The equation is expressed as:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\left(\frac{900}{T+273.16}\right)u2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u2)}$$

where ET_0 is the rate of evapotranspiration in mm/day, R_n is the net radiation at the crop surface in MJ/m^2/day, G is the soil heat flux density in MJ/m^2/day, T is the air temperature at 2 m height in degrees Celsius, u2 is the wind speed at 2 m height in m/s, e_s is the saturation vapor pressure at the air temperature in kPa, e_a is the actual vapor pressure in kPa, Δ is the slope of the saturation vapor pressure–temperature curve in kPa/°C, and γ is the psychrometric constant in kPa/°C.

2.3. Statistical Analysis

SPSS version 25 was used to conduct descriptive statistics on the climatic data, which included measures such as the number of observations, minimum and maximum values, mean, standard deviation, standard error, coefficient of variation, skewness, and kurtosis. In addition to the descriptive statistics, a Pearson correlation matrix was computed to examine the relationships between the various climatic variables. The Pearson correlation coefficient measures the linear relationship between two variables and ranges from −1 to +1, with values closer to +1 indicating a strong positive correlation, values closer to −1 indicating a strong negative correlation, and values close to 0 indicating little or no correlation. The correlation matrix provides a useful tool for identifying potential relationships between variables and can be used to guide further analysis and modelling.

To develop an ET_0 model using climatic variables as predictors, stepwise regression was utilized using XLSTAST version 2016. This method aimed to identify the most significant predictors and enhance the model's predictive power. Stepwise regression involves sequentially entering the climatic variables into the model based on their statistical significance. The most important variables in predicting ET_0 were selected and included in the final model, while variables that did not contribute significantly were removed. The probability of entering a variable into a stepwise regression model is typically set at 0.05, which means that a variable will be considered for inclusion in the model if its *p*-value is less than or equal to 0.05. The probability of removing a variable from the model is 0.1, which means that a variable will be removed if its *p*-value is greater than 0.1. In building the model, 70% of the available data was used for training, while the remaining 30% was used for validation. The goodness-of-fit statistics used to evaluate the performance of the ET_0 model include coefficient of determination (an R^2 value close to 1 indicates a good fit between the model and the data), mean squared error (a low MSE indicates a good fit between the model and the data), root mean squared error (a low RMSE indicates a good fit between the model and the data) and mean absolute percentage error (MAPE indicates the overall accuracy of the model's predictions).

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

$$\text{Adjusted } R^2 = 1 - \frac{\left(1 - R^2\right) \times (n - 1)}{n - k - 1}$$

$$\text{MSE} = \frac{1}{n} \sum (\text{Actual} - \text{Predicted})^2$$

$$\text{RMSE} = \sqrt{\text{MSE}}$$

$$\text{MAPE} = \frac{1}{n} \sum \frac{\text{Actual} - \text{Predicted}}{\text{Actual}} \times 100$$

where R^2 is the coefficient of determination, SS_{res} is the sum of squared residuals (the difference between the predicted and actual values), SS_{tot} is the total sum of squares (the difference between the actual values and the mean of the actual values), n is the number of observations, k is the number of predictors in the model, Actual is the actual value of the variable, and Predicted is the predicted value of the variable.

In addition to stepwise regression, SPSS version 25 was also used for developing an ET_0 model based on an artificial neural network (ANN). An ANN model is a type of machine learning model that is modeled after the structure and function of the human brain. The neural network model comprises multiple layers of interconnected nodes that simulate the neurons in the human brain. Methodological details: (1) Neural network architecture: We employed a feedforward artificial neural network (ANN) model for estimating reference evapotranspiration (ET_0). The neural network consisted of three hidden layers with 128, 64, and 32 nodes, respectively. The activation function used for the hidden layers was the rectified linear unit (ReLU), while a linear activation function was used for the output layer. (2) Learning algorithm: We utilized the backpropagation algorithm to train the neural network. The training process involved minimizing the mean squared error (MSE) between the predicted ET_0 values and the actual ET_0 values in the training dataset. (3) Variables used: We considered a range of meteorological variables as input features for the neural network model. These variables included temperature, relative humidity, wind speed, solar radiation, and atmospheric pressure. A total of six meteorological variables were used as inputs to the neural network model. (4) Data division: The available dataset was divided into three subsets: the learning, validation, and test sets. The learning set comprised 60% of the data, the validation set contained 20%, and the remaining 20% constituted the test set. (5) Model training and validation: The learning set was used to train the neural network model. The model weights were adjusted iteratively during the training process to minimize the MSE on the learning set. The validation set was used to monitor the model's performance and prevent overfitting. We employed early stopping criteria based on the validation set's performance to determine the optimal number of training epochs. (6) Model evaluation: After training, the neural network model was evaluated using the test set, which was held separate from the training and validation sets. The test set allowed us to assess the model's generalization performance on unseen data.

3. Results

3.1. Climatic Variables in the Study Area

Table 1 presents descriptive statistics of the climatic variables used in the study, including minimum temperature, maximum temperature, average temperature, humidity, wind speed, sunshine hours, radiation, reference evapotranspiration (ET_0), rain, and effective rain. The results indicate the minimum temperature ranged from 2.5 to 26.3 °C, with an average of 14.88 ± 5.6 °C. The maximum temperature ranged from 16.6 to 41 °C, averaging 28.09 ± 6.17 °C. Humidity ranged from 7 to 98%, averaging 57.85 ± 17.22%. Wind speed ranged from 17 to 596 km/day, averaging 266.21 ± 122 km/day. Sunshine hours ranged from 4.5 to 11.9 h, with an average of 8.92 ± 1.57 h. Radiation ranged from 9.7 to 27.9 MJ/m^2/day, averaging 19.75 ± 5.23 MJ/m^2/day. In addition, ET_0 ranged from 1.37 to 12.46 mm/day with an average of 5.03 ± 2.27 mm/day. The rain and effective rain variables ranged from 0 to 61 mm and from 0 to 55 mm, respectively. The mean values for rain and effective rain were 4.36 ± 9.25 mm and 4.20 ± 8.66 mm, respectively.

Table 1. Descriptive statistics of climatic variables, ET_0, and precipitations.

Climatic Variables	Obs	Min	Max	Mean	SD	SE	CV	Sk	Kurt
Min temp, °C	384.00	2.50	26.30	14.88	5.60	0.29	37.64	−0.11	−1.10
Max temp, °C	384.00	16.60	41.00	28.09	6.17	0.31	21.97	0.11	−1.04
Humidity, %	384.00	7.00	98.00	57.85	17.22	0.88	29.76	−0.35	−0.24
Wind speed, km/day	384.00	17.00	596.00	266.21	122.04	6.23	45.84	0.42	−0.49
Sunshine, h	384.00	4.50	11.90	8.92	1.57	0.08	17.63	−0.38	−0.62
Radiation, MJ/m^2/day	384.00	9.70	27.90	19.75	5.23	0.27	26.46	−0.26	−1.26
ET_0, mm/day	384.00	1.37	12.46	5.03	2.27	0.12	45.03	0.71	0.16
Rain, mm	384.00	0.00	61.00	4.36	9.25	0.47	211.95	3.39	13.14
Eff rain, mm	384.00	0.00	55.00	4.20	8.66	0.44	206.30	3.22	11.73

3.2. Pearson Correlation Matrix

Figure 2 shows the Pearson correlation matrix between the climatic variables and the ET_0. In the correlation matrix, the reference evapotranspiration (ET_0) is positively correlated with minimum temperature (0.74), maximum temperature (0.84), sunshine (0.81), and radiation (0.82), indicating that as these climatic variables increase, ET_0 also increases. This is because ET_0 is influenced by temperature, solar radiation, and other climatic factors that affect evapotranspiration. There is also a moderately negative correlation between ET_0 and effective rainfall (−0.46), indicating that as effective rain increases, ET_0 decreases. As a result, effective rainfall reduces plants' water demand and reduces the evapotranspiration rate.

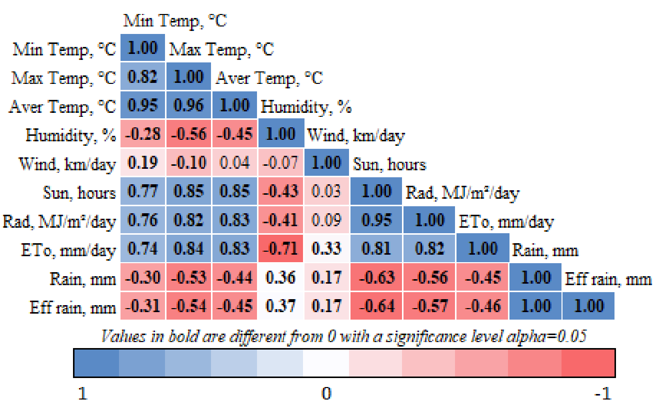

Figure 2. Pearson correlation matrix.

3.3. Stepwise Regression Model

The stepwise regression analysis was conducted to select the most relevant variables for predicting reference evapotranspiration (ET_0) in this study. Table S1 indicate that the maximum temperature (Max Temp, °C) was the most important variable in predicting ET_0, with an R-squared (R^2) value of 0.70. The second variable selected was wind speed (Wind, km/day), which improved the model's performance, resulting in an R^2 value of 0.87. The third variable added to the model was humidity (%), which further improved the model's performance, resulting in an R^2 value of 0.93. Finally, the fourth variable added was radiation (Rad, MJ/m^2/day), resulting in the best-performing model with an R^2 value of 0.95. The adjusted R-squared (Adj. R^2) values for each model also indicate that the model with all four variables fits best, with an Adj. R^2 value of 0.95.

Table 2 provides the goodness of fit statistics for the reference evapotranspiration (ET_0) model used in this study. The statistics are presented for the training and validation sets used to develop and evaluate the model. The R-squared (R^2) and adjusted R-squared (Adj. R^2) values were 0.95 for both the training and validation sets, indicating that the model explains 95% of the variability in the data. MSE and RMSE were 0.24 for the training and validation sets, indicating that the model has a low error and can accurately predict ET_0. The MAPE value was 8.06% for both the training and validation sets, indicating that the model has a low relative error and can accurately predict ET_0. The goodness of fit statistics in Table 2 show that the ET_0 model used in this study has a high level of accuracy and can accurately predict ET_0 based on the climatic variables used in the model.

Table 2. Goodness of fit statistics (ET_0, mm/day).

Statistic	Training Set	Validation Set
Observations	268.00	268.00
Sum of weights	268.00	268.00
DF	263.00	263.00
R^2	0.95	0.95
Adjusted R^2	0.95	0.95
MSE	0.24	0.24
RMSE	0.49	0.49
MAPE	8.06	8.06

Table S2 presents the results of the analysis of variance (ANOVA) for the reference evapotranspiration (ET_0) model used in this study. The ANOVA is a statistical technique used to determine whether the dependent variable (ET_0) variation is significantly explained by the independent variables (climatic variables used in the model). In Table S2, the model is highly significant, with an F-value of 1294.906 and a *p*-value of <0.0001. This indicates that the variation in ET_0 is significantly explained by the climatic variables used in the model. The error term also has a low mean square value, indicating that the model has a low residual error. The ANOVA results in Table S2 provide strong evidence that the ET_0 model used in this study is highly significant and can accurately predict ET_0 based on the climatic variables used in the model.

Table 3 presents the model parameters for the reference evapotranspiration (ET_0) model used in this study. The model parameters represent the estimated coefficients for each of the climatic variables included in the model and the intercept term. In Table 3, all of the variables are highly significant ($p < 0.0001$), indicating that they have a strong relationship with ET_0. The intercept term is negative, indicating that ET_0 decreases when all other variables are constant. The coefficients for maximum temperature, wind speed, and radiation are positive, indicating that ET_0 increases with an increase in these variables. In contrast, the coefficient for humidity is negative, indicating that ET_0 decreases with an increase in humidity. Finally, the equation of stepwise regression of the reference evapotranspiration (ET_0) model used in this study is:

$$ET_0\text{, mm/day} = -1.499 + 0.174 \times \text{TMax} - 0.04 \times \text{Humidity} + 0.006 \times \text{Wind} + 0.118 \times \text{Rad}$$

where:

- TMax is the maximum temperature in °C.
- Humidity is the relative humidity in %.
- Wind is the wind speed in km/day.
- Rad is the radiation in $MJ/m^2/day$.

Table 3. Model parameters (ET_0, mm/day).

Source	Value	SE	T	Pr > |t|	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	−1.499	0.286	−5.236	<0.0001	−2.062	−0.935
Max Temp, °C	0.174	0.010	17.446	<0.0001	0.155	0.194
Humidity, %	−0.040	0.002	−18.821	<0.0001	−0.045	−0.036
Wind, km/day	0.006	0.000	23.086	<0.0001	0.006	0.007
Rad, $MJ/m^2/day$	0.118	0.011	11.132	<0.0001	0.097	0.139

Table S3 presents the standardized coefficients for the reference evapotranspiration (ET_0) model used in this study. The standardized coefficients represent the relative importance of each climatic variable in predicting ET_0, after scaling all variables to have a mean of zero and standard deviation of one. In Table S3, all of the variables are highly significant ($p < 0.0001$), indicating that they have a strong relationship with ET_0. The standardized coefficients indicate that maximum temperature has the highest relative importance in predicting ET_0 (0.462), followed by wind speed (0.332), radiation (0.271), and humidity (−0.310).

Figure 3A,B display scatter plots of the standardized residuals versus the actual ET_0 and predicted ET_0, respectively. These plots can be used to evaluate the performance of the ET_0 model and identify any patterns or trends in the residuals. The absence of any noticeable trend in the scatter plots suggests that the model is effectively correcting for any possible autocorrelation in the residuals. Autocorrelation measures the degree to which the residuals at different time points are correlated, and its presence can indicate that the model does not account for all of the sources of variation in the data. Therefore, the lack of any discernible trend in both Figure 3A,B indicates that the ET_0 model accurately captures the variability in the data without any issues related to autocorrelation. Figure 3C presents a plot showing the predicted ET_0 values plotted against the actual ET_0 values and confidence limits. This plot can compare the model's predictions to the observed values and detect outliers. We can observe values fall within the confidence limits in Figure 3C, which indicates that the model is accurately predicting the ET_0 values for those observations. The confidence limits provide a measure of the uncertainty in the predicted values. If the observed values fall within the limits, it suggests that the predicted values are consistent with the observed values.

3.4. Artificial Neural Network Model

Table S4 provides information on the number and percentage of cases included in the neural network analysis, including the sample size and the division of cases between training and testing samples. In this case, the total number of cases included in the analysis is 384, where 261 cases (or 68.0%) are included in the training sample, and the remaining 123 cases (or 32.0%) are included in the testing sample. The valid percentage is 100%, indicating that all cases were included in the analysis, and none were excluded.

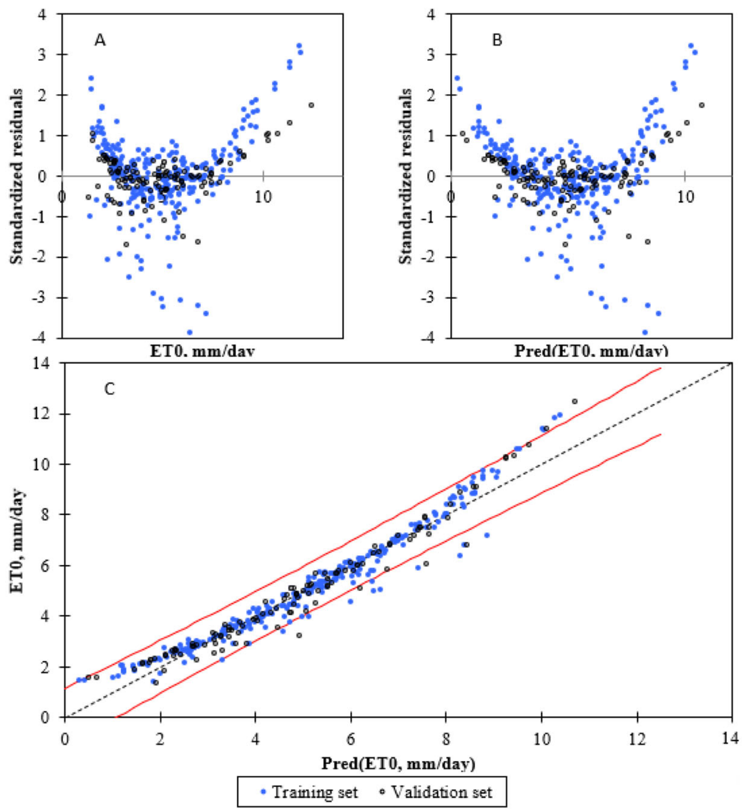

Figure 3. Assessing Model Performance and Detecting Outliers in ET_0 Prediction Using Residual Analysis and Prediction Comparison ((**A**) Standardized Residual versus ET_0, (**B**) Standardized Residual versus Predicted ET_0, and (**C**) Actual values of ET_0 versus Predicted values ET_0).

Table S5 and Figure 4 provide information on the neural network architecture used in the analysis, including details on the input, hidden layer(s), and output layers. The input layer consists of four covariates, including maximum temperature (°C), humidity (%), wind speed (km/day), and radiation (MJ/m^2/day). Each covariate is standardized using the rescaling method for covariates. The number of units in the input layer is four, excluding the bias unit. The bias unit is a constant input added to the model to help adjust the output of each unit. The hidden layer has one layer with three units, excluding the bias unit. The activation function used in the hidden layer is the hyperbolic tangent function, commonly used in artificial neural networks to introduce non-linearity into the model. The output layer consists of one dependent variable, the reference evapotranspiration (ET_0) in units of mm/day (Figure 4). The rescaling method for the scale dependents is standardized, and the activation function used in the output layer is the identity function, which passes the input value through to the output. The error function used to evaluate the performance of the neural network model is the sum of squares, which measures the differences between the predicted and actual values of ET_0 and sums them across all observations.

Table S6 summarizes the neural network model used in the analysis, including information on the training and testing errors, stopping rule used, and training time. The training SSE is 0.623, and the relative error is 0.005, indicating that the model fits the training data reasonably well. The stopping rule used is one consecutive step(s) with no decrease in error, which means that the training process stops if the error does not decrease over a specified number of iterations. The training time for the model is 0:00:00.02, indicating that the model was trained relatively quickly. The testing SSE is 0.238, which is lower than the training SSE, indicating that the model performs well on the testing data, and the relative error for the testing data is 0.004.

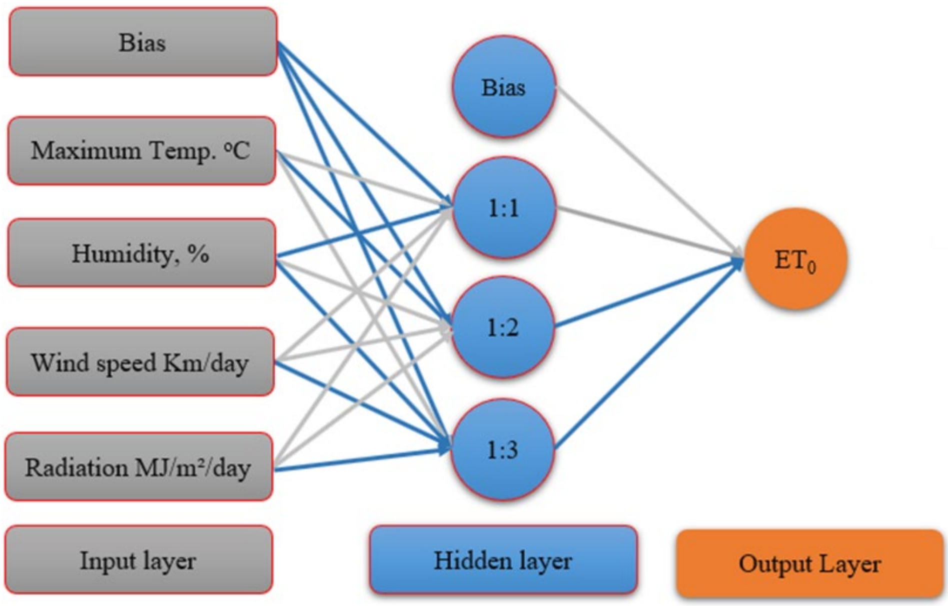

Figure 4. Layers of Artificial Neural Network Model.

Table S7 presents the parameter estimates for the neural network model used in the analysis. The parameter estimates reflect the weights assigned to each predictor variable in the neural network model and provide information on the relative importance of each variable in predicting ET_0. In Table S7, the parameter estimates for the output layer show that the most important unit in predicting the output of the neural network model is the only unit in the output layer, which corresponds to the predicted ET_0. The weight assigned to this unit is the only parameter estimated in the output layer and is given by "ET_0, mm day" with a value of 0.570. For the hidden layer, the parameter estimates indicate the weights assigned to each unit in predicting the output. In this case, the parameter estimate for unit H(1:1) is the largest, with a value of 1.883, indicating that this unit is the most important in predicting ET_0. The estimated parameters for units H(1:2) and H(1:3) are 0.764 and −1.123, respectively, suggesting that these units are likewise significant, but to a smaller degree than unit H(1:1). Therefore, the parameter estimates in Table S7 suggest that unit H(1:1) in the hidden layer is the most important unit in predicting the output of the neural network model, followed by units H(1:2) and H(1:3), with the output unit in the output layer being the most critical in predicting ET_0.

Table 4 presents the independent variable importance for the neural network model used in the analysis. The importance values reflect the contribution of each independent variable to the prediction of ET_0, with higher values indicating greater importance. The importance values in Table 4 show that the most important independent variable in predicting ET_0 is maximum temperature (°C), with an importance value of 0.280, followed by wind (km/day) with an importance value of 0.261, humidity (%) with an importance value of 0.234, and radiation (MJ/m^2/day) with an importance value of 0.225. The normalized importance values in Table 4 show the relative importance of each variable, with the most important variable assigned a value of 100%. Based on the normalized importance values, maximum temperature (°C) is still the most important variable, with a normalized importance value of 100%, followed by wind (km/day) with a normalized importance value of 93.4%, humidity (%) with a normalized importance value of 83.5%, and radiation (MJ/m^2/day) with a normalized importance value of 80.4%.

Table 4. Independent Variable Importance for Neural Network Model.

Variable	Importance	Normalized Importance
Max Temp, °C	0.280	100.0%
Humidity, %	0.234	83.5%
Wind, km/day	0.261	93.4%
Rad, MJ/m^2/day	0.225	80.4%

For the stepwise regression model, the R^2 value is 0.95, which indicates that the independent variables in the model can explain 95% of the variance in the target variable. The adjusted R^2 value is also 0.95, suggesting that the model fits the data well. The MSE value is 0.25, meaning the average squared difference between the actual and predicted values is 0.25. The RMSE value is 0.50, which is the square root of the MSE and represents the typical error in the predictions. For the neural network model, the R^2 value is 0.99, indicating that the model performs strongly and can explain much of the variance in the target variable. The adjusted R^2 value is also 0.99, suggesting that the model fits the data well. The MSE value is much lower than the stepwise regression model at 0.02, indicating that the average squared difference between the actual and predicted values is much lower for the artificial neural network model. The RMSE value is 0.15, which is also much lower than the stepwise regression model, suggesting that the typical error in the artificial neural network model predictions is also much lower.

4. Discussion

The findings indicated that the artificial neural network (ANN) model had the highest level of efficacy in estimating ET_0, as evidenced by a coefficient of determination (R^2) of 0.99 and a mean absolute percentage error (MAPE) of 2.7. Conversely, maximum temperature, humidity, sun radiation, and wind speed were the climatic factors that had the most influence. The prediction of hydro-meteorological parameters, including rainfall, temperature, wind speed, relative humidity, and soil temperature, is commonly performed using data-driven modelling techniques such as multiple linear regression (MLR) and ANNs [12]. The literature includes studies on rainfall, runoff, and sediment modelling [13,14], as well as investigations into groundwater quality parameters and stream discharge [15–17]. Additionally, there are research papers on storm prediction [18], which encompasses the consideration of evapotranspiration [19]. ANNs and MLR techniques have been extensively employed in modelling the potential evapotranspiration (ET) process [20–25].

Kumar et al. [26] constructed an ANN model to forecast reference evapotranspiration (ET_0). Their findings indicated that the ANN model outperformed the conventional approach in accurately predicting ET_0. According to Jain et al. [27], ANNs can effectively estimate ET_0 using the limited temperature and radiation meteorological variables. Additionally, it was discovered that the accuracy of model predictions for estimating ET_0 was lower when utilizing only the minimum parameters in the analysis. However, the accuracy improved as more input parameters were added.

Reddy et al. [28] evaluated weekly reference evapotranspiration using linear regression and ANN Models for several sites in Andhra Pradesh and found that MLR models performed well in weekly ET_0 estimates. Marti et al. [29] outlined different ways to figure out reference evapotranspiration based on multiple linear regression, simple regression, and artificial neural networks (ANNs). They found that the artificial neural network and multiple linear regression algorithms performed similarly, with significantly higher accuracy than simple regression and traditional temperature-based approaches. A study by Patle et al. [18] examined the comparative performance of MLR and ANN techniques in predicting weekly reference evapotranspiration (ET_0) in the semi-humid region of Sikkim, India. The study's findings demonstrated the efficacy of both MLR and ANN models in estimating ET_0 using a limited set of meteorological variables.

Furthermore, the study revealed that increasing the number of parameters enhanced the prediction efficiency of the models. Many researchers have used neural networks to estimate Ep as a function of meteorological data [30]. Several researchers discovered that the neural network methodology produced superior outcomes in Ep estimation than the Priestley–Taylor and Penman methods [31,32]. In line with previous research, the present work successfully illustrated the feasibility of employing the Kohonen Self Organizing Maps (K-SOM) methodology alongside the FAO56-PM and MLR methodologies to model Ep. The comparative analysis revealed that the K-SOM model exhibited more effectiveness overall than the FAO56-PM and MLR techniques. Chang et al. [33] employed various methodologies for estimating pan evaporation, including the K-SOM and the FAO56-PM techniques. Nazari et al. [34] compared 13 reference evapotranspiration equations to the FAO56PM equation in arid regions of Iran. Irmak, Hargreaves-Samani, and Hargreaves equations were the best three approaches. However, these three methods must be calibrated for local conditions and evaluated using multiple weather stations. The findings may aid water management, irrigation, and the efficient use of water resources.

Abdallah et al. [35] compared the performance of the D-vine copula quantile regression model (DVQR) with other models to estimate daily ET_0 during 2000–2015 over two hyper-arid regions in Sudan from 2000 to 2015. The results showed that empirical models performed poorly, while calibration techniques improved the estimation and reduced errors. Wind speed and saturation vapor pressure deficit were the most effective variables for estimating daily ET_0 variation. The three quantile regression models showed the highest accuracy and performed better than empirical and calibrated models. The DVQR model showed performance comparable to that of the extreme gradient boost model in estimating daily ET_0 across all combinations of meteorological variables. However, the study has limitations, such as a short research period and a limited number of weather stations and climate conditions. The DVQR model is strongly suggested as a viable alternative for estimating daily ET_0 in hyper-arid regions. On the other hand, Lee et al. [36] compared 30 empirical models for estimating evapotranspiration (ET_0) over 50 years in South Korea, spanning continental and temperate climate zones. Their results showed that temperature-radiation-based models are consistently effective, while mass-transfer models are poor. This study suggests best practices for estimating ET_0 using empirical models, which can aid water resources management decision-making in regions with limited meteorological data, as long-term data are limited in most developing countries. Sobh et al. [37] investigated the performance of 31 empirical equations and 20 models developed using five artificial intelligence algorithms to estimate reference evapotranspiration (ET_0) in arid regions (Egypt). They proved that empirical equations based on radiation, temperature, and mass transfer perform better in replicating FAO56-PM. The Ritchie equation was the best overall in Egypt, while the random forest (RF) model, which predicts temperatures, wind speed, and relative humidity, outperformed other AI algorithms. The generated ET_0 estimates enabled the detection of a significant increase in the agriculturally dependent Nile Delta.

Research Limitations and Future Directions

Research Limitations: While conducting our research, we encountered several limitations that should be acknowledged. These limitations include the following: (1) Data availability: Acquiring accurate and comprehensive data for reference evapotranspiration (ET_0) calculations in arid regions was challenging. The limited availability of long-term, high-quality meteorological data restricted the scope of our analysis. Future studies should aim to gather more extensive and reliable datasets to enhance the accuracy of ET_0 estimation. (2) Model generalization: Our study focused on a specific arid region, and the findings may not directly apply to other geographical locations. It is important to consider each region's specific climatic and environmental conditions when applying regression models and artificial neural networks for ET_0 calculations. Future research should aim to validate the performance of these models across various arid regions. (3) Model complexity: The artificial neural network models employed in our study can be computationally intensive and require significant computational resources. This complexity may limit their practical application in real-time ET_0 calculations, especially in resource-constrained settings. Future research should explore methods to simplify the models without compromising accuracy, making them more accessible and feasible for widespread implementation.

Future Directions: Building upon our research findings, we propose the following directions for further investigation: (1) Ensemble modeling: Investigate the potential benefits of combining multiple regression models and artificial neural networks to create ensemble models. Ensemble modeling techniques, such as model averaging or stacking, could potentially improve the accuracy and robustness of ET_0 estimations in arid regions. (2) Incorporating remote sensing data: Explore integrating remote sensing data, such as satellite imagery and vegetation indices, into ET_0 calculations. Remote sensing data can provide valuable information on land cover, vegetation dynamics, and soil moisture, which may enhance the accuracy of ET_0 models in arid regions. (3) Climate change impact: Investigate the impact of climate change on reference evapotranspiration in arid regions. Analyzing long-term climate data and projecting future climate scenarios can provide insights into how changing climate patterns may influence ET_0 estimations. This research can help anticipate potential shifts in water demand and inform water resource management strategies.

5. Conclusions

This research endeavor involved developing and evaluating various models to calculate reference evapotranspiration (ET_0) using meteorological variables as predictive factors. The models underwent evaluation using stepwise regression and artificial neural network methodologies. The study's findings suggest that using the artificial neural network model was the most efficient method for calculating ET_0. This model demonstrated high dependability and precision, making it a valuable tool for ET_0 estimation. The climatic factors influencing the model the most were maximum temperature, humidity, sun radiation, and wind speed. The implications of the study's findings extend to other domains, including agriculture, irrigation, and water management, where they can enhance crop productivity and promote efficient water resource utilization.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su152115494/s1, Table S1: Summary of the variable's selection ET0, mm/day, Table S2: Analysis of variance (ET0, mm/day), Table S3: Table S3. Standardized coefficients (ET0, mm/day), Table S4: Case Processing Summary, Table S5: Results of Neural Network Analysis for Predicting ET0, Table S6: Model Summary for Neural Network Model, Table S7: Parameter Estimates for Neural Network Model.

Author Contributions: Conceptualization, M.K.A.-F. and S.K.A.-E.; methodology, M.K.A.-F. and S.K.A.-E.; software, M.K.A.-F. and S.K.A.-E.; validation, M.K.A.-F., S.K.A.-E., A.-R.M.A.M. and Z.Z.; formal analysis, M.K.A.-F.; investigation, M.K.A.-F., S.K.A.-E. and Z.Z.; resources, M.K.A.-F., S.K.A.-E. and Z.Z.; data curation, M.K.A.-F., S.K.A.-E. and Z.Z.; writing—original draft preparation, M.K.A.-F.; writing—review and editing, S.K.A.-E. and Z.Z.; visualization, M.K.A.-F., S.K.A.-E., A.-R.M.A.M. and Z.Z.; supervision, M.K.A.-F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration—Guidelines for Computing Crop Water Requirements—FAO Irrigation and Drainage Paper 56*; Food and Agriculture Organization of the United Nations: Rome, Italy, 1998.
2. Singh, B.K.; Ghosh, S. Predicting reference evapotranspiration using multiple linear regression and artificial neural network models. *J. Water Land Dev.* **2018**, *38*, 89–99. [CrossRef]

3. Zhang, N.; Ma, X.; Zhang, J.; Zhang, S.; Wang, H. Prediction of reference evapotranspiration using artificial neural network in arid region of northwest China. *J. Irrig. Drain. Eng.* **2018**, *144*, 04018022. [CrossRef]
4. Alizadeh, A.; Sepaskhah, A.R.; Borji, M. Prediction of water requirement of grapevine by artificial neural network. *Agric. Water Manag.* **2018**, *208*, 204–213.
5. Ismail, M.H.; Othman, F.B.; Hussin, W.M.A.W. Estimation of reference evapotranspiration using multiple linear regression in selected locations of Malaysia. *J. Water Land Dev.* **2021**, *49*, 19–28.
6. Jia, X.; Zhang, H.; Yang, Q.; Zhang, Y. Estimation of irrigation water requirement for maize in northern China using the FAO Penman-Monteith method. *Agric. Water Manag.* **2018**, *204*, 28–40. [CrossRef]
7. Liu, Y.; Oldekop, M.; Xu, J.; He, J. Evaluation of reference evapotranspiration models under different land-use types in the middle reaches of the Heihe River Basin, China. *J. Hydrol.* **2018**, *559*, 395–406. [CrossRef]
8. Pereira, L.S.; Oweis, T.; Zairi, A. Irrigation management under water scarcity. *Agric. Water Manag.* **2002**, *57*, 175–206. [CrossRef]
9. Dimitriadou, S.; Nikolakopoulos, K.G. Multiple Linear Regression Models with Limited Data for the Prediction of Reference Evapotranspiration of the Peloponnese, Greece. *Water* **2022**, *14*, 124. [CrossRef]
10. Traore, S.; Wang, Y.M.; Kerh, T. Modeling reference evapotranspiration by generalized regression neural network in semi-arid zone of Africa. *WSEAS Trans. Inf. Sci. Appl.* **2008**, *6*, 991–1000–1000.
11. Abdullahi, J.; Elkiran, G.; Nourani, V. Application of artificial neural network to predict reference evapotranspiration in Famagusta, North Cyprus. In Proceedings of the 11th International Scientific Conference on Production Engineering Development and Modernization of Production, Bihać, Bosnia and Herzegovina, 4–7 October 2017; pp. 549–554.
12. Rajendra, P.; Murthy, K.V.N.; Subbarao, A.; Boadh, R. Use of ANN models in the prediction of meteorological data. *Model. Earth Syst. Environ.* **2019**, *5*, 1051–1058. [CrossRef]
13. Chakravarti, A.; Joshi, N.; Panjiar, H. Rainfall runoff analysis using artificial neural network. *Indian J. Sci. Technol.* **2015**, *8*, 14. [CrossRef]
14. Sharma, N.; Zakaullah, M.; Tiwari, H. Runoff and sediment yield modeling using ANN and support vector machines: A case study from Nepal watershed. *Model. Earth Syst. Environ.* **2015**, *1*, 23. [CrossRef]
15. Nathan, S.; Saravanane, N.R.; Sundararajan, T. Application of ANN and MLR Models on Groundwater Quality Using CWQI at Lawspet, Puducherry in India. *J. Geosci. Environ. Prot.* **2017**, *5*, 99–124. [CrossRef]
16. Khan, M.Y.A.; Hasan, F.; Panwar, S.; Chakrapani, G.J. Neural network model for discharge and water-level prediction for Ramganga River catchment of Ganga Basin, India. *Hydrol. Sci. J.* **2016**, *61*, 2084–2095. [CrossRef]
17. Tiwari, A.K.; De Maio, M.; Singh, P.K.; Singh, A.K. Hydrogeochemical Characterization and Groundwater Quality Assessment in a Coal Mining Area, India. *Arab. J. Geosci.* **2016**, *9*, 177. [CrossRef]
18. Patle, I.T.; Mandal, B.P.; Kumar, M.; Jhajharia, D. Evapotranspiration Modelling using Artificial Neural Network and Multiple Linear Regression Approach in Semi-Humid Region of Sikkim. *Res. Sq.* **2022**. [CrossRef]
19. Sriram, A.V.; Rashmi, C.N. Estimation of potential evapotranspiration by multiple linear regression method. *IOSR J Mech. Civ. Eng.* **2014**, *11*, 65–70.
20. Kumar, M.; Bandyopadhyay, A.; Rahguwanshi, N.S.; Singh, R. Comparative study of conventional and artificial neural network based ET0 estimation models. *Irrig. Sci.* **2008**, *26*, 531–545. [CrossRef]
21. Zanetti, S.S.; Sousa, E.F.; Oliveira, V.P.S.; Almeida, F.T.; Bernardo, S. Estimating evapotranspiration using artificial neural network and minimum climatological data. *J. Irrig. Drain. Eng.* **2007**, *133*, 83–89. [CrossRef]
22. Chauhan, S.; Shrivastava, R.K. Performance evaluation of reference evapotranspiration estimation using climate based methods and artificial neural networks. *Water Resour. Manag.* **2009**, *23*, 825–837. [CrossRef]
23. Chattopadhyay, S.; Jain, R.; Chattopadhyay, G. Estimating potential evapotranspiration from limited weather data over Gangetic West Bengal, India: A neuro computing approach. *Meteorol. Appl.* **2009**, *16*, 403–411. [CrossRef]
24. Khedkar, D.D.; Singh, P.K.; Bhakar, S.R. Estimation of evapotranspiration using neural network approach. *J. Agrometeorol.* **2019**, *21*, 233–235. [CrossRef]
25. Mallikarjuna, P.; Kamasani, C.R. Daily Reference Evapotranspiration Estimation using Artificial Neural Networks. *Hydrol. J.* **2012**, *35*, 19. [CrossRef]
26. Kumar, M.; Raghuwanshi, N.S.; Singh, R.; Wallender, W.W.; Pruitt, W.O. Estimating evapotranspiration using artificial neural network. *J. Irrig. Drain. Eng.* **2002**, *128*, 224–233. [CrossRef]
27. Jain, S.K.; Nayak, P.C.; Sudheer, K.P. Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation. *Hydrol. Process.* **2008**, *22*, 2225–2234. [CrossRef]
28. Reddy, K.C.S.; Jyothy, S.A.; Mallikarjuna, P. Estimation of weekly reference evapotranspiration using linear regression and ANN models. *Proc. Int. Conf. Adv. Civ. Eng.* **2010**, *1*, 83–88.
29. Marti, P.; Royuela, A.; Manzano, J.; Palau-Salvador, G. Generalization of ET0 ANN models through data supplanting. *J. Irrig. Drain. Eng.* **2010**, *136*, 161–174. [CrossRef]
30. Keskin, M.E.; Terzi, O. Artificial neural network models of daily pan evaporation. *J. Irrig. Drain. Eng.* **2006**, *11*, 65–70. [CrossRef]
31. Rahimikhoob, A. Estimating daily pan evaporation using artificial neural network in a semi-arid environment. *Theor. Appl. Climatol.* **2009**, *98*, 101–105. [CrossRef]

32. Malik, A.; Kumar, A.; Kim, S.; Kashani, M.H.; Karimi, V.; Sharafati, A.; Ghorbani, M.A.; Al-Ansari, N.; Salih, S.Q.; Yaseen, Z.M.; et al. Modeling monthly pan evaporation process over the Indian central Himalayas: Application of multiple learning artificial intelligence model. *Eng. Appl. Comp. Fluid.* **2020**, *14*, 323–338. [CrossRef]
33. Chang, F.J.; Chang, L.C.; Kao, H.S.; Wu, G.R. Assessing the effort of meteorological variables for evaporation estimation by self-organizing map neural network. *J. Hydrol.* **2010**, *384*, 118–129. [CrossRef]
34. Nazari, M.; Chaichi, M.R.; Kamel, H.; Grismer, M.; Sadeghi, S.M. Evaluation of estimation methods for monthly reference evapotranspiration in arid climates. *Arid Ecosyst.* **2020**, *10*, 329–336. [CrossRef]
35. Abdallah, M.; Mohammadi, B.; Zaroug, M.A.; Omer, A.; Cheraghalizadeh, M.; Eldow, M.E.; Duan, Z. Reference evapotranspiration estimation in hyper-arid regions via D-vine copula based-quantile regression and comparison with empirical approaches and machine learning models. *J. Hydrol. Reg. Stud.* **2022**, *44*, 101259. [CrossRef]
36. Lee, E.; Sadeghi, S.M.M.; Deljouei, A.; Cohen, M.J. A multi-decadal national scale assessment of reference evapotranspiration methods in continental and temperate climate zones of South Korea. *J. Hydrol.* **2023**, *625*, 130021. [CrossRef]
37. Sobh, M.T.; Nashwan, M.S.; Amer, N. High-resolution reference evapotranspiration for arid Egypt: Comparative analysis and evaluation of empirical and artificial intelligence models. *Int. J. Climatol.* **2022**, *42*, 10217–10237. [CrossRef]