



A novel application of transformer neural network (TNN) for estimating pan evaporation rate

Mustafa Abed¹ · Monzur Alam Imteaz¹ · Ali Najah Ahmed² · Yuk Feng Huang³

Received: 12 October 2022 / Accepted: 16 November 2022 / Published online: 30 December 2022
© The Author(s) 2022

Abstract

For decision-making in farming, the operation of dams and irrigation systems, as well as other fields of water resource management and hydrology, evaporation, as a key activity throughout the universal hydrological processes, entails efficient techniques for measuring its variation. The main challenge in creating accurate and dependable predictive models is the evaporation procedure's non-stationarity, nonlinearity, and stochastic characteristics. This work examines, for the first time, a transformer-based deep learning architecture for evaporation prediction in four different Malaysian regions. The effectiveness of the proposed deep learning (DL) model, signified as TNN, is evaluated against two competitive reference DL models, namely Convolutional Neural Network and Long Short-Term Memory, and with regards to various statistical indices using the monthly-scale dataset collected from four Malaysian meteorological stations in the 2000–2019 period. Using a variety of input variable combinations, the impact of every meteorological data on the E_p forecast is also examined. The performance assessment metrics demonstrate that compared to the other benchmark frameworks examined in this work, the developed TNN technique was more precise in modelling monthly water loss owing to evaporation. In terms of predictive effectiveness, the proposed TNN model, enhanced with the self-attention mechanism, outperforms the benchmark models, demonstrating its potential use in the forecasting of evaporation. Relating to application, the predictive model created for E_p projection offers a precise estimate of water loss due to evaporation and can thus be used in irrigation management, agriculture planning based on irrigation, and the decrease in fiscal and economic losses in farming and related industries where consistent supervision and estimation of water are considered necessary for viable living and economy.

Keywords Evaporation · Transformer neural network · Self-attention · Long short-term memory · Convolutional neural network

✉ Yuk Feng Huang
huangyf@utar.edu.my

Mustafa Abed
mustafaabed@swin.edu.au

Monzur Alam Imteaz
mimteaz@swin.edu.au

Ali Najah Ahmed
mahfoodh@uniten.edu.my

- ¹ Department of Civil and Construction Engineering, Swinburne University of Technology, Hawthorn, Melbourne, VIC 3122, Australia
- ² Institute of Energy Infrastructure and Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), 43000 Selangor, Malaysia
- ³ Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Selangor, Malaysia

Introduction

Background

A crucial step in the hydrological cycle is evaporation, which converts liquid water from the surface of earth into steam. Greater evaporation rates are a key marker of global warming (Yizhong Chen et al. 2018). Evaporation also results in considerable water loss, which has an impact on lake and reservoir water levels as well as the water budget. Consequently, accurate measurement and estimation of water loss because of evaporation are crucial for effective management of water resources (Abteu & Melesse 2012). Both indirect and direct methods are used to estimate evaporation, including evaporation pan, water balance, Penman method, energy balance, and mass transfer (L. Wu et al. 2020a). The evaporation pan technique is the most widely used since it

is comparatively easier and less costly (Kisi et al. 2016). The current work attempts to estimate evaporation pan (E_p) with an accuracy comparable to real evaporation considering (Kahler & Brutsaert 2006) demonstration that the pan evaporation technique provides a precise rate of the real alterations in evaporation. For E_p estimation, techniques based on meteorological datasets associated to the experimental evaporation equation, Energy Budget, and Water Budget have been used (L. Wang et al. 2016). The intricate stochastic characteristics of the evaporative procedure, which is not sufficiently represented by the linear modelling method, might cause the predicted errors in these techniques to be rather substantial (M. Abed et al. 2021). Furthermore, empirical models must have their model coefficients calibrated before being applied to various agroclimatic zones because they behave differently under various conditions.

Literature review

Scientists have concentrated their initiatives on machine learning approaches to estimate losses caused by evaporation due to the low performance levels and difficulties with conceptual and practical gauging techniques. These Artificial Intelligence (AI) systems are easier to use, more reliable, and capable of accurately simulating intricate nonlinear procedures (M. M. Abed et al. 2010; Kişi, 2009; Sudheer et al. 2002). Numerous studies have been conducted on utilising AI to estimate various hydrological factors (Ashrafzadeh et al. 2019). Researchers suggest that ANN frameworks offer more accurate projections compared to traditional approaches (Ditthakit et al. 2022; Pham et al. 2022). As a result, AI-based modelling techniques have been effectively applied in a variety of engineering research fields. When comparing the Box & Jenkins approach with ANN for instance, Castellano-Méndez et al. 2004 found that the latter offers higher runoff simulation performance with regards to accuracy. For estimating pan evaporation, numerous research studies have also been carried out by utilising ML techniques with multiple optimisation works (Ashrafzadeh et al. 2020; Malik et al. 2020). Goyal et al. 2014 tested LSSVR, Fuzzy Logic (FL), ANN, and ANFIS strategies for projecting daily E_p , and the results were compared to those of the Stephens-Stewart (SS) and Hargreaves-Samani (HGS) empirical methods. Results of this research have demonstrated that LSSVR and FL approaches are more effective than conventional methods for estimating daily evaporation. In order to calculate pan evaporation over monthly timeframes, Kişi 2013 developed evolutionary neural networks. The findings showed that the models were more accurate compared to empirical techniques. In their study on monthly water loss from evaporation, Deo et al. 2016 employed Multivariate Adaptive Regression Spline (MARS), Extreme Learning Machine (ELM), and Relevance Vector Machine (RVM).

Meteorological factors were used as predictor variables, and RVM was discovered being the most successful strategy. According to Sudheer et al. 2002, ANN approaches could be used to predict evaporation using weather data. They developed an ANN technique for modelling daily evaporation. Falamarzi et al. 2014 examined the application of wavelet ANN and ANN for day-to-day evaporation forecasting. They used measurements of the wind speed and temperature as model predictors. The outcomes showed that the two techniques provided accurate evaporation estimates. These shallow learning techniques have proven successful at forecasting E_p for a variety of climatic situations. However, Deep Learning (DL) algorithm-based modelling has become increasingly popular in many engineering research disciplines to produce predictions that are more precise and trustworthy (Yunzhi Chen et al. 2022).

Since deep learning (DL) approaches, which use improved multi-layered neural networks, are attractive for time series applications, they may open new possibilities for E_p estimations. This is because they are currently gaining popularity among artificial intelligence techniques that are utilised in both commercial and scientific fields owing to their increased precision (Hu et al. 2018). Recurrent neural networks (RNN), which form the foundation of DL approaches and are better candidates for estimating and projecting time series data because to their capacity to maintain and use memory from past network states, are known for their ability to do so (Chang et al. 2016; Daliakopoulos et al. 2005). Although the typical RNN model structures are capable of capturing the patterns of the time series data, they struggle to maintain the variables' longer-term dependence and have problems with vanishing and exploding gradients (Bengio et al. 1994). Because of these two fundamental flaws in the typical RNN, network training might result in unrealistic network weights that are either zero or too large. Practically speaking, remembering vital information, and avoiding redundant or unneeded information among different network states are the two key factors that guarantee improved network training. Long Short-Term Memory (LSTM), an enhanced class of conventional RNN architectures, has been developed as a potent algorithm capable of outclassing the training shortcomings of RNNs (vanishing and exploding gradient issues) by retaining important information for model establishment while preventing needless information from being conveyed to the following states in the model development procedure.

LSTM has been effectively used in research on natural language processing (NLP), financial time series forecasting, travelling period predictions, traffic congestion, and many other areas. Despite its wide applicability in a variety of research domains, LSTM approaches have lately been employed in hydrologic time series forecasting (Hu et al. 2018). Zhang et al. 2018 employed an LSTM method for

predicting water tables in rural areas. Moreover, the authors compared the outcome scheme using LSTM techniques with that of a standard ANN and noted that the former approach outperforms the ANN. Research was conducted by Majhi et al. 2020 to forecast evaporation with use of LSTM-based models. In this research, the LSTM-based approach was contrasted against Multilayer-ANN as well as empirical techniques such as Blaney–Criddle and Hargreaves, to demonstrate its superiority in predicting daily evaporative losses over selected benchmark schemes. Convolutional Neural Network (CNN), an alternative and powerful deep learning technique, has recently drawn widespread attention as a result of its varied application in a range of fields, including object recognition processing (Krizhevsky et al. 2017), time series categorisation (Z. Wang et al. 2017b), robotic haptic and visual data classification (Gao et al. 2016), weather forecasting (Liu et al. 2014), and audio signal classification (Lee et al. 2009). For instance, Ferreira and da Cunha (Ferreira & da Cunha 2020) examined one-dimensional Convolutional Neural Network (1D-CNN) plus a combination of LSTM-CNN, LSTM, as well as ML strategies (ANN and RF), for application in predicting multi step-ahead daily E_p with use of data from weather stations located in Brazil. They established that the developed DL approaches achieved relatively better results than ML strategies. It is notable that numerous researchers have used CNN in several time series forecasting fields such as electrical load estimations, solar energy forecasting, and other modelling scenarios. CNN has largely shown performance superior to that of conventional machine learning models across many studies and achieves state-of-the-art performance in most cases.

Recently, attention-based models have been employed in time series forecasting with effectiveness. Transformer architecture is derived purely from self-attention (intra-attention) mechanisms and the approach has recently become more popular. Transformers were first used in machine translation applications by (Vaswani et al. 2017) and demonstrated an exceptional capability for generalising other key tasks, including sequence modelling and computer vision. In contrast to recurrent networks, a transformer has no vanishing gradient problem and can access all points in the past irrespective of distance. This feature enables the transformer to find long-running dependencies. Unlike with recurrent networks, a transformer forgoes sequential computation and so can run completely in parallel and at higher speeds. In sum, transformer mechanisms do not analyse inputs in a sequential manner for the architecture relies on a self-attention mechanism that overcomes certain issues in recurrent and convolutional sequence-to-sequence modelling. The transformer has been successfully employed in various tasks related to time series forecasting and it outperforms numerous forecasting methods. In this context, certain work has been carried out with the goal of improving recurrent DL

models through use of self-attention mechanisms. As an example, a deep transformer model used in influenza-like illness forecasting was introduced in (N. Wu et al. 2020b) that outperforms sequence-to-sequence and LSTM architectures. The self-attention mechanism of transformer-based models performs better at forecasting than the linear-attention mechanism used in sequence-to-sequence models. Transformer methods therefore have great potential to simulate the complex dynamics found in time series data that are difficult for sequence schemes to handle. The approach may largely resolve the vanishing gradient problem that impedes Recurrent Neural Networks (RNNs) modelling of long-term predictions.

The review of literature confirmed that use of ANN via appropriate learning techniques can properly simulate evaporation in different locations, with results superior to that of relatively complex traditional approaches (Biazar et al. 2019). Nevertheless, identifying and developing efficient, reliable, and properly generalised estimation methods remains challenging due to the nonlinear complex nature of evaporation processes. Of the various ANN techniques employed recently, the innovative DL model has great potential for resolving prediction problems and is known to outperform more complex techniques. In particular, the literature shows that of the various DL methods, CNN and LSTM offer the strongest performance potential and therefore will be considered as modelling benchmarks. In recent research, attention-based models have similarly been used in time series forecasting with much success in overcoming the problems found in standard RNN and convolutional sequence-to-sequence modelling. This research presents a novel approach in the field of evaporative losses, for it is the first to attempt use of a transformer model that relies on a self-attention mechanism for E_p predictions. Successful development would be of high significance particularly for water resources management towards sustainability of farming.

Objectives

The current study aims to evaluate the applicability, predictability, and accuracy of Transformer Neural Network (TNN) schemes in predicting monthly E_p levels in four regions across Malaysia, using climatological data sets for the duration from 2000 to 2019. The TNN model performance is contrasted with two well-known deep learning approaches, Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN). Both methods are effective and compete well in DL modelling. The predictive accuracy of the models is investigated across a range of input combination scenarios to attain the highest possible levels of precision. Efficiency values for the model are analysed and evaluated using conventional statistical performance metrics that establish their

suitability for forecasting evaporation rates. Moreover, adequate analysis was performed in this research to prove TNN modelling reliability with the objective of developing a reliable approach to forecasting evaporation, a task which is particularly essential to agricultural planning and water resource management. This study presents a novel approach in the field of evaporative water loss studies, for it is the first attempt in which transformer-based architecture is used for E_p predictions.

Study area and data

Study area

Malaysia is in the tropics and thus receives a substantial amount of rain. Nonetheless, development has led to growing water demand in recent years. In addition, climate change appears to have extended the dry season all while increasing evaporation rates in reservoirs. Many researchers believe that drought is a very complex and inadequately understood natural disaster, one which affects populations far more than other threats (Shaaban & Low 2003). Precise evaporation forecasting is therefore key to developmental efforts. This study's intent is to devise accurate schemes for forecasting E_p that would be particularly beneficial in agricultural planning and water resource management. Monthly climate data is recorded across four meteorological sites located as follows: Alor Setar station (longitude 100° 24' E, latitude 6° 12' N, elevation 3.4 m), Kota Bharu station (longitude 102° 18' E, latitude 6° 10' N, elevation 4.4 m), KLIA Sepang station (longitude 101° 42' E, latitude 2° 44' N, elevation 16.1 m), and Kuantan station (longitude 103° 13' E, latitude 3° 46' N, elevation 15.2 m). These locations have been selected as a case study due to the existence of good-quality daily evaporation data as well as the importance of these cities in the region. All sites are run by the Malaysian Meteorological Department (MMD) and their outputs are relied on to both calibrate and validate recommended prediction models. Climatic variables were gathered across a variety of regions across Malaysia, as shown in Fig. 1. As well, Google Maps was utilised to map and describe the areas under study.

Data description

All proposed prediction models were constructed using seven meteorological variables, which include minimum, maximum, and mean air temperatures (T_{\min} , T_{\max} , T_a), relative humidity (RH), wind speed (S_w), solar radiation (R_s), and open pan evaporation (E_p). Data sets comprised some 19 years of daily statistics from 2000 until 2019. Various meteorological parameters logged every month that pertain

to the quantified weather data, as collected by the four previously mentioned stations, are displayed in Table 1. Additionally, Fig. 2 displays average monthly variations for all meteorological parameters during the period from 2000 until 2019.

In the table provided, X_{\min} , X_{\max} , X_{mean} , C_x , C_v , and S_x correspond to the minimum, maximum, mean, skewness, coefficient of variation, and standard deviation values of the modelled weather indicators. From the table data, the minimum value of E_p was recorded at Kuantan station, while the greatest value was logged at KLIA Sepang station. This trend might relate to site variation in the value of relative humidity, which is conversely proportional to evaporation. Kuantan station recorded the highest rate of relative humidity, whereas KLIA Sepang station recorded the lowest rate. On the other hand, the maximum skewness of E_p was logged in KLIA Sepang station, whereas the minimum skewness was logged at Kuantan. Positive values of skewness imply that the attached information is not proportional and does not follow the standard dispersion.

Model development procedure

Input combination scenarios

Selection of suitable predictors is a key step in the development of robust predictive models (Tofiq et al. 2022); various input sets of weather parameters were considered to devise successful input–output schemes and enhance the predictive capability of the ML model. This approach allows for more pragmatic understanding of how all input parameters influence evaporation estimates for a region (M. Abed et al. 2021). Input variables (predictors) were selected based on Pearson correlation coefficient considerations (Freedman et al. 2007). Pearson correlation produces bound test that quantify the statistically significant correlation, or correlation, between two continuous factors. It is recognised as a magnificent means of measuring correlations among parameters under study, for it derives from the autocorrelation method (Hauke & Kossowski 2011). The technique provides data on the association (the magnitude of correlation) and the path of the trend relationship. Both variables can be positively or negatively associated, with no relationship between the two variables determinable if the correlation coefficient equals 0. Regarding the relevant features of the meteorological parameters used in estimating monthly E_p , ranges and interpretations of the Pearson correlation coefficient results were discussed in the previous research of Mustafa et al. (M. Abed et al. 2022). Pearson correlation coefficient was used to determine the meteorological parameters that most affect evaporation estimates, as listed in Table 2. The

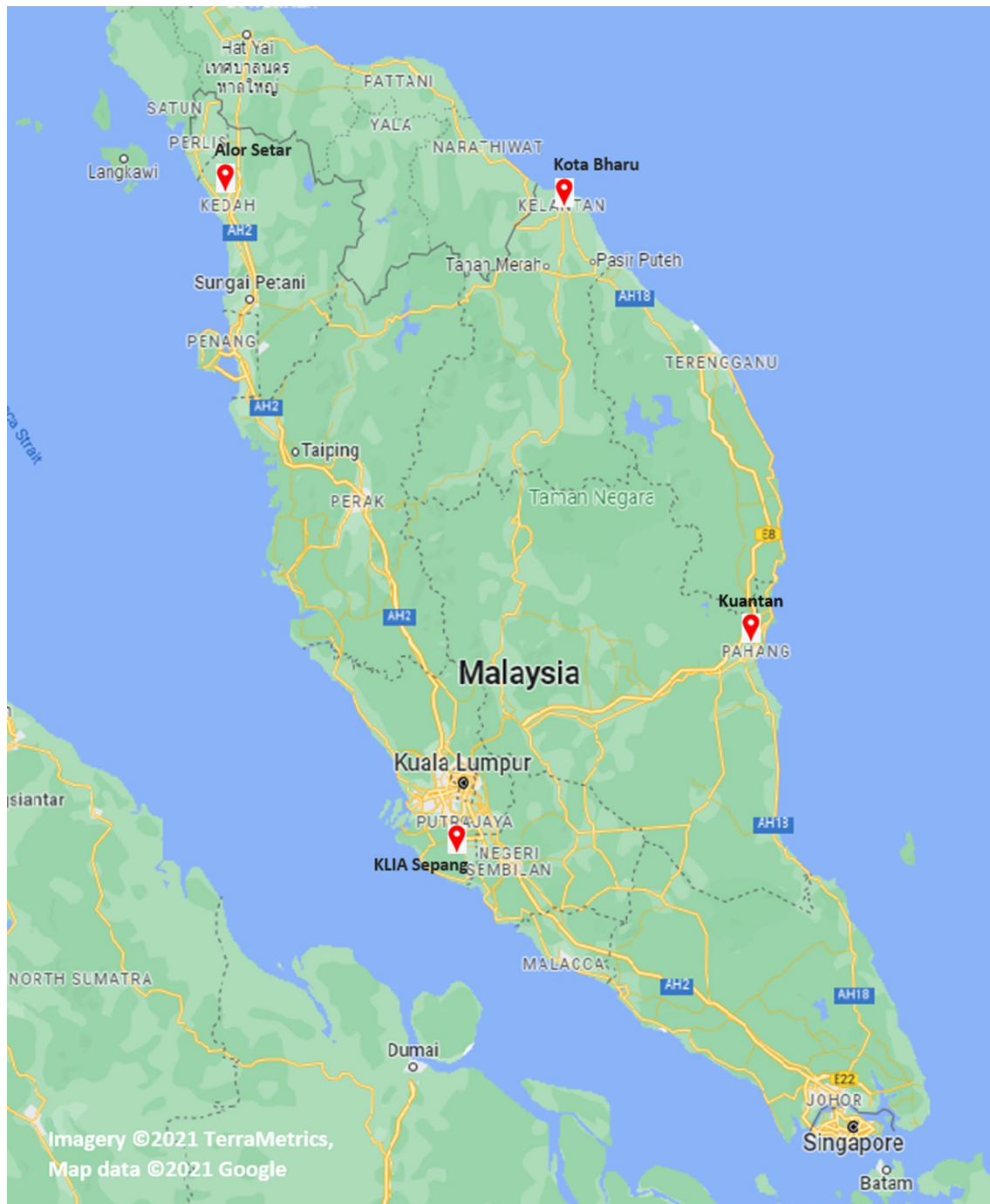


Fig. 1 Meteorological stations location [Imagery ©2021 TerraMetrics, Map data ©2021 Google]

results displayed in Table 2 indicate that T_{\max} , T_{\min} , RH, R_s , and S_w were each related to some degree with E_p , and therefore may have a key role in forecasting evaporation parameters for data from all stations. In particular, the RH and T_{\max} parameters for every site show the strongest association with E_p . RH and T_{\max} therefore will be applied in every input combination to strengthen E_p estimation accuracy. Previous studies have also implied that

T_{\max} , T_{\min} , RH, R_s , and S_w are among the key predictors of evaporation (Dalkiliç et al. 2014; L. Wang et al. 2016).

The current study also investigated the effect of the input parameter E_p on enhancing the model's performance in evaporation prediction. In this context, data were selected based on correlations with previous records and their relation with the forecasted outcome. As depicted in Fig. 3, for every station, the autocorrelation assessment concerning the

Table 1 Weather variables and descriptive statistics

Station	Dataset	Unit	X_{mean}	S_x	C_v	C_x	X_{min}	X_{max}
Alor Setar	T_{max}	°C	32.82	1.24	3.79	1.09	30.61	37.34
	T_{min}	°C	24.18	0.69	2.87	-0.40	21.58	25.98
	RH	%	80.87	5.89	7.28	-1.13	63.88	89.98
	S_w	m/s	1.66	0.41	24.66	0.10	0.51	2.95
	R_s	MJ m ⁻²	18.45	2.17	11.76	0.32	13.95	24.19
	E_p	mm	4.44	0.98	22.11	0.93	2.61	6.10
Kota Bharu	T_{max}	°C	31.33	1.29	4.14	-0.25	28.04	34.54
	T_{min}	°C	24.22	0.60	2.49	-0.20	21.81	26.48
	RH	%	80.58	2.90	3.60	0.29	72.97	87.51
	S_w	m/s	2.32	0.52	22.42	1.07	1.48	3.93
	R_s	MJ m ⁻²	19.01	2.70	14.22	-0.48	10.75	24.64
	E_p	mm	4.22	0.69	16.36	0.14	2.66	6.08
KLIA Sepang	T_{max}	°C	32.20	0.81	2.54	0.63	30.29	34.77
	T_{min}	°C	24.42	0.49	2.02	0.10	23.24	25.69
	RH	%	79.62	4.13	5.19	-0.82	63.51	87.40
	S_w	m/s	1.87	0.27	14.62	0.45	1.15	2.81
	R_s	MJ m ⁻²	17.55	2.38	13.59	0.59	11.12	24.76
	E_p	mm	4.17	0.48	11.66	1.13	3.20	6.12
Kuantan	T_{max}	°C	32.17	1.24	3.88	-0.63	28.52	34.89
	T_{min}	°C	23.71	0.64	2.70	-0.64	21.14	25.53
	RH	%	84.29	3.01	3.58	0.33	77.33	92.39
	S_w	m/s	1.64	0.30	18.49	0.64	0.91	2.65
	R_s	MJ m ⁻²	17.26	2.16	12.56	-0.47	11.73	22.39
	E_p	mm	3.79	0.53	13.64	-0.26	2.28	5.13

evaluated monthly time series for E_p levels indicated that the relationship reduced notably when it crossed the second lag-period record. It indicates that the second prior evaporation characteristics impacted evaporation at a given time. Hence, previous pan evaporation data can be used along with correlation evaluation. Consequently, using historical evaporation pan rate data with the benefit of correlation examination, the model was built using the most significant time lag concerning the past two records.

Therefore, the present study evaluated six distinct inputs for the proposed TNN framework (Table 3). Every climatic dataset was split as 80% and 20%, representing the training and testing (calibration and validation) sets. Initial years from the split dataset were used for training the model, while the remaining were used for testing.

Data pre-processing

Considering the time series aspect of this problem, data concerning specific predictors were normalised in the (0,1) range to eliminate variance before the framework was devised and trained. Since this process comprises regression and forecasting, the max–min scaling technique is employed as per the following equation.

$$\hat{X}_i = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (1)$$

where \hat{X}_i and x_i denote the normalised and observed numbers, x_{min} and x_{max} are the minimum and maximum observed values. The normalised predictor and predicted variable values were divided into training and validation sets. As specified previously, the training set comprised 80% of the observations, while the remaining 20% were used for testing.

Machine learning algorithms

The proposed TNN model and the other contrasting benchmark approaches (i.e. LSTM and CNN) were developed using an Intel Core i7-1195G7 CPU @ 2.90 GHz and 16 GB of RAM computer, built using TensorFlow in Python 3.9.0.

Transformer neural network (TNN) model for E_p prediction

The transformer framework comprises a self-attention (intra-attention) mechanism that attempts to address several concerns encountered in recurrent and convolutional sequence-to-sequence approaches. The transformer approach employs the self-attention technique to retain only the critical data from the preceding token by selecting vital details

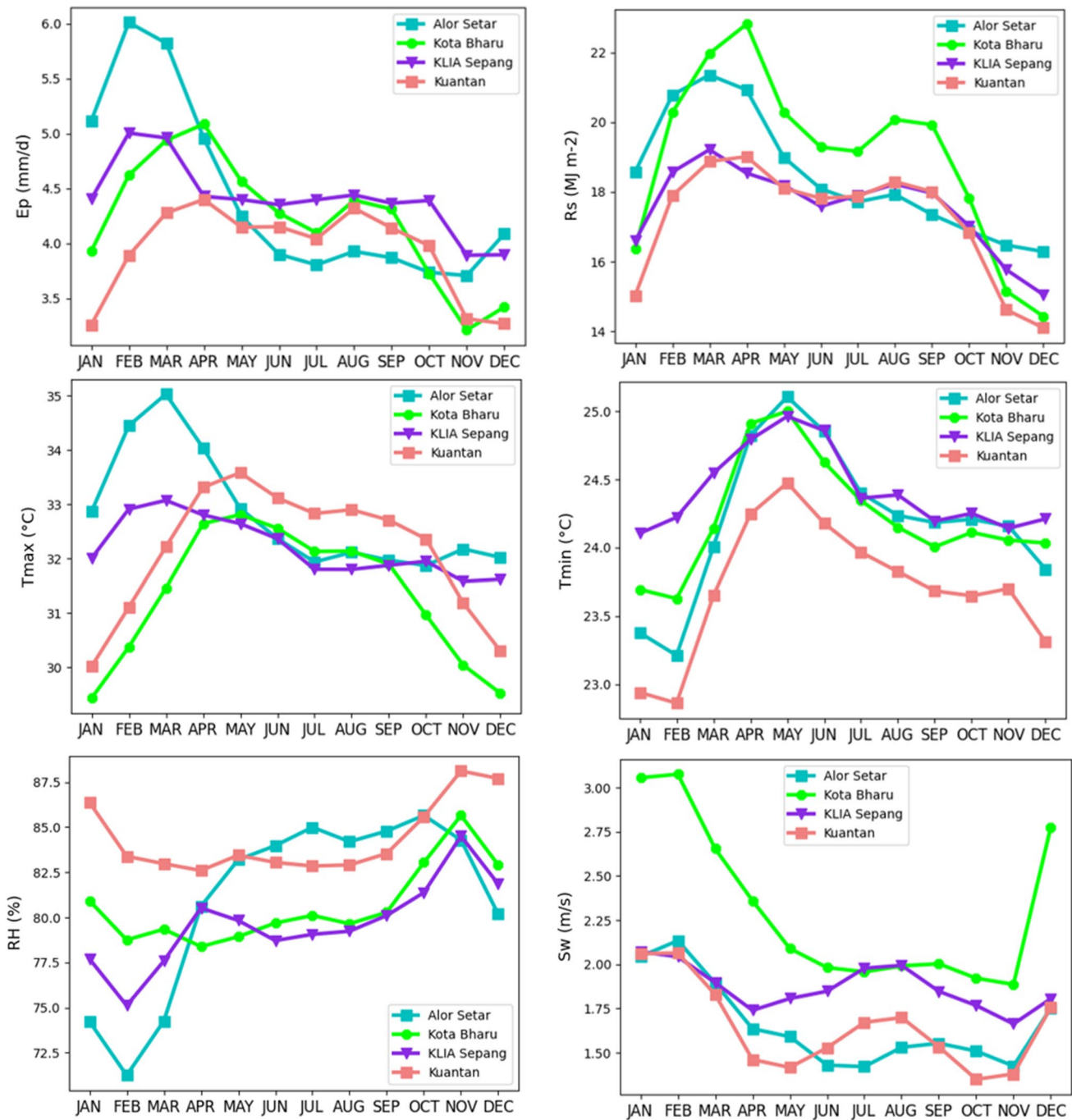


Fig. 2 Monthly variations of E_p and related meteorological indicators

concerning the present token’s encoding. Put differently, the attention approach is modified to calculate the latent space equivalent for the encoder and decoder. Nevertheless, the loss of recurrence requires positional encoding to be integrated with the inputs and outputs. Likewise, considering the recurrent time-step, positional data offers the transformer system sequences of the inputs and outputs. The encoding layer comprises two components: multi-head

self-attention (MSA) and the feed-forward layer. The attention system creates a one-to-one correlation concerning the time-specific moments. Aspects of human attention have stimulated the attention layers; however, essentially, it comprises a weighted mean reduction. Three inputs are fed to the attention layer: values, query, and keys. Every sub-layer comprises residual associations; subsequently, the layers are normalised. The objective behind several heads is typically

Table 2 Pearson's correlation coefficient matrix at each selected station

	T_{max}	T_{min}	RH	S_w	R_s	E_p
<i>Pearson's correlation coefficient array based on data from the Alor Setar station</i>						
T_{max}	1	-0.0797	-0.7183	0.4671	0.7904	0.8257
T_{min}		1	0.4625	-0.1716	0.0322	-0.3519
RH			1	-0.7365	-0.5843	-0.8773
S_w				1	0.5610	0.6562
R_s					1	0.7339
E_p						1
<i>Pearson's correlation coefficient array based on data from the Kota Bharu station</i>						
T_{max}	1	0.6820	-0.4975	-0.4272	0.7011	0.6029
T_{min}		1	-0.2089	-0.1485	0.2871	0.3957
RH			1	-0.2386	-0.6435	-0.7009
S_w				1	0.0413	0.3071
R_s					1	0.5818
E_p						1
<i>Pearson's correlation coefficient array based on data from the KLIA Sepang station</i>						
T_{max}	1	0.6557	-0.6245	0.2781	0.5006	0.6887
T_{min}		1	-0.3692	0.1567	0.4022	0.4104
RH			1	-0.4481	-0.4582	-0.7096
S_w				1	0.3911	0.5025
R_s					1	0.6653
E_p						1
<i>Pearson's correlation coefficient array based on data from the Kuantan station</i>						
T_{max}	1	0.7072	-0.5687	-0.4319	0.7011	0.7021
T_{min}		1	-0.1759	-0.4913	0.3202	0.4156
RH			1	-0.1455	-0.6274	-0.7961
S_w				1	0.0415	-0.6826
R_s					1	0.6720
E_p						1

contrasted against using several CNN filters, where every filter extracts latent features from the input. Likewise, several latent features concerning the input are extracted by several heads in the multiheaded attention approach. The outputs from every head are combined using the concatenation function. In contrast to recurrent networks, the transformer approach is free from the vanishing gradient issue and can reference any previous point, notwithstanding its distance. This aspect permits the transformer system to identify long-term dependencies. Moreover, in contrast to recurrent systems, the transformer does not require sequential computing, allowing faster speed using parallel processing. Put differently, transformer inputs are not assessed sequentially. Hence, the vanishing gradient issue is inherently eliminated. On the other hand, Recurrent Neural Networks (RNNs) suffer from this issue for long-term forecasts. Figure 4 presents the fundamental difference concerning how information is handled in an RNN vs the self-attention system. Comparatively, transformers preserve direct associations to every previous timestamp so that data can be moved over extended sequences. Nevertheless, there is a new concern:

the framework directly correlates with massive data inputs. The self-attention mechanism is used in the transformer framework to segregate non-essential information.

TNN model customization Several studies have attempted to establish that DL frameworks are superior to other machine learning techniques when measured using forecasting accuracy. Nevertheless, there is no mention of self-attention frameworks in the literature for evaporation forecast modelling. Hence, this review aims to assess the efficacy of the transformer approach for evaporation estimation when measured on the efficiency and accuracy metrics. Nevertheless, the transformer model suitable for machine transliteration cannot be directly employed for estimating time series. The following section specifies the modifications applied to the transformer model to allow its use for predicting time series. Hence, the embedding layers concerning the framework input associated with NLP are disregarded, and the time series magnitude at a specific time is provided as the input to the system. The soft-max classification layer at the output is also disregarded. Moreover,

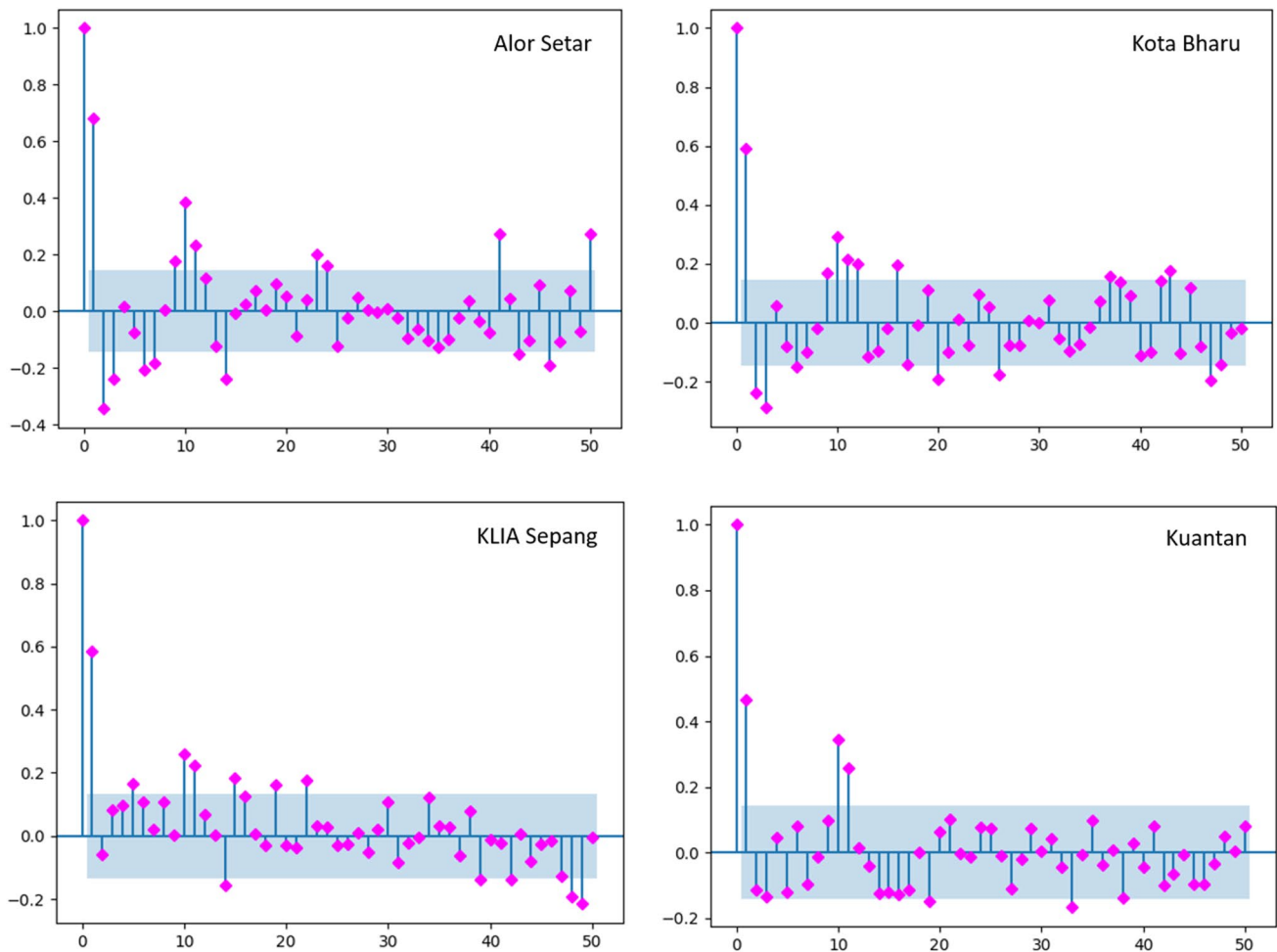


Fig. 3 Partial Autocorrelation analysis for the studied stations (Monthly)

Table 3 Meteorological variable input combinations for TNN model

TNN predictive model	Input combinations	No. of inputs
TNN-1	RH, T_{max}	2
TNN-2	RH, T_{max} , T_{min}	3
TNN-3	RH, T_{max} , T_{min} , S_w	4
TNN-4	RH, T_{max} , T_{min} , R_s	4
TNN-5	RH, T_{max} , T_{min} , R_s , S_w	5
TNN-6	RH, T_{max} , T_{min} , R_s , S_w , E_p	6

the output layer is modified to provide a linear activation function. The regression-based mean square error (MSE) expression is employed as the loss function. The original transformer framework’s encoder is employed for the training scheme. Every encoding later comprises two sub-layer: the self-attention and a fully connected feed-forward. This study comprises a one-dimensional convolutional system as a substitute for the fully connected layer to identify high-

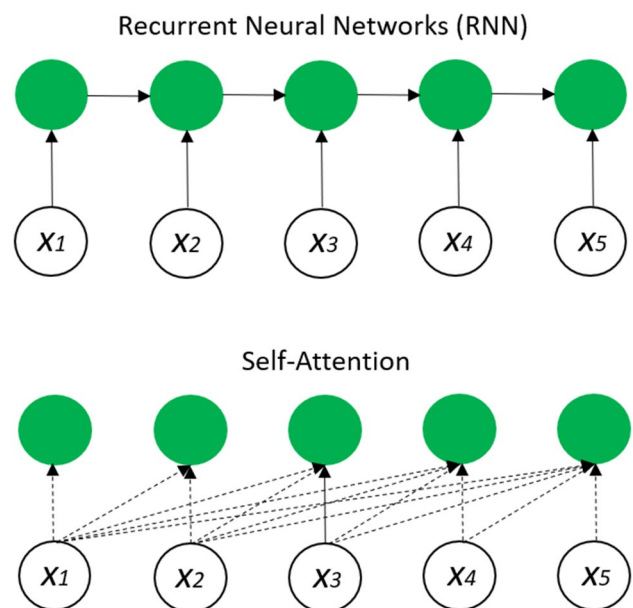


Fig. 4 Self-attention mechanism versus RNN

level characteristics. Moreover, the convolutions lack dense connections; all output nodes are not affected by all input nodes. Such an arrangement provides convolutional layers additional versatility in learning data attributes. Briefly, the encoding system comprises the transformer input comprising a specific time series, as depicted in Fig. 5. The data is then used as input for the self-attention layer comprising the encoding system; subsequently, layer normalisation is implemented. Further, a feed-forward layer is used with another layer normalisation process. This study uses a TNN framework where four similar encoding blocks are associated in a feed-forward manner. Figure 5 presents the TNN framework setup.

This work used an exhaustive search process for the system design and training hyperparameters to build optimal structures for the studied TNN model. Consequently, many differently configured models have been assessed to determine the optimal architecture. The optimal hyperparameters of this work are listed in Table 4. Based on the outcomes presented in the table, the optimal TNN framework is devised using four similar transformer encoders whose output is processed using a one-dimensional Global Average Pooling layer. Global average pooling is beneficial because it is fundamentally close to the convolutional architecture by implementation of communication between feature maps and classifications. Hence, it is simple to understand feature maps as category confidence indicators. Further, global average pooling lacks parameter optimisation need, thereby avoiding overfitting (Lin et al. 2013). Global average pooling is followed by adding a 128-neuron dense layer comprising the ELU activation function. Also, a dropout layer is introduced to control overfitting (Ferreira & da Cunha 2020). Lastly, one more single-output fully connected linear layer (i.e. forecast E_p numbers) is used. Model training was based on several iterations comprising 16 batches and 200 epochs that followed the system configurations discussed above. Network weights were regulated using the Adam algorithm (Kingma & Ba 2014) for loss function reduction. Moreover, network performance was evaluated at a $1e-2$ learning speed. After devising framework architectures, the system was trained by validating it based on the data in the training set. The subsequent step evaluated the model's forecasting ability based on fresh data. The forecast numbers were denormalised to facilitate visual representation, followed by a comparison with actual values. Figure 6 depicts the development steps concerning the TNN prediction framework.

Baseline models used for performances assessment of the proposed TNN model

This study benchmarked model performance based on Long Short-Term Memory Neural Networks (LSTM) and Convolutional Neural Networks (CNN) to contrast the proposed

model's performance. These two benchmark frameworks have different architectures and belong to distinct families in the DL framework. The optimal hyperparameters concerning the CNN and LSTM models are listed in Table 5. The hyperparameters are values that control the learning process and determine the values of model parameters that a learning algorithm learns.

The LSTM network is an adapted and enhanced form of RNN that can learn extended correlations between several time-steps comprising a data sequence. LSTMs are appropriate for forecasting sequence information since they control the exploding and vanishing gradient challenges faced by conventional RNNs. These problems are addressed by implementing gating expressions and state information. Using a specifically devised structure, the LSTM approach exhibited enhanced modelling ability for distinct time series problems. The LSTM system comprises several memory blocks connected using layers that comprise multiple recurrently associated cells. The primary blocks of a simple LSTM system comprise an input layer for feeding sequences (i.e. time series data); the model layer is employed for training the system for long-term use of the sequence (time series) data. To address a fundamental regression issue, four layers are used in the LSTM system: the network originates using the 'sequence input layer' and the 'LSTM layer'. The terminal side comprises the 'fully connected layer' and the 'regression output layer'. Theoretical notions of the LSTM are detailed in (Hochreiter & Schmidhuber 1997).

The CNN approach belongs to the deep learning paradigm. This neural network is primarily unlike a traditional ANN (i.e. MLP) since it comprises convolutional layers (of filters). Automated feature identification is implemented in such layers, where critical input data features are mapped to the required input–output association. Hence, CNN can process raw data, eliminating pre-processing or manual feature identification. CNN is typically employed for image processing. Hence, two-dimensional (2D CNN) convolutional filters have been employed (representation matches an image). Nevertheless, sequential or temporal data evaluation uses single-dimension (1D CNN) convolutional filters (Li et al. 2017). Such filters crawl the inputs to record probable temporal patterns for time series sequences. Hence, this paper uses 1D CNN. The conceptual framework of the CNN is explained by LeCun et al. (LeCun et al. 1998).

The primary objective for choosing these benchmark frameworks is to enhance precision and validity for performance assessment. Hence, the models' distinct architectures and commendable performance were the primary selection reasons; several recent papers have employed these for forecasting evaporation (M. Abed et al. 2021, 2022). The target TNN model's performance can be assessed from a wider perspective since these benchmark frameworks have distinct architectures and belong to a broad spectrum of deep learning approaches.

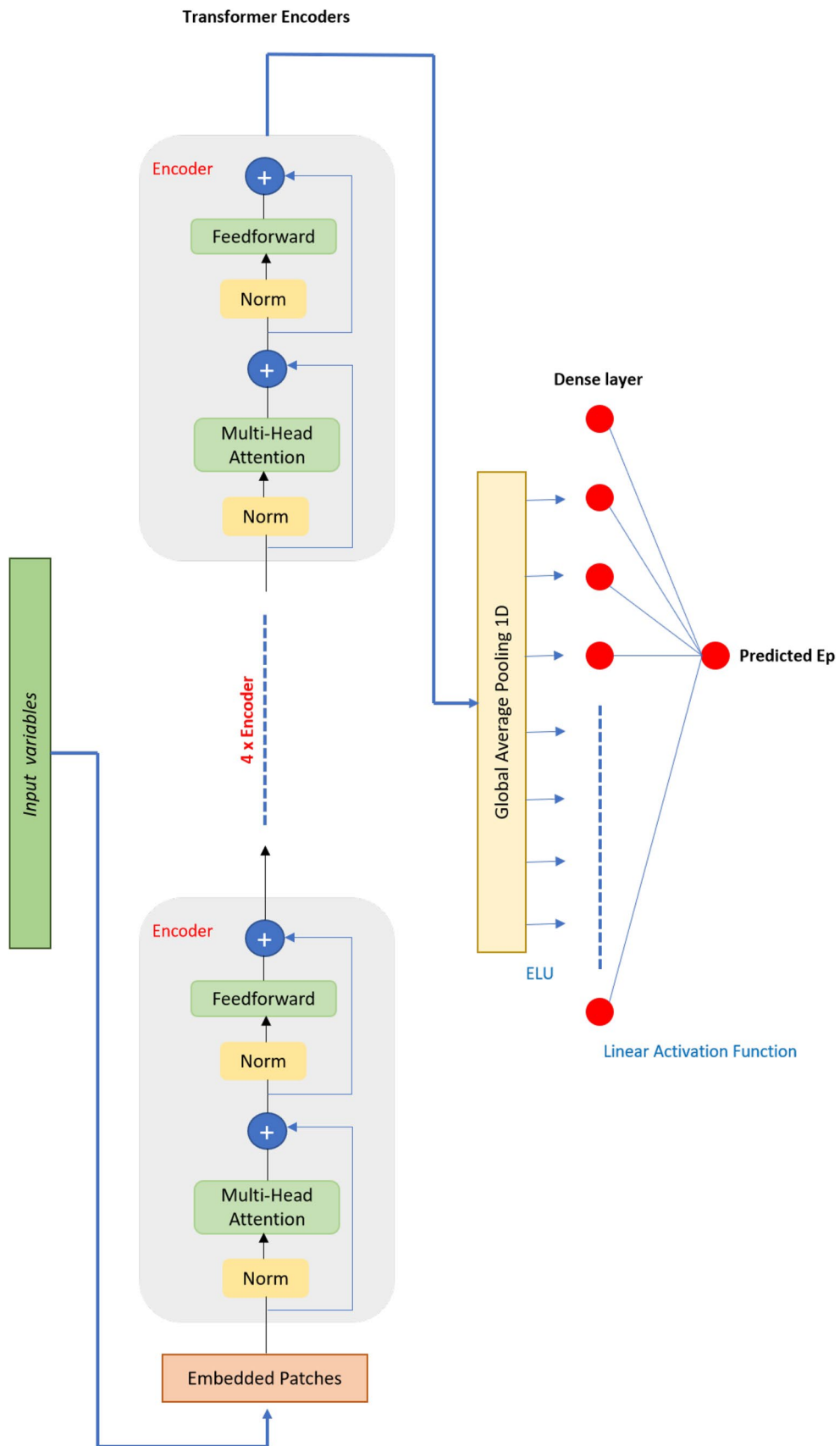


Fig. 5 The proposed TNN model architecture

Table 4 Hyperparameter tuning of Transformer model

Element	Hyperparameter	Selected
TNN Architecture	Transformer Encoders	4
	Number of neurons in MLP	128
	Activation function	ELU
	Dropout ratio	0.2
	Output activation function	Linear
Encoder	Number of heads	4
	Head size	64
	Dropout ratio	0.25
	Number of output filters in Convolutional 1D	16
	Kernel size (size of filter)	1
Training	Loss function	MSE
	Optimizer	Adam
	Learning rate	1e-2
	Number of epochs	200
	Batch size	16

This study does not employ empirical techniques for benchmarks. It is better to choose sophisticated machine learning benchmark frameworks that perform better than empirical techniques specified in the literature. Moreover, employing models with relatively poor performance than superior non-conventional machine learning approaches might highlight significant performance gaps, causing TNN framework overestimation.

Performance evaluation

It is crucial to choose suitable performance indicators since each indicator has its own set of attributes. Furthermore, the way a model performs can be better understood by knowing the properties of every statistical indicator. Thus, in this research, the predictive performance of the model was assessed by using several statistical indicators, which are described below:

(1) Coefficient of determination (R^2): The coefficient of determination represents the relationship between the estimated and real outputs; its value has a span of 0–1 (including both limits). A value of zero signifies a stochastic framework, while a value of one signifies perfect fit. R^2 is very widespread and makes model comparison more consistent and easier. It aims to assess how well a prediction model fits a dataset, providing researchers with instant feedback on the model's performance.

$$R^2 = \frac{\sum_{i=1}^n (y - \bar{y})(\hat{y} - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y - \bar{y})^2 \sum_{i=1}^n (\hat{y} - \bar{\hat{y}})^2}} \quad (2)$$

(2) Root mean square error (RMSE): RMSE represents the square root of the error squares average with respect to the estimated and real values. In regression model performance assessment, RMSE is more widely used compared to MSE. Moreover, RMSE is simple and easy to determine. Additionally, RMSE penalises huge errors, and thus become more acceptable.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y - \hat{y})^2}{n}} \quad (3)$$

(3) Mean absolute error (MAE): The MAE is the absolute difference between the estimated and actual output. MAE does not penalise high errors caused due to outliers.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y - \hat{y}| \quad (4)$$

(4) Nash–Sutcliffe efficiency (NSE): NSE is a normalised metric that determines the intensity of residual variance (noise) compared to that of the computed variability (information). It is still extensively used in hydrologic modelling, partly because it normalises the precision to a more understandable level.

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (y - \hat{y})^2}{\sum_{i=1}^n (y - \bar{y})^2} \quad (5)$$

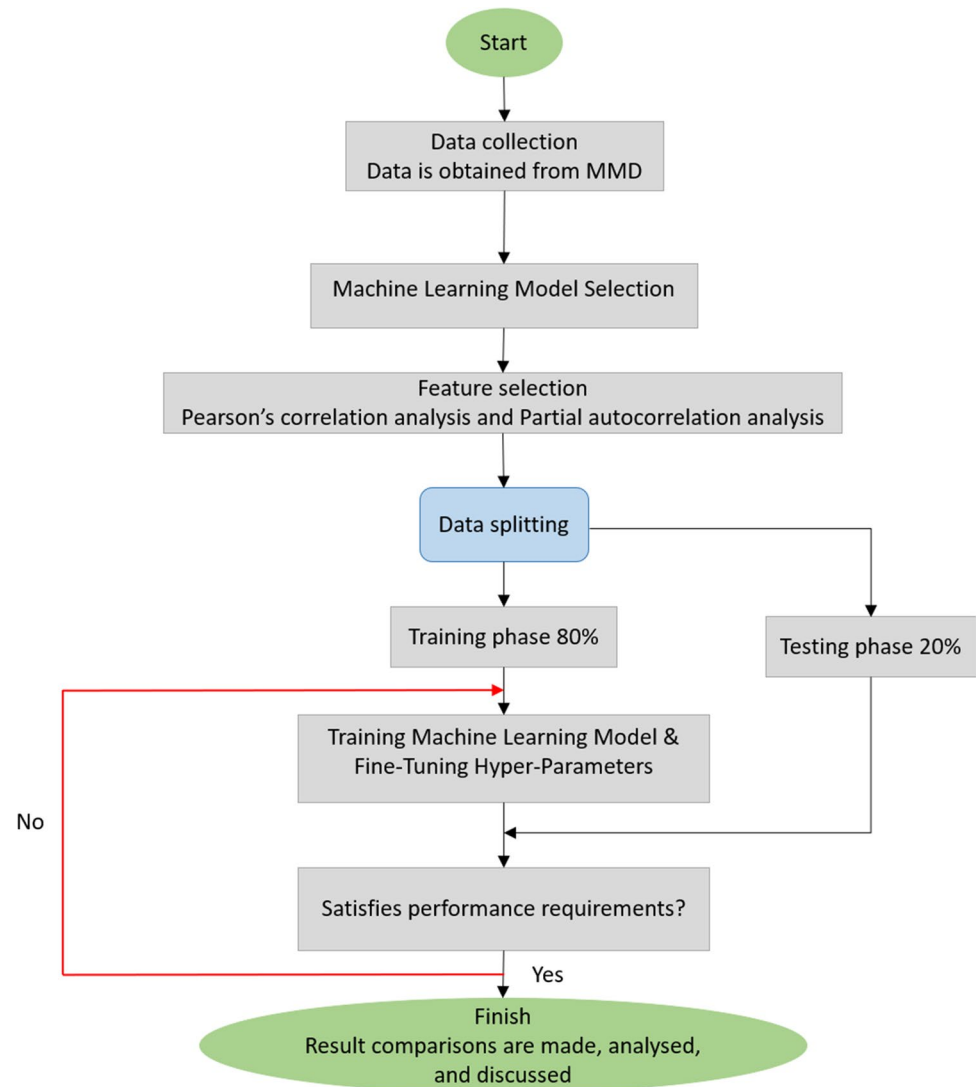
where n represents sample count, y represents the real output, \hat{y} represents the predicted values, and \bar{y} represents the real output average.

Results and discussion

Results

To indicate that the TNN model for evaporation prediction is robust, this portion provides complete analysis of the empirical outcomes derived from experimenting using this model and comparative performance assessment of various models. In this study, in all, three models were used including the model presented by the author, i.e. TNN model and standard models; CNN and LSTM were used in monthly E_p forecasting task at four sites, namely Kuantan- 103° 13' E, 3° 46' N; KLIA Sepang- 101° 42' E, 2° 44' N; Kota Bharu- 102° 18' E, 6° 10' N; and Alor Setar- 100° 24' E, 6° 12' N in Malaysia. The performance of the TNN forecasting model was first studied under different input combinations to obtain the highest accuracy of the forecast. Then, comparison of the best model was carried out against the other two standard models. All models, including the proposed model, were assessed using the outcomes of the following performance

Fig. 6 The TNN prediction model development procedure



indicators: RMSE, MAE, NSE, and R^2 in the testing duration for all research sites.

The value of R^2 is utilised to evaluate the effectiveness of all models investigated in this research in respect of degree of correlation between observed (E_p) and forecasted (E_p) values. For each model, the best statistical metrics have been displayed in bold. As can be noticed in Table 6, there is indeed a significant difference in the monthly E_p prediction accuracy determined by the input combinations. It had been possible to recognise the best prediction accuracy through the model by using the entire meteorological dataset (RH, E_p , T_{max} , T_{min} , R_s , and S_w) with respect to all sites, when compared with combination of inputs concerning other inadequate data input. In general, it further indicated that the accuracy of prediction models enhanced with extended input variables, which was consistent with the findings of previous research (Fan et al. 2016; L. Wang et al. 2017a). Four input combinations that did not include R_s or S_w were sufficient

to achieve reasonable accuracy with respect to monthly E_p prediction. When only T_{max} and RH data were available, the TNN model's prediction accuracy was found to be inadequate for all stations. This indicated that utilising advanced capacities, such as that of AI, may not enhance the predictive performance of the ML model, especially when there are limited number of meteorological inputs. In addition, there was a slight improvement in the prediction accuracy when E_p was used as an input.

In comparison with the TNN and the benchmark models, the values of R^2 for TNN tested at the four studied stations are recorded to be 0.977 for Alor Setar, 0.989 for Kota Bharu, 0.972 for KLIA Sepang, and 0.974 for Kuantan. For all the stations, the R^2 values produced by the standard models considered for this study are lower than that of the TNN model (see Table 7). Thus, it can be said that the TNN model shows the greatest degree of collinearity between estimated (E_p) and observed (E_p) values. Moreover, R^2 values produced

Table 5 Hyperparameter tuning of LSTM and CNN models

Model	Hyperparameter	Selected
LSTM	Number of hidden layers	2
	Number of neurons in the first hidden layer	512
	Number of neurons in the second hidden layer	64
	Number of epochs	500
	Batch size	8
	Training algorithm	Adam
	Dropout ratio	0.4
	Activation function	ReLU
	Learning rate	0.001
	Loss function	MAE
CNN	Number of Conv1D layer	1
	Number of filters in the Conv1D layer	32
	Kernal size	2
	Dropout ratio in the Conv1D layer	0.2
	Number of fully connected layers (MLP)	2
	Number of neurons in the first hidden layer	128
	Number of neurons in the second hidden layer	256
	Dropout ratio in the fully connected layers	0.1
	Number of epochs	500
	Batch size	16
	Training algorithm	Adam
	Activation function	ReLU
	Learning rate	0.001
	Loss function	MSE

Table 6 Statistical performance of the TNN model (testing period) used for estimating monthly E_p under six input scenarios for the studied stations

Station	Predictive Model	R^2	MAE	RMSE	NSE
Alor Setar	TNN-1	0.875	0.219	0.319	0.876
	TNN-2	0.914	0.187	0.265	0.915
	TNN-3	0.935	0.168	0.230	0.936
	TNN-4	0.954	0.142	0.193	0.955
	TNN-5	0.968	0.128	0.163	0.969
	TNN-6	0.977	0.097	0.138	0.978
Kota Bharu	TNN-1	0.865	0.177	0.221	0.866
	TNN-2	0.878	0.160	0.209	0.879
	TNN-3	0.880	0.159	0.205	0.881
	TNN-4	0.929	0.120	0.155	0.930
	TNN-5	0.959	0.019	0.117	0.960
	TNN-6	0.989	0.016	0.023	0.987
KLIA Sepang	TNN-1	0.857	0.135	0.196	0.878
	TNN-2	0.912	0.111	0.145	0.913
	TNN-3	0.929	0.096	0.132	0.930
	TNN-4	0.940	0.091	0.107	0.941
	TNN-5	0.952	0.087	0.103	0.953
	TNN-6	0.972	0.075	0.091	0.968
Kuantan	TNN-1	0.875	0.131	0.174	0.876
	TNN-2	0.889	0.124	0.164	0.890
	TNN-3	0.920	0.100	0.130	0.929
	TNN-4	0.919	0.107	0.148	0.920
	TNN-5	0.966	0.071	0.089	0.967
	TNN-6	0.974	0.067	0.082	0.975

by the TNN model for all stations are rather near to 1. It deserves to be remarked that the value of the coefficient of determination (R^2) achieves 1 in case of ideal performance of the model. Therefore, the TNN model apparently shows a better performance in comparison with the other benchmark models with respect to the value of R^2 determined for the estimated and observed values.

The current results suggest that the values of the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) become lower with higher model performance. For instance, for all study sites, the TNN model gives RMSE values of 0.138, 0.023, 0.091, and 0.082, and MAE values of 0.097, 0.016, 0.075, and 0.067, respectively. Importantly, the benchmark models yield higher RMSE and MAE values versus the TNN model for all the study stations, as seen in Table 7. For example, with regards to LSTM, RMSE, and MAE values were 0.166 and 0.135, and for CNN, they were 0.175 and 0.143, respectively. At the Alor Setar site, all values are greater when compared with the RMSE (0.138) and MAE (0.097) values with regards to the TNN model. Such evaluation criteria evaluate the models based on forecast errors, and thus it is presumed that these forecasting models will yield lower values. As per the RMSE and MAE values, the TNN

Table 7 Statistical performance of the proposed TNN and benchmark models (testing period) used for predicting monthly E_p for the studied stations

Station	Predictive Model	R^2	MAE	RMSE	NSE
Alor Setar	LSTM*	0.970	0.135	0.166	0.971
	CNN	0.960	0.143	0.175	0.961
	TNN-6	0.977	0.097	0.138	0.978
Kota Bharu	LSTM*	0.986	0.058	0.074	0.986
	CNN	0.973	0.072	0.095	0.974
	TNN-6	0.989	0.016	0.023	0.987
KLIA Sepang	LSTM	0.947	0.047	0.121	0.948
	CNN**	0.965	0.079	0.091	0.966
	TNN-6	0.972	0.075	0.091	0.968
Kuantan	LSTM	0.957	0.080	0.102	0.958
	CNN**	0.962	0.084	0.103	0.962
	TNN-6	0.974	0.067	0.082	0.975

*results obtained from (M. Abed et al. 2021)

**results obtained from (M. Abed et al. 2022)

model showed significantly improved performance when compared with all other comparable models as it demonstrated lesser forecasting errors. As another measure pertaining to model accuracy, Fig. 7 displayed the radar plots in terms of the RMSE for the TNN model and the benchmark models pertaining to all study sites. Moreover, the Nash–Sutcliffe efficiency (NSE) value, which can be regarded as another metric employed for evaluating the put forward deep learning model’s efficacy, seems to be near to a value of oneness for all study regions versus the benchmark models. To emphasise this metric with regards to the TNN model, for instance, $NSE \geq 0.968$ has been noted for all the study regions. More importantly, these values were found to be higher when compared with the comparative models with regards to all the study stations, as presented in Table 7. Overall, the current investigation offers convincing proof supporting that a significant potential is associated with the TNN model to predict monthly E_p , and this performance was found to be higher versus

comparative models pertaining to all the study sites in Malaysia.

It is worthy to note that the TNN model showed lower predictive error for all the study stations versus the benchmark models employed in the testing phase. Figures 8 and 9 display the scatter plots as well as time series that have been observed compared with the predicted monthly E_p with regards to the TNN model as well as the benchmark models with regards to the testing phase. These scatter plots give the coefficient of determination (R^2), which can inform a reader how effectively the variability in observed E_p compares to that of modelled E_p , where this value is between 0 and 1. If the model fits well and there is good resonance in the comparative correlations between the observed and forecasted E_p , the R^2 can be deemed to be nearer to oneness. As per Figs. 8 and 9, all the models showed high R^2 values closer to unity. However, the TNN model gave the highest R^2 value versus the other models with regards to all study sites. This deduction demonstrated that the TNN model could yield better accuracy with regards to E_p versus

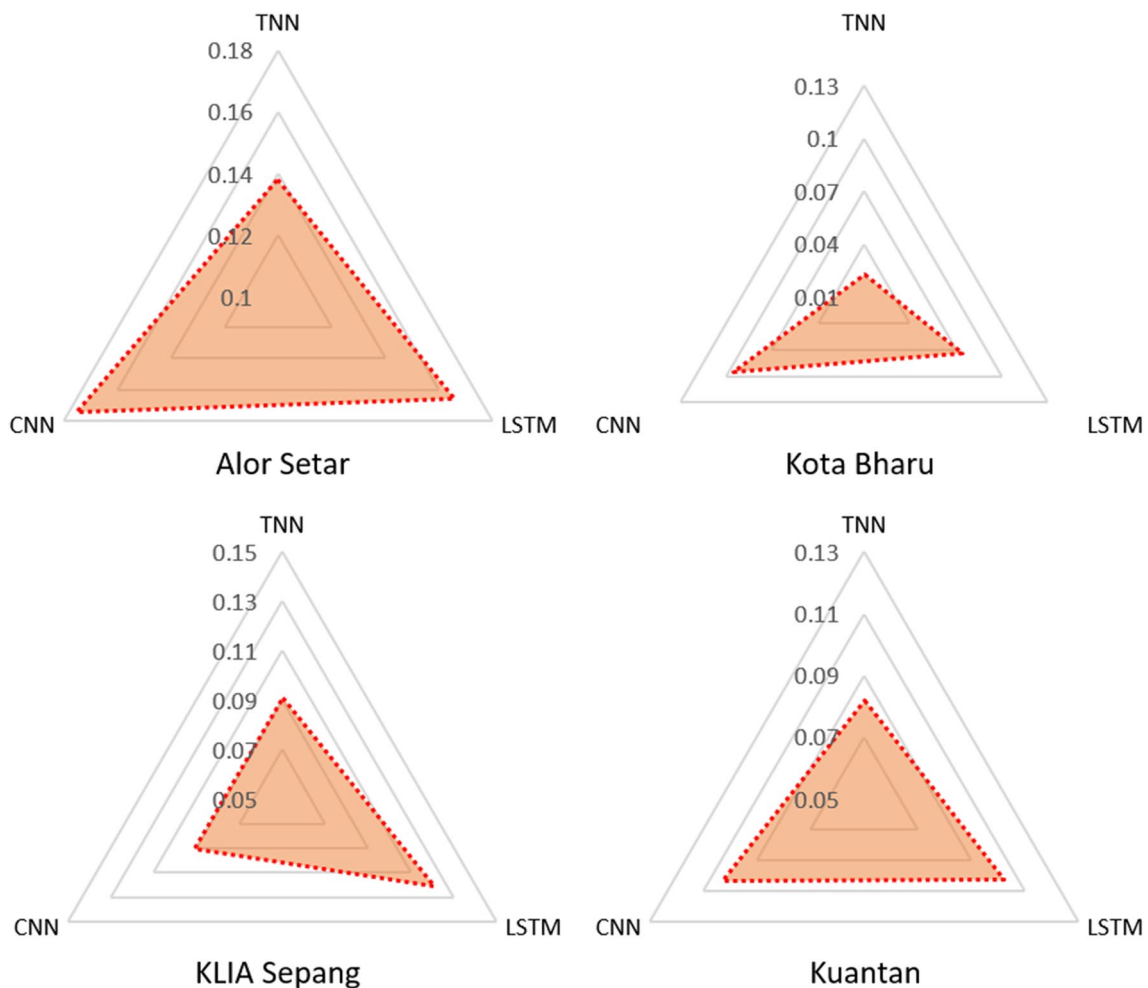


Fig. 7 Spider chart of the Root Mean Square Error (RMSE) for the developed TNN and the comparative benchmark models

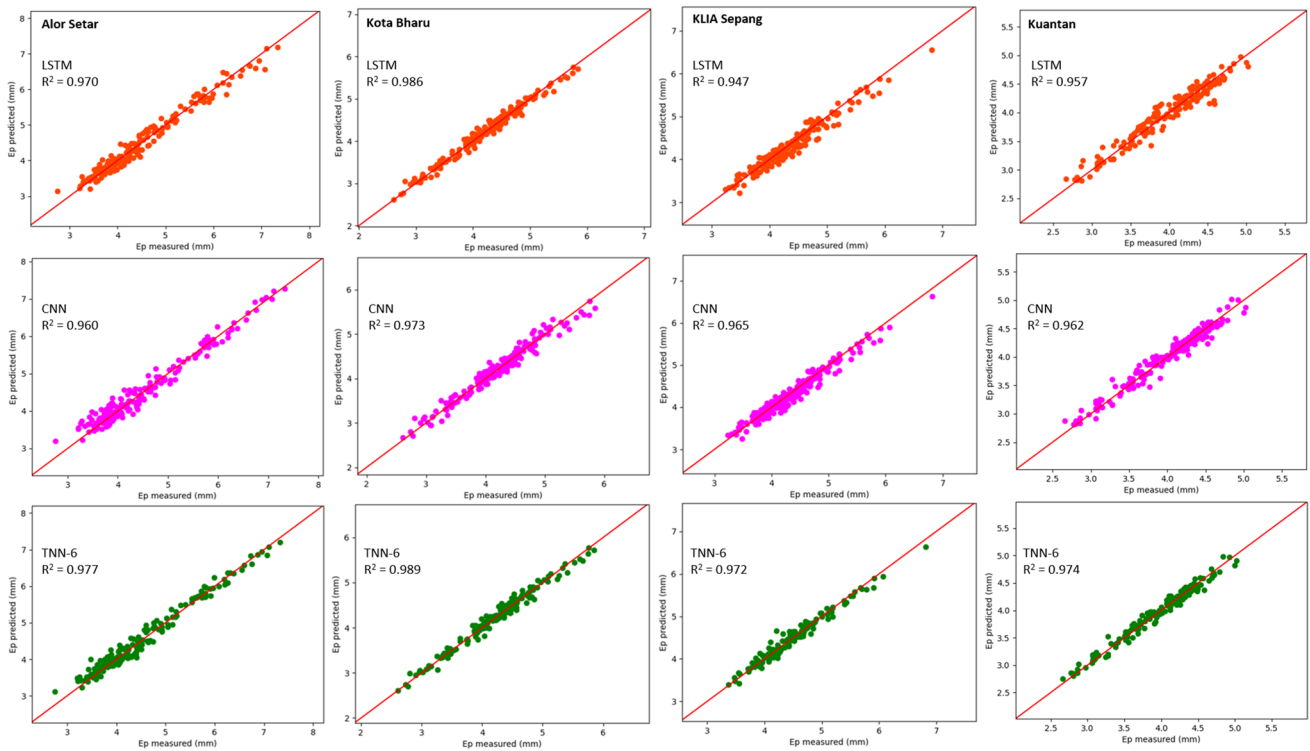


Fig. 8 Scatter plot of measured E_p versus predicted E_p of TNN and benchmark models at all stations

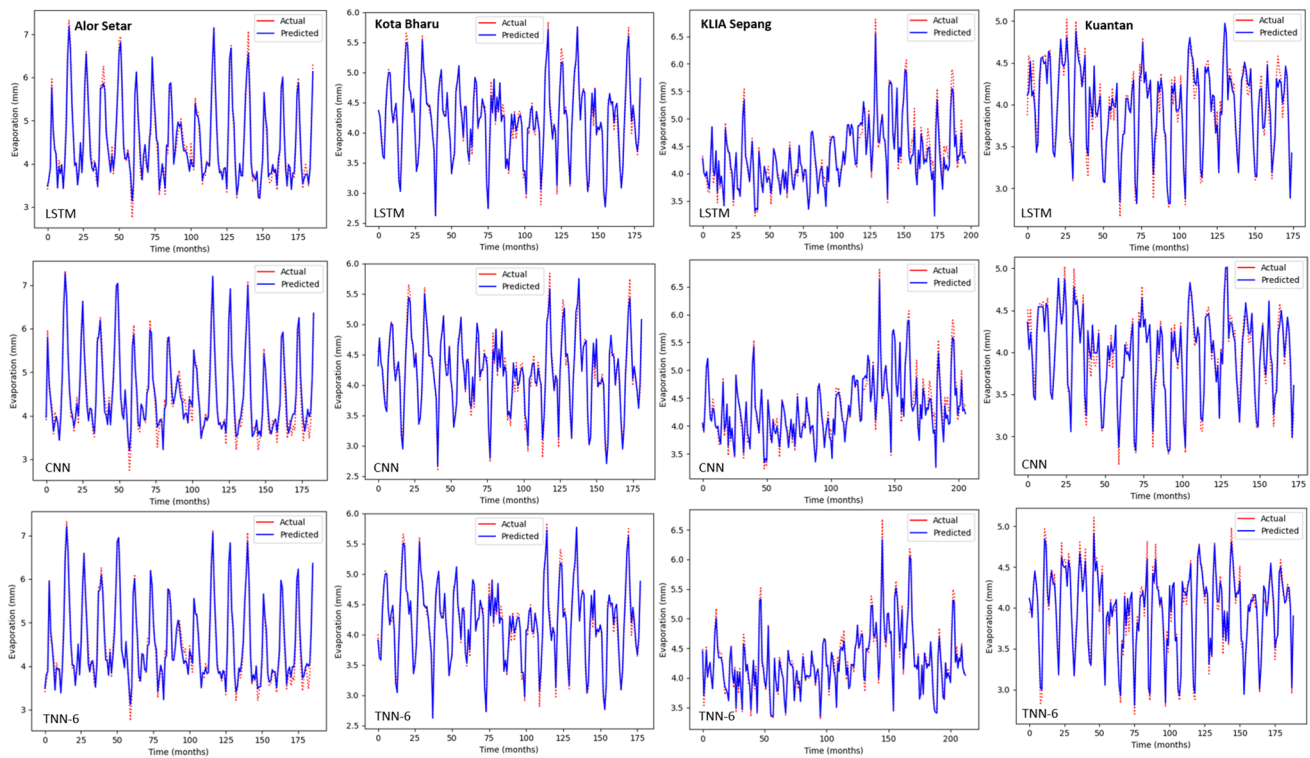


Fig. 9 Time series of measured E_p versus predicted E_p of TNN and benchmark models at all stations

the other models examined in this study and proved that the putative DL model could be regarded as an important tool for predicting E_p . As per the results, the put forward TNN model was found to be better when compared with all benchmark models in all sites, demonstrating the positive impact cast by self-attention mechanism on improving the accuracy of prediction.

Discussion

For deep learning, transformers are regarded to be a state-of-the-art approach. The adoption of attention has been revolutionised by the transformer model via dispensing with convolution and recurrence and, alternatively, depending exclusively on a self-attention mechanism. When it comes to time series forecasting, transformers are not able to analyse their input in a sequential manner. Thus, this has helped deal with the vanishing gradient problem that hampers the recurrent neural networks (RNNs) with regards to long-term prediction. The present study applies a novel application pertaining to self-attention algorithms with regards to the evaporative prediction domain. The put forward TNN model has demonstrated good performances with regards to E_p prediction for all the four selected stations, whereas other models investigated in this study have been placed in different positions by considering the E_p prediction performances pertaining to various sites. The consistency demonstrated by the put forward TNN model with regards to the E_p prediction performances by being the best E_p prediction model for all study sites validates the impact of the self-attention mechanism. Thus, the architecture of the transformer, which is solely based on self-attention mechanisms (intra-attention), has demonstrated potential to enhance the DL predictive models' performance.

With regards to the study limitations, data were gathered and modelled based on just four study regions in Malaysia (as a case study) to accomplish the goals. While this pioneering research has yielded a new modelling structure with regards to the E_p prediction, despite its restricted context, further research could choose a broader range of regions elsewhere, signifying different weather conditions. Nevertheless, the deep learning seems to have considerable implications with regards to managing irrigations as well as other water resource systems by monitoring the changes pertaining to monthly E_p .

One of the research's practical implications is that the E_p modelling approach, which can give a quite close estimate of the real water loss because of evaporation as well as its relation to managing water resources, could be employed as a science-based strategic approach that can be applied to irrigation and other agricultural tasks. When E_p values is multiplied by the surface area pertaining to the irrigation water resources, the amount of water shortages resulting

from evaporation (a primary component of water loss common for the existing water asset volume) can be evaluated. Thus, it becomes easy to estimate the total amount of existing water that can be used for irrigation, and planning and implementation of a toolset of intelligent irrigation schedules. These schedules can also help dodge unnecessary water losses since irrigation practices become more relaxed. Thus, the current TNN model employed for predicting E_p is also believed to provide considerable economic advantages to the agriculturists, especially in areas where farming is influenced by water resource issues, droughts, and other types of hydrological disparities. In addition, this study offers optimal advice for hydrologists on how to effectively analyse non-stationary and nonlinear behaviours pertaining to hydrological cycles employing soft computing.

Conclusion

This study concentrated on developing a transformer-based architecture (TNN) model that can be employed practically for prediction of monthly E_p losses and provide a detailed comparison with other benchmark DL models, including CNN and LSTM approaches. To evaluate the capabilities pertaining to the designed DL models, monthly data from four meteorological stations in Malaysia were considered for predicting the E_p rates. Time series data with regards to the monthly E_p , like T_{\min} , T_{\max} , RH, R_s , S_w , and E_p , in the years 2000–2019 were employed for testing (validation) and training (calibration) for the designed models. The input parameters (predictors) were chosen based on the Pearson's correlation coefficient values to detect the most effective input combinations pertaining to the TNN model. Based on standard statistical measures, the performance of each model and its efficacy with regard to evaporation forecasting were evaluated.

The investigation provided the following results:

- A high level of prediction accuracy for the monthly E_p was observed with the three developed DL models.
- The TNN model delivered enhanced performance when it comes to predicting monthly E_p versus the benchmark models with regards to all study sites.
- Models that considered complete meteorological datasets (T_{\min} , T_{\max} , S_w , R_s , E_p , and RH) were found to achieve the best prediction accuracy at all stations, versus other combinations employing limited data input.
- As evident in the results, the performance of the developed TNN model was significantly better versus other benchmark models at all study sites. This supported the fact that the TNN model can be employed in an efficient manner for predicting monthly E_p data series.

- In terms of application, the TNN model provides a precise estimate of water loss due to evaporation and can thus be used in irrigation management, agriculture planning based on irrigation, and the reduction in fiscal and economic losses in farming and related industries where consistent supervision and estimation of water are considered necessary for viable living and economy.
- In the future, the applicability of the proposed technique can be tested for different areas in Malaysia or elsewhere using different data sets to develop a reliable, generalizable model that can predict evaporation.

Acknowledgements The author would like to thank the Australian Government Research Training Programme Scholarship (RTP) for its support. Also, the Malaysian Meteorological Department (MMD) for providing this study with the data.

Author Contributions MA contributed to methodology, formal analysis, visualisation, and writing—review and editing; MAI contributed to writing—review and editing and supervision; ANA contributed to writing—review and editing and supervision; YFH contributed to data curation and writing—review and editing.

Funding The APC was covered by Universiti Tunku Abdul Rahman (UTAR), Malaysia, via Research Publication Scheme (Project Number: UTARRPS 6251/H03) and UTARFSJPP.

Data availability The datasets used during the current study are available from the first author on reasonable request.

Declarations

Conflict of interests The authors declare no conflict of interest.

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

References

- Abed MM, El-Shafie A, Osman SAB (2010) Creep predicting model in masonry structure utilizing dynamic neural network. *J Comput Sci* 6(5):597
- Abed M, Imteaz MA, Ahmed AN, Huang YF (2021) Application of long short-term memory neural network technique for predicting monthly pan evaporation. *Sci Rep* 11(1):1–19
- Abed M, Imteaz MA, Ahmed AN, Huang YF (2022) Modelling monthly pan evaporation utilising random forest and deep learning algorithms. *Sci Rep* 12(1):1–29
- Abteu W, Melesse A (2012) *Evaporation and evapotranspiration: measurements and estimations*. Springer Science & Business Media, Berlin
- Ashrafzadeh A, Ghorbani MA, Biazar SM, Yaseen ZM (2019) Evaporation process modelling over northern Iran: application of an integrative data-intelligence model with the krill herd optimization algorithm. *Hydrol Sci J* 64(15):1843–1856
- Ashrafzadeh A, Malik A, Jothiprakash V, Ghorbani MA, Biazar SM (2020) Estimation of daily pan evaporation using neural networks and meta-heuristic approaches. *ISH J Hydraulic Eng* 26(4):421–429
- Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Networks* 5(2):157–166. <https://doi.org/10.1109/72.279181>
- Biazar SM, Ghorbani MALI, Shahedi K (2019) Uncertainty of artificial neural networks for daily evaporation prediction (Case study: Rasht and Manjil Stations). *J Watershed Manag Res* 10(19):1–12
- Castellano-Méndez M, González-Manteiga W, Febrero-Bande M, Prada-Sánchez JM, Lozano-Calderón R (2004) Modelling of the monthly and daily behaviour of the runoff of the Xallas river using box-jenkins and neural networks methods. *J Hydrol* 296(1–4):38–58
- Chang F-J, Chang L-C, Huang C-W, Kao I-F (2016) Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. *J Hydrol* 541:965–976. <https://doi.org/10.1016/j.jhydrol.2016.08.006>
- Chen Y, He L, Li J, Zhang S (2018) 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* 109:216–235
- Chen Y, Chen W, Janizadeh S, Bhunia GS, Bera A, Pham QB, Linh NTT, Balogun A-L, Wang X (2022) Deep learning and boosting framework for piping erosion susceptibility modeling: spatial evaluation of agricultural areas in the semi-arid region. *Geocarto Int* 37(16):4628–4654
- Daliakopoulos IN, Coulibaly P, Tsanis IK (2005) Groundwater level forecasting using artificial neural networks. *J Hydrol* 309(1):229–240. <https://doi.org/10.1016/j.jhydrol.2004.12.001>
- Dalkılıç Y, Okkan U, Baykan N (2014) Comparison of different ANN approaches in daily pan evaporation prediction. *J Water Resour Protect* 6:319–326
- Deo RC, Samui P, Kim D (2016) Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. *Stoch Env Res Risk Assess* 30(6):1769–1784
- Ditthakrit P, Pinthong S, Salaeh N, Weekaew J, Tran TT, & Pham QB (2022) Comparative study of machine learning methods and GR2M model for monthly runoff prediction. *Ain Shams Eng J* 101941
- Falamarzi Y, Palizdan N, Huang YF, Lee TS (2014) Estimating evapotranspiration from temperature and wind speed data using artificial and wavelet neural networks (WNNs). *Agric Water Manag* 140:26–36
- Fan J, Wu L, Zhang F, Xiang Y, Zheng J (2016) Climate change effects on reference crop evapotranspiration across different climatic zones of China during 1956–2015. *J Hydrol* 542:923–937
- Ferreira LB, da Cunha FF (2020) Multi-step ahead forecasting of daily reference evapotranspiration using deep learning. *Comput Electron Agric* 178:105728. <https://doi.org/10.1016/j.compag.2020.105728>
- Freedman D, Pisani R, Purves R, Adhikari A (2007) *Statistics*. WW Norton & Company, New York
- Gao Y, Hendricks LA, Kuchenbecker KJ, Darrell T (2016) Deep learning for tactile understanding from visual and haptic data. *IEEE Int Conf Robot. Autom (ICRA)* 2016:536–543

- Goyal MK, Bharti B, Quilty J, Adamowski J, Pandey A (2014) Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. *Expert Syst Appl* 41(11):5267–5276
- Hauke J, Kossowski T (2011) Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data. *Quaest Geogr* 30:87
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Hu C, Wu Q, Li H, Jian S, Li N, Lou Z (2018) Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water*. <https://doi.org/10.3390/w10111543>
- Kahler DM, Brutsaert W (2006) Complementary relationship between daily evaporation in the environment and pan evaporation. *Water Resour Res* 42(5)
- Kingma DP, & Ba J (2014). Adam: a method for stochastic optimization. ArXiv Preprint <http://arxiv.org/abs/1412.6980>
- Kisi Ö (2009) Daily pan evaporation modelling using multi-layer perceptrons and radial basis neural networks. *Hydrol Process Int J* 23(2):213–223
- Kisi Ö (2013) Evolutionary neural networks for monthly pan evaporation modeling. *J Hydrol* 498:36–45
- Kisi O, Genc O, Dinc S, Zounemat-Kermani M (2016) Daily pan evaporation modeling using chi-squared automatic interaction detector, neural networks, classification and regression tree. *Comput Electron Agric* 122:112–117
- Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
- LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- Lee H, Pham P, Largman Y, Ng A (2009) Unsupervised feature learning for audio classification using convolutional deep belief networks. *Adv Neural Inf Process Syst* 22:1096–1104
- Li D, Zhang J, Zhang Q, Wei X (2017) Classification of ECG signals based on 1D convolution neural network. In: *2017 IEEE 19th International Conference on E-Health Networking, Applications and Services (Healthcom)*, 1–6. Doi: <https://doi.org/10.1109/HealthCom.2017.8210784>
- Lin, M., Chen, Q., & Yan, S. (2013). Network in network. Preprint <http://arxiv.org/abs/1312.4400>
- Liu JNK, Hu Y, You JJ, Chan PW (2014). Deep neural network based feature representation for weather forecasting. In: *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, 1
- Majhi B, Naidu D, Mishra AP, Satapathy SC (2020) Improved prediction of daily pan evaporation using Deep-LSTM model. *Neural Comput Appl* 32(12):7823–7838
- Malik A, Kumar A, Kim S, Kashani MH, Karimi V, Sharafati A, Ghorbani MA, Al-Ansari N, Salih SQ, Yaseen ZM (2020) Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model. *Eng Appl Comput Fluid Mech* 14(1):323–338
- Pham QB, Kumar M, Di Nunno F, Elbeltagi A, Granata F, Islam ARM, Talukdar S, Nguyen XC, Ahmed AN, & Anh DT (2022). Groundwater level prediction using machine learning algorithms in a drought-prone area. *Neural Comput Appl* 1–23
- Shaaban AJ, Low KS (2003) Droughts in Malaysia: a look at its characteristics, impacts, related policies and management strategies. In: *Water and Drainage 2003 Conference*, 28–29
- Sudheer KP, Gosain AK, Mohana Rangan D, Saheb SM (2002) Modelling evaporation using an artificial neural network algorithm. *Hydrol Process* 16(16):3189–3202
- Tofiq YM, Latif SD, Ahmed AN, Kumar P, El-Shafie A (2022) Optimized model inputs selections for enhancing river streamflow forecasting accuracy using different artificial intelligence techniques. *Water Resour Manag* 1–18
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, & Polosukhin I (2017) Attention is all you need. *Adv Neural Inform Process Syst* 30
- Wang L, Kisi O, Zounemat-Kermani M, Li H (2017a) Pan evaporation modeling using six different heuristic computing methods in different climates of China. *J Hydrol* 544:407–427
- Wang L, Kisi O, Zounemat-Kermani M, Gan Y (2016) Comparison of six different soft computing methods in modeling evaporation in different climates. *Hydrol Earth Syst Sci Discuss* 1–51
- Wang Z, Yan W, Oates T (2017b) Time series classification from scratch with deep neural networks: a strong baseline. In: *2017b International Joint Conference on Neural Networks (IJCNN)*, 1578–1585
- Wu L, Huang G, Fan J, Ma X, Zhou H, Zeng W (2020a) Hybrid extreme learning machine with meta-heuristic algorithms for monthly pan evaporation prediction. *Comput Electron Agric* 168:105115
- Wu N, Green B, Ben X, O'Banion S (2020b). Deep transformer models for time series forecasting: The influenza prevalence case. Preprint <http://arxiv.org/abs/2001.08317>
- Zhang J, Zhu Y, Zhang X, Ye M, Yang J (2018) Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas. *J Hydrol* 561:918–929

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.