Proceedings of the International MultiConference of Engineers and Computer Scientists 2023
IMECS 2023, July 5 - 7 July, 2023, Hong Kong

Utilising Machine Learning for Pan Evaporation Prediction - A Case Study in Western Australia

Mustafa Abed*, Monzur Alam Imteaz, Ali Najah Ahmed

Abstract— Evaporation has a significant impact on the management of water resources, irrigation system designs, and hydrological modelling due to its complex and nonlinear nature. This is because evaporation is a result of the interactions of various climatic factors. In Australia, research suggests that evaporation causes about 40% of the water in open water lakes to be lost each year. Given the potential consequences of climate change, this water loss could become a major issue. This paper presents efficiency of Transformer Neural Network (TNN) approach in predicting monthly pan evaporation (Ep) through a case study in Perth, the capital of Western Australia. Daily meteorological data from a weather station in Perth was deployed for testing and training the model by utilising weather parameters, including maximum temperature, minimum temperature, solar radiation, relative humidity, and wind speed for the period 2009–2022. The Pearson correlation coefficient was used to determine the optimal ML model input parameters. Several models have been developed by combining different input combinations and other model parameters. To evaluate the ML model's performance, it was compared to Stephens and Stewart, a widely used empirical technique. The model's performance was subsequently assessed using standard statistical measures. The results of the performance evaluation criteria suggest that the Transformer model proposed in this study can effectively predict the monthly evaporation rate, benefiting from its self-attention mechanism. The proposed model performed admirably (R^2=0.986, RMSE=0.031, MAE=0.025, and NSE=0.987). Additionally, it was demonstrated that the transformer model was more accurate than the empirical method for the same input sets, leading to a notable improvement in the estimation of monthly evaporation rates.

***Index Terms*—Evaporation, Stephens and Stewart Model, Self-Attention, Transformer Model**

I. Introduction

Evaporation is an essential part of the hydrological cycle, transforming liquid water from the earth's surface into atmospheric water vapour. Higher rates of evaporation are a key marker of global warming [1]. Consequently, keeping track of evaporation patterns is essential for manage and regulate resources of water [2]. Evaporation results in considerable loss of water, which affects lake and reservoir water levels and the water budget. Thus, evaporation losses must be projected prior putting water resource policies into practise and planning watering systems. Evaporation rates are influenced by the vapour pressure differentials and heat availability; these decisive elements are influenced by meteorological elements like humidity, solar radiation, wind speed, air pressure, and air temperature [3]. Other criteria including location, kind of climate, seasonal effect, and time of day are also strongly correlated with, such aspects. As a result, evaporation is a complicated phenomenon with highly nonlinear properties.

Evaporation is projected using direct as well as indirect techniques, such as the evaporation pan, Penman approach, water balance, energy balance and mass transfer [4]. Kisi [5] demonstrated that the most popular technique employed is the evaporation pan method due to its affordability and simplicity. This method also provides a precise assessment of the changes in evaporation [6]. The present work aims to project evaporation pan (Ep) with a precision comparable to real evaporation. Approaches based on weather data linked to the energy budget, water budget, and experimental evaporation equation have been employed for Ep estimation [7]. The linear modelling method does not adequately capture the subtle stochastic aspects of the evaporative process, which could result in very significant inaccuracies [8]. Moreover, because empirical models perform differently under diverse circumstances, the parameters of those models must be adjusted to suit different agroclimatic zones before they can be used.

Recently, using ML techniques and different optimization methods, a variety of studies have been undertaken to predict evaporation pan, taking into account the difficulties of practical and theoretical measurement methodologies stated above [9]. Kişi [10] devised evolutionary neural networks to forecast monthly Ep. The results suggested that the developed approaches offered excellent accuracy compared to empirical methods. Deo et al. [11] assessed monthly evaporation and associated water loss; they employed three machine learning approaches: Multivariate Adaptive Regression Spline (MARS), Extreme Learning Machine (ELM), and Relevance Vector Machine (RVM). RVM was found to be the most successful out of the three approaches when using meteorological variables as predictor variables.

With the aid of data collected from a meteorological station located in Perth, Australia, this research makes an important contribution by assessing the effectiveness of the self-attention transformer scheme in projecting monthly

Manuscript received March 25, 2023; revised April 15, 2023.

This work was supported by the Australian Government Research Training Program Scholarship (RTP).

Mustafa Abed is a PhD candidate in the Department of Civil and Construction Engineering at Swinburne University of Technology, Hawthorn, Melbourne, VIC 3122, Australia (corresponding author, phone: 0415 758 177; e-mail: mustafaabed@swin.edu.au).

Monzur Alam Imteaz is an Associate Professor and Head of the Urban Environmental & Transport Systems within the Department of Civil & Construction Engineering at Swinburne University of Technology, Melbourne. (e-mail: mimteaz@swin.edu.au).

Ali Najah Ahmed is a Senior Lecturer in the Department of Civil Engineering, College of Engineering, National Energy University (UNITEN), (e-mail: mahfoodh@uniten.edu.my).

evaporation (Ep). Under different input combination conditions, the estimation precision of the models is investigated. The suggested TNN model is compared to Stephens & Stewart, a popular empirical approach [12]. Furthermore, to determine the success of the suggested model in the field of evaporation estimation, its performance is examined and evaluated using a variety of standard performance metrics.

II. Study Area and Data

Like most semi-arid regions, Australia depends on stored water reservoirs to meet food production and drinking water needs. Unfortunately, evaporation rates in these countries could be extremely high. Australia loses about 40% of its overall stored water annually owing to high evaporation rates [13]. Therefore, devising a precise prediction approach to estimate the water deficit is strategic for creating hydrological and water resource planning approaches in this drought-affected country. The present study used data from a weather station in Perth, Western Australia (longitude 115.9742173, latitude -31.9410266, elevation 20 m), to develop a transformer based neural network for evaporation prediction.

Data pertaining to the 01 January 2009 to 31 October 2022 timeframe was acquired from the Bureau of Meteorology (BoM). Figure 1 illustrates the location of the case study. Several parameters like relative humidity (RH), solar radiation (R_s), mean, minimum, maximum air temperatures (T_a, T_{min}, T_{max}), wind speed (S_w), and evaporation pan (E_p) were recorded for the selected station. The weather parameters recorded each month concerning quantified weather information gathered from the selected station are specified in Table 1. The table displays Cv, Cx, Xmean, Xmin, Xmax, Sx, and CC-Ep which are the coefficient of variation, skewness, mean, minimum, maximum, standard deviation, and the Pearson correlation coefficient respectively for the studied meteorological parameters.

III. Methodology

A. Input Combination

Choosing the appropriate predictors is critical to devising a robust predictive model [14]; this study assessed different sets of inputs concerning meteorological variables to successfully develop the suggested TNN input–output model and enhance its predictive characteristics. There are several conscious decisions behind selecting these sets. First, to enable comparison, the TNN model's inputs were selected according to the essential weather factors in the suggested conventional framework (Stephens & Stewart). Additionally, predictors were selected after examining the Pearson Correlation Coefficient (PCC) [15]. The PCC technique is a statistical indicator that indicates the mathematical association, or correlation, between two constant variables.

Table 1 illustrates that RH, Sw, Rs, Tmax, and Tmin had all been significantly associated with Ep, implying that they might be essential for estimating evaporation parameter. Specifically, T_{max} and RH were most profoundly associated with Ep. Hence, Tmax and RH will be considered in all datasets to improve the accuracy of Ep forecasts. This study assessed the impact of the parameter Ep on evaporation prediction. Data points were selected based on how they correlated with the predicted output. As shown in Figure 2, the autocorrelation evaluation of the monthly time series of Ep rate demonstrated that correlation decreased significantly after the second lagging. Which indicates that the prior record of the second Ep magnitude had a notable effect on the predicted Ep magnitude for any period. Thus, to improve the accuracy of the TNN model predictions, two previous pan evaporation records were used as model input.

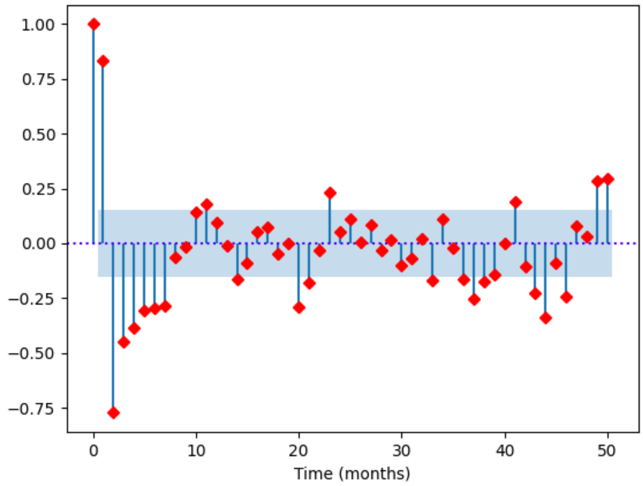

Fig. 2. Perth stations' partial autocorrelation (Monthly)

Correspondingly, the present work considered five distinct input combinations for building the TNN model (Table 2). Climate dataset was partitioned into two parts: 80% was used to train (calibrate) the model, while 20% was used to test (validate). Hence, the dataset was split into two sections based on the initial years, with the first section being used for training and the second section used for validation. This research attempts to conduct a comprehensive assessment to test AI capabilities and employ empirical framework to estimate Ep magnitudes at a monthly timescale at Perth, Western Australia.

TABLE II
The Input Combinations for the TNN Model

Model	Scenario of inputs
T.N.N-1	Ta, Rs
T.N.N-2	RH, T_{max}, T_{min}
T.N.N-3	T_{max}, T_{min}, R_s, RH
T.N.N-4	T_{max}, T_{min}, R_s, S_w, RH
T.N.N-5	T_{max}, T_{min}, R_s, S_w, RH, Ep

ISBN: 978-988-14049-4-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
DATA DESCRIPTIVE ANALYSIS

Station	Dataset	Unit	X_{min}	X_{max}	X_{mean}	S_x	C_v	C_x	*Cc-Ep*
Perth	T_{max}	°C	17.4	35.7	25.5	5.37	21.04	0.17	0.95
	T_{min}	°C	5.8	20.4	12.5	3.85	30.62	0.24	0.9
	RH	%	39.9	73.9	60.7	8.26	13.6	-0.22	-0.93
	S_w	m/s	3.15	6.72	4.8	0.79	16.2	0.04	0.91
	R_s	MJ m^{-2}	8.4	34.81	19.4	7.24	37.2	0.23	0.92
	E_p	mm	1.75	12.98	5.7	3.07	53.4	0.34	1

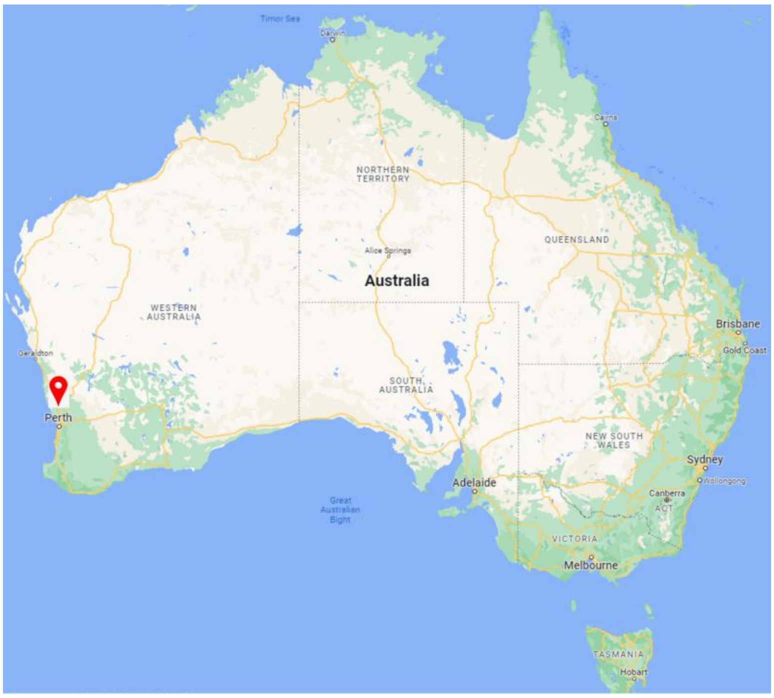

Fig. 1. Location of the case study

TABLE III
EVALUATION PERFORMANCE OF THE T.N.N AND SS MODELS FOR PREDICTING MONTHLY EP

Station	**Model**	**R^2**	**RMSE**	**MAE**	**NSE**
Perth	SS	0.705	0.536	0.430	0.706
	T.N.N-1	0.859	0.100	0.081	0.860
	T.N.N-2	0.949	0.058	0.047	0.953
	T.N.N-3	0.956	0.056	0.048	0.957
	T.N.N-4	0.979	0.039	0.030	0.980
	T.N.N-5	0.986	0.031	0.025	0.987

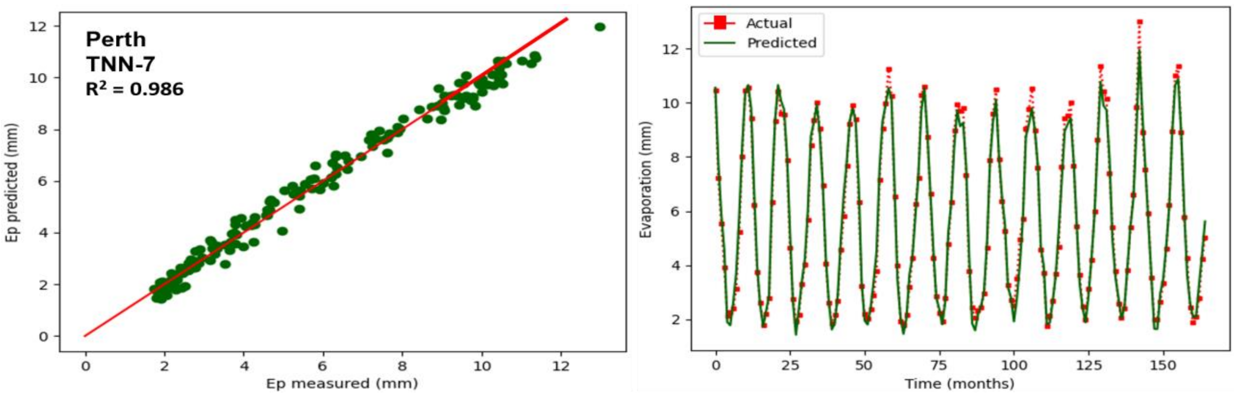

Fig. 3. Scatter plot and time series of the TNN-5 model

ISBN: 978-988-14049-4-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

B.Empirical Model

This study used Stephens & Stewart to contrast the practical method since it is an extensively employed approach [16], considering the count of meteorological inputs and data availability.

Stephens and Stewart (SS)

As specified in equation 1, Stephens and Stewart [17] recommended more precise outcomes were obtained when measured radiation Qs was used, assuming that data were available. Further, it permits temperature correlations too:

$$Ep = (0.0082Ta - 0.19)\left(\frac{Qs}{1500}\right) \times 25.4$$

Where Qs, Ep, and Ta represent solar radiation, (Cal cm^{-2} day^{-1}), evaporation (mm), and mean air temperature (F) respectively.

C.Self-attention Transformer model for Ep prediction

To overcome the limitations of recurrent and convolutional sequential techniques, the framework of Transformer employs a self-attention component [18]. The transformer architecture utilizes self-attention to determine which data is essential to the encoding of the current token by selectively preserving only the most relevant information from the previous token. Otherwise stated, tweaking is conducted for the attention approach for calculating the equivalent of latent space pertaining to the decoder and encoder.

An exhaustive search process is employed in this work with regards to the designing the system as well as hyperparameter training to construct best architectural for the put forward TNN model. Thus, to determine the optimal architecture, several differently configured models were evaluated. The optimal hyperparameters pertaining to the put forward TNN model have been devised by employing four identical transformer encoders. Each of the encoder includes 4 heads of size 64, whose output is processed by employing a 1-D Global Average Pooling layer that includes 16 output filters along with a Kernel size of 1. Global average pooling is advantageous because it works in tandem with the convolutional architecture, connecting feature maps and classifications.

The model was trained in assorted cycles, each with 200 epochs and 16 batches. In order to regulate network weights pertaining to decrease of MSE loss function, the Adam algorithm was employed [19]. To evaluate the effectiveness of the predictive model, several indices of performance were used, including the mean absolute error (MAE), coefficient of determination (R^2), root mean square error (RMSE), and Nash-Sutcliffe efficiency (NSE). For more information and specifics on these performance indices, see [8].

IV.Results and Discussion

In Table 3, a substantial distinction was observed in the accuracy of Ep forecast as determined based on the combination of inputs. In fact, the accuracy of a model's prediction could be improved through the use of the full climatological dataset (Ep, RH, Tmin, Tmax, Rs, and Sw) rather than combining inputs with incomplete data. Current findings show that the accuracy of prediction models increases as the number of input variables increases, which corresponds to previous study of [20]. This demonstrated that employing advanced capabilities like AI may not enhance the ML model's predictive performance, especially when a limited number of meteorological inputs exist. To attain good agreement in monthly Ep prediction, four input combinations were adequate. When five input parameters were employed, adequate results were attained; however, when Ep was used as an input, there was a slight enhancement in the accuracy of the prediction. Figure 3 shows the scatter plot and time series of the TNN-5 model.

Table 3 shows the outcomes of the empirical model used to forecast monthly Ep. In the initial observation for the input combination of R_s and T_a, the lowest prediction accuracy in terms of the R^2 values (which was 0.705) was offered by the radiation-based model (Stewart and Stephens) versus the TNN-1 model. The results of Table 3 show that the TNN model exhibits significantly higher performance than the empirical model. It demonstrated a remarkable capability to accurately predict monthly Ep, even when using the same input parameters, due to its ability to handle non-linear and complex functions. This could be due to the TNN self-attention feature that can detect concealed characteristics, which suggests that the transformer architecture is a more powerful approach for evaporation prediction.

The Ep modelling approach, demonstrated by this research, provides a reliable estimate for water losses caused by evaporation with R^2 value of 0.986, which is essential for managing water resources efficiently. Multiplying Ep values by the land area of watering resources offers an effective scientific-based method to assess the quantity of evaporative water loss, a major factor in determining the existing water asset volume. This calculation simplifies the task of estimating the total amount of existing water available for watering and enables the use of a set of intelligent irrigation schedules. These also help reduce unessential water loss and make the irrigation process more efficient. Thus, this research suggests that the application of the TNN model to predict Ep comes with considerable economic benefits for farmers, especially in areas suffering from drought, water scarcity, or other hydrological imbalances. Additionally, it offers valuable insight for hydrologists about how to incorporate soft computing into their analysis of non-stationary and non-linear hydrological variables.

V.Conclusion

The objective of this study was to assess the effectiveness of a transformer-based neural network (TNN) predictive model for monthly evaporation losses and compare it with the radiation-based model (Stephens and Stewart). The DL model's effectiveness was evaluated by predicting Ep rates utilising data on a monthly scale from Perth station in Western Australia. The monthly Ep from 2009–2022 was used as time series data for training (calibration) and testing (validation) of the designed model. The PCC was used to choose the appropriate input parameters (predictors) for the TNN model in terms of Ep forecasting. Conventional evaluation metrics were used to determine the effectiveness

ISBN: 978-988-14049-4-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

of models.

The investigation led to the following findings:

- The developed TNN model demonstrated a remarkable level of accuracy when used to forecast monthly Ep values at the site selected in this study.
- The developed TNN model proved to be superior to the empirical method. Furthermore, when using the same set of inputs as that method, the accuracy of TNN's monthly Ep projections was significantly improved.
- To establish a reliable and generalisable Ep predictive model, the applicability of the developed technique can be evaluated by applying it to various regions, which can be a part of future study.

REFERENCES

[1] Y. Chen, L. He, J. Li, and S. Zhang, "Multi-criteria design of shale-gas-water supply chains and production systems towards optimal life cycle economics and greenhouse gas emissions under uncertainty," *Comput. Chem. Eng.*, vol. 109, pp. 216–235, 2018.

[2] Q. Wang, W. Wang, Z. Zhong, H. Wang, and Y. Fu, "Variation in glomalin in soil profiles and its association with climatic conditions, shelterbelt characteristics, and soil properties in poplar shelterbelts of Northeast China," *J. For. Res.*, vol. 31, no. 1, pp. 279–290, 2020.

[3] J. Fan, B. Chen, L. Wu, F. Zhang, X. Lu, and Y. Xiang, "Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions," *Energy*, vol. 144, pp. 903–914, 2018.

[4] L. Wu, G. Huang, J. Fan, X. Ma, H. Zhou, and W. Zeng, "Hybrid extreme learning machine with meta-heuristic algorithms for monthly pan evaporation prediction," *Comput. Electron. Agric.*, vol. 168, p. 105115, 2020.

[5] O. Kisi, O. Genc, S. Dinc, and M. Zounemat-Kermani, "Daily pan evaporation modeling using chi-squared automatic interaction detector, neural networks, classification and regression tree," *Comput. Electron. Agric.*, vol. 122, pp. 112–117, 2016.

[6] D. M. Kahler and W. Brutsaert, "Complementary relationship between daily evaporation in the environment and pan evaporation," *Water Resour. Res.*, vol. 42, no. 5, 2006.

[7] L. Wang, O. Kisi, M. Zounemat-Kermani, and Y. Gan, "Comparison of six different soft computing methods in modeling evaporation in different climates," *Hydrol. Earth Syst. Sci. Discuss.*, pp. 1–51, 2016.

[8] M. Abed, M. A. Imteaz, A. N. Ahmed, and Y. F. Huang, "Application of long short-term memory neural network technique for predicting monthly pan evaporation," *Sci. Rep.*, vol. 11, no. 1, pp. 1–19, 2021.

[9] A. Ashrafzadeh, A. Malik, V. Jothiprakash, M. A. Ghorbani, and S. M. Biazar, "Estimation of daily pan evaporation using neural networks and meta-heuristic approaches," *ISH J. Hydraul. Eng.*, vol. 26, no. 4, pp. 421–429, 2020.

[10] Ö. Kişi, "Evolutionary neural networks for monthly pan evaporation modeling," *J. Hydrol.*, vol. 498, pp. 36–45, 2013.

[11] R. C. Deo, P. Samui, and D. Kim, "Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models," *Stoch. Environ. Res. Risk Assess.*, vol. 30, no. 6, pp. 1769–1784, 2016.

[12] K. P. Sudheer, A. K. Gosain, D. Mohana Rangan, and S. M. Saheb, "Modelling evaporation using an artificial neural network algorithm," *Hydrol. Process.*, vol. 16, no. 16, pp. 3189–3202, 2002.

[13] I. Craig, A. Green, M. Scobie, and E. Schmidt, "Controlling evaporation loss from water storages," 2005.

[14] A. Malik *et al.*, "Deep learning versus gradient boosting machine for pan evaporation prediction," *Eng. Appl. Comput. Fluid Mech.*, vol. 16, no. 1, pp. 570–587, 2022, doi: 10.1080/19942060.2022.2027273.

[15] D. Freedman, R. Pisani, R. Purves, and A. Adhikari, "Statistics." WW Norton & Company New York, 2007.

[16] D. O. Rosenberry, D. I. Stannard, T. C. Winter, and M. L. Martinez, "Comparison of 13 equations for determining evapotranspiration from a prairie wetland, Cottonwood Lake area, North Dakota, USA," *Wetlands*, vol. 24, no. 3, pp. 483–497, 2004.

[17] J. C. Stephens and E. H. Stewart, "A comparison of procedures for computing evaporation and evapotranspiration," *Publication*, vol. 62, pp. 123–133, 1963.

[18] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv Prepr. arXiv1412.6980*, 2014.

[20] M. Abed, M. A. Imteaz, A. N. Ahmed, and Y. F. Huang, "Modelling monthly pan evaporation utilising Random Forest and deep learning algorithms," *Sci. Rep.*, vol. 12, no. 1, pp. 1–29, 2022.