
*Journal of Agrometeorology*

*ISSN : 0972-1665 (print), 2583-2980 (online)*
*Vol. No. 27 (2) :210 - 215 (June - 2025)*
*https://doi.org/10.54386/jam.v27i2.2877*
*https://journal.agrimetassociation.org/index.php/jam*

***Research Paper***

# Estimation of climatological parameters using ANN and WEKA models in Diyala Governorate, Iraq

**DHER I. BAKR[1], JASIM Al-KHALIDI[2] and HAZIM NOMAN ABED[3]**

*[1]Department of Petroleum Geology and Mineral, College of Sciences, University of Diyala, Diyala, Iraq.*
*[2]Department of Physics, College of Sciences, University of Diyala, Diyala, Iraq.*
*[3]Department of Computer, College of Sciences, University of Diyala, Diyala, Iraq*
**Corresponding Author: dher@uodiyala.edu.iq*

**ABSTRACT**

Artificial Neural Networks (ANN) and Waikato Environment for Knowledge Analysis (WEKA) model were used to estimate the climatic parameters viz. minimum temperature ($T_{min}$), maximum temperature ($T_{max}$), relative humidity (RH), wind velocity (WV) using the time series of monthly data for the period of 1980 to 2022. It was found that the estimation of the climate parameters using the two methods (WEKA and ANN) obtained acceptable values of correlation ($R^2$) and error standards (RMSE and MAE) between the observed and estimated values, but they differed in accuracy. The WEKA method obtained better values in the estimation of the $T_{min}$ component than ANN while the estimation of the .$T_{max}$, RH, WV, the ANN method was better than the WEAK model in the estimation of the parameters

***Keyword**:* WEKA, Artificial neural networks, Parameters estimation, Iraq

Weather predicting is one of the extreme problems faced by atmospheric scientists. Weather forecasts are made by accumulating quantifiable information about the current state of the atmosphere and use of scientific knowledge about atmospheric processes to predict their future development (Al-Khalidi *et al.*, 2024). As the time difference between current time and the time for which the forecast is being made increases, forecast accuracy decreases due to the chaotic nature of the atmosphere, the enormous computational power needed to solve the measurements and equations that describe the atmosphere. Further the spatial resolutions also affect the accuracy of forecast. Weather predictions are used for several purposes, because they are intended to safeguard people and property, including agricultural production (Kumar *et al.,* 2024).

The time series data are often used to develop model to predict the climatic parameters in future using statistical techniques. However, the observational datasets often suffer from data gaps within their time series, necessitating imputation to ensure dataset integrity for further analysis (Qaraghuli *et al*., 2024). Kaur *et al.,* (2022) and Kothiyal *et al.,* (2025) applied seasonal ARIMA model to predict the rainfall and temperature at different locations of India. Singh *et al.,* (2024) used regression and ANN to predict the reference evapotranspiration ($ET_0$). The present study is aimed to test the performance of Artificial Neural Networks (ANN) and Waikato Environment for Knowledge Analysis (WEKA) models for the estimation of climate parameters of Diyala governorate of Iraq and to know how sensitive each model is to the amount of input data and its impact on the accuracy of the estimate.

## MATERIAL AND METHODS

### *Study area and data source*

Diyala is a governorate in Iraq, located in the middle-eastern region of Iraq, which is located between latitudes 33.34 to 35.26°N and longitude 44.22 to 45.56°E. This governorate is characterized by having a terrain with diverse characteristics, the northeastern part is mountainous, the land is flat in the southern and western parts, and it is an agricultural governorate that possesses wide agricultural areas. The data for the Diyala region on the monthly averages of climatic elements i.e. minimum temperature ($T_{min}$), maximum temperature ($T_{max}$), relative humidity (RH), wind velocity (WV) for the time period of 1981 to 2022 were extracted

***Article info - DOI**: https://doi.org/10.54386/jam.v27i2.2877*
*Received:2 January 2025; Accepted:22 March 2025; Published online : 1 June, 2025*

from the NASA website (NASA/POWER CERES/MERRA2). The data were analyzed using the artificial neural network (ANN) and Waikato Environment for Knowledge Analysis (WEKA) techniques.

The artificial neural network (ANN) technique has unique features-such as its strong capacity for self-learning, flexibility, actual fault tolerance and operation, ease of implementation, and cost effectiveness-make it popular for many different purposes. (Madhiarasan, 2020). An artificial neuron consists of five basic components: input signals, synaptic weights, summing function, activation function and output. The input and weight vectors are multiplied to obtain the activation function. Waikato Environment for Knowledge Analysis (WEKA) is a collection of free software for data analysis and machine learning, approved by the GNU General Public License. WEKA contains an assembly of imagining algorithms and tools for predictive analytics and data analysis forming (Bakr *et al*., 2024; Witten *et al.*, 2017). The WEKA experience is altering stability via the use of criteria artificial intelligence systems to a variety of plant and growing problems (Kohavi *et al.,* 1994).

The climate parameters of $T_{min}$ , $T_{max}$, RH and WV of Diyala city were calculated for 12 months in a year by two different methods viz. WEKA and ANN. The 40 years data were entered into the time series that comes before the estimation. Between the estimated and real values, statistical standards were also computed. While using artificial neural networks, we have designed the best artificial neural network structure that can be used to estimate climate parameters. The monthly data for weather variables used in this study consisted of three parts the training, the testing set, and the estimation set. The parameters included in the network were $T_{min}$, $T_{max,}$ RH, WV and month sequence in year. In training the ANN, a data set of 468 months were used. The testing phase used 24 months dataset and the 12-month data set were used for the estimation stage. This data was used to compare it with the estimation results of the parameters.

The statistical measures viz. root mean square error (RMSE), mean absolute error (MAE) and co-efficient of determination ($R^2$) were used to compare the estimated and observed values for each climate parameter.

## RESULTS AND DISCUSSION

### *Estimating climate parameters using ANN*

Using artificial neural networks and relying on monthly values of climate parameters directly or indirectly related to the parameter to be guessed, five parameters ($T_{min}$, $T_{max,}$ RH, WV and month sequence in year) were entered into the input layer. While changing the number of layers that are hidden and the number of iterations. The best artificial neural network that could be utilized in estimating the monthly minimum temperature values was that it was structured with five inputs, with two hidden layers, and achieved the best values for the error criteria between estimated and actual values (Table 1). The best artificial neural network used to estimate monthly maximum temperature values was in the input layer with two hidden layers, each containing five neurons, and element in the output layer. Each of the two contains hidden levels. Five neurons, with a parameter in the output layer. The best values of error criteria between the real values and the estimated values were achieved.

The appropriate architecture for an artificial neural network in estimating monthly values of relative humidity is five parameters in the input layer with three hidden layers and an output layer containing one neuron that satisfies the error criteria between the observation and the estimated values.

The appropriate neural network structure in estimating monthly values of wind velocity was five parameters in the input layer, three hidden layers in each layer, and five neurons with one neuron in the output layer, and achieved good values of the error parameters between observation and estimation values (Table 1). The convergence between the detected and predictable values (Fig. 1).

### *Estimate the weather parameters using WEKA*

By entering data representing the monthly values of atmospheric parameters into the time series for a period of 480 months for the purpose of estimating the values of atmospheric parameters for a period of 12 months, using the WEKA model, it achieved good values in the error standards and relationship between estimated and actual values. for all climate parameters at a velocity Wind, and the reason is that the wind velocity is highly fluctuating in addition to the fluctuation in wind velocity being subject to many climatic factors (Table 1), and the convergence was important between the observed values and the estimated values (Fig. 1).

### *Comparison ANN and WEKA estimates*

After the atmospheric parameters were estimated using the ANN and WEKA methods, estimation values were achieved that were close to the real values measured in (Table 1), and both methods achieved good values in terms of error standards and the relationship between the estimated and the actual values. It is necessary to compare the two methods for indicating the best and most appropriate method. For estimating weather parameters and the possibility of adopting it as an estimation method, it can be added to the known and widely used estimation methods.

It was found that estimating the monthly values of the minimum temperature using the WEKA program achieved better values of correlation and error standards between the observation

**Table 1:** Correlation and error criteria between observed and estimated values using ANN and WEKA

| Parameter | ANN | | | WEKA | | |
|---|---|---|---|---|---|---|
| | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ |
| Minimum temperature | 2.6288 | 0.6152 | 0.966 | 1.676 | 0.2952 | 0.987 |
| Maximum temperature | 2.5028 | 0.0538 | 0.976 | 2.65 | 0.0509 | 0.98 |
| Relative humidity | 6.8838 | 0.1493 | 0.958 | 8.13 | 0.218 | 0.93 |
| Wind velocity | 0.3121 | 0.0747 | 0.92 | 0.016 | 0.0768 | 0.811 |

**Table 2:** Correlation and error criteria between observed and estimated values using ANN 30-35-40 years input

| ANN Structure | Parameter | years input 30 | | | years input 35 | | | 40 years input | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ |
| 5661 | T min | 0.989 | 1.484 | 0.947 | 0.725 | 0.885 | 0.969 | 0.759 | 0.53 | 0.966 |
| 5551 | T max | 1.743 | 0.144 | 0.855 | 1.188 | 0.108 | 0.94 | 0.723 | 0.054 | 0.976 |
| 54561 | R H | 2.122 | 0.124 | 0.957 | 2.1718 | 0.161 | 0.937 | 1.987 | 0.149 | 0.958 |
| 55551 | W V | 0.098 | 0.073 | 0.800 | 0.101 | 0.074 | 0.78 | 0.09 | 0.075 | 0.92 |

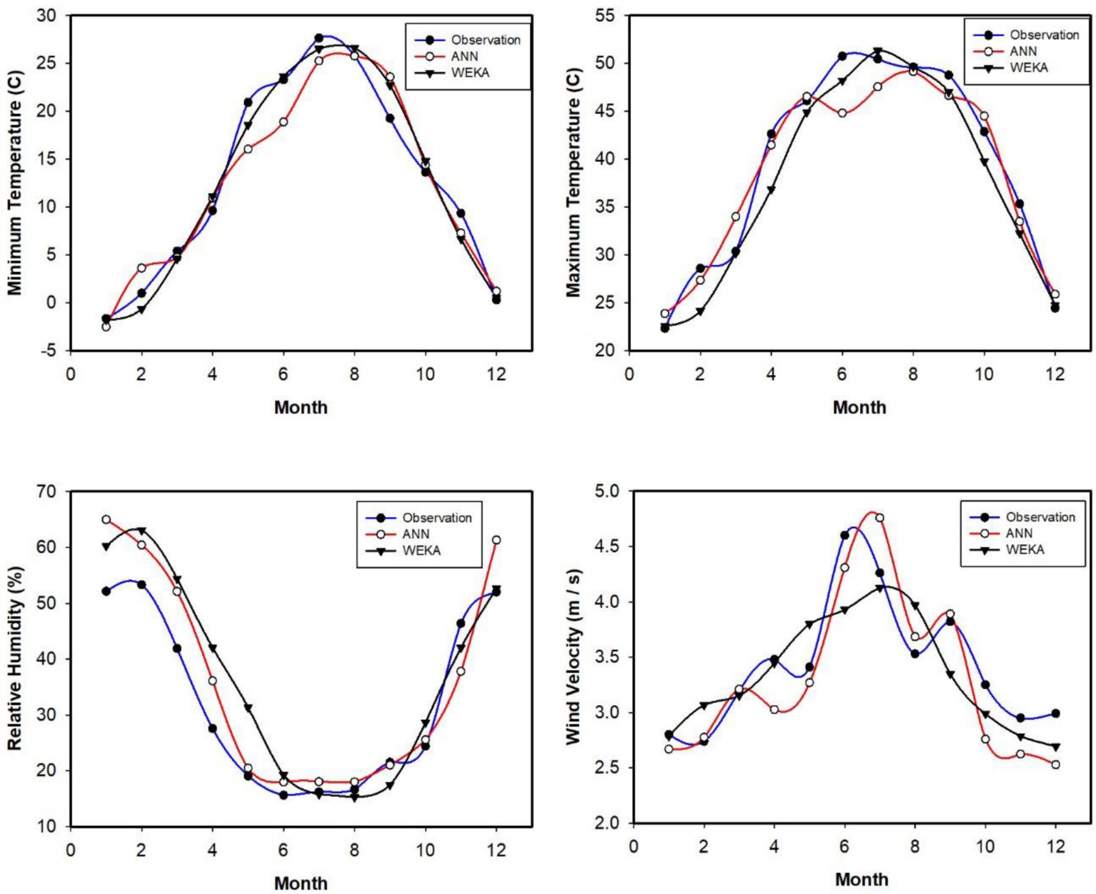


**Fig. 1:** Observed and estimated values using ANN and WEKA

values and estimated values. The correlation value in the WEKA method was greater than that of ANN method, while the values of MAE and RMSE in the WEKA method were lower than in ANN. Thus, estimation using the WEKA program is the best in estimating the monthly minimum temperature values (Table 1).

As for estimating the monthly maximum temperature values, the two methods achieved close and good values in terms of error standards and the correlation between the Forecasted and the observation values, as the correlation in estimating the two methods is very close. As for RMSE, its value was greater in the WEKA method than in the artificial neural networks. Thus, estimating the monthly maximum temperature values using the ANN method is better than the WEKA method because the error criterion is given priority over the correlation in the case of comparison.

Estimating the monthly average relative humidity using the ANN method achieved better values in terms of error criteria and correlation between the estimated values and observation values. The correlation value was greater than in the WEKA method, in addition to the values of the error criteria ANN being lower than in the WEKA method. Therefore, the ANN method is the best in estimating this parameter.

As for estimating the monthly values of wind velocity, it was found that using the two methods achieved different values of error standards. The value of (RMSE) in (ANN) was greater than its value in WEKA, while the value of MAE, $R^2$ was the opposite. Therefore, estimating the monthly values of wind velocity using the ANN method is better than WEKA (Fig. 1).

***The effect of time series length on estimation accuracy***

This part of the research is concerned with studying the

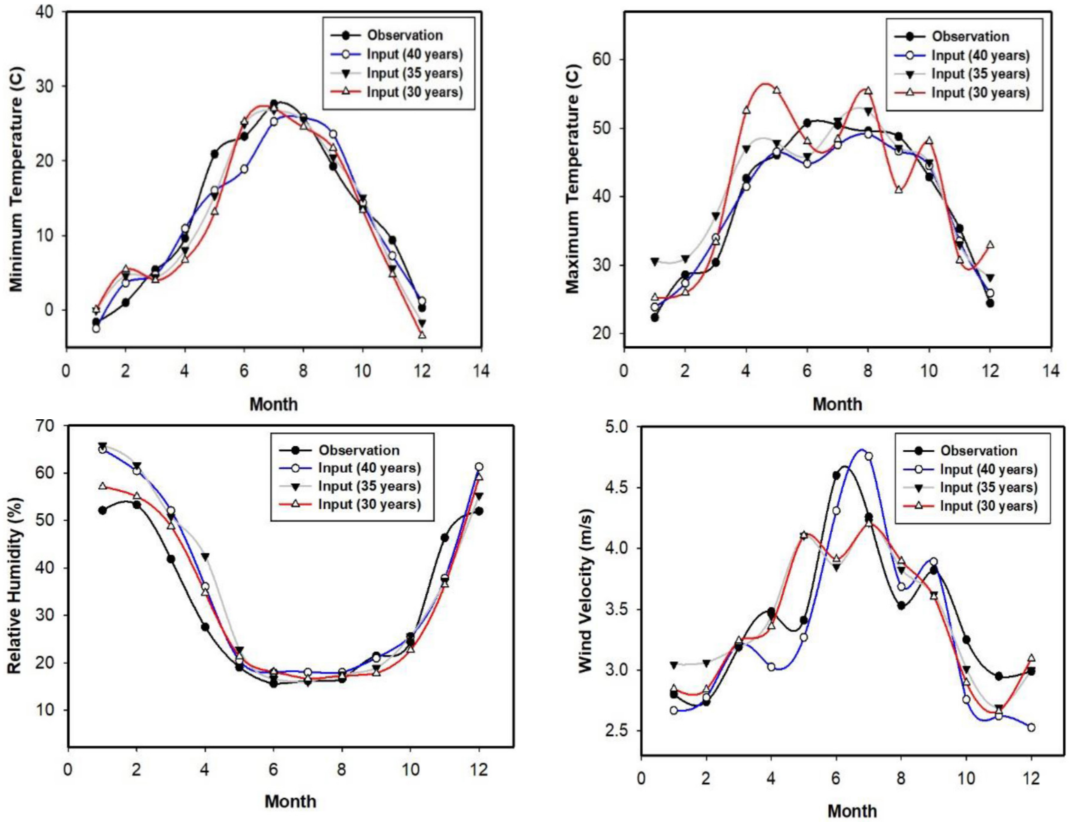


**Fig. 2:** Observed and estimated values using ANN with 30-35-40 years input

**Table 3:** Correlation and error criteria between observed and estimated values using WEKA (30-35-40) years input

| Parameter | 30 years input | | | 35 years input | | | 40 years input | | |
|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ |
| T min | 0.414 | 0.47 | 0.991 | 0.486 | 0.644 | 0.9867 | 0.484 | 0.281 | 0.987 |
| T max | 0.488 | 0.043 | 0.988 | 0.558 | 0.042 | 0.991 | 0.765 | 0.051 | 0.98 |
| R H | 1.103 | 0.102 | 0.972 | 1.324 | 0.153 | 0.971 | 2.278 | 0.218 | 0.93 |
| W V | 0.053 | 0.047 | 0.950 | 0.064 | 0.052 | 0.93 | 0.096 | 0.077 | 0.81 |

effect of the size of time series on accuracy of the estimation, by entering data into two models (ANN and WEKA) with time series data of 30, 35 and 40 years. It has been shown that lengthening the time series results in an increase in accuracy of estimation of the climate component in artificial neural networks. This is due to the fact that increasing the series entering the model leads to increased training of the ANNs, and thus an increase in the mechanism for dealing with data in the case of estimation (Table 2). The increase in estimation accuracy in the event of an increase for data entering the model ANN is fluctuating and irregular for different climate elements or even for a single element. The reason is due to the nature of the time series data for the element, in addition to the increase in extreme values entering the model with an increase for data entered, other than the accuracy of wind speed estimation and this. This results from the fact that wind velocity is characterized by high fluctuation and asymmetry in the data (Fig. 2).

As for increasing the length of the time series included in the model WEKA, it has been shown that the greater the amount of data entering this model, this leads to a decrease in the accuracy of the estimation for all the estimated elements (Table 3), this decrease in the accuracy of the estimation varies between the estimated elements and this is a result of the nature of the data in time series for a single element, which differs between elements, and this differs from the ANN model. The reason is that the WEKA model depends only on the time series for the same element only in estimation, while the ANN model depends in estimation on data of elements that can affect, it is affected by the element to be estimated Fig. 3. This indicates that the WEKA model is more sensitive and affected by extreme values than the ANN model.

## CONCLUSIONS

The methods for estimating climate parameters are different in terms of inputs as well as in terms of accuracy between actual and estimated values. The ANN and WEKA method achieved convergence between the predicted and the actual values. It has been shown that the accurateness of the model varies depending on the climatic parameter to be estimated, in addition to the data included

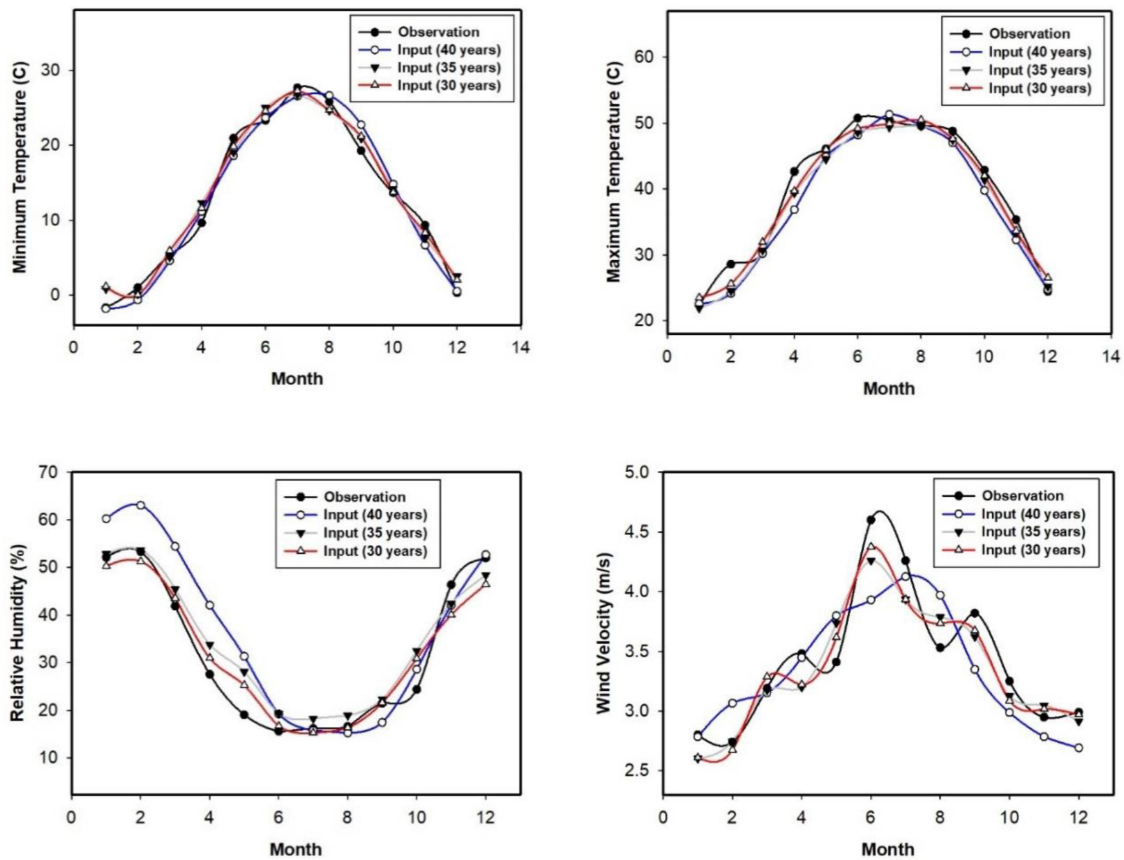


**Fig. 3:** Observed and estimated values using WEKA with 30-35-40 years input

with the nature of this data. It has been shown that the WEKA method is the best in estimating the monthly minimum temperature values, as it achieved better values in the error criteria and correlation value. The ANN method was the best in estimating the monthly values of maximum temperature, relative humidity, and wind velocity, as the correlation values and error standards were the best.

**ACKNOWLEDGEMENT**

We thank the University of Diyala and NASA for supplying the information.

***Funding:*** Funding was not taken.

***Conflict of interest*:** The authors do not disclose any conflicts of interest.

***Data availability*:** The data utilized in study is satellite data, which represents the monthly averages of climatic elements ($T_{min}$, $T_{max}$, RH and WV) for the period (1981 to 2022).

***Author contribution*:** **D. I. Bakr**: collected data and wrote the paper, **J. Al-Khalidi**: Data analysis and drawing diagrams, **H. N. Abed**: manuscript review.



**REFERENCES**

Al-Khalidi, J., Bakr, D., Sedeeq, R., and Faleh, L. (2024). 850 hPa Geopotential Height Link with Iraqi Precipitation and Temperature. *Revista Brasil. Meteorol.*, 39: e39240020.

Bakr, D. I., Al-Khalidi, J., Abas, A. W. M., and ALHak, L. A. (2024). Estimation of some Climatological Parameters by WEKA Software for Selective Regions in Iraq. *Iraqi J. Sci.*, 65(10): 5948-5958. https://doi.org/10.24996/ijs.2024.65.10(SI).5

Kaur, Baljeet, Navneet Kaur, K. K. Gill, Jagjeevan Singh, S. C. Bhan, and Sunayan Saha. (2022). Forecasting mean monthly maximum and minimum air temperature of Jalandhar district of Punjab, India using seasonal ARIMA model. *J. Agrometeorol.,* 24(1): 42-49. https://doi.org/10.54386/jam.v24i1.777

Kothiyal, S., Goel, S., Singh, R. K., Chilwal, A., and Ranjan, R. (2025). SARIMA-based time series analysis of rainfall and temperature for the Tarai region of Uttarakhand. *J. Agrometeorol.*, 27(1): 73-76. https://doi.org/10.54386/jam.v27i1.2838

Kohavi, R., John, G., Long, R., Manley, D., and Pfleger, K. (1994). ALC++: A Machine Learning Library in C++. *Proceedings - International Conference on Tools with*

*Artificial Intelligence, ICTAI*, *December*, 740-743. https://doi.org/10.1109/TAI.1994.346412

Kumar, A., Sarangi, A., Singh, D., Mani, I., Bandhyopadhyay, K. K., Dash, S., and Khanna, M. (2024). Evaluation of soft-computing techniques for pan evaporation estimation. *J. Agrometeorol.*, 26(1):56-62. https://doi.org/10.54386/jam.v26i1.2247

Madhiarasan, M. (2020). Accurate prediction of different forecast horizons wind speed using a recursive radial basis function neural network. *Prot. Control Modern Power Syst.*, *5*(1). https://doi.org/10.1186/s41601-020-00166-8

Singh, Y. P., Singh, P., and A.S. Tomar. (2024). Performance comparison of linear regression and ANN models in estimating monthly reference evapotranspiration (ET0). *J. Agrometeorol.*, 26(3):383-386. https://doi.org/10.54386/jam.v26i3.2641

Qaraghuli, K., Murshed, M. F., M. Said, M. A., Mokhtar, A., and Rousta, I. (2024). Univariate and multivariate imputation methods evaluation for reconstructing climate time series data: A case study of Mosul station-Iraq. *J. Agrometeorol.*, 26(3): 318-323. https://doi.org/10.54386/jam.v26i3.2657

Witten, I. H., Frank, E., A. Hall, M., and J. Pal, C. (2017). Data Mining - Practical machine learning tools and techniques. Elsevier Inc.