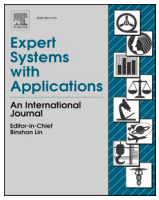
Expert Systems With Applications 235 (2024) 121160

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Data-driven multi-step prediction and analysis of monthly rainfall using explainable deep learning

Renfei He [a,b], Limao Zhang [c,*], Alvin Wei Ze Chew [d]

[a] School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore
[b] Environmental Process Modelling Centre, Nanyang Environment and Water Research Institute (NEWRI), Interdisciplinary Graduate Programme, Nanyang Technological University, 1 Cleantech Loop, 637141, Singapore
[c] School of Civil and Hydraulic Engineering, National Center of Technology Innovation for Digital Construction, Huazhong University of Science and Technological, Wuhan, Hubei 430074, China
[d] Bentley Systems (Singapore) Pte Ltd, 3 Harbourfront Pl, #05-01 to 02 HarbourFront Tower Two, 099254, Singapore

ARTICLE INFO

Keywords:
Multi-step prediction
Monthly rainfall
Explainable deep learning
Encoder-decoder with an attention mechanism
Expected gradient

ABSTRACT

Monthly rainfall prediction is a crucial topic for the management of water resources and prevention of hydrological disasters. To make a multi-step monthly rainfall prediction and discover the primary factors affecting rainfall, this study presents an explainable deep learning approach that integrates four consecutive modules, namely the Gated-Recurrent-Unit-based (GRU-based) encoder module, attention mechanism module, GRU-based decoder module, and expected-gradient-based explanation module, respectively. The first three modules constitute an encoder-decoder with an attention mechanism which could predict monthly rainfall in multiple continuous months in the future, while the explanation module computes attribution values for the input weather and climate features to quantify their importance. A case study is conducted on monthly rainfall data collected from Darwin and Perth, Australia, from Jan 1921 to Dec 2020. The results mainly indicate that: (1) Compared with the baseline methods, the multi-step monthly rainfall predictions made by the encoder-decoder with an attention mechanism show better agreement with the ground truth. (2) For January rainfall in Darwin, the most important weather factors are last January and February rainfall, while the most critical climate index is last February's Southern Oscillation Index, with its high values inhibit while low values promote January rainfall in Darwin. (3) The most significant feature for June rainfall in Perth is last June's solar radiation whose feature value is weakly negatively correlated with its attribution value. The study's significance lies in enhancing the accuracy of multi-step rainfall prediction and providing interpretability through identification of influential factors, which facilitates long-term planning of water resources and deeper understanding of complex weather systems.

1. Introduction

Rainfall is one of the most essential weather phenomena that impacts human activities in all scales, including urban constructions, crop irrigation, hydroelectric power management, and so forth (Ponnoprat, 2021). As a critical variable describing rainfall, monthly rainfall is not only employed to mark the beginning, duration, and end of the rainy season (Bagirov et al., 2017), but also often used as a significant indicator of hydrological disasters including floods and droughts (Mehdizadeh et al., 2018). Specifically, intensive rainfall often enhances the flood risk and leads to loss of life and severe damage to infrastructures (He et al., 2023), while lack of rainfall may result in extreme drought, destroying the regional hydro-climatological functions and biodiversity (Marengo et al., 2021). Therefore, an accurate prediction of monthly rainfall will be beneficial to not only water resource management but also hydrological disaster control.

In the past decades, many approaches have been developed by researchers to forecast rainfall. These predictive models can be generally classified into two categories: physical models and data-driven models. The physical models employ a series of geophysical laws and equations to describe the relevant atmospheric or oceanic processes that contribute to rainfall and hence give a rainfall estimation (Bagirov et al., 2017). However, due to the randomness and uncertainty of rainfall processes (Wang et al., 2022), such physical models are prone to large errors at the local scale (Khan et al., 2019). The data-driven models, by

* Corresponding author.
E-mail addresses: RENFEI001@e.ntu.edu.sg (R. He), zlm@hust.edu.cn (L. Zhang), Alvin.Chew@bentley.com (A.W.Z. Chew).

https://doi.org/10.1016/j.eswa.2023.121160
Received 8 March 2023; Received in revised form 28 July 2023; Accepted 7 August 2023
Available online 9 August 2023

Nomenclature

List of abbreviations

ABOM	Australian Bureau of Meteorology
AI	Artificial Intelligence
BLSTM-GRU	Bidirectional Long Short Term Memory with Gated Recurrent Unit
DeepLIFT	Deep Learning Important FeaTures
DMI	Dipole Mode Index
ENSO	El Niño–Southern Oscillation
GRU	Gated Recurrent Unit
IOD	Indian Ocean Dipole
JSD	Jensen-Shannon divergence
LSTM	Long Short Term Memory
LSTM-AM	Long Short Term Memory with Attention Mechanism
MAA	Mean Absolute Attribution
MAE	Mean Absolute Error
MSE	Mean Square Error
PDO	Pacific Decadal Oscillation
QBD	Quantile-Based Distance
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SHAP	SHapley Additive exPlanations
SILO	Scientific Information for Land Owners
SOI	Southern Oscillation Index

List of symbols

$a_{q,k}$	Normalized weight of h_k with respect to the query h'_q
c_q	Attention value of h'_q
$F(h'_q, h_k)$	Original weight of h_k with respect to the query h'_q
φ_i^{EG}	Expected gradient of the ith dimension
h_{n-p}	Hidden state of the encoder module at time $t-p$
h'_q	Hidden state of the decoder module at time $t+q$
n	Number of past months used as model input
N	Number of future months to predict
N_S	Number of samples
$P_{\min}$	Minimum of the N predictions for one specific month
$P_{\max}$	Maximum of the N predictions for one specific month
P_{Q1}	The first quantile of the N predictions for one specific month
P_{Q3}	The third quantile of the N predictions for one specific month
p_1	Proportions of months satisfying $P_{Q1} < T < P_{Q3}$
p_2	Proportions of months satisfying $P_{\min} < T < P_{Q1}$ or $P_{Q3} < T < P_{\max}$
p_3	Proportions of months satisfying $T < P_{\min}$ or $T > P_{\max}$
r_{t-p}	Output of the reset gate of the encoder module at time $t-p$
R^2	Coefficient of determination
σ	Sigmoid activation function
T	True rainfall of one specific month
v_{atn}	Trained vector of the attention mechanism module
W_r	Trained weight matrix of the reset gate of the encoder module
W_z	Trained weight matrix of the update gate of the encoder module
W_{atn}	Trained matrix of the attention mechanism module
x	input at hand to calculate the expected gradient
x'	Baseline input of the expected gradient
$\mathscr{X}_{t-p}$	Input feature arrays (consisting of weather and climate features) at time $t-p$
X1	Monthly rainfall [mm]
X2	Daily average maximum temperature [℃]
X3	Daily average minimum temperature [℃]
X4	Daily average solar radiation [MJ/m^2]
X5	Daily average relative humidity at the time of maximum temperature [%]
X6	Daily average relative humidity at the time of minimum temperature [%]
X7	Southern Oscillation Index
X8	Nino 1 + 2 [K]
X9	Nino 3 [K]
X10	Nino 3.4 [K]
X11	Nino 4 [K]
X12	Dipole Mode Index [℃]
X13	Pacific Decadal Oscillation
z_{t-p}	Output of the update gate of the encoder module at time $t-p$

contrast, rely on historical meteorological data to predict future rainfall through statistical or artificial intelligence (AI) approaches, which have been proved to have superior performances to the physical models in certain cases (Abbot & Marohasy, 2012). In recent years, with the development of deep learning techniques, a number of neural network models have been developed and utilized in this area (Venkatachalam et al., 2023). For example, Chhetri et al. (2020) combined the bidirectional Long Short Term Memory (LSTM) network with Gated Recurrent Unit (GRU), proposing a BLSTM-GRU-based deep learning model to predict monthly rainfall. Tao et al. (2021) developed a multiscale LSTM network with an attention mechanism (MLSTM-AM), which showed a satisfying predictive capability on monthly rainfall in 129 stations over Yangtze River basin. Fahad et al. (2023) presented an optimized GRU network to forecast monthly rainfall in Pakistan based on 30 years of climate data from 1991 to 2020.

Although demonstrating the superiority of deep learning in rainfall prediction, most of these studies could only realize a single-step prediction, i.e., giving the predicted rainfall for the next month. In practice, however, people often need a multi-step monthly rainfall prediction, which offers several practical advantages compared with a single-step prediction. On the one hand, multi-step monthly rainfall predictions allow for longer-term planning and decision-making in fields related to rainfall such as agriculture, water resource management, and urban planning. With a reliable multi-step-ahead prediction of monthly rainfall, farmers can schedule their planting seasons based on future rainfall conditions, while water resource managers can devise more effective water supply strategies. On the other hand, multi-step monthly rainfall predictions play a crucial role in rainfall-related disaster warning and risk management. By gaining insights into rainfall patterns for the upcoming months, proactive measures can be taken to mitigate potential risks of floods, droughts, or other rainfall-induced hazards. This aids in improving emergency preparedness and thus reducing the impact of disasters on people. Admittedly, multi-step monthly rainfall prediction is more challenging than single-step prediction due to the complexity and uncertainty of weather systems, which demands the predictive models to fully capture long-term dependencies, trends, and periodicities in rainfall time series. Previous research has also indicated a potential loss of accuracy in multi-step prediction compared to single-step prediction (Ben Taieb et al., 2012; Rodriguez et al., 2020; W. Zhang et al., 2021). Nonetheless, motivated by the aforementioned advantages, a reliable multi-step monthly rainfall prediction is still necessary to support the long-term planning and disaster risk management. However, there are currently few studies that employ deep learning to investigate multi-step prediction of monthly rainfall. Therefore, this study leverages

an encoder-decoder model with an attention mechanism, which consists of a GRU-based encoder module, an attention mechanism module, and a GRU-based decoder module, to conduct the multi-step monthly rainfall prediction, attempting to make improvements in this area.

Deep learning has demonstrated remarkable capabilities in rainfall prediction. However, a significant challenge associated with deep neural networks is their lack of interpretability, often referred to as the "Black Box" problem. Despite achieving satisfactory rainfall prediction results, deep learning models operate as complex systems whose internal mechanisms remain unknown to model designers. Consequently, it becomes difficult to explain why these models make specific predictions and identify the factors influencing their outputs. Fortunately, recent advances in research have introduced various feature attribution methods, such as DeepLIFT (Shrikumar et al., 2017), Deep SHAP (Lundberg & Lee, 2017), expected gradient (Erion et al., 2021), etc., which allow for analyzing the feature importance in a deep learning model. These methods have now been used in many areas for model explanation and knowledge discovery (Chew & Zhang, 2022; Shojaei et al., 2023). However, the application of these feature attribution methods to rainfall prediction remains largely unexplored. In fact, the atmospheric system comprises a multitude of variables that interact with one another, influencing rainfall patterns. Relying solely on physical models to discern the underlying relationships between rainfall and other factors may introduce biases or limitations. In this case, feature attribution methods could serve as a powerful supplement, which is beneficial to uncovering the effects of other meteorological features on rainfall in a data-driven manner and thus improving people's insight into the complex weather systems. Therefore, this study incorporates an advanced feature attribution method termed expected gradient as a post-predictive module for the encoder-decoder, aiming to achieve better model interpretability and discover influential factors of monthly rainfall.

The main research purposes of this study are to (1) make accurate multi-step monthly rainfall predictions and (2) provide reasonable explanations for the prediction results and hence extract the main factors that affect rainfall. To achieve these two purposes, this study proposes an explainable deep learning approach integrating four consecutive modules, i.e., GRU-based encoder module, attention mechanism module, GRU-based decoder module, and expected-gradient-based explanation module. Among the four modules, the first three constitute an encoder-decoder with an attention mechanism for multi-step monthly rainfall prediction. The explanation module employs the expected gradient to calculate the attribution value of each input feature to the output target, finding out the primary factors affecting monthly rainfall. The applicability and validity of the proposed approach are well verified by the case study in Darwin and Perth in Australia. The novelty of this study lies in two aspects. Firstly, from a methodological perspective, this study proposes an integrated approach that combines the encoder-decoder with an attention mechanism and expected gradient, achieving feature attribution analysis for multi-step time series prediction. Previous studies on time series prediction and interpretation mainly focused on feature importance analysis for single output prediction (Su & Iliadou, 2022; Yang et al., 2022). However, in this study, feature attribution values are calculated for monthly rainfall in each predicted future time step in the multi-step prediction results. Additionally, to evaluate the performance of the multi-step predictive model, this study constructs a novel metric named quantile-based distance (QBD) using the Jensen-Shannon divergence (JSD). Secondly, from an application perspective, multi-step prediction of monthly rainfall is a novel application area for explainable deep learning. Advanced deep learning techniques enable the multi-step prediction of monthly rainfall, while feature attribution methods enhance the interpretability of deep learning models, facilitating the analysis of the impact mechanism of other variables in the weather system on rainfall. Applying explainable deep learning to the multi-step prediction of monthly rainfall is a novel and valuable attempt with broad application potential.

The rest of the paper is organized below. Section 2 presents the detailed methodology of the explainable deep learning approach. Section 3 introduces the studied area and the basic process of data preparation. Section 4 shows the prediction performance of the encoder-decoder with an attention mechanism through comparisons with some baseline methods on several evaluation metrics. Section 5 implements the feature attribution analysis for the wettest calendar months of the studied cities. Section 6 draws some concluding remarks and proposes some future works.

2. Methodology

This section aims to present the details of the methodology proposed in this study, which consists of four consecutive modules, namely GRU-based encoder module, attention mechanism module, GRU-based decoder module, and expected-gradient-based explanation module. The general framework is illustrated in Fig. 1. Firstly, a series of features (including both weather and climate features) in chronological order is input into the GRU-based encoder module to obtain the corresponding hidden states. Secondly, based on the hidden states of the encoder and decoder modules, the attention mechanism module is exploited to compute a series of attention values. Thirdly, in the decoder module, each attention value is concatenated with the predicted rainfall value at the last time step, and then input into the corresponding GRU cell to acquire the current hidden state. Each hidden state should further pass through a fully connected layer (FC layer in Fig. 1) to generate the corresponding predicted rainfall value. In particular, the decoder module consists of several GRU cells, which means it can predict the rainfall values in multiple continuous months in the future. The first three modules constitute an encoder-decoder with an attention mechanism, which is a novel sequence-to-sequence deep learning model. Finally, for better interpretability of the model, one novel feature attribution method, namely expected gradient, is adopted to quantify the importance of each input feature to the output target. The expected gradient values are computed through the gradient explainer in the SHAP package. Through in-depth analysis of the calculation results, the effects of the past weather factors and climate indices on the future monthly rainfall are fully revealed, and the main factors that affect monthly rainfall can be correspondingly identified. The details of the four modules are elaborated in the following sections.

2.1. GRU-based encoder module

In this study, the input of the encoder module is considered a sequence of previous weather or climate features in chronological order. Under this circumstance, recurrent neural network (RNN) is an applicable and effective model that can capture the dynamics of input sequences via cycles in the network of nodes (Lipton et al., 2015). However, traditional RNN has the problem of gradient vanishing or explosion which is caused by the backpropagation algorithm and the long-term dependency; hence, it is difficult for traditional RNN to capture the dependency of a large time step in a time series in practice (He et al., 2022; Lei, 2021). To overcome this drawback, Cho et al. (2014) proposed GRU as a variant of RNN, which can not only overcome the gradient vanishing or explosion problem but also prevent overfitting by using fewer training parameters. In recent studies, GRU has been widely utilized in time series prediction in many areas and shows high accuracy and good robustness (Fu et al., 2020; Shen et al., 2018; Y.-G. Zhang et al., 2021). Therefore, in this study, GRU is employed as the basic cell to construct the encoder module.

As shown in Fig. 2, a series of GRU cells are connected in sequence to form the encoder module. Assume that the features in a total of n previous months are used as the model input; thus, one layer of n GRU cells should be constructed to encode the input information. The input feature arrays, i.e., $\mathscr{X}_{t-n+1}$, $\mathscr{X}_{t-n+2}$, …, $\mathscr{X}_{t-1}$, $\mathscr{X}_t$, would consist of monthly weather factors and climate indices at time $t-n+1$, $t-n+2$,

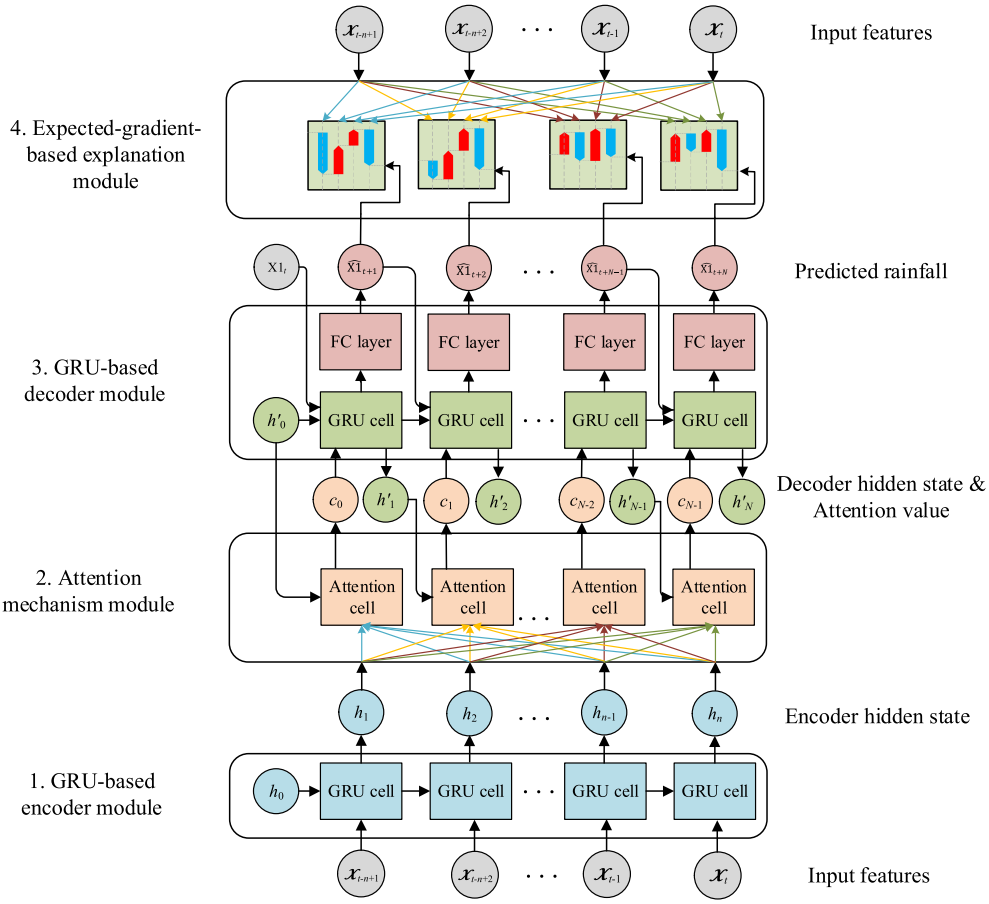


Fig. 1. General framework of the explainable deep learning approach comprising GRU-based encoder module, attention mechanism module, GRU-based decoder module, and expected-gradient-based explanation module.

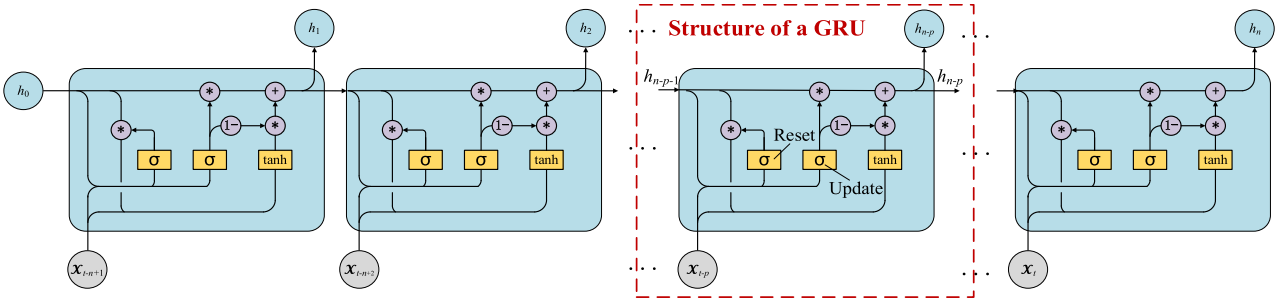


Fig. 2. Detailed structure of the GRU-based encoder module (the structure of a GRU is shown in the red frame).

…, $t-1$, t, respectively. A typical GRU cell with a reset gate and an update gate can be seen in the red frame in Fig. 2. As illustrated, at time $t-p$, the input of the current GRU and the previous hidden state are denoted as $\mathscr{X}_{t-p}$ and h_{n-p-1}, respectively. The input $\mathscr{X}_{t-p}$ will pass through both the reset gate whose output r_{t-p} is computed as Eq. (1) and the update gate whose output z_{t-p} is computed as Eq. (2). Then, the current hidden state h_{n-p}, which is also the output of the current GRU, can be computed by Eq. (3), in which $\widetilde{h}_{n-p}$ is expressed as Eq. (4).

$$r_{t-p} = \sigma\left(W_r \cdot \left[h_{n-p-1}, \mathscr{X}_{t-p}\right]\right) \tag{1}$$

$$z_{t-p} = \sigma\left(W_z \cdot \left[h_{n-p-1}, \mathscr{X}_{t-p}\right]\right) \tag{2}$$

$$h_{n-p} = z_{t-p} * h_{n-p-1} + \left(1 - z_{t-p}\right) * \widetilde{h}_{n-p} \tag{3}$$

$$\widetilde{h}_{n-p} = \tanh\left(W \cdot \left[r_{t-p} * h_{n-p-1}, \mathscr{X}_{t-p}\right]\right) \tag{4}$$

where the sign * denotes the element-wise product between matrices; W_r, W_z, and W are the weight matrices of the reset gate, update gate and $\widetilde{h}_{n-p}$, respectively; σ is the sigmoid activation function $\sigma = \frac{1}{1+e^{-x}}$ that maps any real number to (0,1).

In a GRU, the reset gate could allow the hidden state to drop any information that is irrelevant later in the future, and the update gate controls how much information from the previous hidden state will carry over to the current hidden state. Specifically, with sigmoid σ as the activation function, the output values of the reset gate and update gate, i.e., r_{t-p} and z_{t-p}, are both from 0 to 1. Based on Eq. (4), the reset gate can drop more information of the previous hidden state, h_{n-p-1}, if it is irrelevant to $\mathscr{X}_{t-p}$ and thus r_{t-p} is closer to 0; otherwise, the reset gate

will drop less information. For the update gate, according to Eq. (3), the current hidden state, h_{n-p}, will carry less information from the previous hidden state if z_{t-p} is closer to 0, while carry more information if z_{t-p} is closer to 1. Through multiple activation functions and various pathways of gradient backpropagation, the GRU can greatly avoid the problem of gradient vanishing or explosion and hence well capture the long-term dependency in the monthly meteorological time series.

As shown in Fig. 2, with each GRU cell taking one feature vector and the previous hidden state to implement the calculations expressed by Eqs. (1)–(4), the output of the encoder module is a series of hidden states, i.e., $h_1, h_2, \ldots, h_{n-1}, h_n$, each of which will contain the information of all the input features at all previous time steps. In particular, the last output h_n theoretically carries all the information from $\mathscr{X}_{t-n+1}$ to $\mathscr{X}_t$. These output hidden states will then serve as the input of the following attention mechanism module.

2.2. Attention mechanism module

The encoder output, $h_1, h_2, \ldots, h_{n-1}, h_n$, after being obtained through a series of GRU cells, should be further utilized to make the multi-step prediction, i.e., simultaneously generate multiple output values. As mentioned above, all the input information is compressed into h_n; hence, one possible approach is to decode only h_n into a series of predicted values through neural networks (Cho et al., 2014). However, this method tends to make it difficult for the model to cope with the case where there is a long input sequence (i.e., n is large) because part of the information may get lost when compressed into a fixed-length vector. In order to solve this problem, an attention mechanism (Bahdanau et al., 2014) is added to the encoder-decoder model, which enables the model to focus on and fully learn that important information among all input features. Specifically, the attention mechanism computes the weight of each "key" from the encoder for the current "query" from the decoder, so that it can get the relevance of each input feature to the current output prediction. Therefore, when the model generates an output value, it will search for certain positions in the input sequence where the most relevant information is concentrated, thus contributing to a good multi-step prediction accuracy. Due to its outstanding advantages, the attention mechanism has been incorporated into many end-to-end models such as Transformer and used for time series prediction (Li et al., 2019; Lim et al., 2021).

The detailed structure of the attention cell is shown in the red frame in Fig. 3. As illustrated, the input of the attention cell consists of two parts: one is the encoder output, $h_1, h_2, \ldots, h_{n-1}, h_n$, which are referred to as "key"; the other is the last hidden state of the decoder, h'_q, which is referred to as "query" (Vaswani et al., 2017) and will be elaborated in Section 2.3. Firstly, each key h_k is concatenated to the query h'_q, and the operation written as Eq. (5) is then performed to obtain $F(h'_q, h_k)$ which represents the original weight of the key h_k. Afterward, all the original weights are normalized by a softmax layer to generate the normalized weight of each key, i.e., $a_{q,k}$. The softmax layer implements the computation expressed by Eq. (6), from which it is seen that the summation of all normalized weights is equal to 1. Finally, based on the normalized weights, the attention value c_q is calculated by Eq. (7), which is also the output of the attention cell.

$$F\left(h'_q, h_k\right) = v_{atn}{}^T \cdot \tanh\left(W_{atn} \cdot \left[h'_q, h_k\right]\right) \quad k = 1, 2, \ldots, n \tag{5}$$

$$a_{q,k} = \frac{\exp F\left(h'_q, h_k\right)}{\sum_{i=1}^{n} \exp F\left(h'_q, h_i\right)} \quad k = 1, 2, \cdots, n \tag{6}$$

$$c_q = \sum_{k=1}^{n} a_{q,k} h_k \tag{7}$$

where $h'_q \in R^{128\times1}$ is the query vector; $h_k \in R^{128\times1}$ is the kth key vector; $W_{atn} \in R^{128\times256}$ and $v_{atn} \in R^{128\times1}$ are the matrix and vector that should be trained; $a_{q,k}$ is the normalized weight of h_k with respect to the query h'_q; $c_q \in R^{128\times1}$ is the attention value.

According to Eqs. (5)—(7), the attention value c_q is actually a weighted average of all the encoder output, i.e., h_k ($k = 1, 2, \ldots, n$). Thus, those h_k with higher weight will contribute more to the attention value. Set aside the weight normalization process by the softmax layer, it can be found from Eq. (5) that the weight essentially reflects the correlation between the key h_k and the query h'_q. That is to say, if containing the information that is more relevant to the current h'_q, some certain h_k will have greater weight during the process of model training, hence accounting for a larger proportion of the attention value c_q. It is this ingenious mechanism that makes the model selectively learn the information from the encoder.

Reusing the attention cell for N times (N is the length of the monthly rainfall sequence to be predicted), we structure the attention mechanism module, which is shown in Fig. 3. Initially, the attention cell brings in the initial hidden state of the decoder, h'_0, with the encoder output h_1, h_2, …, h_{n-1}, h_n, and then calculate the output attention value c_0. Afterward, c_0 will be processed by the GRU cell in the decoder module (elaborated in Section 2.3), and the next hidden state h'_1 will be generated. The above procedures will continue until the last attention value c_{N-1} is computed and input into the decoder module. Note that all the attention cells illustrated in Fig. 3 share the same training parameters W_{atn} and v_{atn}. The interaction between the attention mechanism model and decoder module enables each decoder output to focus on and fully utilize the most relevant information among all the encoder output, which is conducive to a multi-step prediction.

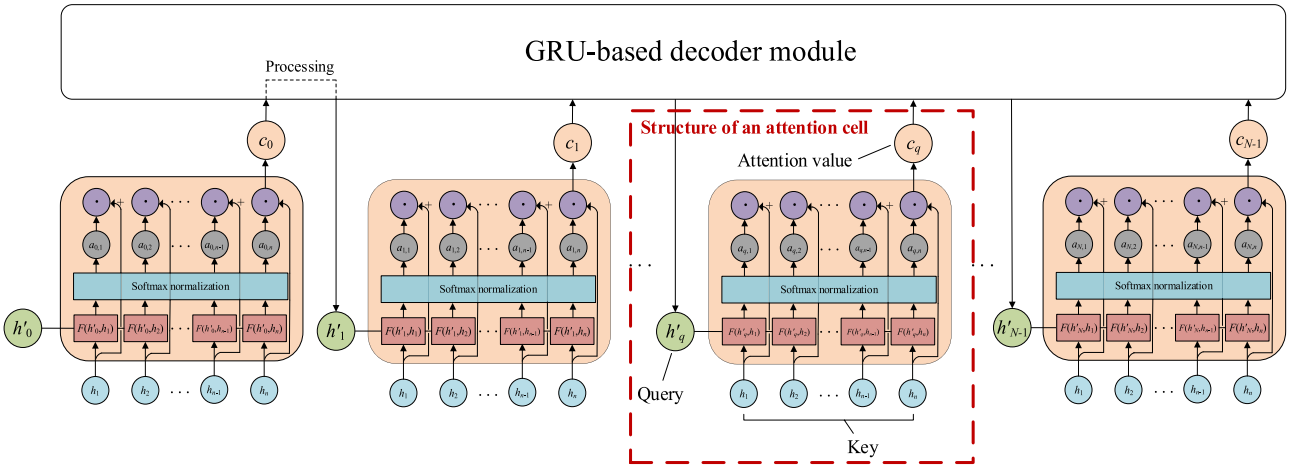


Fig. 3. Detailed structure of the attention mechanism module (the structure of an attention cell is shown in the red frame).

2.3. GRU-based decoder module

Fig. 4 shows the detailed structure of the GRU-based decoder module, whose basic unit is given in the red frame. Although the decoder module also employs GRU as the basic cell for model output, there are two main differences between the encoder and decoder modules, which are elaborated below:

Firstly, in the decoder module, not only the current attention value but also the last prediction is taken as the input of one GRU cell. It is seen from Fig. 4 that the input of each GRU cell consists of two elements: one is the current attention value c_q obtained from the attention mechanism module; the other is the predicted monthly rainfall at the last time step, which is denoted by $\widehat{\mathrm{X1}}_{t+q}$. The two elements are concatenated as one vector and then input into the GRU cell, in which the computations between the input vector and the last hidden state h'_q are conducted based on Eqs. (1)–(4). The output of the GRU cell is the current hidden state h'_{q+1}. Through the combination of c_q and $\widehat{\mathrm{X1}}_{t+q}$, the information from both the input meteorological features and output prediction can be fused into h'_{q+1}. This will be potentially beneficial to monthly rainfall prediction at the current time step. On the one hand, the attention value c_q includes the encoded metrological information that is the most relevant to the current prediction. On the other hand, the last predicted monthly rainfall, $\widehat{\mathrm{X1}}_{t+q}$, is also considered instructive to the current prediction since the rainfall in two consecutive months is generally close to each other.

Secondly, a fully connected layer, which is used to reduce the dimension of the hidden state h'_{q+1}, is added following the GRU layer. In the end, Rectified Linear Unit (ReLU) (Glorot et al., 2011) is adopted as the activation function. The reasons why ReLU is used are: (1) It can realize an effective gradient backpropagation by a simple calculation process with less computational cost; (2) It can convert the predicted value to a non-negative number, which is consistent with the monthly rainfall value.

Similar to the attention cell, the basic unit of the decoder module will be repeatedly used for N times to generate the predicted monthly rainfall in the future N months (see Fig. 4). In the beginning, there is no predicted rainfall value; hence, the ground truth of monthly rainfall at time $t-1$ is utilized by the first GRU. The initial hidden state h'_0 is input into the first GRU; simultaneously, it will enter the attention mechanism module, and hence the first attention value c_0 can be obtained through Eqs. (5)–(7). The calculation process within the first GRU yields the hidden state h'_1, which will be used for (1) the dimension reduction by the fully connected layer to acquire the prediction at time $t+1$, i.e., $\widehat{\mathrm{X1}}_{t+1}$; (2) the calculation of the next attention value c_1 by the attention mechanism module; (3) the operation in the next GRU. The above procedure will repeat for N times until the predicted monthly rainfall values from time $t+1$ to $t+N$ are all generated.

Till now, an encoder-decoder with an attention mechanism, which is a novel sequence-to-sequence machine learning model, has been constructed. In the present model, each predicted rainfall value will fully utilize the information in every regard, which contributes to its excellent performance in multi-step rainfall prediction. Specifically, when generating the current prediction, the encoder-decoder will combine the last hidden state of the decoder, the last predicted rainfall value, and the current attention value. Among the three sources, the last hidden state passes the information across sequence steps of the decoder; the last predicted rainfall value serves as "context" which could guide the current prediction; the attention value selectively integrates all the encoder output that carries past weather and climate information. Through the encoder-decoder with an attention mechanism, the non-linear temporal relationship between past weather and climate features (from $\mathscr{X}_{t-n+1}$ to $\mathscr{X}_t$) and future monthly rainfall values (from $\mathrm{X1}_{t+1}$ to $\mathrm{X1}_{t+N}$) can be eventually established.

2.4. Expected-gradient-based explanation module

Following the encoder-decoder with an attention mechanism, an explanation module is further constructed as a post-predictive tool to conduct the feature attribution analysis (Baehrens et al., 2010; Shrikumar et al., 2017), that is, to identify which input features are the most responsible for the output predictions. Through the explanation module, the primary factors affecting monthly rainfall can be identified, and the internal connecting links between the past weather factors or climate indices and the future monthly rainfall can be discovered.

In this study, one novel feature attribution method, expected gradient (Erion et al., 2021), is adopted to quantify the importance of each input feature to the output prediction. In the expected gradient method, an integrated gradient (Sundararajan et al., 2017) needs to be calculated first. Suppose that a function $F: R^{n\times 1}\to[0,1]$ represents a deep network. Let $x\in R^{n\times 1}$ be the input at hand and another $x'\in R^{n\times 1}$ be the baseline input. Then, the integrated gradient is calculated by cumulating the gradients at all points along the path from the baseline x' to the input x. In other words, the integrated gradient is defined as the path integral of the gradient along the straight-line path from the baseline x' to the input x, written as Eq. (8):

$$IntegratedGradient_i\left(x,x'\right)=\left(x-x'\right)\cdot\int_{\alpha=0}^{1}\frac{\delta F(x'+\alpha(x-x'))}{\delta x_i}d\alpha \tag{8}$$

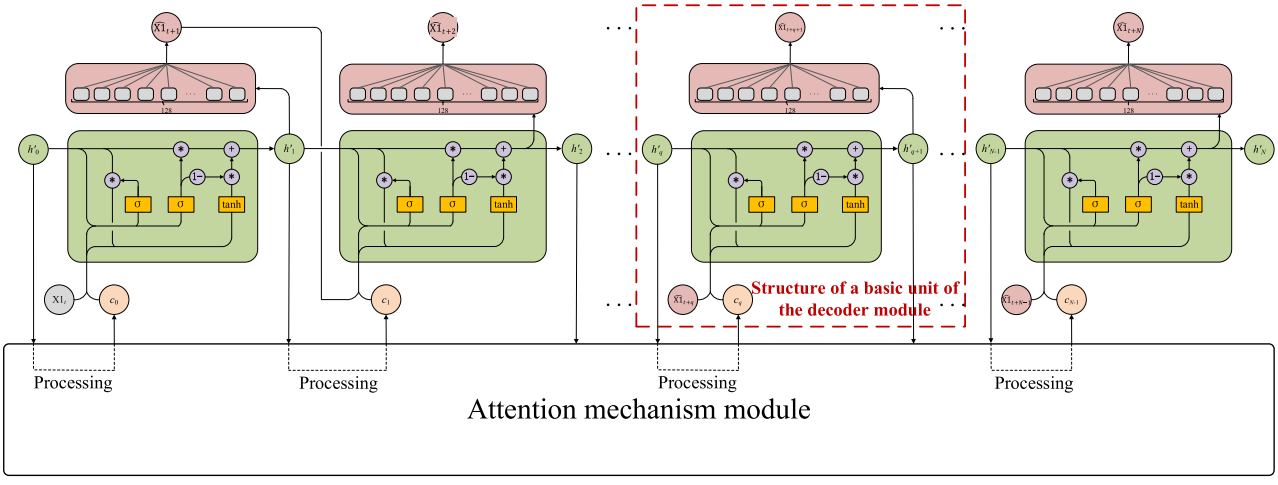


Fig. 4. Detailed structure of the GRU-based decoder module (the structure of a basic unit is shown in the red frame).

where x_i and $x_i^{'}$ are the values of the ith dimension of the input x and the baseline $x^{'}$, respectively.

Subsequently, the integrated gradient is computed over multiple baselines to generate the expected gradient. Specifically, the integral is reformulated as an expectation that is combined with baselines from the background dataset. The expected gradient of the ith dimension of the input x with respect to the background dataset is expressed as Eq. (9), which can be considered as the attribution value to measure the importance of the ith dimension of the input x:

$$\begin{aligned}&\varphi_i^{EG}(F,x,D)\\&=\int_{x'} IntegratedGradient_i\left(x,x'\right)p_D\left(x'\right)dx'\\&=\int_{x'}\left[(x_i-x_i')\cdot\int_{\alpha=0}^{1}\frac{\delta F(x'+\alpha(x-x'))}{\delta x_i}d\alpha\right]p_D\left(x'\right)dx'\\&=\mathop{\mathbb{E}}_{x'\sim D,\alpha\sim U(0.1)}\left[\left(x_i-x_i'\right)\cdot\frac{\delta F(x'+\alpha(x-x'))}{\delta x_i}\right]\end{aligned}\tag{9}$$

where $\varphi_i^{EG}(F,x,D)$ is the expected gradient of the ith dimension of the input x; F represents the deep network; D represents the distribution of the background dataset.

To perform a quick computation of Eq. (9), sampling is employed to collapse the multiple integral. The sampling-based approximation that numerically computes Eq. (9) is written as Eq. (10):

$$\widehat{\varphi}_i^{EG}(F,x,D)=\frac{1}{k}\sum_{j=1}^{k}\left(x_i-x_i^{'j}\right)\cdot\frac{\delta F(x^{'j}+\alpha^j(x-x^{'j}))}{\delta x_i}\tag{10}$$

where $\widehat{\varphi}_i^{EG}$ is the approximation of φ_i^{EG}; $x^{'j}$ is the jth sample from D; α^j is the jth sample from the uniform distribution between 0 and 1; k is the sample number.

Used as the feature attribution value, the expected gradient can quantify the importance of each input feature to the output prediction for all samples in the dataset. In this study, the expected gradient values are computed through the gradient explainer in the SHAP package (Lundberg & Lee, 2017). In the SHAP package, SmoothGrad (Smilkov et al., 2017) is available to add normally distributed noise with a certain standard deviation to the input during the expectation calculation, which can create smoother feature attributions that better capture correlations within the input features. Finally, through a careful investigation into the calculation results, some weather or climate features can be identified as the influential factors of monthly rainfall and some inherent laws of the complex weather system can be correspondingly revealed.

3. Studied area and data preparation

This section pertains to the introduction of the study area including Darwin and Perth and the elucidation of the procedure for data preparation. The climate features of both cities are introduced, and reasons for their selection are explained. Additionally, the dataset's 6 weather factors and 7 climate indices are delineated, and the process of sample generation is illustrated.

In this study, Darwin and Perth in Australia are taken as the studied area. Darwin is the capital city of the Northern Territory, Australia, located in the north of the Australian mainland. Darwin experiences a tropical savanna climate (Sturman & Tapper, 1996) with distinct wet and dry seasons. Associated with tropical cyclones and monsoon rains, the wet season in Darwin is between December and March, during which the majority of rainfall occurs. The dry season runs from about May to September, during which nearly every day is sunny with no precipitation (ABOM, 2022d). The second city of the case study, Perth, is the capital city of Western Australia, which is situated on the Indian Ocean coast in the southwest of Australia. It has a hot-summer Mediterranean climate (Sturman & Tapper, 1996) characterized by hot dry summer and cool wet winter. In Perth, during the winter running from May to September, significant rainfall occurs as a result of frontal systems moving across the region (Simmonds & Richter, 2000; SM, 2021). Fig. 5 shows the contour map of the average annual rainfall in Australia from 1981 to 2010 (ABOM, 2022b), on which the locations of the two cities are marked out. Overall, based on the statistics confirmed by Fig. 5, the average annual rainfall values in Darwin and Perth are 1811.4 mm and 725.0 mm, respectively (ABOM, 2022c).

The reasons why Darwin and Perth are selected for the present case study can be concluded as the two points below. Firstly, as the capitals of the respective states, both Darwin and Perth are relatively densely populated, which means that the possible floods induced by the extreme rainfall events in the wet seasons could cause considerable loss of life and property. Therefore, it is necessary to make accurate predictions on monthly rainfall in both cities for better water disaster warning and watershed management. Secondly, as mentioned above, the climate features of Darwin and Perth are entirely distinct. Using the two cities as the studied cases can better confirm the applicability and effectiveness of the proposed approach in forecasting monthly rainfall and identifying influential factors under diverse rainfall patterns.

To start with, preparing a comprehensive dataset is the fundamental step of the training and testing of a machine learning model. As verified by Bagirov et al. (2017), there exists a great dependency of monthly rainfall on other weather factors. Therefore, in this study, a total of 6 monthly weather factors regarding rainfall, temperature, solar radiation, and humidity are incorporated into the dataset. The selected 6 monthly weather factors are (1) monthly rainfall; (2) daily average maximum temperature; (3) daily average minimum temperature; (4) daily average solar radiation; (5) daily average relative humidity at the time of maximum temperature; (6) daily average relative humidity at the time of minimum temperature. The reasons for using the 6 weather factors are: (1) Monthly rainfall series is a time series with distinct autocorrelation, which means the rainfall in the current month is partially correlated with that in previous months (Papalaskaris et al., 2016). (2) The dependence between rainfall and temperature shows different patterns in different regions, which is confirmed by previous studies (Trenberth & Shea, 2005). (3) As the original driving force of water cycle, solar radiation is also adopted as a predictive indicator for monthly rainfall prediction in previous literature (Bagirov & Mahmood, 2018). (4) Relative humidity and rainfall are both weather factors regarding atmospheric water which may be potentially correlated with each other; hence, some studies on monthly rainfall prediction also incorporate relative humidity information as the model input (Chhetri et al., 2020; Gu et al., 2022). The original data of these factors in Darwin and Perth can be freely downloaded from the Scientific Information for Land Owners (SILO) [https://www.longpaddock.qld.gov.au/silo/], which is an enhanced climate database that includes the historical data of several meteorological variables in Australia from 1889 to the present year. In this study, a total of 100 years from 1921 to 2020 are taken as the studied period. It should be noted that SILO only provides daily weather data; hence, the summation of daily rainfall and the average of another five daily factors within a month should be computed to eventually acquire the monthly data.

In addition to the weather factors, some large-scale climate variations would also exert significant impacts on local monthly rainfall. Therefore, to obtain a reliable monthly rainfall prediction, it is also necessary to consider some relevant climate indices as the model input (Abbot & Marohasy, 2014; Haidar & Verma, 2018; Johny et al., 2022). In this study, the following 7 monthly climate indices are used as the model input: (1) Southern Oscillation Index (SOI); (2) Nino 1 + 2; (3) Nino 3; (4) Nino 3.4; (5) Nino 4; (6) Dipole Mode Index (DMI); (7) Pacific Decadal Oscillation (PDO). The reasons for choosing the 7 climate indices are: (1) SOI, Nino 1 + 2, Nino 3, and Nino 3.4 are five indices gauging El Niño–Southern Oscillation (ENSO) (ABOM, 2022a), which is a crucial climate phenomenon referring to an irregular periodic variation in winds and sea surface temperatures over the tropical eastern Pacific Ocean. ENSO could exert certain impacts on rainfall in Australia.

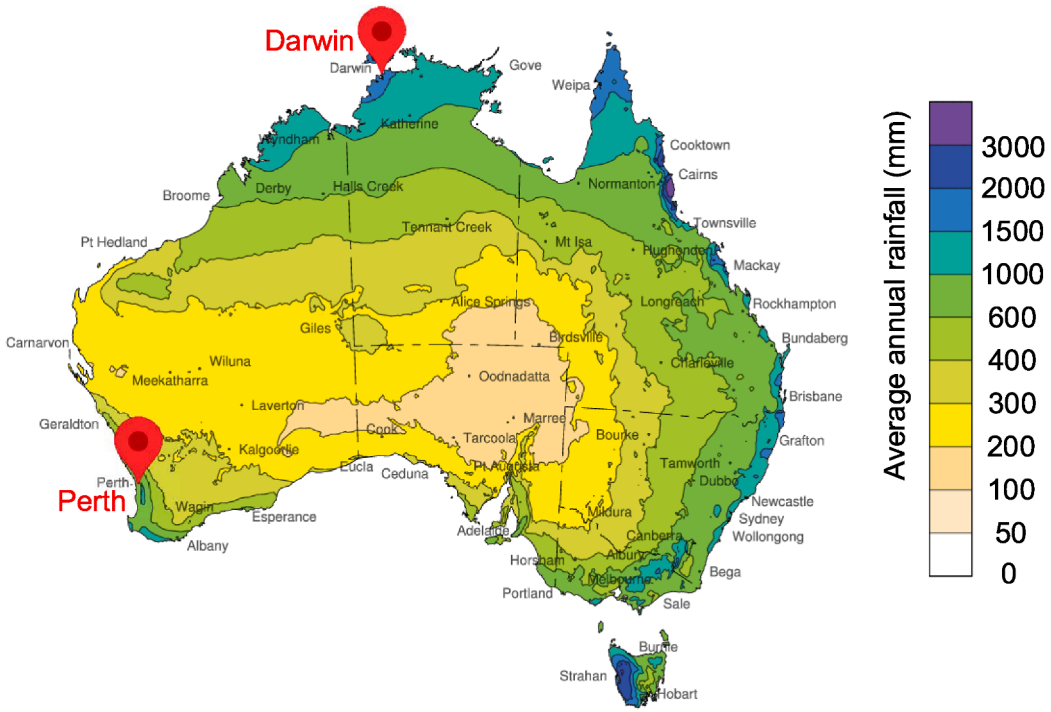


Fig. 5. Contour map of average annual rainfall in Australia from 1981 to 2010 (the locations of Darwin and Perth are marked out).

Specifically, during an ENSO warm event (also known as the El Niño event), drought occurs in regions of northeastern Australia, leading to lower than average winter-spring rainfall over eastern and northern Australia (ABOM, 2022e; Cai et al., 2001). By contrast, during an ENSO cool event (also known as the La Niña event), higher than average rainfall or even floods will occur across much of Australia, particularly inland eastern and northern regions (ABOM, 2022e). (2) DMI is the difference between sea surface temperature anomalies in two regions of the tropical Indian Ocean (ABOM, 2022a), which could measure the Indian Ocean Dipole (IOD) (ABOM, 2022a), a critical climate event influencing rainfall in some regions of Australia (Ashok et al., 2003; Pepler et al., 2014). (3) PDO is described as the dominant year-round pattern of monthly North Pacific Sea surface temperature variability (Newman et al., 2016), which could also potentially affect the rainfall patterns in Australia (Cai & van Rensch, 2012; Mantua & Hare, 2002). The monthly records of the abovementioned 7 climate indices are available at Royal Netherlands Meteorological Institute Climate Explorer [https://climexp.knmi.nl/selectindex.cgi?id=someone@somewhere]. Corresponding to the weather factors, all the 7 monthly climate indices in 1200 months from Jan 1921 to Dec 2020 are collected.

Table 1 gives the 6 monthly weather factors and 7 monthly climate indices included in the prepared dataset. The name, unit, variable symbol, mean, standard deviation, and interval of each factor or index are all listed. Note that the statistical characteristics of the 6 weather factors in the two studied cities are different and thus listed separately in Table 1.

Then, to appropriately feed the data to the encoder-decoder model, a sliding window is designed herein to generate the samples for model training and testing. As shown in Fig. 6, all the 13 monthly features in the 1200 months from Jan 1921 to Dec 2020 are represented by the gray blocks. When generating the first sample, a sliding window shown as the orange frame is placed at the left end of the dataset. The first $n \times 13$ features masked by the purple rectangle are taken as the input, while the following N rainfall values are used as the output. Then, the sliding window moves a single time step to the right to generate the second sample, and the input and output are taken the same way as the first sample. This procedure continues until the sliding window moves to the right end of the dataset, and a total of $N_S = 1201 - n - N$ samples can be eventually produced.

4. Model prediction performance

This section aims to demonstrate the prediction performance of the encoder-decoder with an attention mechanism. Firstly, three metrics are given to evaluate the model prediction performance. Secondly, five baseline methods, including BLSTM-GRU (Chhetri et al., 2020), LSTM-AM (Tao et al., 2021), GRU (Fahad et al., 2023), separate method, and recursive method, are introduced for comparison with the encoder-decoder with an attention mechanism. Lastly, the predicted results are shown and discussed in detail.

4.1. Evaluation metrics

In this study, three metrics are employed to evaluate the model prediction performance. The first two metrics are the coefficient of determination (R^2) and the mean absolute error (MAE), which are calculated by Eqs. (11) and (12), respectively. R^2 measures the goodness of fit in comparison with the average. MAE is the average magnitude of the errors, which shares the same unit with the prediction target, i.e., monthly rainfall. Generally, the presence of higher R^2 closer to 1 and smaller MAE closer to 0 indicates a more precise model.

$$R^2 = 1 - \frac{\sum_{j=1}^{N_s}\sum_{i=1}^{N}\left(\widehat{\mathrm{X1}}_{t+i,j} - \mathrm{X1}_{t+i,j}\right)^2}{\sum_{j=1}^{N_s}\sum_{i=1}^{N}\left(\overline{\mathrm{X1}} - \mathrm{X1}_{t+i,j}\right)^2} \tag{11}$$

$$MAE = \frac{1}{N_s \cdot N}\sum_{j=1}^{N_s}\sum_{i=1}^{N}\left|\widehat{\mathrm{X1}}_{t+i,j} - \mathrm{X1}_{t+i,j}\right| \tag{12}$$

where $\widehat{\mathrm{X1}}_{t+i,j}$ and $\mathrm{X1}_{t+i,j}$ are the predicted and true rainfall values at time $t+i$ of the jth sample; N_S is the number of samples in the whole dataset; N is the length of the monthly rainfall sequence that needs to be predicted; $\overline{\mathrm{X1}}$ is the mean rainfall value of all $N_s \cdot N$ samples.

Besides the above two metrics, a quantile-based distance (QBD) is proposed to measure the model's multi-step prediction performance.

Table 1
Statistical characteristics of the 6 weather factors and 7 climate indices in the prepared dataset.

Weather factor/Climate index	Unit	Symbol	Mean	Standard Deviation	[Min, Max]
Monthly rainfall	mm	X1	Darwin: 139.012 Perth: 68.689	Darwin: 178.544 Perth: 70.855	Darwin: [0, 1110.2] Perth: [0, 476.1]
Daily average maximum temperature	℃	X2	Darwin: 32.112 Perth: 23.806	Darwin: 1.181 Perth: 4.715	Darwin: [28.653, 35.261] Perth: [15.965, 35.089]
Daily average minimum temperature	℃	X3	Darwin: 23.173 Perth: 13.348	Darwin: 2.321 Perth: 3.524	Darwin: [15.890, 27.297] Perth: [5.623, 22.739]
Daily average solar radiation	MJ/m^2	X4	Darwin: 19.895 Perth: 18.417	Darwin: 2.499 Perth: 6.620	Darwin: [12.014, 26.290] Perth: [6.767,30.3]
Daily average relative humidity at the time of maximum temperature	%	X5	Darwin: 53.440 Perth: 46.491	Darwin: 10.143 Perth: 8.114	Darwin: [26.852, 74.743] Perth: [28.819, 66.461]
Daily average relative humidity at the time of minimum temperature	%	X6	Darwin: 87.448 Perth: 84.083	Darwin: 8.615 Perth: 9.170	Darwin: [56.417, 99.890] Perth: [60.479, 98.967]
Southern Oscillation Index (SOI)	/	X7	0.056	9.902	[-33.3, 31.6]
Nino 1 + 2	K	X8	−0.149	0.925	[-2.530, 4.379]
Nino 3	K	X9	−0.078	0.797	[-2.098, 3.237]
Nino 3.4	K	X10	0.005	0.771	[-2.159, 2.647]
Nino 4	K	X11	0.034	0.578	[-1.769, 1.749]
Dipole Mode Index (DMI)	℃	X12	0.072	0.334	[-1.193, 1.535]
Pacific Decadal Oscillation (PDO)	/	X13	0.043	1.084	[-3.60, 3.51]

Detailed descriptions of QBD are given below. In this study, according to the sample generation process shown in Fig. 6, the rainfall values in the last N_S months will be predicted N times. Assume that $n = N = 12$ and take Jan 1923 as an example. As shown in Fig. 7, when the 12 months from Feb 1921 to Jan 1922 are used as the model input, the corresponding output is the 12 months from Feb 1922 to Jan 1923. In this sample, the rainfall in Jan 1923 is predicted with a lead month equal to 12 (note: the lead month is defined as the time interval between the last input month and a certain output month). In the next sample, the rainfall in Jan 1923 is predicted again with a lead month equal to 11. Analogously, the rainfall in Jan 1923 will be predicted with a lead month of 10, 9, …, and 1 in the following samples. Therefore, suppose $P_{\min}$, $P_{\max}$, P_{Q1}, and P_{Q3} are the minimum, maximum, first quantile, and third quantile of the N predictions for one specific month, there may be three results comparing the true rainfall (denoted as T) with the corresponding $P_{\min}$, $P_{\max}$, P_{Q1}, and P_{Q3}: (1) Result 1: $P_{Q1} < T < P_{Q3}$; (2) Result 2: $P_{\min} < T < P_{Q1}$ or $P_{Q3} < T < P_{\max}$; (3) Result 3: $T < P_{\min}$ or $T > P_{\max}$. For the N_S months with N predictions, suppose p_1, p_2, and p_3 are the proportions of months with Results 1, 2, and 3, respectively ($p_1 + p_2 + p_3 = 1$). It is expected that a large proportion of months could have true rainfall falling within the interval of $[P_{Q1}, P_{Q3}]$, while few months have true rainfall higher than $P_{\max}$ or lower than $P_{\min}$, that is, p_1 is large while p_3 is small, in which case the range where the true rainfall falls can be inferred. Hence, given the array $p = (p_1, p_2, p_3)$, the best scenario is $p_{best} = \left(p_{best,1}, p_{best,2}, p_{best,3}\right) = (1,0,0)$ while the worst is $p_{worst} = \left(p_{worst,1}, p_{worst,2}, p_{worst,3}\right) = (0, 0, 1)$. As mentioned, it is expected that the obtained array p is closer to p_{best} while farther from p_{worst}. Because $p_1 + p_2 + p_3 = 1$, p can be treated as a probability mass function, and thus the Jensen-Shannon divergence (JSD) can be introduced to measure the distance between p and p_{best} or p_{worst}. Note that a larger JSD means a greater distance of two arrays. Based on JSD, QBD can be further constructed, written as Eq. (13). It can be seen that QBD is a normalized value varying from 0 to 1, and a larger QBD closer to 1 implies a better prediction made by the model.

$$QBD = \frac{JSD(p||p_{worst})}{JSD(p||p_{best}) + JSD(p||p_{worst})} \tag{13}$$

where $JSD(p||p_{best})$ is the JSD between p and p_{best} calculated by $JSD(p||p_{best}) = \frac{1}{2}\sum_{i=1}^{3} p_i \log \frac{p_i}{\frac{p_i + p_{best,i}}{2}} + \frac{1}{2}\sum_{i=1}^{3} p_{best,i} \log \frac{p_{best,i}}{\frac{p_i + p_{best,i}}{2}}$; $JSD(p||p_{worst})$ is the JSD between p and p_{worst} calculated by $JSD(p||p_{worst}) = \frac{1}{2}\sum_{i=1}^{3} p_i \log \frac{p_i}{\frac{p_i + p_{worst,i}}{2}} + \frac{1}{2}\sum_{i=1}^{3} p_{worst,i} \log \frac{p_{worst,i}}{\frac{p_i + p_{worst,i}}{2}}$.

4.2. Baseline methods

To better demonstrate the predictive capability of the encoder-decoder with an attention mechanism, this study applies five baseline methods to the same dataset for multi-step monthly rainfall prediction and compares the results with the results acquired by the encoder-decoder with an attention mechanism. The first three baselines come from some recently developed monthly rainfall prediction methods, including BLSTM-GRU (Chhetri et al., 2020), LSTM-AM (Tao et al., 2021), and GRU (Fahad et al., 2023). Besides, multi-step monthly rainfall prediction can also be carried out based on single-step prediction in two ways: separate method and recursive method. To specify, the separate method trains a series of single-output models to predict rainfall in the future several months separately, while the recursive method iterates a single-step predictive model forward for N times to generate the predicted rainfall in the next N months. Descriptions of the five baseline methods are given below:

(1) BLSTM-GRU. The architecture of the BLSTM-GRU model is like an encoder-decoder without the attention mechanism. Specifically, a series of weather and climate features are first fed into a bidirectional LSTM layer with 14 neurons, which can read data in both forward and backward directions and creates an appropriate encoding. Subsequently, the output of the LSTM layer will enter a GRU layer with 7 neurons and then a fully connected layer to be decoded as the predicted value. Note that batch normalization is performed between the BLSTM and GRU layers as well as between the GRU and fully connected layers. In the original paper, BLSTM-GRU is used for a single-step rainfall prediction; hence, the final output layer has only one neuron. In this study, its number of final output neurons is modified to fit for a multi-step prediction.

(2) LSTM-AM. LSTM-AM is, in essence, the combination of an attention mechanism module and an LSTM-based decoder module. The weather and climate features are directly input into the attention mechanism module to generate the attention values, which are then used as the input of the LSTM layer and decoded as the predicted rainfall.

(3) GRU. A single GRU network has also been used for monthly rainfall prediction (Fahad et al., 2023). In this study, the GRU network used as the baseline incorporates two GRU layers, as well as a fully connected layer following the second GRU reduce the output dimension and generate the final prediction. The GRU baseline can also be

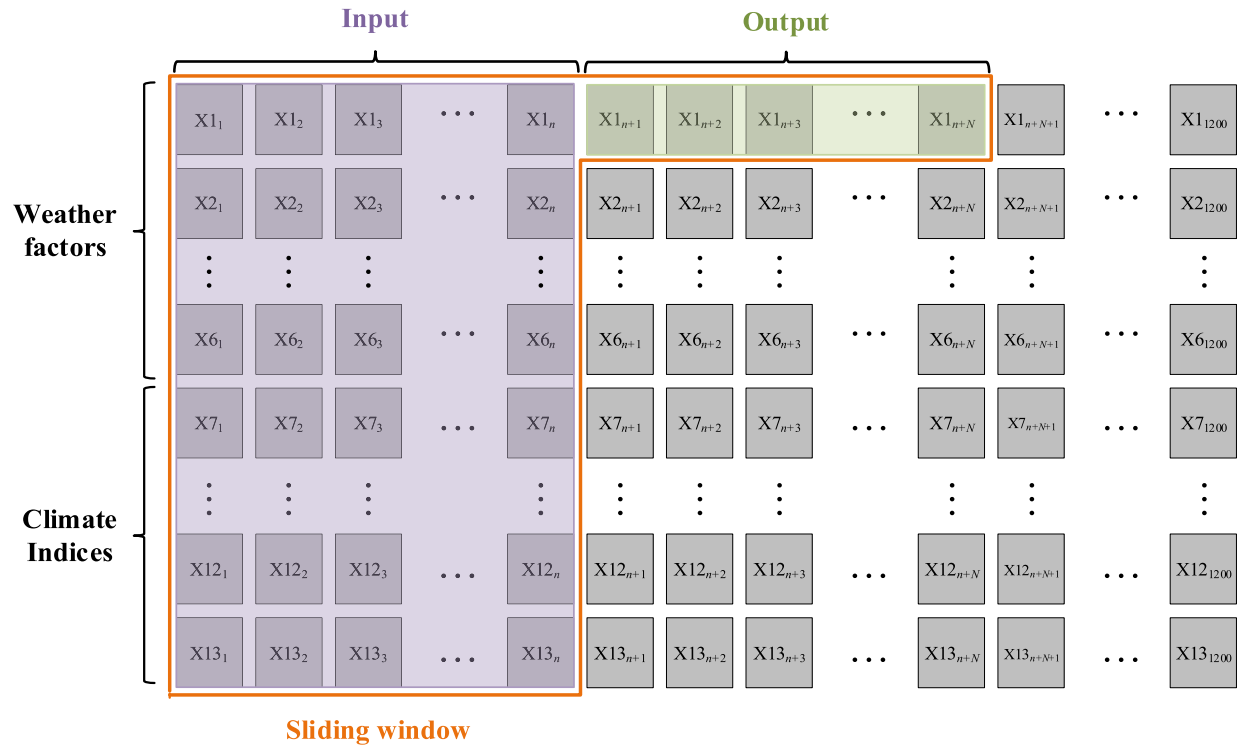


Fig. 6. Sample generation using a sliding window.

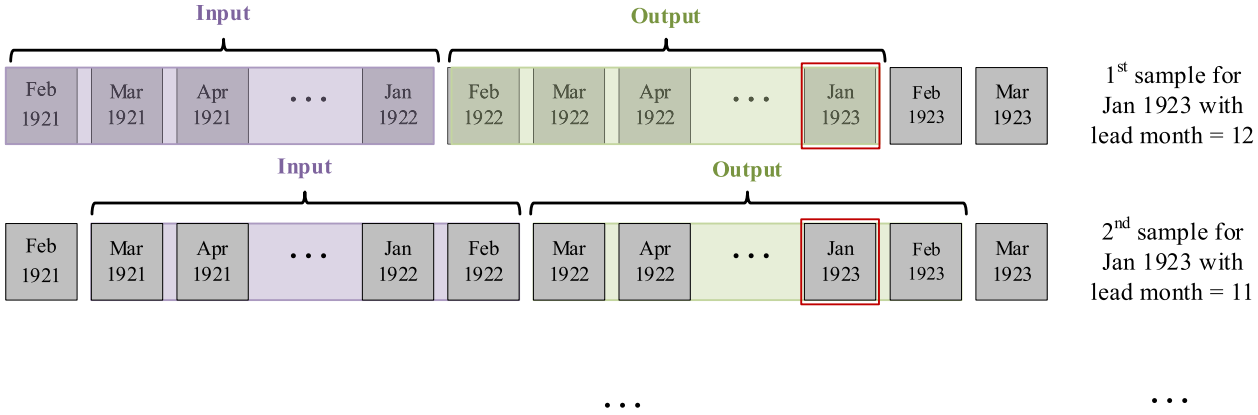


Fig. 7. Samples for predicting rainfall in Jan 1923 with different lead months.

considered as an encoder-decoder model, in which the encoder is the two GRU layers while the decoder is the following fully connected layer.

(4) Separate method. The separate method, also known as the direct method, trains different models for predicting monthly rainfall with different lead months. To be more specific, it develops a model to predict rainfall in the next month and a separate model to predict rainfall in the month after next, and so on. Hence, each model is a single-output model that produces rainfall for one future month, and the multi-step prediction can be gained by aggregating the predicted results of all separate models. For fair comparison, the encoder-decoder with an attention mechanism is used here as each single-output model. The difference is that the fully connected layer is only added following the last GRU cell since only one predicted value needs to be output.

(5) Recursive method. This method recursively uses the predicted values as input to the model for future prediction. To begin with, a single-step predictive model based on the encoder-decoder with an attention mechanism is developed to predict rainfall and other weather or climate features in the next month using the previous n months' weather factors and climate indices. The predicted values are then used as input to the model to further produce the subsequent month's values. This process is repeated until N future months' weather and climate features are predicted, and the N predicted monthly rainfall values can be extracted. Like the separate method, the single-step predictive model here also uses the encoder-decoder with an attention mechanism and only adds the fully connected layer following the last GRU cell of the decoder module. The output dimension is 13 because a total of 6 weather factors and 7 climate indices in the next month should be forecasted.

4.3. Results analysis

Before formal experiments, two critical parameters, i.e., the input sequence length n and the output sequence length N, should be determined. Table 2 provides the experimental results of different settings of n and N. According to Table 2, it is seen that setting both n and N to 12, i.e., using only one-year historical data to predict rainfall in the next year,

is the optimal choice. Under this setting, the input and output are well aligned, which means that the input and output in the same position will share the same calendar month. This can be convenient for the attention mechanism module to fully extract the seasonal periodicity of rainfall. Moreover, compared with the cases where n and N are equal to 18 or 24, setting n and N to 12 can reduce model complexity and improve training speed and efficiency while achieving the best prediction accuracy. Some other hyper-parameters like learning rate and batch size of the proposed model and the five baseline methods are tuned by grid search. The mean square error (MSE) is calculated as the loss function. Adam (Kingma & Ba, 2014) is used as the optimizer and Cosine Annealing (Loshchilov & Hutter, 2016) is employed as the approach to adjusting the learning rate. The coding environment is Python 3.8.8. All algorithms are implemented in Windows 11 system with an AMD 3600X CPU (6 cores and 12 processing threads) and 16 GB RAM.

The model performance after 3-fold cross validation (Pan et al., 2021) is demonstrated in Table 3 and Figs. 8–13. Specifically, Table 3 gives the comparisons of R^2, MAE, and QBD between the encoder-decoder with an attention mechanism and the five baseline methods. Fig. 8 shows the scatter plots of the predicted and true rainfall values in both cities generated by the encoder-decoder with an attention mechanism. Comparing the true rainfall with the P_{min}, P_{max}, P_{Q1}, and P_{Q3} of the corresponding 12 predictions in each month, Fig. 9 shows the histograms of true rainfall values versus the predicted P_{min}, P_{max}, P_{Q1}, and P_{Q3} in 1164 months from Jan 1923 to Dec 2019 in both cities. To demonstrate the model performance in abnormally wet or dry months, Figs. 10 and 11 give the prediction results in the five months with the highest and lowest rainfall in both cities, respectively. Finally, all the prediction results are classified by their lead month (from 1 to 12) and calendar month (from January to December), and the R^2 and MAE are calculated for each combination of lead month and calendar month, as shown in Figs. 12 and 13. The results are analyzed below.

(1) The encoder-decoder with an attention mechanism outperforms the baseline methods with higher R^2 and QBD while lower MAE in both studied cities. As listed in Table 3, in Darwin, the encoder-decoder with an attention mechanism could achieve an R^2, MAE, and QBD of 0.835, 39.903 mm, and 0.684, respectively, while in Perth 0.796, 20.884 mm, and 0.668, respectively. The values of all these metrics obtained by the proposed model are superior to those obtained by the five baseline methods. The excellent multi-step performance can also be seen from Figs. 8 and 9. To specify, most points in Fig. 8 gather around the line $y = x$; in Fig. 9, 51.5 % months in Darwin and 42.1 % months in Perth have the predicted $[P_{Q1}, P_{Q3}]$ including the observed rainfall, both of which further confirm the good predictive capability of the encoder-decoder with an attention mechanism. Among the five baselines, the best and worst baselines are the separate method and recursive method, respectively. While the separate method produces comparable results to the encoder-decoder with an attention mechanism, it incurs significant time and computational cost as 12 separate models need to be trained. Moreover, a possible reason for the lower accuracy of the separate method compared to the encoder-decoder with an attention mechanism is the model's lack of knowledge about the rainfall information in nearby months that it is predicting. This impedes its ability to capture short-term dependencies in the rainfall series. For instance, the model does not know the rainfall information in Month $t+1$ and Month $t+2$ when predicting the rainfall in Month $t+3$ using the weather and climate features from Month $t-11$ to Month t. Conversely, the decoder module of the encoder-decoder with attention mechanism generates multiple outputs simultaneously, where the predicted rainfall values in Month $t+1$ and Month $t+2$ serve as inputs when predicting the rainfall in Month $t+3$ (as mentioned in Section 2.3). This enhances the model's prediction performance to some extent. For the recursive method, it is prone to error accumulation during the iterative prediction process, which leads to its worst performance. Specifically, to predict the rainfall in a far future month, near future rainfall predictions must be fed as inputs to the model. However, there exist errors in the near future predictions; hence, with the progression of iterations, the deviation between the predicted rainfall and the ground truth will increase, resulting in the poorest accuracy of the recursive method. Besides, the comparisons between the proposed model and the first three baselines could demonstrate the indispensability of each module in the model architecture to an accurate multi-step prediction. To be specific, among the first three baselines, BLSTM-GRU and GRU can be considered as an encoder-decoder without attention mechanism, both of which have inferior performances on all the three evaluation metrics. LSTM-AM is actually an attention mechanism module connecting a decoder. Although better than BLSTM-GRU and GRU, LSTM-AM cannot fully extract the underlying information within the input features due to the lack of the encoder module; hence, its predictions are not so accurate as the predictions made by the encoder-decoder with an attention mechanism.

(2) In those abnormally wet or dry months, the encoder-decoder with an attention mechanism could still show satisfactory predictive performance. Fig. 10 compares the true rainfall with the corresponding 12 predictions in the five wettest months from Jan 1923 to Dec 2019 in the two cities. The red dots denote the true rainfall values, each of which is overlapped by the box plot of 12 predictions for that month. In both cities, three out of the five wettest months have true rainfall values close to the predicted P_{Q3} (Mar 1977, Jan 1995, and Jan 1991 in Darwin; Jun 1945, Jul 1958, and Jul 1926 in Perth), while the other two cases have the ground truths approaching P_{max} (Feb 2011 and Jan 1981 in Darwin; Jun 1923 and Aug 1945 in Perth). Therefore, although a single prediction with a certain lead month might not be that accurate, all 12 predictions can be aggregated and the P_{Q3} or P_{max} can be used as a sufficiently conservative estimation of monthly rainfall even for those wettest months. Similarly, Fig. 11 shows the prediction results in the five driest months which have no rainfall at all. In Darwin, most predictions

Table 2
Experimental results of different settings of n and N.

City	Setting of n and N	R^2	MAE (mm)	QBD
Darwin	$n = 12, N = 12$	**0.835**	**39.903**	**0.684**
	$n = 18, N = 12$	0.771	45.102	0.583
	$n = 24, N = 12$	0.763	47.539	0.554
	$n = 18, N = 18$	0.816	42.588	0.679
	$n = 24, N = 18$	0.750	50.691	0.571
	$n = 24, N = 24$	0.801	43.73	0.653
Perth	$n = 12, N = 12$	**0.796**	**20.884**	**0.668**
	$n = 18, N = 12$	0.768	21.175	0.579
	$n = 24, N = 12$	0.744	22.844	0.557
	$n = 18, N = 18$	0.790	20.93	0.662
	$n = 24, N = 18$	0.752	22.203	0.593
	$n = 24, N = 24$	0.781	20.991	0.648

Table 3
Comparisons of R^2, MAE, and QBD between the encoder-decoder with an attention mechanism and the five baseline methods.

City	Model	R^2	MAE (mm)	QBD
Darwin	Encoder-decoder with an attention mechanism	**0.835**	**39.903**	**0.684**
	BLSTM-GRU	0.667	60.435	0.510
	LSTM-AM	0.742	51.980	0.603
	GRU	0.675	60.261	0.483
	Separate method	0.799	44.085	0.671
	Recursive method	0.601	69.031	0.372
Perth	Encoder-decoder with an attention mechanism	**0.796**	**20.884**	**0.668**
	BLSTM-GRU	0.679	27.287	0.308
	LSTM-AM	0.730	23.791	0.505
	GRU	0.698	26.874	0.255
	Separate method	0.778	21.713	0.597
	Recursive method	0.603	32.510	0.194

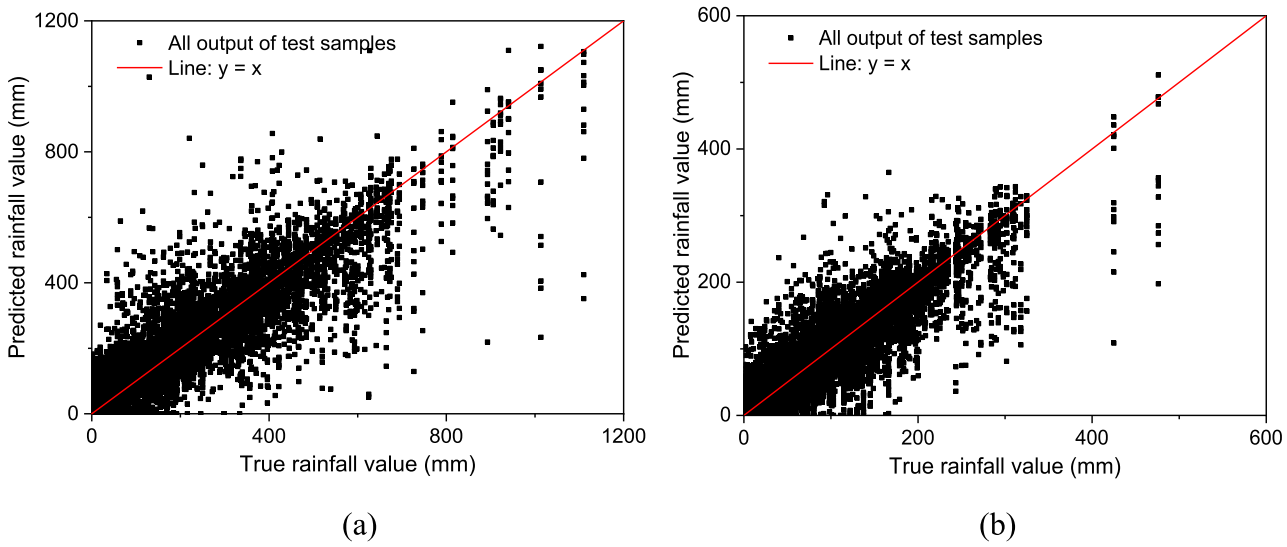


Fig. 8. Scatter plots of the prediction results: (a) Darwin; (b) Perth.

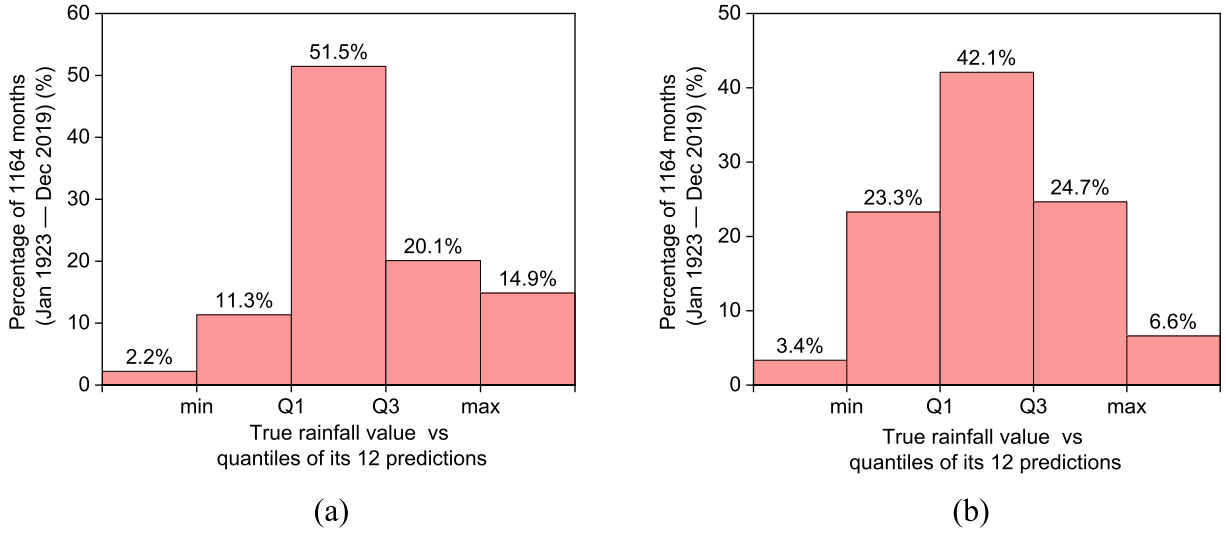


Fig. 9. Histogram of true rainfall values compared with the quantiles of predicted values in 1164 months from Jan 1923 to Dec 2019: (a) Darwin; (b) Perth.

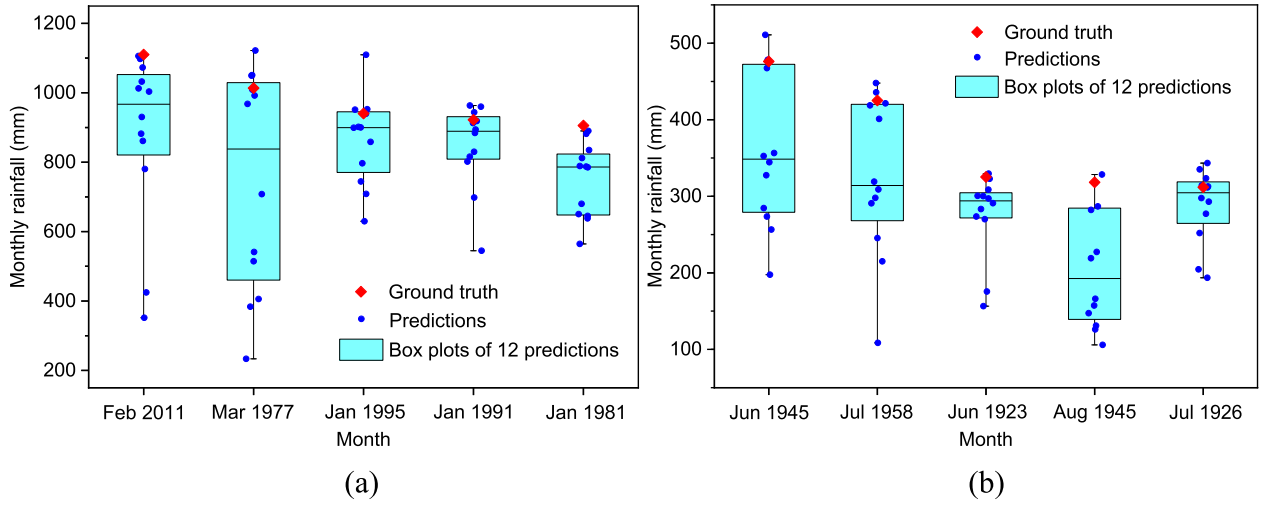


Fig. 10. Comparison between the true rainfall and the corresponding 12 predictions in the five wettest months from Jan 1923 to Dec 2019: (a) Darwin; (b) Perth.

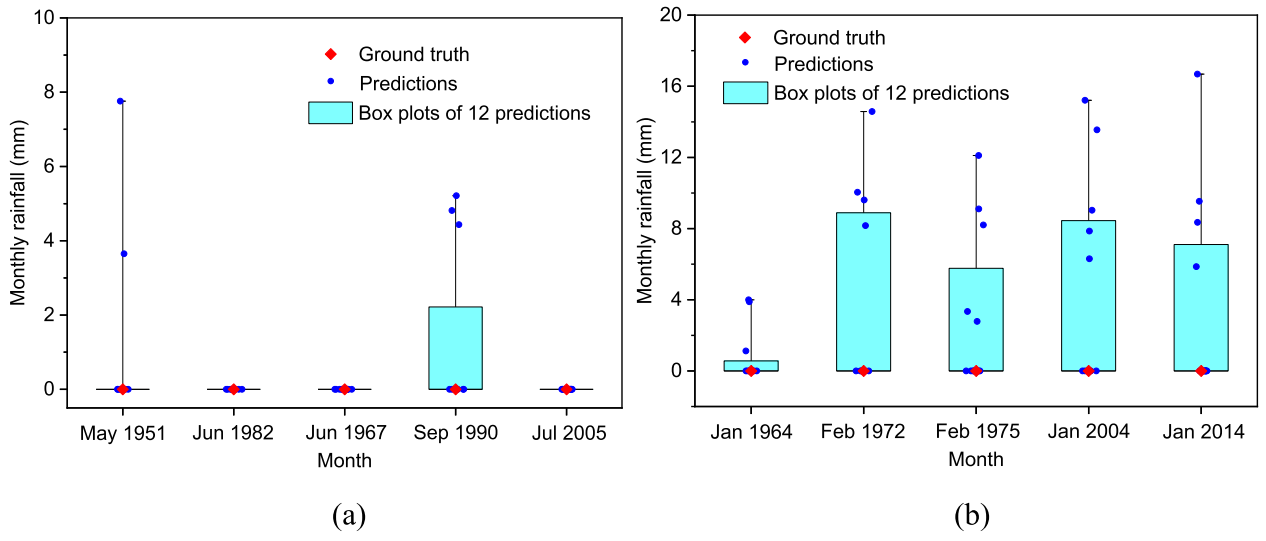


Fig. 11. Comparison between the true rainfall and the corresponding 12 predictions in the five driest months from Jan 1923 to Dec 2019: (a) Darwin; (b) Perth.

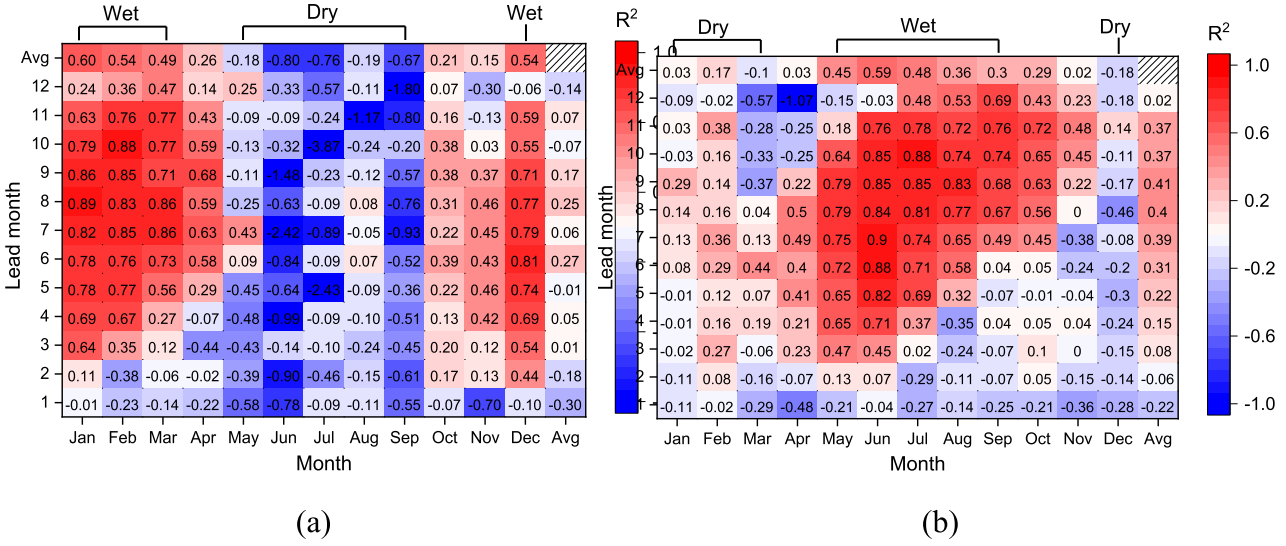


Fig. 12. R^2 of rainfall prediction in different calendar months with different lead months (wet and dry months are marked out): (a) Darwin; (b) Perth.

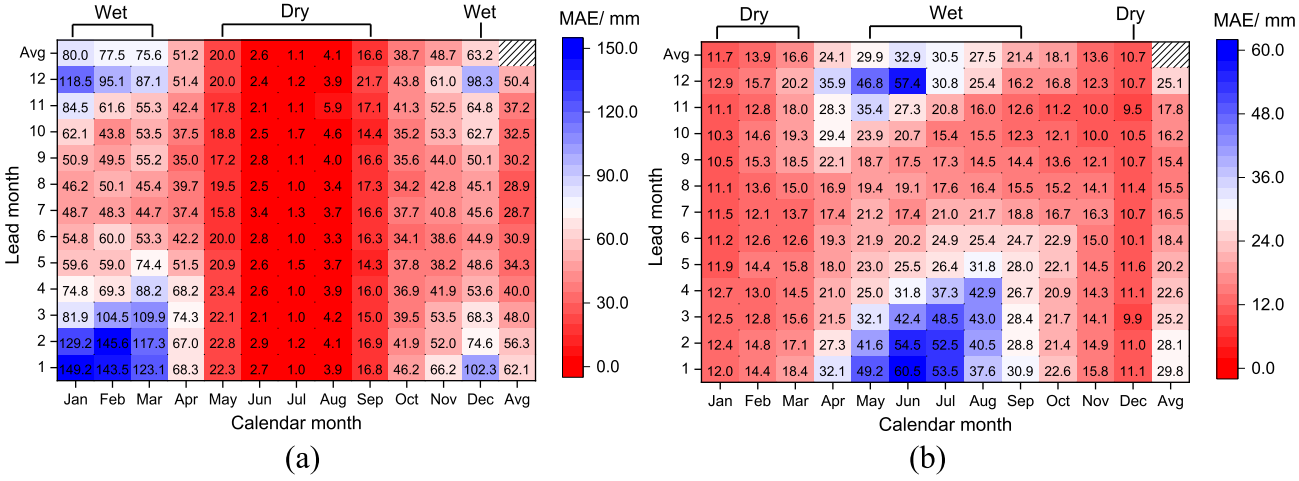


Fig. 13. MAE of rainfall prediction in different calendar months with different lead months (wet and dry months are marked out): (a) Darwin; (b) Perth.

regardless of the lead month are also zero, and the absolute errors of those few non-zero predictions for May 1951 and Sep 1990 are all smaller than 8 mm. In Perth, the predicted medians, P_{Q1}, and P_{min} of all five cases are zero, and the maximum absolute error is lower than 20 mm. These results demonstrate the excellent predictive capability of the encoder-decoder model to reproduce abnormally high or low monthly rainfall.

(3) With a lead month of 6–10, the encoder-decoder with an attention mechanism could produce a good prediction of large rainfall in wet seasons; for small rainfall in dry seasons, the predictive model can generate accurate estimation regardless of the lead month. As mentioned in the previous analysis, the predictive model will generate 12 predictions with different lead months for a specific month. However, which lead month could produce the most accurate prediction? Moreover, as mentioned in Section 3, the wet seasons in Darwin and Perth are from December to March and from May to September, respectively. The distinction between wet and dry seasons within a year implies that the best lead month for rainfall prediction could be different in different calendar months. To solve the above question, all the prediction results are classified by their lead month (from 1 to 12) and calendar month (from January to December), and the R^2 and MAE for each combination of lead month and calendar month (each combination contains 97 samples) are computed. The results are shown in Figs. 12 and 13. It is seen that the best lead month for wet seasons is about 6–10. In Darwin from December to March, the average R^2 and MAE with a lead month of 6–10 are 0.79 and 50.7 mm, respectively. In Perth from May to September, the average R^2 and MAE with a lead month of 6–10 are 0.74 and 19.0 mm, respectively. This not only demonstrates the good performance of the proposed model in predicting large rainfall in wet seasons, but also indicates that the monthly rainfall is potentially related to the weather or climate conditions several months ago rather than recently. In dry months, by contrast, the model could well predict the small rainfall regardless of the lead month. This can be easily seen from the relatively low MAE (see Fig. 13) in the dry months of both cities. Specifically, the average MAE in Darwin from May to September is 12.6 mm, while in Perth from December to March is 13.2 mm. However, as illustrated in Fig. 12, the R^2 values in the dry months of both cities are close to or even lower than 0, which seems to indicate poor predictive capability. This is caused by the calculation process of R^2. Written as Eq. (11), R^2 measures the model prediction error (the numerator of the second term) compared with the error caused by using the mean rainfall value ($\overline{X1}$) as the prediction (the denominator of the second term). In dry months, the rainfall values are generally rather small, which makes the mean rainfall value close to the ground truth and thus leads to a low R^2. Therefore, under this circumstance, it will be more suitable to use MAE as the evaluation metric since it can directly reflect the prediction error in dry months.

5. Feature attribution analysis

This section aims to conduct a detailed feature attribution analysis on two cases: (1) January rainfall in Darwin with a lead month of 8; (2) June rainfall in Perth with a lead month of 7. As mentioned in Section 2.4, the explanation module could identify the input features that contribute most to the model output, while such explanations are meaningful only when the model output coincides with the ground truth. Because extreme rainfall in wet months can induce flood hazards that lead to massive property losses and casualties, this study focuses on discovering the main factors affecting rainfall in the wettest calendar month of the two cities. According to the monthly rainfall record in the two cities, January and June are the calendar months with the most rainfall in Darwin and Perth, respectively. As shown in Figs. 12 and 13, for January rainfall in Darwin, the prediction with a lead month of 8 is the most accurate; for June rainfall in Perth, the prediction with a lead month of 7 is the most accurate. Hence, a detailed feature attribution analysis is implemented on the above two cases in this section.

As a result, Fig. 14 shows the results of feature attributions for January rainfall in Darwin with a lead month of 8. Specifically, Fig. 14 (a) gives a bar chart where the top 20 important features are sorted in descending order according to their mean absolute attribution (MAA) values, each of which is calculated by averaging the absolute values of the attribution values of that feature among all 97 samples. Fig. 14(b) shows a bee swarm plot of the top 20 features, in which each point in a line represents the corresponding feature attribution value in a specific sample, and the point color represents the value of the feature in that sample. Note that the default output target is X1 (monthly rainfall) at time t, and each input feature is written as the weather factor or climate index connected with the time step by an underline. For example, the most important weather feature for January rainfall in Darwin with a lead month of 8 is X1_t −12, which means that last January's rainfall is the most important to the coming January's rainfall. As a complement to Fig. 14, Fig. 15 further illustrates the variations of the attribution values with normalized values for each of the top 20 features. Similarly, Fig. 16 shows the results of feature attributions for June rainfall in Perth with a lead month of 7, and Fig. 17 illustrates the corresponding variations of the feature attribution values with normalized feature values. The results are analyzed below.

(1) In Darwin, the two most important weather features to the coming January's rainfall are last January's and February's rainfall, which tend to exert enhancement effects on the coming January's rainfall. From Fig. 14 (a), it is seen that the top two important features are X1 (monthly rainfall) at time t −12 and t −11, with an MAA of 20.62 mm and 20.28 mm, respectively. This indicates that the rainfall values in last January and February are the most significant factor to predict the rainfall in the coming January. It is worth mentioning that the two important features are more correlational than causal factors. That is, the encoder-decoder with an attention mechanism estimates the January rainfall in Darwin by leveraging the autocorrelation in the monthly rainfall series, or more specifically, the seasonal periodicity in this case. From Fig. 14(b) and Fig. 15, most attribution values of X1 at time t −12 and t −11 are positive, and the attribution values tend to go up with the increase of the values of the features. This demonstrates that last January's and February's rainfall could exert enhancement effects on the coming January's rainfall, and such enhancement effects will be greater if their values are large. In other words, if extremely large rainfall in January and February is observed in this year, it can be judged that next January's rainfall will be also larger than usual. Besides, X3 (daily average minimum temperature) is also a weather factor that ranks high in Fig. 14; however, based on Fig. 15, there is no clear relationship between the attribution value and normalized value of X3.

(2) In Darwin, the most important climate feature to the coming January's rainfall is last February's SOI value, which exerts an inhibitory effect on the coming January's rainfall when its value is high while having a promoting effect when its value is low. According to Fig. 14(a), the X7 (SOI) at time t −11, with an MAA of 11.82 mm, is the climate feature that ranks highest. Among all input features, it also ranks at the forefront, which verifies the necessity of incorporating the climate indices into the model input. Besides, from Fig. 14(b) and Fig. 15, a strong negative correlation can be identified between the attribution value and normalized value of X7 at time t −11. To be more specific, when the last February's SOI is high, the corresponding attribution value is negative, which means it will exert an inhibitory effect on the next January's rainfall, and the inhibitory effect tends to be greater with the increase of the feature's value. By contrast, when last February's SOI is low, the corresponding attribution value is positive, indicating a promoting effect on the next January's rainfall, and such promoting effect will be higher with the decrease of the feature's value. This result establishes a distinct relationship between the past SOI and future monthly rainfall in a data-driven manner. In fact, the relationship between the past SOI and future rainfall could also be partially explained from the perspective of climatology. A continuously lower than usual SOI

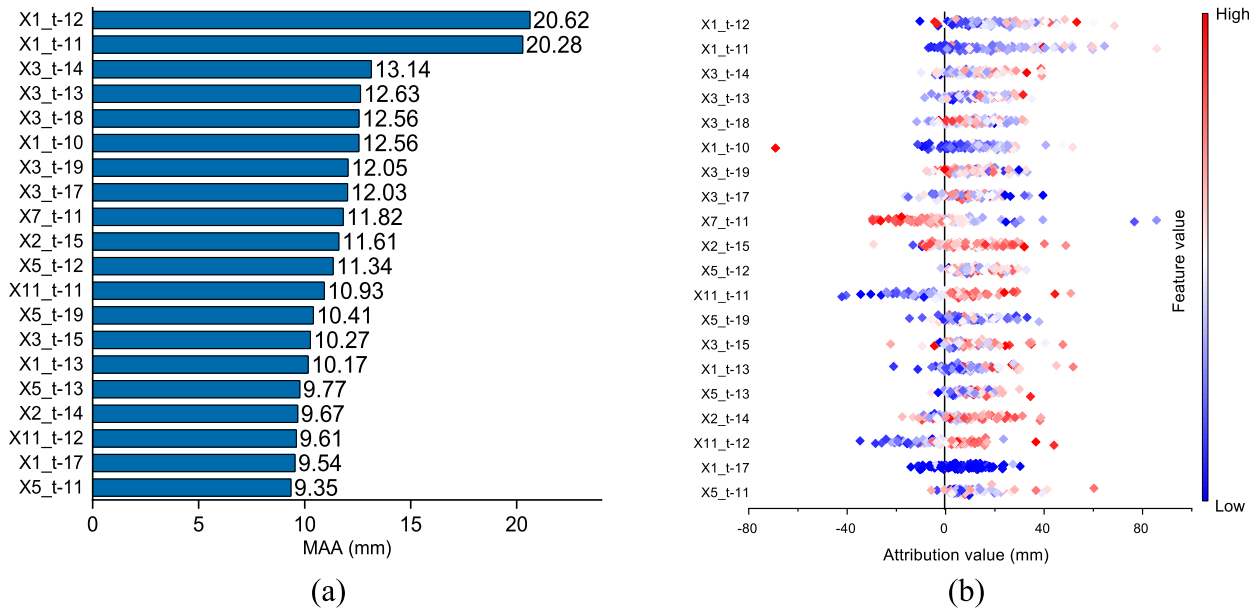


Fig. 14. Attributions of the top 20 important features for January rainfall in Darwin (lead month = 8): (a) Bar chart of MAA. (b) Bee swarm plot.

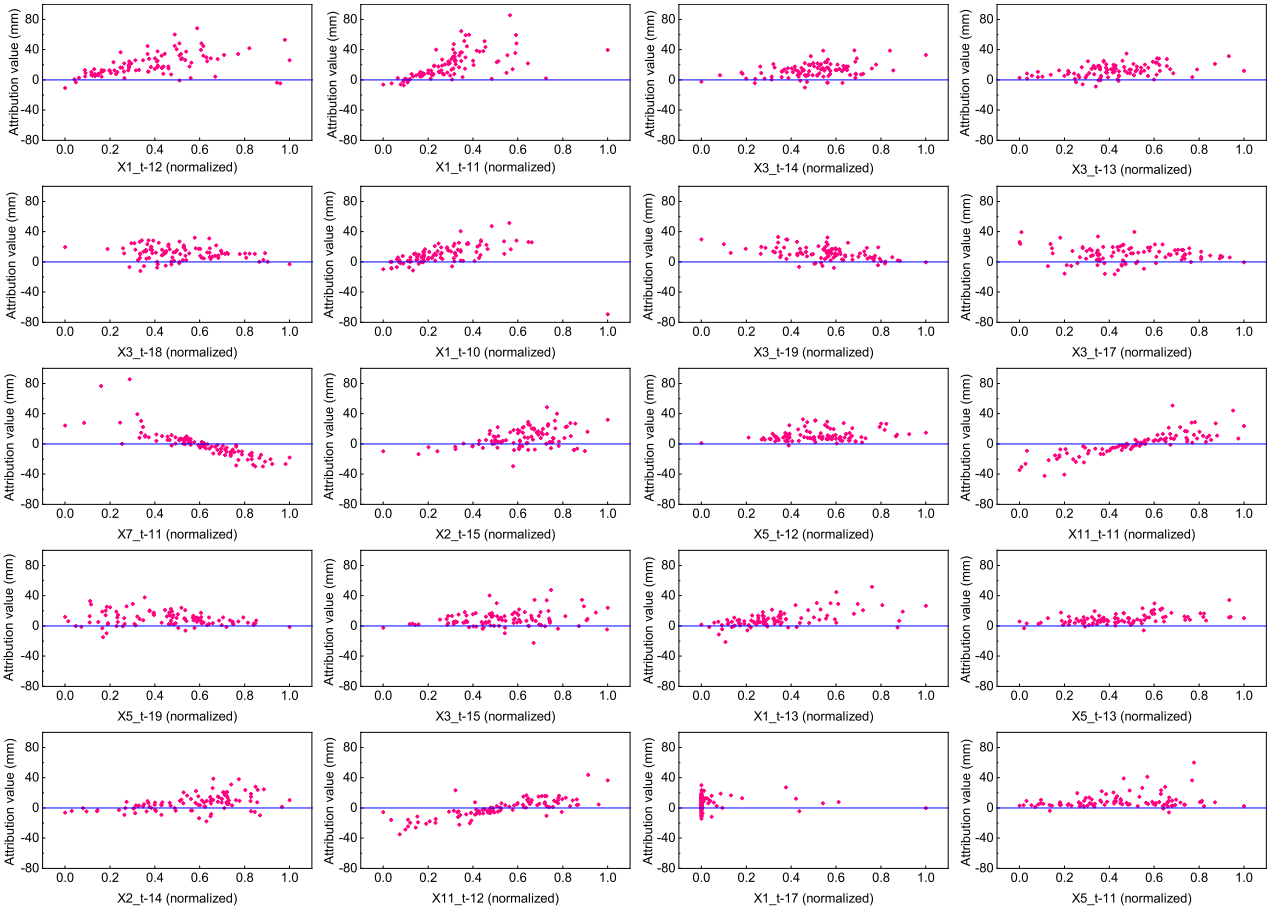


Fig. 15. Variations of the attribution values with normalized values of the top 20 important features for January rainfall in Darwin (lead month = 8).

indicates a strong El Nino event in that year, which also means that a La Nina event will generally occur in the next year (Suarez & Schopf, 1988). During a La Nina event, the rainfall in Australia, particularly across the east and north (Darwin is in the north of Australia), will significantly increase, sometimes resulting in widespread flooding (ABOM, 2022f). The consistency between the data-driven model and domain knowledge exhibits the great potential of the explainable deep learning approach in knowledge discovery.

(3) In Perth, compared with the climate features, the weather features are more significant to the June rainfall, and the most important

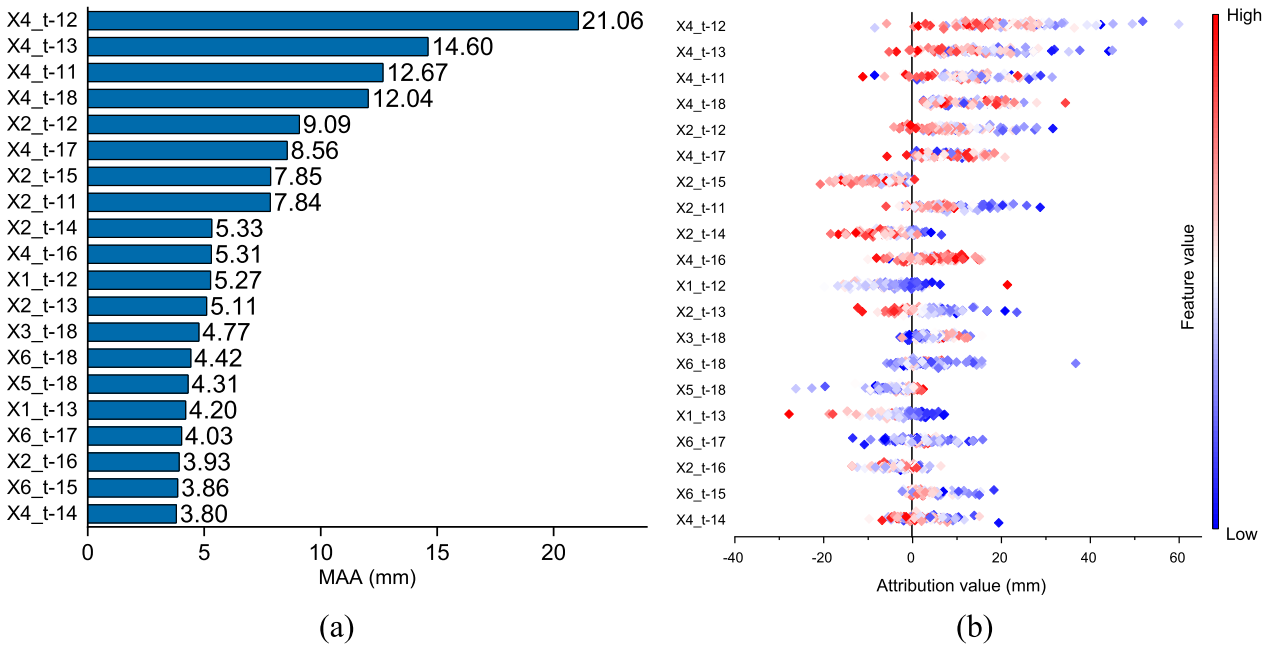


Fig. 16. Attributions of the top 20 important features for June rainfall in Perth (lead month = 7): (a) Bar chart of MAA. (b) Bee swarm plot.

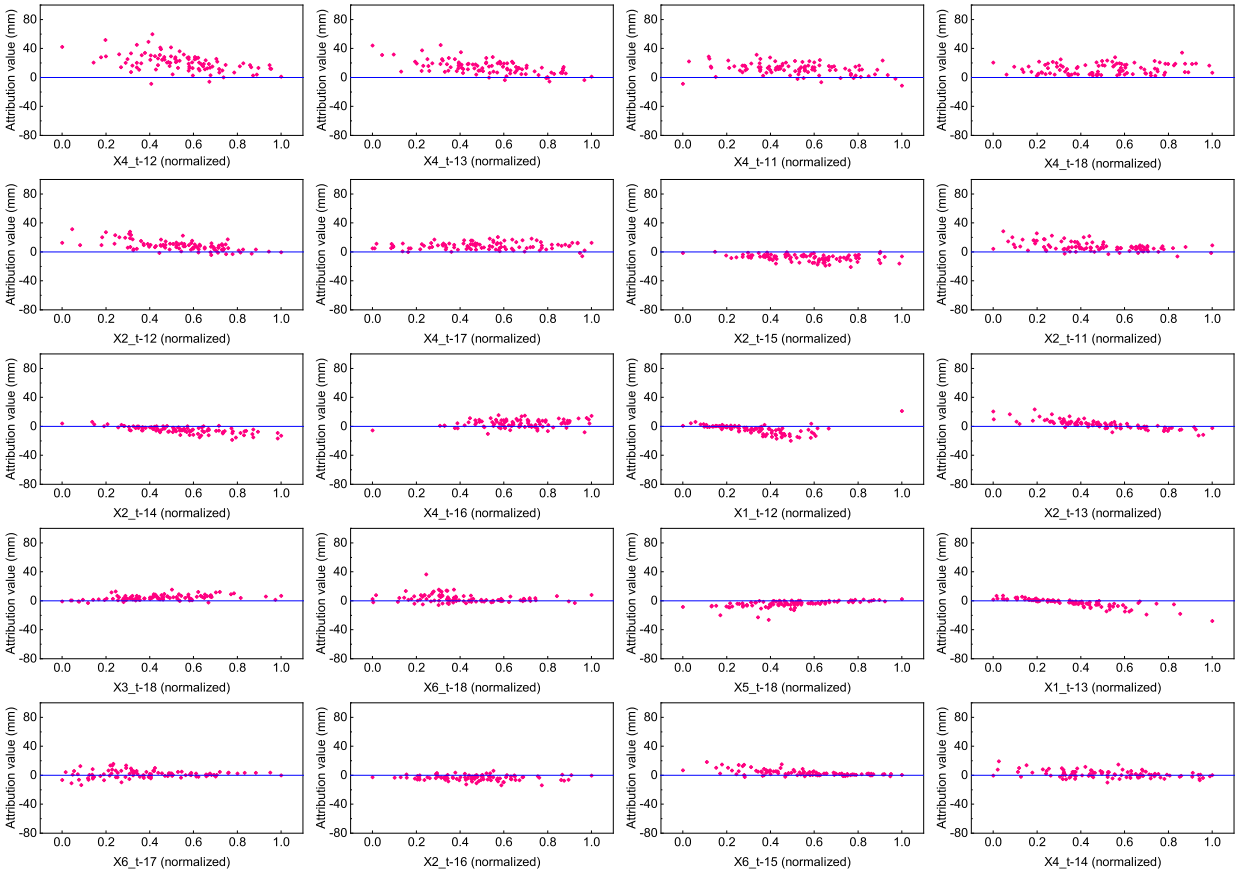


Fig. 17. Variations of the attribution values with normalized values of the top 20 important features for June rainfall in Perth (lead month = 7).

weather feature to the coming June's rainfall is last June's solar radiation. To be specific, according to Fig. 16, there are all weather factors and no climate indices among the top 20 features, which means that the effects of the adopted climate phenomenon are not as significant as those of the weather factors in Perth. The three features that matter the most are X4 (daily average solar radiation) at time $t-12$, $t-13$ and $t-11$, with an MAA of 21.06 mm, 14.60 mm, and 12.67 mm, respectively, among which X4 at time $t-12$, i.e., last June's daily average solar radiation, is the most significant. From Fig. 16(b) and Fig. 17, a weak negative correlation can be identified between the attribution value and

the feature value of the last June's solar radiation, which demonstrates that a higher last June's solar radiation will exert an inhibitory effect on the coming June's rainfall and hence make the rainfall lower. This may be explained by the climate characteristics of Perth. With a Mediterranean climate, Perth has a hotter and dryer summer controlled by subtropic anticyclone (Williams & Stone, 2009) while a cooler and wetter winter controlled by westerly winds (Dey et al., 2019). June is one winter month in Perth; hence, a higher solar radiation in June indicates a stronger subtropic anticyclone that lasts longer in that year. Such an event may have an influence on subsequent alternations between subtropic anticyclones and westerly winds. Therefore, it is highly possible for the coming June to be also controlled by the after-effects of a stronger subtropic anticyclone, which leads to less monthly rainfall. The importance of the last June's solar radiation can also be discussed from the perspective of seasonal periodicity. As the sunniest capital city in Australia, Perth possesses a strong solar radiation, especially in its dry summer due to the Mediterranean climate. Therefore, there exists a distinct seasonality of solar radiation in Perth. Coincidently, the rainfall in Perth, although moderate, is also highly seasonal with more rainfall in winter. Hence, a clear synchronized appearance of lower solar radiation and more monthly rainfall lies in Perth's winter months, especially in June. It is this synchronized appearance property that the predictive model utilizes to forecast the June rainfall in Perth, which further leads to the high and positive attribution value of last June's solar radiation.

6. Concluding remarks and future works

Monthly rainfall prediction is an essential weather forecast task in water resource management and hydrological disaster control. To conduct a multi-step prediction and feature attribution analysis for monthly rainfall, this study proposes an explainable deep learning approach that combines four consecutive modules: GRU-based encoder module, attention mechanism module, GRU-based decoder module, and expected-gradient-based explanation module. The first three modules compose an encoder-decoder with an attention mechanism that could predict the monthly rainfall in multiple continuous months in the future. The explanation module based on expected gradient could compute the attribution values of the input features so that those important factors can be identified and analyzed. Through the case study in Darwin and Perth, the applicability and validity of the proposed approach is well demonstrated.

Based on the case study, the conclusions are as follows: (1) The encoder-decoder with an attention mechanism accurately predicts multi-step monthly rainfall with strong agreement with the ground truth. Compared to the five baselines, the proposed method outperforms in R^2, MAE, and QBD. (2) The encoder-decoder with an attention mechanism can provide precise forecasts of rainfall in both wet and dry seasons, even in extreme weather months. Specifically, it achieves optimal accuracy for wet seasons with a lead month of 6–10 and accurate estimates for dry months regardless of the lead month. (3) For predicting January rainfall in Darwin, last January and February's rainfall are the most critical weather features with relatively strong correlation to the predicted target, while last February's SOI being the most significant climate index. Higher values of these features have a strengthening promotion effect, whereas the inverse remains true for SOI. (4) For forecasting June rainfall in Perth, last June's daily average solar radiation plays a crucial role with a weak negative correlation between its attribution value and feature value. This may because a higher solar radiation in June indicates a stronger subtropic anticyclone, which could have lasting effects on subsequent alternations between subtropic anticyclones and westerly winds, leading to lower monthly rainfall.

By introducing four common modules, this study proposes an explainable deep learning approach that contributes to two aspects. Firstly, this study extends the application of deep learning in the field of rainfall prediction from single-step to multi-step prediction. Through the encoder-decoder with an attention mechanism, this study achieves accurate and reliable prediction of future consecutive months of rainfall, providing valuable guidance for various applications such as water resource management, agriculture, and disaster preparedness. Secondly, this study explores the main factors that affect monthly rainfall from a data-driven perspective, rather than relying on physical knowledge, offering an alternative solution for studying complex weather systems. By using the expected gradient to address the issue of deep model interpretability, the influential factors behind precipitation variations are identified, which enables people to gain deeper insights into the dynamics of rainfall patterns.

The contributions of this study benefit both the theory and practice. From a theoretical perspective, this research enriches the field of explainable deep learning for multi-step rainfall prediction. By shedding light on the main factors affecting monthly rainfall patterns, this study advances the understanding of the complex relationships between meteorological variables and rainfall patterns. Regarding practical applications, compared with other precipitation products based on satellites or radars that can only forecast rainfall for the next few hours or days, the encoder-decoder with an attention mechanism can provide reliable predictions of rainfall for a longer period in the future, thus facilitating longer-term planning and decision-making in fields that are susceptible to rainfall fluctuations such as water resource management and agriculture. Some critical water management actions like optimal irrigation, water allocation, and water infrastructure operations can be arranged beforehand with the proposed predictive model. Implementing multi-step monthly rainfall predictions is also valuable in rainfall-related disaster control. With forthcoming rainfall patterns predicted, potential hazards like droughts or floods can be proactively countered, which improves emergency preparedness and contributes to the overall safety of people. If combined with city-level digital twin (Pan & Zhang, 2023), the predictive model can also effectively promote local resilience management and provide guidance for the government's resilience strategy.

Furthermore, the proposed rainfall predictive model enables the early estimation of some economic or societal phenomena associated with rainfall. For instance, variations in rainfall patterns can greatly impact commodity prices, particularly in the agricultural sector. Insufficient rainfall can lead to crop reductions or even failure, while abnormally high rainfall can result in an oversupply of agricultural products. The fluctuation in the supply of agricultural products can cause price volatility, potentially affecting the consumption capacity of rural areas and investment decisions of related companies, ultimately influencing the overall stability of the financial market. Such reductions in agricultural productivity and income due to abnormal rainfall can also increase the likelihood of civil unrest and recruitment for rebel activity in some regions (Nillesen & Verwimp, 2009). Through multi-step prediction of rainfall, relevant authorities can anticipate the future supply–demand relationship at the agricultural market and prepare in advance for potential social risks. This illustrates the broader implications of the proposed research, extending its practical significance beyond meteorology and providing valuable insights for changes in socioeconomic conditions.

There are still some limitations and challenges in this study. On the one hand, the limited availability of meteorological variables from the weather stations used in this study restricts the accuracy of the predictive model to some extent. Furthermore, compared to remote sensing data or echo data that directly reflect precipitation obtained from satellite or radar precipitation products, the input features used in this study for rainfall prediction may be less informative. On the other hand, this study only focuses on the case study of two cities in Australia, and the applicability and effectiveness of the proposed explainable deep learning approach in other regions still need to be verified. Therefore, to address these existing limitations, subsequent studies will incorporate more weather and climate features as input for the model to further enhance the prediction accuracy. The ERA5-Land reanalysis dataset

(Muñoz-Sabater et al., 2021) is a high-quality global climate reanalysis dataset that provides historical records of various meteorological elements. Relevant data from the ERA5-Land reanalysis dataset will be incorporated and combined with satellite remote sensing data and ground weather station measurements to achieve more accurate multi-step forecasting of monthly rainfall with more informative features. Moreover, by using these global precipitation data products, the scope of the case study will be expanded from Australia to other parts of the world with various climate types and rainfall patterns, thereby strengthening the generalizability and effectiveness of the proposed methodology. A more general framework is expected to be created to predict rainfall on a global scale and provide deep insights into the primary factors affecting rainfall worldwide.

CRediT authorship contribution statement

Renfei He: Writing – original draft, Methodology, Visualization, Data curation, Investigation, Formal analysis. **Limao Zhang:** Conceptualization, Supervision, Methodology, Writing – review & editing, Funding acquisition. **Alvin Wei Ze Chew:** Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data, Python code, and some result files of the model in this study are available at https://github.com/RenfeiHe/Monthly-rainfall-prediction-and-analysis.

Acknowledgment

This research is supported by the National Research Foundation, Singapore, and PUB, Singapore's National Water Agency under its RIE2025 Urban Solutions and Sustainability (USS) (Water) Centre of Excellence (CoE) Programme, awarded to Nanyang Environment & Water Research Institute (NEWRI), Nanyang Technological University, Singapore (NTU).

Disclaimer

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore and PUB, Singapore's National Water Agency.

References

Abbot, J., & Marohasy, J. (2012). Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Advances in Atmospheric Sciences, 29*, 717–730.

Abbot, J., & Marohasy, J. (2014). Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmospheric Research, 138*, 166–178.

ABOM. (2022a). Australian Bureau of Meteorology: About ENSO and IOD indices. http://www.bom.gov.au/climate/enso/indices/about.shtml#:~:text=The%20Indian%20Ocean%20Dipole%20indices&text=The%20IOD%20is%20commonly%20measured,S%20to%2010%C2%B0N (accessed 16 Feb 2022).

ABOM. (2022b). Australian Bureau of Meteorology: Average annual, seasonal and monthly rainfall. http://www.bom.gov.au/jsp/ncc/climate_averages/rainfall/index.jsp?period=an&area=oz#maps (accessed 16 Feb 2022).

ABOM. (2022c). Australian Bureau of Meteorology: Climate Data Online. http://www.bom.gov.au/climate/data/index.shtml (accessed 16 Feb 2022).

ABOM. (2022d). Australian Bureau of Meteorology: Climate statistics for Australian locations–Darwin Airport. http://www.bom.gov.au/climate/averages/tables/cw_014015_All.shtml (accessed 16 Feb 2022).

ABOM. (2022e). Australian Bureau of Meteorology: ENSO impacts – rainfall. http://www.bom.gov.au/climate/enso/history/ln-2010-12/ENSO-rainfall.shtml (accessed 16 Feb 2022).

ABOM. (2022f). Australian Bureau of Meteorology: What is La Niña and how does it impact Australia? http://www.bom.gov.au/climate/updates/articles/a020.shtml (accessed 16 Feb 2022).

Ashok, K., Guan, Z., & Yamagata, T. (2003). Influence of the Indian Ocean Dipole on the Australian winter rainfall. *Geophysical Research Letters, 30*, 1821.

Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., & Müller, K.-R. (2010). How to explain individual classification decisions. *The Journal of Machine Learning Research, 11*, 1803–1831.

Bagirov, A. M., & Mahmood, A. (2018). A comparative assessment of models to predict monthly rainfall in Australia. *Water Resources Management, 32*, 1777–1794.

Bagirov, A. M., Mahmood, A., & Barton, A. (2017). Prediction of monthly rainfall in Victoria, Australia: Clusterwise linear regression approach. *Atmospheric Research, 188*, 20–29.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Ben Taieb, S., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications, 39*, 7067–7083.

Cai, W., & van Rensch, P. (2012). The 2011 southeast Queensland extreme summer rainfall: A confirmation of a negative Pacific Decadal Oscillation phase? *Geophysical Research Letters, 39*, L08702.

Cai, W., Whetton, P. H., & Pittock, A. B. (2001). Fluctuations of the relationship between ENSO and northeast Australian rainfall. *Climate Dynamics, 17*, 421–432.

Chew, A. W. Z., & Zhang, L. (2022). Data-driven multiscale modelling and analysis of COVID-19 spatiotemporal evolution using explainable AI. *Sustainable Cities and Society, 80*, Article 103772.

Chhetri, M., Kumar, S., Pratim Roy, P., & Kim, B.-G. (2020). Deep BLSTM-GRU Model for Monthly Rainfall Prediction: A Case Study of Simtokha, Bhutan. *Remote Sensing, 12*, 3174.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Dey, R., Lewis, S. C., Arblaster, J. M., & Abram, N. J. (2019). A review of past and projected changes in Australia's rainfall. *Wiley Interdisciplinary Reviews: Climate Change, 10*, e577.

Erion, G., Janizek, J. D., Sturmfels, P., Lundberg, S. M., & Lee, S.-I. (2021). Improving performance of deep learning models with axiomatic attribution priors and expected gradients. *Nature Machine Intelligence, 3*, 620–631.

Fahad, S., Su, F., Khan, S. U., Naeem, M. R., & Wei, K. (2023). Implementing a novel deep learning technique for rainfall forecasting via climatic variables: An approach via hierarchical clustering analysis. *Science of The Total Environment, 854*, Article 158760.

Fu, X., Luo, W., Xu, C., & Zhao, X. (2020). Short-term traffic speed prediction method for urban road sections based on wavelet transform and gated recurrent unit. *Mathematical Problems in Engineering, 2020*, 3697625.

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. In G. Geoffrey, D. David & D. Miroslav (Eds.), *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (Vol. 15, pp. 315–323). Proceedings of Machine Learning Research: PMLR.

Gu, J., Liu, S., Zhou, Z., Chalov, S. R., & Zhuang, Q. (2022). A stacking ensemble learning model for monthly rainfall prediction in the Taihu Basin, China. *Water, 14*, 492.

Haidar, A., & Verma, B. (2018). Monthly rainfall forecasting using one-dimensional deep convolutional neural network. *IEEE Access, 6*, 69053–69063.

He, R., Zhang, L., & Chew, A. W. Z. (2022). Modeling and predicting rainfall time series using seasonal-trend decomposition and machine learning. *Knowledge-Based Systems, 251*, Article 109125.

He, R., Zhang, L., & Tiong, R. L. K. (2023). Flood risk assessment and mitigation for metro stations: An evidential-reasoning-based optimality approach considering uncertainty of subjective parameters. *Reliability Engineering & System Safety, 238*, 109453.

Johny, K., & Pai, M. L. (2022). A multivariate EMD-LSTM model aided with Time Dependent Intrinsic Cross-Correlation for monthly rainfall prediction. *Applied Soft Computing, 123*, Article 108941.

Khan, N., Shahid, S., Juneng, L., Ahmed, K., Ismail, T., & Nawaz, N. (2019). Prediction of heat waves in Pakistan using quantile regression forests. *Atmospheric Research, 221*, 1–11.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lei, C. (2021). RNN. In *Deep Learning and Practice with MindSpore* (pp. 83-93). Singapore: Springer Singapore.

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., & Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems, 32*.

Lim, B., Arık, S.Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting, 37*, 1748–1764.

Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.

Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NIPS), 30*, 4765–4774.

Mantua, N. J., & Hare, S. R. (2002). The Pacific Decadal Oscillation. *Journal of Oceanography, 58*, 35–44.
Marengo, J. A., Cunha, A. P., Cuartas, L. A., Leal, K. R. D., Broedel, E., Seluchi, M. E., … Bender, F. (2021). Extreme drought in the Brazilian Pantanal in 2019–2020: Characterization, causes, and impacts. *Frontiers in Water, 3*, Article 639204.
Mehdizadeh, S., Behmanesh, J., & Khalili, K. (2018). New Approaches for Estimation of Monthly Rainfall Based on GEP-ARCH and ANN-ARCH Hybrid Models. *Water Resources Management, 32*, 527–545.
Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., … Hersbach, H. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data, 13*, 4349–4383.
Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., … Smith, C. A. (2016). The Pacific Decadal Oscillation, Revisited. *Journal of Climate, 29*, 4399–4427.
Nillesen, E. E., & Verwimp, P. (2009). Grievance, commodity prices and rainfall: A village-level analysis of rebel recruitment in Burundi.
Pan, Y., Zhang, L., Yan, Z., Lwin, M. O., & Skibniewski, M. J. (2021). Discovering optimal strategies for mitigating COVID-19 spread using machine learning: Experience from Asia. *Sustainable Cities and Society, 75*, Article 103254.
Pan, Y., & Zhang, L. (2023). Integrating BIM and AI for Smart Construction Management: Current Status and Future Directions. *Archives of Computational Methods in Engineering, 30*, 1081–1110.
Papalaskaris, T., Panagiotidis, T., & Pantrakis, A. (2016). Stochastic monthly rainfall time series analysis, modeling and forecasting in Kavala City, Greece, North-Eastern Mediterranean Basin. *Procedia Engineering, 162*, 254–263.
Pepler, A., Timbal, B., Rakich, C., & Coutts-Smith, A. (2014). Indian Ocean dipole overrides ENSO's influence on cool season rainfall across the Eastern Seaboard of Australia. *Journal of Climate, 27*, 3816–3826.
Ponnoprat, D. (2021). Short-term daily precipitation forecasting with seasonally-integrated autoencoder. *Applied Soft Computing, 102*, Article 107083.
Rodriguez, H., Medrano, M., Rosales, L. M., Peñuñuri, G. P., & Flores, J. J. (2020). Multi-step forecasting strategies for wind speed time series. In *2020 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)* (Vol. 4, pp. 1-6).
Shen, G., Tan, Q., Zhang, H., Zeng, P., & Xu, J. (2018). Deep learning with gated recurrent unit networks for financial sequence predictions. *Procedia Computer Science, 131*, 895–903.
Shojaei, S., Saniee Abadeh, M., & Momeni, Z. (2023). An evolutionary explainable deep learning approach for Alzheimer's MRI classification. *Expert Systems with Applications, 220*, Article 119709.
Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. In *International conference on machine learning* (pp. 3145-3153): PMLR.
Simmonds, I., & Richter, T. (2000). Synoptic comparison of cold events in winter and summerin Melbourne and Perth. *Theoretical and Applied Climatology, 67*, 19–32.
SM. (2021). Shanghai Metro: Important announcement of operation information: service suspension. https://m.weibo.cn/1742987497/4662834027562911 (in Chinese, accessed 4 Oct 2022).
Smilkov, D., Thorat, N., Kim, B., Viégas, F., & Wattenberg, M. (2017). Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*.
Sturman, A. P., & Tapper, N. J. (1996). *The weather and climate of Australia and New Zealand*. USA: Oxford University Press.
Su, Q., & Iliadou, E. (2022). Predicting and explaining hearing aid usage using encoder-decoder with attention mechanism and SHAP. In *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (pp. 308–315).
Suarez, M. J., & Schopf, P. S. (1988). A delayed action oscillator for ENSO. *Journal of Atmospheric Sciences, 45*, 3283–3287.
Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic Attribution for Deep Networks. In P. Doina & T. Yee Whye (Eds.), *Proceedings of the 34th International Conference on Machine Learning* (Vol. 70, pp. 3319–3328). Proceedings of Machine Learning Research: PMLR.
Tao, L., He, X., Li, J., & Yang, D. (2021). A multiscale long short-term memory model with attention mechanism for improving monthly precipitation prediction. *Journal of Hydrology, 602*, Article 126815.
Trenberth, K. E., & Shea, D. J. (2005). Relationships between precipitation and surface temperature. *Geophysical Research Letters, 32*.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NIPS), 30*, 5998–6008.
Venkatachalam, K., Trojovský, P., Pamucar, D., Bacanin, N., & Simic, V. (2023). DWFH: An improved data-driven deep weather forecasting hybrid model using Transductive Long Short Term Memory (T-LSTM). *Expert Systems with Applications, 213*, Article 119270.
Wang, Y., Yuan, Z., Liu, H., Xing, Z., Ji, Y., Li, H., … Mo, C. (2022). A new scheme for probabilistic forecasting with an ensemble model based on CEEMDAN and AM-MCMC and its application in precipitation forecasting. *Expert Systems with Applications, 187*, Article 115872.
Williams, A. A., & Stone, R. C. (2009). An assessment of relationships between the Australian subtropical ridge, rainfall variability, and high-latitude circulation patterns. *International Journal of Climatology: A Journal of the Royal Meteorological Society, 29*, 691–709.
Yang, E., Zhang, H., Guo, X., Zang, Z., Liu, Z., & Liu, Y. (2022). A multivariate multi-step LSTM forecasting model for tuberculosis incidence with model explanation in Liaoning Province, China. *BMC Infectious Diseases, 22*, 490.
Zhang, W., Xie, J., Wan, G., & Tong, M. (2021). Single-step and multi-step time series prediction for urban temperature based on LSTM model of TensorFlow. In *2021 Photonics & Electromagnetics Research Symposium (PIERS)* (pp. 1531–1535).
Zhang, Y.-G., Tang, J., He, Z.-Y., Tan, J., & Li, C. (2021). A novel displacement prediction method using gated recurrent unit model with time series analysis in the Erdaohe landslide. *Natural Hazards, 105*, 783–813.