Engineering Applications of Artificial Intelligence 165 (2026) 113426

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A robust artificial intelligence informed over complete rational dilation wavelet transform technique coupled with deep learning for long-term rainfall prediction

Mohammed Diykh [a,b], Mumtaz Ali [a,b,e,*], Aitazaz Ahsan Farooque [b,f], Anwar Ali Aldhafeeri [c,**], Mehdi Jamei [b], Abdulhaleem Labban [d]

[a] UniSQ College, University of Southern Queensland, QLD, Australia
[b] Canadian Centre for Climate Change and Adaptation, University of Prince Edward Island, St Peters, PE, Canada
[c] Department of Mathematics and Statistics, Faculty of Science, King Faisal University, P.O. Box 400, Al-Ahsa, 31982, Saudi Arabia
[d] Department of Meteorology, Faculty of Environmental Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[e] Scientific Research Centre, Al-Ayen Iraqi University, Thi-Qar, 64001, Iraq
[f] Faculty of Sustainable Design Engineering, University of Prince Edward Island, Charlottetown, PE, Canada

ARTICLE INFO

Keywords:
Artificial intelligence
Overcomplete rational dilation discrete wavelet transform
Rainfall
Prediction
Drought
Climate warming

ABSTRACT

The intensity of heavy rainfall, driven by climate change, has significant effects worldwide, including flash flood, droughts, water degradation, landslides and crop damages. To ameliorate these impacts, accurate forecasting is crucial to address the dynamic nature of rainfall for sustainable utilization. But the non-linearity inherited within the rainfall significantly influence the model precision. Artificial Intelligence (AI) models have shown promising results in detecting complex rainfall patterns. This paper proposed a hybrid model using overcomplete rational dilation discrete wavelet transform (ORDWT) integrated with autoregressive integrated moving average (ARIMA) and long-short-term memory (LSTM), constructing ORDWT-ARIMA-LSTM to forecast one-month ahead rainfall. The ORDWT provides multi-scale decomposition and better shift-invariance, while ARIMA with LSTM captures complementary dynamics across ORDWT coefficients, lowering errors. Aiming to extract more representative features, the ORDWT coefficients are investigated, and then sent to the ARIMA-LSTM for prediction. The ORDWT–ARIMA–LSTM achieved highest performance for Melbourne Airport: Root Mean Square Error (RMSE) = 2.9, Mean Absolute Error (MAE) = 1.93, RSE = 0.215, Willmott's Index (WI) = 0.990, Nash–Sutcliffe Index (ENI) = 0.970; Melbourne Botanical Gardens: RMSE = 3.84 MAE = 2.65, RSE = 0.287, WI = 0.710, ENI = 0.962; and Preston Reservoir: RMSE = 3.94 MAE = 2.87 RSE = 0.310, WI = 0.973, ENI = 0.971. The ORDWT–ARIMA–LSTM reduced RMSE by 4.5 % and MAE by 5.3 % on average across stations against comparing models. Results confirmed the efficiency of ORDWT–ARIMA–LSTM in rainfall forecasts, providing valuable support in weather, water management, droughts and floods.

1. Introduction

Forecasting accurate rainfall is crucial, specifically in most Australian states due to substantial climate variability (Bagirov et al., 2017; Abbot and Marohasy, 2018). The agriculture sector is primarily relies on the availability of water resources, therefore, accurate rainfall forecasting is essential for efficient policy planning to secure a sustainable management of water systems (Bagirov et al., 2017; Acworth et al., 2024).

Victoria, Australia, experienced three consecutive years of severe drought from 2017 to 2019. During this period, rainfall across the state was substantially below average. The drought damaged agriculture threatened water supplies and ecosystems and became known as the "Tinderbox Drought" for its link to heightened fire risk. Fig. 1 shows its effects across the Murray Darling Basin (MDB), a vital region for Australia's food security and export economy: it accounts for about 40 % of Australia's agricultural production and is home to approximately 2.3 million people. This region was significantly affected by the Tinderbox

* Corresponding author. UniSQ College, University of Southern Queensland, QLD, Australia.
** Corresponding author.
E-mail addresses: mumtaz.ali@usq.edu.au (M. Ali), aaaldhafeeri@kfu.edu.sa (A.A. Aldhafeeri).

https://doi.org/10.1016/j.engappai.2025.113426
Received 15 July 2025; Received in revised form 5 November 2025; Accepted 1 December 2025
Available online 11 December 2025

Drought (Devanand et al., 2024).

The monthly rainfall is considered to be used because it provides better understanding about shape, patterns, and the challenges during a year as compared to the seasonal (Chhetri et al., 2020). This information is crucial to describe the rainy season's duration, onset, and the amount of rainfall is vital for agricultural decision-makers to improve resource management and optimization (Tao et al., 2021). Generally, rainfall prediction is categorized into physical and data-driven approaches. The physical approaches are based on the physics for processing rainfall, while the data-driven based approach works on historical data to predict future values (Abbot and Marohasy, 2014). Moreover, the data-driven methods have been increasingly adopted due to their advantages over physical methods to provide critical information for seasonal rainfall and improved prediction accuracy for practical applications (El-Shafie et al., 2012).

Decomposition techniques have been widely adopted to predict rainfall, for example (Paul et al., 20025) applied complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose time series and the outputs of CEEMDAN were sent to a combination of RNN, LSTM and GRU based on boosting ensemble model. That study was tested on annual level rainfall dataset collected from a limited set of areas. In addition, authors did not specify an explicit hyperparameter-tuning protocol which makes the configuration hard to reproduce (Cramer et al., 2018). designed a decomposition Genetic Programming (DGP) based on SVM. The study did not report computational requirements (Jamei et al., 2023). suggested a time varying filter-based empirical mode decomposition model, with a singular Valued Decomposition based on the Encoder Decoder Bidirectional. Combining multi-models could add considerable complexity, increase overfitting risk while provide only point forecasts without calibrated uncertainty (He et al., 2022). employed seasonal-trend decomposition based on GRU, Light-GBM. Their study tested on daily meteorology data collected in Cairns, Australia, from January 1, 2000 to December 31, 2020. Although the study obtained promising results, there was a risk of information leakage into validation or testing sets as the decomposition was applied on the full dataset. However, the proposed model provides a redundant, shift-invariant, and adaptive time frequency representation that captures finer signal variations while maintaining robustness to noise and irregularities. Compared with the mentioned methods the ORDWT offers lower computational complexity, better stability, and superior preservation of temporal dynamics in nonstationary data. Furthermore, while the ARIMA, and LSTM have been used previously, our novelty lies in integrating ORDWT with hybrid ARIMA-LSTM. This improves ARIMA's ability to model linear trends and enhances LSTM's learning of nonlinear dependencies, reducing sensitivity to outliers and improving generalization. These innovations collectively highlight the unique contribution of the proposed ORDWT–ARIMA–LSTM model, distinguishing it from existing hybrid approaches.

Advanced deep learning are data-driven models, designed and utilised for rainfall forecast (Basha et al., 2020). For example (Chhetri et al., 2020), suggested a hybrid deep learning model named BLSTM-GRU to combine bidirectional LSTM with GRU to predict monthly rainfall. Tao et al. (Tao et al., 2021) applied a multiscale LSTM integrated with attention mechanism for improving monthly precipitation prediction (Fahad et al., 2023). designed an optimized GRU model for rainfall forecasting using climatic variables on Pakistan. Nayak et al. (Nayak and Ghosh, 2013), established a support vector machine (SVM) based model to forecast rainfall where they showed the weather patterns exhibited different characteristics before the rainfall at nighttime as compared to the day time. Gupta et al. (Gupta and Gupta, 2019), proposed ensemble-based techniques to remove noise from rainfall time series data. De Vos et al. (De Vos and Rientjes, 2005) designed multi-layer feedforward artificial neural networks (ANNs) for rainfall forecasting while Aderyani et al. (Aderyani et al., 2022), in another study integrated support for long-short-term memory (LSTM), vector regression (PSO-SVR), and convolutional neural network (CNN) to predict 5- and 15-min ahead rainfall. (Trivedi et al., 2025) suggested a robust model for rainfall prediction. In that study, high-resolution (3 km) time series data from the Weather Research and Forecasting (WRF) model were used. Two models, U-Net (an attention-based U-Net architecture) and KU-Net (an attention-based kernelized U-Net architecture), were adopted. The authors trained their models on 24 monsoon cases between 2007 and 2018 and tested them on two cases from July 2023 and August 2023 (Sharma et al., 2024). suggested that deep learning models can deliver accurate categorical and spatial rainfall predictions up to 96 h lead time. Authors trained their model on historical monsoon cases (2007–2018) and test on two real-time cases (July 2023 and August 2023). Their results showed that deep learning models reduced mean absolute error by ~8 mm relative to the deterministic WRF baseline across districts (Trivedi et al., 2025). introduced a novel U-Net approach for rainfall prediction in the Assam region. In that study, two

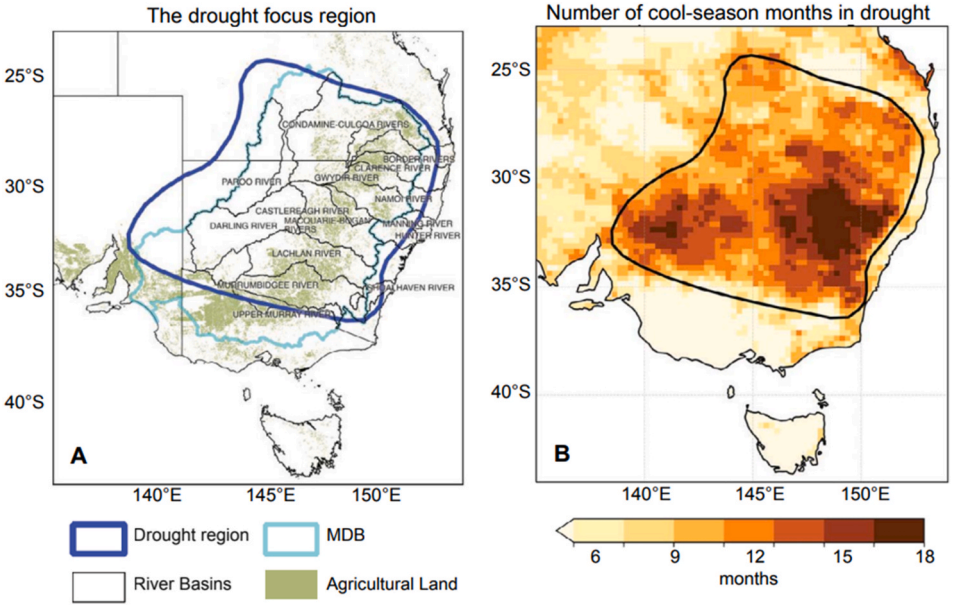

Fig. 1. Regional affected by drought in Australia during the period from 2017 to 2029 (Devanand et al., 2024).

encoders were involved in the designed model to individually extract features correlated with both district-scale and overall Assam rainfall patterns. Waqas et al., 2023 designed an advanced Seasonal precipitation prediction approach for Eastern Thailand. In that study LSTM was integrated with Recurrent Neural Networks (RNNs) and WT. their model obtained excellent results compared to traditional LSTM models across all seasons. Another study by Waqas et al., 2023 combined WT with Radial Basis Function Neural Networks and with Long Short-Term Memory Recurrent Neural Networks. In that study, authors compared their model with several state-of-the-art models, and their findings showed that the combination of WT with Radial Basis Function Neural Networks and with Long Short-Term Memory Recurrent Neural Networks delivered superior prediction results compared with the previous models.

Linear models-based approaches have also been examined for prediction tasks. (George et al., 2016) It applied bivariate generalized linear model. In that study, A logistic regression model was tested to predict rainfall amount on wet days. (Patel et al., 2025) designed liquid state machine learning model. (Shin et al., 2025) applied linear model-based feature engineering based on CEEMD to improve prediction accuracy. (Cui et al., 2023) suggested an ensemble feature selection model. Their model integrated multiple statistical learning approaches to select the most appropriate features. Wang et al., 2017 demonstrated the benefit of feature selection for prediction tasks. The most informative features were identified using the maximal information coefficient then, decomposed with improved variational mode decomposition and aggregated using similar sample entropy.

Rainfall prediction is a challenging task as it contains a complex relationship among numerous climatic variables. Traditional machine learning and deep learning-based models often fail to capture complex patterns and relationships, leading to increased prediction errors and limitations. Much rainfall prediction research (Waqas et al., 2023; Baig et al., 2024) has been tested at regional or national scales which ignores local differences. However, our study targets small-scale, site-specific prediction at three Victorian locations: Melbourne Airport, Melbourne Botanical Gardens, and Preston Reservoir regions that are only tens of kilometres apart yet differ in exposure, land cover, and local circulation. These micro-scale differences generate distinct rainfall behaviours that regional models often miss. In addition, most of hybrids models applied such as BLSTM, GRU, LSTM did not separate linear and nonlinear behaviours and patterns. As a result, they suffer overfit, lack generalisation, and yield unstable errors during extremes. We address these limitations by integrating ORDWT with hybrid ARIMA–LSTM model. The ORDWT produces shift-invariant and sub bands that isolate scales, while the combination of ARIMA–LSTM capture the linear autocorrelation within sub bands, and the nonlinear residual dynamics providing a simple, complementary design to drought forecasting. This study leverages the more novel decomposition technique with an advanced deep learning architecture to improve model interpretability and uncover hidden factors affecting monthly rainfall. Additionally, there is no work performed on the designing of a developed intelligent model managing the issues of non-linearity and non-stationarity associated within the rainfall data.

The ORDWT model improves prediction at three stations by producing shift-invariant, non-decimated sub bands that were robust to monthly prediction and boundary effects. This produced stable linear components for the ARIMA and clean nonlinear residuals for the LSTM. The proposed ORDWT is compared to classical decomposition methods empirical mode decomposition (EMD) (Zeiler et al., 2010), and discrete wavelet transform (DWT) (Othman and Zeebaree, 2020). The ORDWT overcomes limitations of these techniques, for example, the DWT halves the sample at each level making coefficients shift-variant and degrading sub-band modelling. While EMD produced mode mixing and endpoint sensitivity, especially at stations with heavier tails, which blurred scale separation and weakened the linear, and nonlinear split. Cross three stations, ORDWT scored the largest error reductions because its redundancy preserves peak energy without aliasing at the smoother site (Melbourne Airport). Together, these factors explain why ORDWT provides more accurate, more stable sub-band features, enabling the hybrid ARIMA–LSTM to generalize better across stations.

The novelty of this paper lies in two aspects. Firstly, employing ORDWT, an effective technique for analysing nonstationary, nonlinearity and noisy data. Secondly, the construction of hybridised ORDWT with the ARIMA-LSTM model where the ARIMA is an effective tool for linear feature extraction, and LSTM can capture nonlinear features to predict one month ahead rainfall. Based on the authors knowledge, this work is among the first to employ ORDWT with explainable deep learning to predict monthly rainfall in Australia to produces accurate predictions and improves understanding of how hidden patterns impact rainfall.

The main objectives of this work comprise; (1) decompose the rainfall time series into several sub-bands (i.e., coefficients) using ORDWT to resolve non-linearity and the non-stationarity issues; (2) combine ORDWT with ARIMA model to extract linear features; (3) hybridized with LSTM model to capture nonlinear features and predict monthly rainfall; and (4) examine the prediction ability of the proposed ORDWT-ARIMA-LSTM model in rainfall prediction under different climatic regions in Australia. For comparison, We benchmarked the proposed model with SVR, LSTM, BPNN, BiLSTM, BiGRU and GRU models and hybrid counterpart models EMD-ARIMA-LSTM, EMD-SVR, EMD-GRU, EMD-BPNN, EMD-BILSTM, EMD-BI-GRU, DWT-ARIMA-LSTM, DWT-SVR, DWT-GRU DWT-BPNN, DWT-BILSTM, DWT-BI-GRU, ORWDT-SVR, ORWDT-GRU, ORDWT-BPNN, ORDWT-BILSTM, ORDWT-BI-GRU, and ORDWT-ARIMA-LSTM. The developed approach is assessed to forecast monthly rainfall for Melbourne Airport, Melbourne Botanical Gardens, and Preston Reservoir stations in Victoria State, Australia. The next section of the paper outlines the materials and methods (section 2), followed by the proposed modelling framework and strategy (section 3), Results and discussion (section 4), Further Discussion, limitations, and future work (section 5), and Conclusions (section 6).

2. Materials and methods

2.1. Study area and data description

Rainfall is critical for nourishing life and ecosystems, providing freshwater for agriculture, environment, industry, and human consumption. It also affecting and changing temperatures and maintaining different plant and animal life. Rainfall is facing challenges in Victoria, Australia from climate change, comprising decreasing cool season and increased intensity of extreme rainfall events. The agricultural sector profoundly depends on rainfall for crop and livestock production. The farmers facing sever challenges due to rainfall patterns variations, including increased irregularity and potential for extreme events. Moreover, the extreme rainfall events can precedent to flash flooding, specifically in urban areas, while prolonged dry periods increase the risk of bushfires. Agriculture sector substantially contributes to the Victorian economy, export food and fibre products, approximately $10.8 billion in 2022-23. Therefore, Melbourne Botanical Gardens, Preston Reservoir and Melbourne Airport stations in Victoria were chosen as three pilot study locations where the relevant dataset were acquired to forecast monthly rainfall. The topographical visualization of these stations is presented in Fig. 2.

As described, the rainfall data were gathered at monthly basis for Melbourne Botanical Gardens, Preston Reservoir and Melbourne Airport stations during the period from January 01, 1960 until December 12, 2023. These datasets were obtained from Bureau of Meteorology, Australia (BoM) (Webb, 2010). The input predictors consist of rainfall time-series with the statistical description recorded in Table 1 which comprises of minimum, maximum, mean, range, standard deviations, skewness, and kurtosis. Fig. 3 shows time series of rainfall trends at

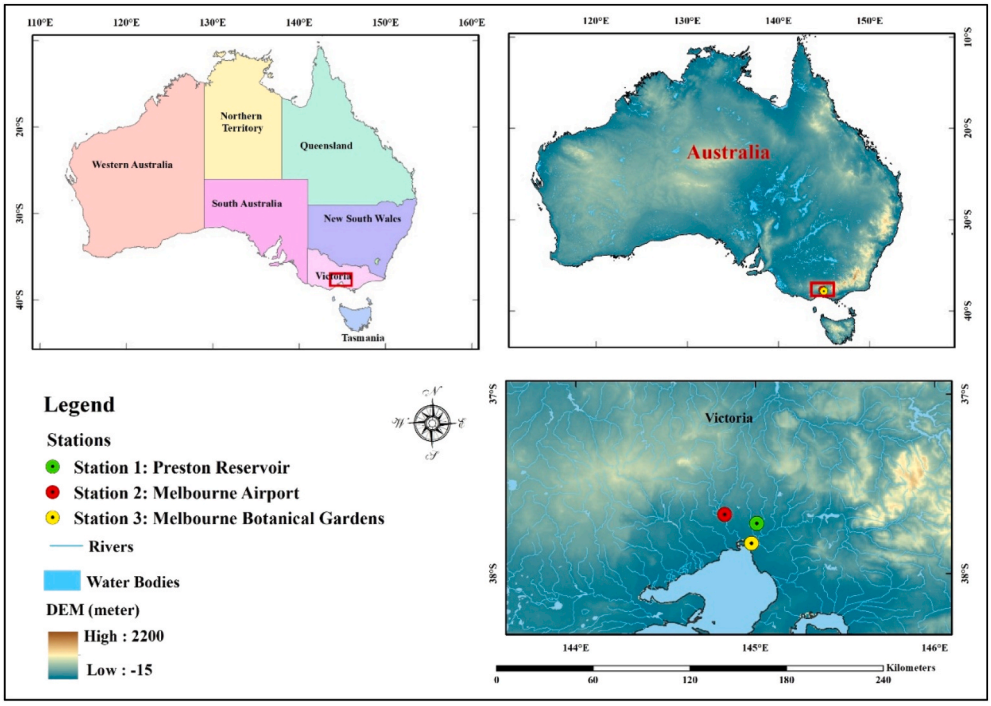

Fig. 2. Map of the stations.

Table 1
Geographic and statistical characteristics of the monthly rainfall data.

Geographic and Statistical index	Melbourne Botanical Gardens	Preston Reservoir	Melbourne Airport
Longitude (°E)	144.98° E	145.01° E	144.83° E
Latitude (°S)	37.83° S	37.72° S	37.67° S
Elevation (m)	29.0	94.0	113.0
Minimum	0.8	0.8	1
Maximum	15.4	215.4	200.6
Range	157.7	214.6	117.9
Mean	98.0	56.3	47.4
Std. Deviation	33.6	34.4	28.9
Skewness	1.0	1.2	1.2
Kurtosis	1.0	1.8	1.9

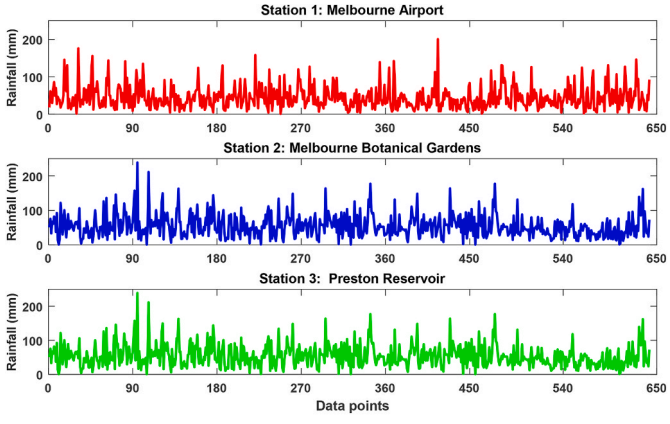

Fig. 3. Time series of rainfall trends at Melbourne Airport, Melbourne Botanical Gardens, and Preston Reservoir stations.

Melbourne Airport, Melbourne Botanical Gardens, and Preston Reservoir stations. The indicators show that the Preston Reservoir stations has recorded the maximum 215.4-mm (mm) rainfall in January 1963, while the maximum rainfall in Melbourne Airport stations was 200.6 mm during February 2005 and Melbourne Botanical Gardens with 238.7 mm rainfall in February 1972.

2.2. Dyadic wavelet theory

A wavelet basis includes a fast-decaying function in the time domain. A scaling of mother wavelet function $\varphi(t)$ with a and translating it by b is employed to obtain time-frequency atoms of (wavelet theory) WT (Bayram and Selesnick, 2007). The wavelet transforms of a signal X is defined as

$$WT(a,b) = \frac{1}{\sqrt{a}}\left(x(t),\ \psi_{a,b}(t)\right) = \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-b}{a}\right)dt \tag{1}$$

Dyadic wavelet theory used translation and scaling factors to ease invertibility and computation. Scaling factors are selected to be integers of power two. The wavelet function $\psi(t)$ and scaling function $\varphi(t)$ are represented by $h(n)$ and $g(n)$ respectively (Bayram and Selesnick, 2009). The wavelet function $\psi(t)$ and scaling faction $\varphi(t)$ should be satisfied 'the following conditions:

$$\varphi(t) = \sum\nolimits_{n\in z} h(n)\varphi(2t-n) \tag{2}$$

$$\psi(t) = \sum\nolimits_{n\in z} g(n)\psi(2t-n) \tag{3}$$

The set $\left\{\psi_{j,m}(t) = 2^{-j/2}\left(\frac{\psi\left(t-2^{j}n\right)}{2^{j}}\right)\right\}_{j,m\in z}$ is formed an orthonormal basis for $I^2(R)$. Mallat's iterated model is employed to form Dyadic wavelet. The decomposition of Dyadic wavelet theory is described as recursively convolving the low pass filter $h(n)$ and bandpass filter $g(n)$ with a subsequently down sample the filter series in each channel by factor of 2.

The frequency resolution of Dyadic wavelet theory sub-bands is gradually processed via splitting the passband of its previous sub-bands. The produced filter frequency responses are not ultimate and cause frequency aliasing problems. As an improvement, dual-tree complex wavelet transform (DTCWT) is designed as an alternative of DWT

(Bayram and Selesnick, 2007). It uses wavelet functions with two separate orthonormal Dyadic wavelet theory bases to construct an approximate Hilbert transferor pair.

$$\psi^{\xi m}(t) \approx H\left[\psi^{Re}(t)\right] \tag{4}$$

where $H[.]$ refers to the Hilbert transform.

2.3. Over-complete rational dilation wavelet transform (ORDWT)

Over complete-rational dilation wavelet transform (ORDWT) is a type of wavelet transform. The ORDWT is designed to improve the limitations of empirical mode decomposition and Discrete wavelet transform, and dyadic wavelet (Bayram and Selesnick, 2007; Bayram and Selesnick, 2009, Chen et al., 2012, Kamble et al., 2023). It reveals adjustable and finer frequencies to reduce nearly shift-invariant and frequency-aliasing. The ORDWT model utilizes two filter banks. The dilation factor relies on the rational value of p/q, with the following condition $q < p$. This study set the dilation factor to $1 < q/p < 2$. The wavelet function and scaling function are associated with the scaling relation and rational dilation, as shown in Eq. 2, and 3 as following.

$$\varphi(r) = \left(\frac{q}{p}\right)^{1/2} \sum_{i\in Z}^{1} h_0(i)\varphi\left(\frac{q}{p}r - i\right) \tag{5}$$

$$\psi(r) = \left(\frac{q}{p}\right)^{1/2} \sum_{i\in Z}^{1} h_1(i)\psi\left(\frac{q}{p}r - i\right) \tag{6}$$

where $\varphi(r)$ refers to the scaling function, $\psi(r)$ denotes the wavelet function ,$h_0(i)$ refers a scaling function filter, and $h_1(i)$ indicates the wavelet function filter. The ORDWT decomposes the input time series into bands according to p, q and r parameters. The input time series is decomposed at each decomposition level into a low pass band and a stop band. The functions $H_1(\omega)$ and of $H_0(\omega)$ are the frequency outcomes of $h_0(n)$ and$h_1(n)$ as shown in Fig. 4. To compute the frequency outcomes, $H_0(\omega)$ and $H_1(\omega)$ are employed as shown in Eqs. (5) and (6):

$$H_0(\omega) = \begin{cases} \sqrt{pq}, \omega \in \left[0, \left(1 - \frac{1}{r}\right)\frac{1}{p}\pi\right] \\ \sqrt{pq}\,\theta_{H_0}\left(\frac{(\omega - a)}{b}\right), \omega \in \left[\left(1 - \frac{1}{r}\right)\frac{1}{p}\pi, \pi\right] \\ 0, \omega \in \left[\frac{1}{q}\pi, \pi\right] \end{cases} \tag{7}$$

$$H_1(\omega) = \begin{cases} 0, \omega \in \left[0, \left(1 - \frac{1}{r}\right)\pi\right] \\ \sqrt{r}\theta_{H_1}\left(\frac{(\omega - pa)}{pb}\right), \omega \in \left[\frac{r-1}{r}\pi, \frac{p}{q}\pi\right] \\ \sqrt{r}, \quad \omega \in \left[\frac{p}{q}\pi, \pi\right] \end{cases} \tag{8}$$

where a, b and θ are defined as:

$$a = \frac{\left(1 - \frac{1}{r}\right)\pi}{p} \tag{9}$$

$$b = \frac{1}{q} - \frac{\left(1 - \frac{1}{r}\right)1}{p} \tag{10}$$

$$\theta(\omega) = \frac{(1 + cos(\omega))}{2}\sqrt{2 - cos(\omega)}, for\ \omega \in [0, \pi] \tag{11}$$

We employed function $\theta(\omega)$ to obtain the down pass filter $H_0(\omega)$ and top pass bands $H_1(\omega)$ using Daubechies wavelet filter. Orthonormal filters are essential to design any sharp edges, however, they can cause slow decaying rate of the $\psi(t)$ and $\varphi(t)$ which should be prevented.

To rebuild the decomposed time series data, the conditional holds of the ORDWT $|H_0(\omega/p)|^2/(pq) + |H_1(\omega)|^2/r = 1$ are checked and indicates $x(n) = \overline{x}(n)$ as presented in Fig. 3. The reconstruction process of the decomposed time series is defined as follows (Bayram and Selesnick, 2007, 2009).

1. The length of the stop band series is cut off as the decomposition proceeds. To recapture the length of the stop band, the single branch reconstruction (SBR) is employed to rebuild the sequences of the decomposition levels into equivalent wavelet sub-bands, as described in Fig. 5.

2. After decomposing the input time series to the jth level using the SBR, the time series is rebuilt into the jth^{-1} level band as described in Fig. 4 using Eq. (9).

$$\overline{x}(n) = \{A_j(n), D_j(n), D_{j-1}(n), \ldots, D_1(n)\} \tag{12}$$

where $A_j(n)$ denotes the rebuild approximation sub-bands of the time series and $D_j(n)$ refers to the jth rebuild sub-bands of the time series.

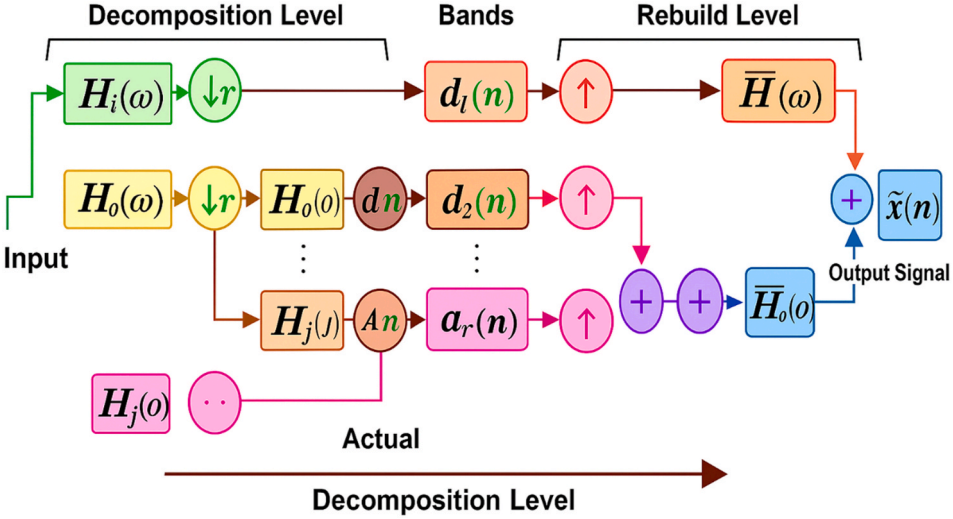

Fig. 4. Frequency component of input time series.

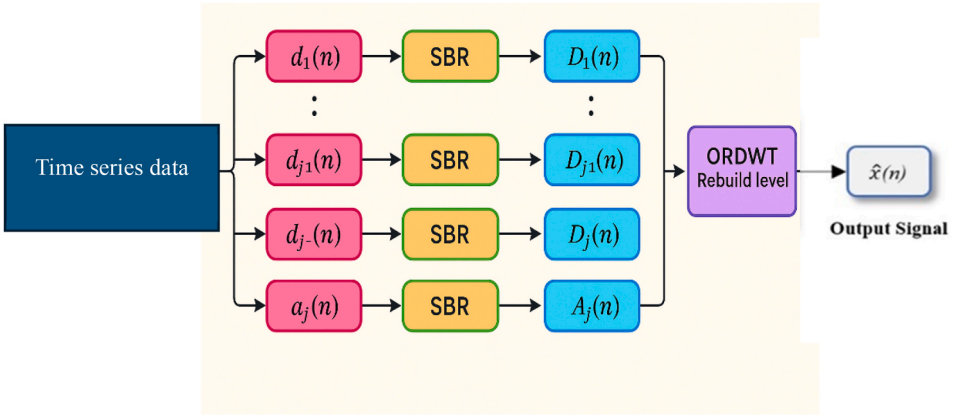

Fig. 5. Frequency component of input time series.

3. Two functions $\overline{H}_0(\omega)$ and $\overline{H}_1(\omega)$ are used to obtain the frequency responses of $\overline{h}_0(n)$ and$\overline{h}_1(n)$.

$$\overline{H}_0(\omega) = \frac{1}{p} h_0\left(\frac{\omega}{p}\right) \tag{13}$$

$$\overline{H}_1(\omega) = h_1(\omega) \tag{14}$$

where $\overline{h}_0(n)$ and$\overline{h}_1(n)$ are employed during the decomposition process.

In addition, for the jth decomposition level, the frequency responses, $\overline{H}_0(\omega)$ and $\overline{H}_1(\omega)$ are extended to:

$$\overline{H}_{0_j}(\omega) = \frac{1}{p_j} H_{0_j}\left(\frac{\omega}{p_j}\right), j \geq 1 \tag{15}$$

$$\overline{H}_{1_{j-1}}(\omega) = \begin{cases} H_1(\omega), j = 0 \\ \frac{1}{p_j} H_{0j}\left(\frac{\omega}{p_j}\right) H_1\left(\left(\frac{q}{p}\right)^j \omega\right), j \geq 1 \end{cases} \tag{16}$$

where

$$\overline{H}_{0_j}(\omega) = \begin{cases} \prod_{n=0}^{j-1} H\left(q_{j-1-n} p_n \omega\right), \omega \in \left[0, \frac{\pi}{q_j}\right] \\ 0, \omega \in \left[\frac{\pi}{q_j}, \pi\right] \end{cases}, j \geq 1 \tag{17}$$

To rebuild the decomposed time series using the ORDWT, the parameters of p, q, and r should satisfy the following conditions:

$$\left(1 - \frac{1}{r}\right)\frac{1}{p}\pi < \frac{\pi}{q} \Rightarrow \frac{p}{q} + \frac{1}{r} > 1 \tag{18}$$

This condition does not allow the passbands to be overlapping between bands.

2.4. Discrete wavelet transform (DWT)

DWT is one of the powerful signal processing techniques that is used to analyse and reveals hidden features of non-stationary signals. Due to the rainfall time series has a dynamic nature, we utilised DWT to comprise rainfall time series into various frequencies. DWT offers multi-resolution analysis, time-frequency localization, efficient computation, and better compression with less artifact as compared to other methods.

DWT decomposes the input data into low and high frequency bands. DWT utilizes low-pass filters (*LPF*) and high pass filter (*HPF*) to obtain DWT coefficients. Suppose a signal, the DWT can be defined as

$$C(t, e) = \sum_{m \in z} x(m) \Psi t, e(m) \tag{19}$$

where t denotes to scale, e for translation, and Ψ is the discrete wavelet function. Ψ is defined as

$$\Psi t, e = \left(\frac{1}{\sqrt{t}}\right) \times \Psi\left(\frac{m - e}{t}\right) \tag{20}$$

2.5. Empirical mode decomposition (EMD)

The EMD is a signal decomposition method to handle the non-stationarity and non-linearity issues in the data (Torres et al., 2011). The EMD method decompose the input data into resolved factors named intrinsic mode functions (IMFs) and a residual component. A short description of the CEEMD process for an undecomposed input $x(t)$, is presented below (Torres et al., 2011).

1) Decompose $s(t)$ via EMD technique with an operator $w_j(.)$ at jth mode where $u(.)$ Construct the local mean of $s'(t)$.
2) Calculate the mean of IMF_1:

$$IMF_1(t) = \frac{1}{\tau} \sum_{k=1}^{\tau} \psi_1(s(t) + \varepsilon . w_k(t)) \tag{23}$$

Where ψ_1 is the computation of kth IMF and ε indicates the amplitude regulation, required to acquire a suitable signal.

3) Hence, the first residual component or the remaining component results as follows:

$$R_1(t) = s(t) - IMF_1(t) \tag{24}$$

4) Next, the extraction of IMF_2 from the original signal, $s(t)$, can be extracted as:

$$IMF_2(t) = \frac{1}{\tau} \sum_{k=1}^{\tau} \psi_1[s(t) + \varepsilon . \psi_1(w_k(t))] \tag{25}$$

5) The above steps are repeated to obtain the $(\beta + 1)^{th}$ IMF component as:

$$IMF_{\beta+1}(t) = \frac{1}{\tau} \sum_{k=1}^{\tau} \psi_1\left[s(t) + \varepsilon . \psi_\beta(w_k(t))\right] \tag{26}$$

Lastly, the residuals are averaged which generally is the trend, demonstrate a gradual deviation around the long-term average.

2.6. Autoregressive integrated moving average (ARIMA)

The ARIMA model design to construct time series from multivariant time series. The model is known as Box Jenkins model (Zhang et al., 2024), consists of units: integrated (I), autoregressive (AR), and moving average (Ma). The ARIMA model works on the time series to analyse patterns and stationary behaviour (Fan et al., 2021). The general structure of ARIMA is defined as (p, d, q), where p refers to the AR, q denoted the moving average, and d denotes the difference. AR and MA are combined for non-stationary time series [20]. The AR, I, and MA terms are equal to or greater than to zero if they satisfy the following condition:

$$\varphi^d X_t = (1-\beta)^d \tag{27}$$

where β refers to backshift parameter, then we defined the ARIMA model as

$$\theta(\beta)(1-\beta)^d X_t = \theta(B)\varepsilon_t \tag{28}$$

$\theta(\beta) = 1 - \theta_1\beta - \theta_2\beta^2 - \ldots - \theta_p\beta^p, p$ refers to AR term

$\theta(\beta) = 1 - \theta_1\beta - \theta_2\beta^2 - \ldots - \theta_q\beta^q, q$ refers to MA term

$\varepsilon_t \sim WN(c, \sigma^2)$, refers to the difference $d > 0$, AIRMA model follows ARMA (p, q) when the difference $d = 0$.

2.7. Long short-term memory (LSTM)

The Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) model that is commonly utilised in several complex prediction issues. It consists of several integrated unites including input gate, output gate, forget gate, and memory cell (Yu et al., 2019). The LSTM has the ability to forget and remember information from previous timestep makes it more suitable for applications that involve long-term dependencies modelling.(Staudemeyer and Morris, 2019). To resolve the vanishing gradient problem associated with RNN, the LSTM learn from the data gathered over a long-term period. The first layer of the memory gate is responsible for eliminating redundant information, and it is described as follows:

$$f_t = \sigma\left(w_f \times X_t + Y_f \times h_{t-1} + p_f\right) \tag{29}$$

Where f_t refers to the prediction threshold at time t, w_f and Y_f denote to the weights, σ refers to the sigmoid activation function, X_t is the input time series, and p_f is the bias parameter, and h_{t-1} is the outcome value at time t.

The second input gate controls which the information should be moved from the current input set to the cell state (Sherstinsky, 2020). This involves i_t which defined *tanh* layer, as well as the value to create a new state value C_t. It is defined as

$$i_t = \sigma(w_i \times X_i + Y_i \times h_{t-1} + p_i) \tag{30}$$

$$C_t = \sigma(w_c \times X_c + Y_c \times h_{t-1} + p_c) \tag{31}$$

Where i_t refers to the input threshold at time T, p_i, p_c are bias parameters, and w_i, w_c, Y_i, Y are defined weights. As a result, the state of the cell can be modified at time t as:

$$C_t = f_t \times C_t + i_t \times p_o \tag{32}$$

The third layer is employed as output data in the current time step, and it is defined as

$$O_t = \sigma(w_0 \times X_c + Y_o \times h_{t-1} + p_o) \tag{33}$$

Where O_t denotes to the output threshold at time t.

Then the output value of the cell is expressed as

$$h_t = O_t \times \tanh(C_t) \tag{34}$$

where tanh is the activation function, and h_t is to the output value of the cell at time t. The input data is passed through all three gates. As a result, the significant information is output, and the unacceptable information is removed.

2.8. Bidirectional LSTM (BiLSTM)

Bidirectional LSTM (BiLSTM) is a modified version of the LSTM proposed by Schuster and Paliwal (1997). The BiLSTM has two LSTM operating in parallel where one works in backward direction $\overleftarrow{H}_{t_p}$ and the second as works in forward direction $\overrightarrow{H}_{t_p}$ (Yildirim, 2018; Jaseena and Kovoor, 2021). As the BiLSTM works in two directions, it has been used to understand the behaviour and context of complex time series data (Bian et al., 2020):

$$u_{t_p} = \psi\left(\overrightarrow{H}_{t_p}, \overleftarrow{H}_{t_p}\right) \tag{35}$$

2.9. Gated recurrent unit (GRU)

The GRU model was designed by (Bahdanau et al., 2014). Then the GRU model was modified by adding decoder, and encoder structures (Cho et al., 2014). It includes two gates update and reset gates. The update gate is responsible for determining the amount of data that should be retained from previous hidden layer for the next layer while the reset gate quantifies how much of past hidden state should be disregarded (Chung et al., 2014). The fundamental procedure is as follows:

$$r_t = \sigma(W_{ir}x_t + W_{hr}h_{t-1} + b_{ir} + b_{hr}) \tag{36}$$

$$z_t = \sigma(W_{iz}x_t + W_{hz}h_{t-1} + b_{iz} + b_{hz}) \tag{37}$$

$$n_t = \varnothing_h(W_{in}x_t + r_t \odot (W_{hn}h_{t-1+} + b_{hn}) + b_{in}) \tag{38}$$

$$h_t = (1 - z_t) \odot n_t + z_t \odot h_{t-1} \tag{39}$$

Where r_t, z_t, n_t, h_t represents reset gate vector, update gate vector, candidate activation vector and hidden states respectively, W, b, are weights and bias, $\varnothing_h$ is the hyperbolic tangent, σ is the sigmoid function.

2.10. Bidirectional GRU (BiGRU)

The Bidirectional GRU (BiGRU) is bidirectional version of the original GRU model. It processes sequential data in two ways forward and backward (Rana, 2016). The BiGRU was designed to solve complex problems. The structure of BiGRU allows to control the follow of data via network while preserving the complexity of the model (Wang, Skerry-Ryan et al. 2017; Wang et al., 2018). Fig. 6 shows the structural diagrams of LSTM, BiLSTM, GRU, and BiGRU models.

2.11. Support vector regressor (SVR)

SVR is a modified version of Support Vector Machines, designed for regression problems (Awad et al., 2015). Compared to classic deep learning model, SVR aims at reducing the error directly (Anand et al., 2020). Suppose a time series data $X = \{(x_1, y_1), (x_2, y_2), \ldots, (x_3, y_3)\}$

Where x_1 denotes to feature, and y_2 refers to the corresponding target. SVM model employs a function $F(x_i)$ that estimate y_i to an acceptance margin.

$$F(x_i) = w^t\sigma(x) + b \tag{40}$$

Where $w, and\ b$ are weight vector and bias term, $\sigma(x)$ maps the input features into high dimensional space.

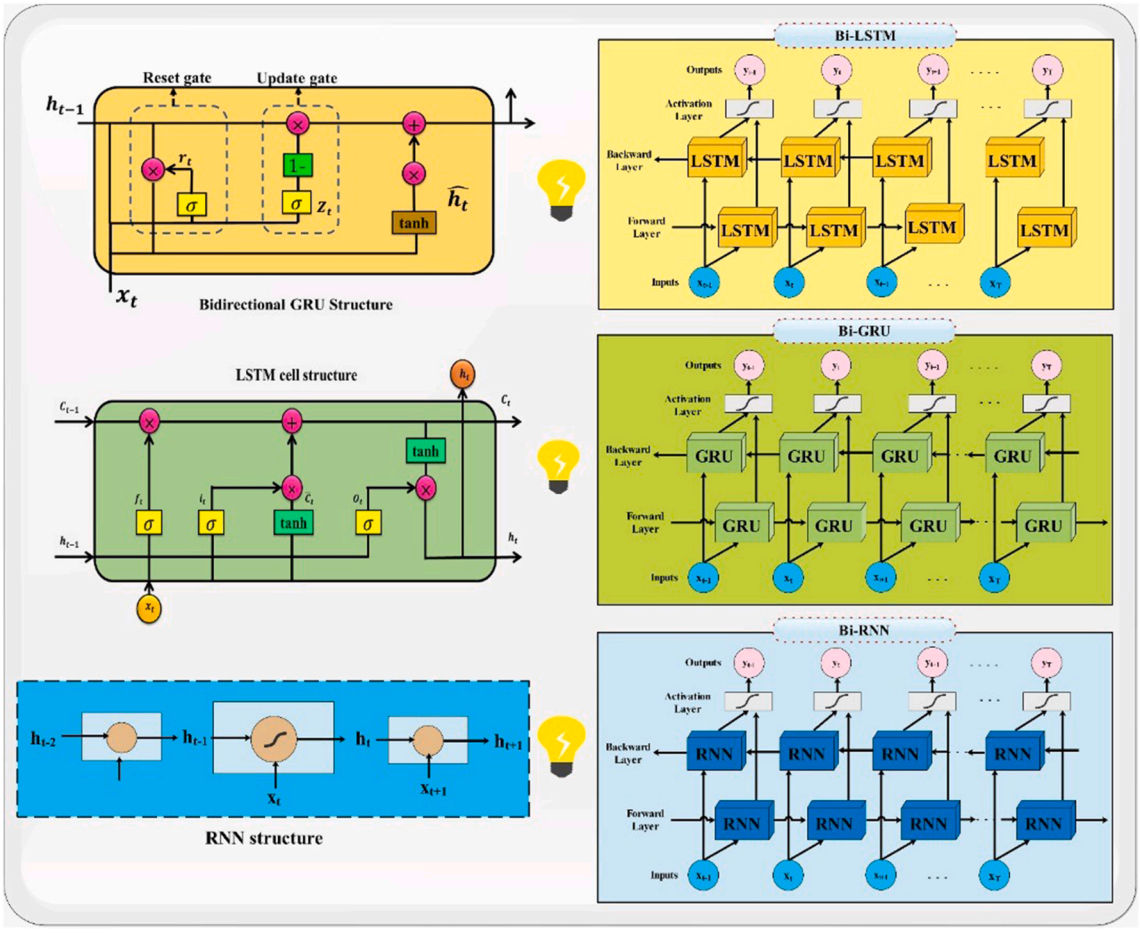

Fig. 6. Structural diagrams of LSTM, BiLSTM, GRU, and BiGRU models.

2.12. Backpropagation neural network (BPNN)

The BPNN is a multilayer feedforward neural network model which includes three layers named the hidden layers (HL), output layer (OL) and input layer (IL) (Buscema, 1998; Cilimkovic, 2015). Each layer responsible a special task for example, IL collects the raw features, HL computes the weighted while and the OL produces the final prediction (Buscema, 1998). BPNN is passed through two phases during the training process.

- Forward Propagation phase: in this phase, the input features are transferred through the model layers in which each neuron assigns a weighted. Then, an activation function is employed to creäte non-linearity. The last layer is used to compute the final output.
- Backward Propagation phase: It's also called Error Correction phase. It is compared the computed output with target output using a loss. The error rates are broadcasted backward via the network by the chain rule of differentiation. All weights are modified using gradient descent as

$$w_{ij}^{t+1} = w_{ij}^{t+1} - \rho \frac{\sigma E}{\sigma w_{ij}} \tag{41}$$

where w_{ij} refers to the weight, represents the weight, ρ denotes to the learning rate, and E represents the error function.

3. The proposed modelling framework and strategy

3.1. Development of the proposed ORDWT-LSTM-ARIMA model

3.1.1. Stage 1: the ORDWT method

The proposed model was designed to predict monthly ahead accumulated rainfall at each station (units: mm per month), not daily values, rapid rain rate, or intensity. All reported results based on one-month lead. Given the last L monthly totals up to month t, the proposed model forecasts the next month $t+1$ and thus the prediction is one month ahead. The proposed model strictly predicts, not reflecting estimation. The collected data were split chronologically into training, validation, and testing sets. All preprocessing and the ORDWT decomposition were fit on the training window only and applied forward to validation, and testing, so at each test time the model sees data only up to t and produces $\ddot{y}_t + 1$. While the proposed model reports the $h = 1$ month horizon, it can be extended to longer lead times using multi-output strategies.

At stage 1, the ORDWT is applied to decompose the input predictor data into several sub-bands to extract the hidden nonlinear and linear features in rainfall time series. The coefficients (i.e., sub-bands) of ORDWT are investigated and examined to identify the most essential characteristics where the ORDWT separates the high-frequency component of the input data. Fig. 7 shows the ORDWT coefficients at 5 levels for the Melbourne Airport, Preston Reservoir; and Melbourne Botanical Gardens stations.

To comprehensively evaluate the performance of the ORDWT method, it is vital to compare with discrete wavelet transform (DWT) and Empirical Mode Decomposition (EMD). This comparison can

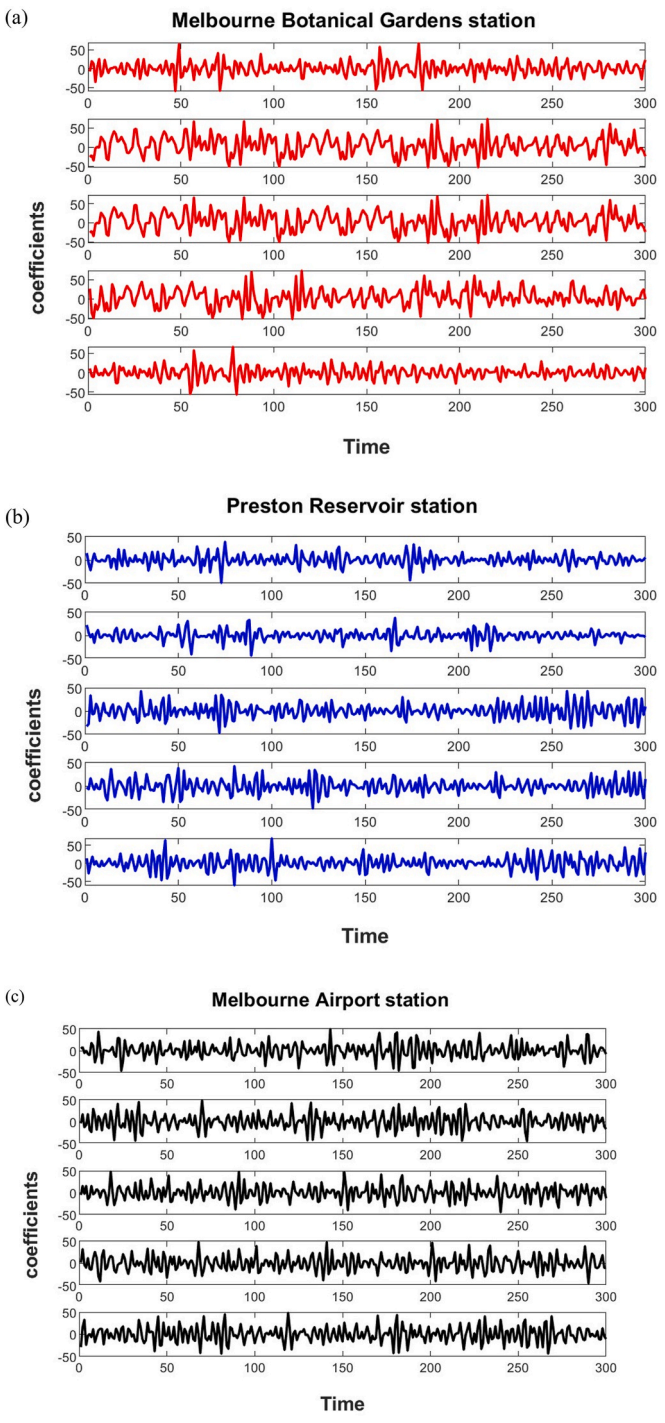

Fig. 7. The decomposed sub-bands based on ORDWT methods of rainfall for (a) Melbourne Airport; (b) Preston Reservoir; and (c) Melbourne Botanical Gardens.

highlight the strengths and limitations of the proposed model, enhancing the reliability and clarity of our proposed ORDWT-based ARIMA-LSTM approach for rainfall prediction.

3.1.2. Stage 2: hybridization of ARIMA and LSTM models

The dataset was pre-processed using z-score normalization as it was proved to be robust to outliers. To avoid shuffling and ensure no temporal overlap, the time series was divided strictly in chronological order into 60 % training, 20 % validation, and 20 % testing sets. It is not needed to randomly partition the data for time series forecasting; in fact, it is strongly discouraged. Randomly splitting the data would disrupt the temporal order, causing data leakage where future information is used to train the model, leading to unrealistic and inaccurate predictions. Instead, data should be split chronologically, with a portion for training and a later portion for testing. The decomposed sub-bands (i.e., coefficients) of rainfall time series produce linear and nonlinear features in terms of patterns and behaviour, and it can be defined as

$$X_t = L_t + N_t + \varepsilon_t \tag{42}$$

Where L_t and N_t denote to the linearity and non-linearity patterns at time t, respectively, and ε_t is the error term. The ARIMA is used to model the linear patterns in the rainfall time series, while the LSTM is employed to model the nonlinear features. The ORDWT based sub-bands are passed into the ARIMA and LSTM model to design the hybrid ORDWT-ARIMA-LSTM model for monthly rainfall prediction. The AIRMA is applied to extract the linear patterns L_t from ORDWT coefficient and returns the residuals ε_L, and autoregression order p, which are used as the input to the next step. The following Fig. 8 describes the steps involved in the ARIMA and LSTM model.

- First, the residuals ε_L from ARIMA model is used as the input for the LSTM model. In this step, the nonlinear patterns are predicted as $N_\varepsilon = f(\varepsilon(t-1), \varepsilon(t-2), \ldots, \varepsilon(t-p))$.
- Second, the output of ORDWT and the residuals ε_L are used as inputs into the LSTM. The nonlinearity of the time series is expressed as

$$N_\varepsilon = f(\varepsilon(t-1), \varepsilon(t-2), \ldots, \varepsilon(t-p), h(t-2), \ldots, h(t-p))$$

- Combining the ARIMA and LSTM results to form the final prediction.

Fig. 9 demonstrates the schematic view of the proposed ORDWT-ARIMA-LSTM model.

3.2. Benchmark models

The experiments were run on a Windows 64-bit workstation with 64

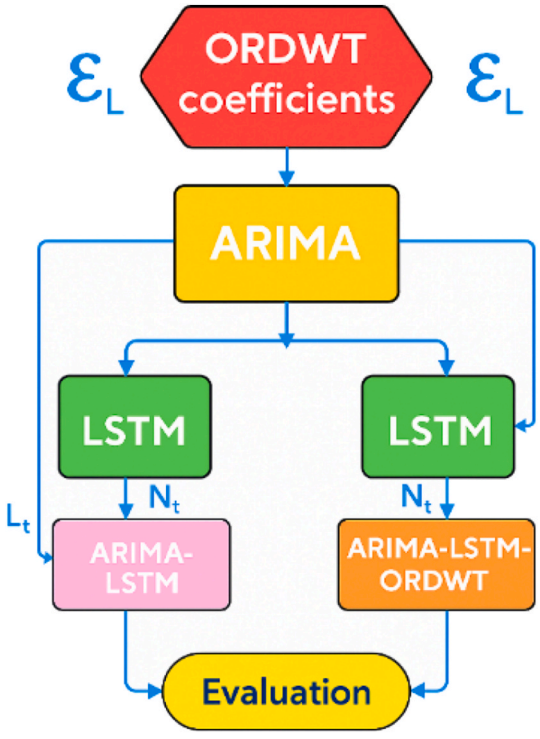

Fig. 8. The proposed ARIMA-LSTM model.

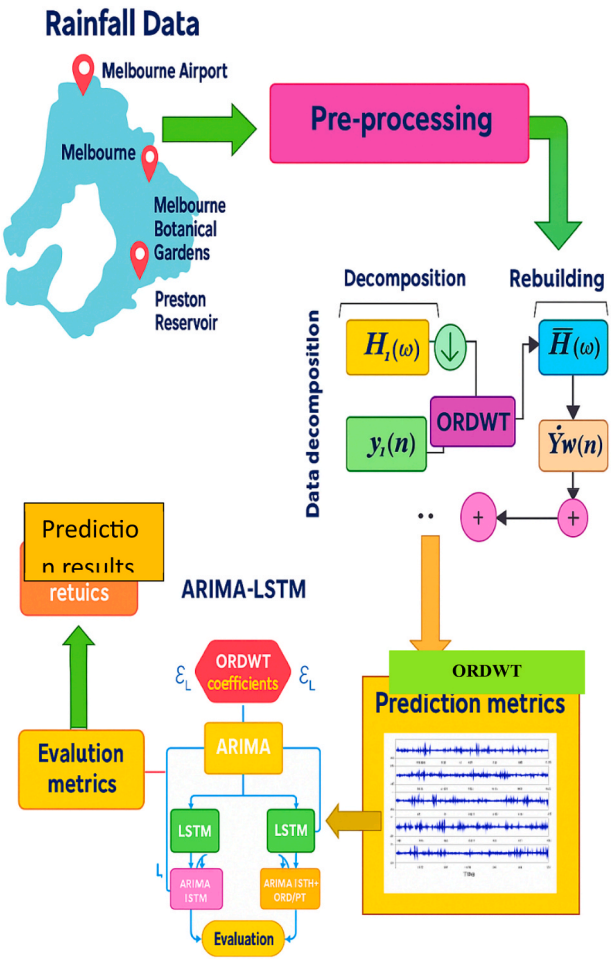

Fig. 9. The proposed model for rainfall prediction.

GB RAM; therefore, results are not dependent on GPU acceleration or GPU memory limits. We have added the specification table and fixed random seeds to improve reproducibility. We fixed the MATLAB random number generator using rng(2024) to ensure reproducibility; all data preprocessing, signal decomposition, and statistical diagnostics were performed in MATLAB. The efficiency of the proposed ORDWT-ARIMA-LSTM model was assessed against several algorithms, including SVR, LSTM, BPNN, BiLSTM, GRU, BiGRU, and their hybrid versions. Table 2 describes the model requirement.

3.3. Parameter settings

The parameters of ORDWT-ARIMA-LSTM were selected carefully using grid search method. The validation set was employed to choose the model's parameters. Table 3 lists the selected parameters of the proposed ORDWT-ARIMA-LSTM model for all three stations.

Table 2
Hardware and software requirements used for implementing the proposed model.

Hardware	Model/Specs	Software	Version
OS	Windows 11 pro, 64-bit operating system, x64-based processor	Origin	2025 Pro
CPU	12th Gen Intel(R) Core (TM) i7-1260U (1.10 GHz)	MATLAB	R2024b, Toolboxes: Signal Processing, Wavelet, Statistics & Machine Learning
RAM	64 GB		
Storage	SSD 1 TB		

Table 3
Parameters setting of all decomposition models.

Model	Parameters
EMD-ARIMA-LSTM	$LSTM = [batch\ size\ 32, Epoch = 100, and\ hidden\ layers = 256]$; $ARIMA = [P = 2, d = 4, q = 4]$; $EMD = [IMF = 5, MaxSiftIterations = 50]$
EMD-SVR	$SVR = [Kernel='rbf', C = 1000, Epsilon = 0.1, Gamma = 0.001]$; $EMD = [IMF = 5, MaxSiftIterations = 50]$
EMD-GRU	$BI-GRU = [Learning\ Rate = 0.001 - 0.0001, batch\ size = 128, Epochs = 100, loss\ function = MSE]$; $EMD = [IMF = 5, MaxSiftIterations = 50]$
EMD-BPNN	$BPNN = [Batch\ Size = 128, Learning\ Rate = 0.001 - 0.0001, Epoc = 200, loss\ function = MSE]$; $EMD = [IMF = 5, MaxSiftIterations = 50]$
EMD-BILSTM	BILSTM = [Batch Size = 128, Epochs = 100, Learning Rate = 0.001–0.0001], EMD = [IMF = 5, MaxSiftIterations = 50]
EMD-BI-GRU	BI-GRU = $[batch\ size = 128, Learning\ Rate = 0.001 - 0.0001, Epochs = 100, loss\ function = MSE]$; EMD = [IMF = 5, MaxSiftIterations = 50]
DWT-ARIMA-LSTM	$LSTM = [batch\ size\ 32, Epoch = 100, and\ hidden\ layers = 256]$; $ARIMA = [P = 2, d = 4, q = 4]$; $Wavelet\ function = db2, No.\ of\ level = 5$
DWT-SVR	$SVR = [Kernel='rbf', C = 1000, Epsilon = 0.1, Gamma = 0.001]$; $Wavelet\ function = db2, No.\ of\ level = 5$
DWT-GRU	$GRU = [Batch\ Size = 128, Learning\ Rate = 0.002 - 0.0001, Epochs = 100, loss\ function = MSE]$; $Wavelet\ function = db2, No.\ of\ level = 5$
DWT-BPNN	BPNN = $[Batch\ Size = 128, Epochs = 200, loss\ function = MSE, Learning\ Rate = 0.002 - 0.0001]$; Wavelet function = db2, No. of level = 5
DWT-BILSTM	$BILSTM = [Batch\ Size = 128, Epochs = 100, Learning\ Rate = 0.002 - 0.0001,]$; $Wavelet\ function = db2, No.\ of\ level = 5$
DWT-BI-GRU	BI-GRU = $[batch\ size = 128, Epochs = 100, loss\ function = MSE, Learning\ Rate = 0.002 - 0.0001]$; Wavelet function = db2, No. of level = 5
ORWDT-SVR	$SVR\ model = [C = 1000, Epsilon = 0.1, Kernel='rbf', Gamma = 0.001]$; $RDWT = [j = 3, q = 4, p = 5, s = 2]$
ORWDT-GRU	$GRU = [Batch\ Size = 128, Learning\ Rate = 0.001 - 0.0001, Epochs = 100, loss\ function = MSE]$; $RDWT = [j = 3, q = 4, p = 5, s = 2]$
ORDWT-BPNN	$BPNN = [Batch\ Size = 128, Learning\ Rate = 0.001 - 0.0001, Epochs = 200, loss\ function = MSE]$; $RDWT = [p = 5, j = 3, q = 4, s = 2]$
ORDWT-BILSTM	$BILSTM = [Batch\ Size = 128, Learning\ Rate = 0.001 - 0.0001, Epochs = 100]$; $RDWT = [j = 3, q = 4, p = 5, s = 2]$
ORDWT-BI-GRU	$BI-GRU = [batch\ size = 128, Epochs = 100, loss\ function = MSE, Learning\ Rate = 0.001 - 0.0001]$; $RDWT = [j = 3, q = 4, p = 5, s = 2]$
ORDWT-ARIMA-LSTM	LSTM = [batch size 32, Epoch = 100, and hidden layers = 256]; ARIMA = [P = 2, d = 4, q = 4]; ORDWT = [j = 3, q = 4, p = 5, s = 2]

For ARIMA, we evaluated a fixed grid $(p, q) \in \{0, 1, 2, 3, 4\}$, $d \in \{0, 1, 2, 3, 4\}$ on the training phase and selected the optimal parameters $(2, 4, 4)$ by Akaike Information Criterion (AIC). Model adequacy was checked via with residual diagnostics Autocorrelation Function (ACF). Although $d = 4$ exceeds the common $d = 1 - 2$, our tests indicated persistent low-frequency drift in the rainfall series after lower-order differencing. We therefore increased d sequentially until stationarity tests were passed and residual long-memory was removed. We also checked for over-differencing (inflated high-frequency variance), which was not detected in the selected specification. Details have been added to Methods.

For LSTM, the hidden size was selected via a small grid search on the validation split (128, 256, 512), with 256 giving the lowest validation RMSE. The LSTM uses its standard internal gating activations (*sigmoid* / *tanh*), and we also tested a *ReLU* activation in the post-LSTM dense layer (no consistent gain). Optimization uses Adam (initial $lr = 1 \times 10^{-}3$) with step decay ($\times$ 0.5 *every* 20 *epochs*), early stopping (patience = 10), and gradient clipping (global norm = 1.0). To prevent overfitting, we employed *dropout* = 0.2 on recurrent outputs and

L2 weight decay $= 1 \times 10^{-5}$ on all trainable weights. We used *batch size* $= 32$ for up to 100 epochs

Both baselines EMD-GRU, DWT-GRU were adjusted under the same protocol. We searched initial learning rates in $\{2\times10^{-3}, 1\times10^{-3}, 5\times10^{-4}\}$ with the same scheduler (*decay to* 1×10^{-4}), Adam optimiser, *batch size* $= 128$, *epochs* $= 100$, and MSE loss. The reported started reflecting the best-validation choices from this shared grid. EMD–BI-GRU converged best from $1\times\ 10^{-}3 \rightarrow 1\times10^{-4}$ (EMD settings: IMF = 5, Max Sift Iterations = 50), while DWT–GRU converged best from $2\times10^{-}3 \rightarrow 1\times10^{-4}$ (wavelet: db2, level = 5). The difference happened from the feature scales and gradient magnitudes formed by the two decompositions. Our results showed that the EMD IMFs show higher variance than wavelet coefficients making a slightly larger starting rate stable and faster for DWT–GRU. We confirmed that using a common starting *LR* $(1\times10^{-}3)$ yields similar accuracy and does not change model rankings.

3.4. Evaluation metrics

To perform objective and subjective evaluations, this paper employed several statistical metrics to assess the proposed ORDWT-ARIMA-LSTM model against the benchmark comparing models. In this paper, we adopted Coefficient of Determination (R) (Di Bucchianico, 2008), Mean Absolute Error (MAE) (Chai and Draxler, 2014), Root Mean Square Error (RMSE) (Chai and Draxler, 2014), Root Square Error (RSE), Willmott's Index (WI) (Willmott et al., 2012), Nash–Sutcliffe Index (ENI) (McCuen et al., 2006), Legates and McCabe's Index (LMI) (Legates and McCabe, 2013), Kling Gupta Efficiency (KGE) (Gupta et al., 2009), and Absolute Percentage Bias (APB) (Lam et al., 2001). Subjective model performance was assessed using various plots such as Taylor diagram, forecasting error histograms, scatter plots and box plots (Al-Saadi et al., 2022; Diykh et al., 2024, 2025). We also employed POD (Probability of Detection), ETS (Equitable Threat Score), CRA (Correct Rate/Accuracy), and FAR (False Alarm Ratio) were employed.

Table 4 provides a mathematical explanation for all metrics used for the performance evaluation.

Table 4
Evaluation metrics used for performance evaluation.

Metric	Mathematical explanation
R	$R^2 = \left[\frac{\sum_{i=1}^{n}(X- \prec X \succ)(Y- \prec Y \succ)}{\sqrt{\sum_{i=1}^{n}(X- \prec X \succ)^2}\sqrt{\sum_{i=1}^{n}(Y- \prec Y \succ)^2}}\right]^2$
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert X - Y\rvert$
RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X-Y)^2}$
RSE	$RSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X-Y)^2}}{Y}.100$
RMAE	$RMAE = \frac{1}{n}\sum_{i=1}^{n}\frac{\lvert X-Y\rvert}{Y}.100$
WI	$WI = 1 - \frac{\sum_{i=1}^{n}(X-Y)^2}{\sum_{i=1}^{n}(\lvert X- \prec Y \succ\rvert + \lvert X- \prec Y \succ\rvert)^2}$
ENI	$ENI = 1 - \frac{\sum_{i=1}^{n}\lvert X-Y\rvert}{\sum_{i=1}^{n}\lvert X- \prec Y \succ\rvert}$
LMI	$LMI = 1 - \frac{\sum_{i=1}^{n}\lvert X-Y\rvert}{\sum_{i=1}^{n}\lvert Y-Y\rvert}$
KGE	$\text{KGE} = 1-(r-1)^2 + \left(\sqrt{\frac{\overline{X}}{Y}}-1\right)^2 + \left(\frac{CV}{CV}\right)^2$
APB	$APB = \left\lvert\frac{\sum_{i=1}^{n}(X-Y)}{\sum_{i=1}^{n}X}\right\rvert.100$
POD	$POD = \frac{H}{H+M}$
FAR	$FAR = \frac{F}{H+F}$
ETS	$ETS = (H - Hrand)/(H + M + F - Hrand)$
CRA	$CRA = (H + C)/N$

The ENI measures how much better the model is than a naïve, and LMI highlights systematic error, and it is less sensitive to outliers than WI (McCuen et al., 2006). LMI a robust complement to ENI when extremes or skewed rainfall distributions were present (Legates and McCabe, 2013). While WI summarizes overall pattern agreement between predicted and observed values (Willmott et al., 2012). Together, all metrics determine whether the model better than climatology, and they check that improvements were not driven by a few extremes. Using those metrics alongside with MAE, RMSE provide a balanced view of magnitude accuracy, relative skill, and robustness.

4. Results and discussion

4.1. The influence of the integration ORDWT-ARIMA-LSTM on rainfall prediction

In this experiment, the performance of the proposed ORDWT-ARIMA-LSTM model was evaluated for one month ahead rainfall prediction. Initially, the ORDWT was integrated separately with ARIMA and LSTM models to demonstrate the individual benefits of combining ARIMA and LSTM with ORDWT.

4.1.1. Prediction based on ORDWT-ARIMA and ORDWT-LSTM model

The ORDWT is integrated with the ARIMA model to construct ORDWT-ARIMA. The ORDWT was also integrated with LSTM to establish the ORDWT-LSTM model. Fig. 10 reports the prediction results in terms of MSE for selection p, d, and q. To further assess the effectiveness of the ORDWT decomposition method, the DWT and EMD methods were also hybridised with ARIMA and LSTM to create the DWT-ARIMA, EMD-ARIMA, EMD-LSTM and DWT-LSTM models. The DWT and EMD have been widely used in time series prediction and analysis. This comparative analysis was crucial for highlighting the strengths and weaknesses of the proposed ORDWT against DWT and EMD techniques, showing the precision and reliability of our ORDWT-ARIMA-LSTM approach for rainfall prediction.

After selecting the parameters of the ARIMA (p, D, q) model, the monthly rainfall was predicted for three stations. The prediction results were presented in Table 4 in terms of RMSE, MAE, RSE, LMI, ENI, WI. It was found that the hybrid ORDWT-ARIMA model performed better than EMD-ARIMA and DWT-ARIMA in all three stations. The ORDWT-ARIMA scored the lowest values of MAE = 6.51, RSME = 7.59 in Melbourne Airport, MAE = 6.31, RSME = 7.75 for Melbourne Botanical Gardens, and MAE = 6.45, RSME = 7.93 for Preston Reservoir station. In addition, the LMI, ENI, and WI were also reported in Table 5 showed that ORDWT-ARIMA outperforms the other models examined. It significantly improved prediction accuracy compared to other models based on *LMI* = 0.851, *ENI* = 0.842, and *WI* = 0.851 for Melbourne Airport station, *LMI* = 0.832, *ENI* = 0.831, and *WI* = 0.842 for Melbourne Botanical Gardens, and *LMI* = 0.851, *ENI* = 0.850, and *WI* = 0.863 for Preston Reservoir station.

Table 6 reports the prediction results in terms of terms of RSME, MAE, RSE, *LMI*, *ENI*, *WI*. The ORDWT-LSTM scored the highest *LMI*, *ENI*, *WI* compared to EMD-LSTM, and DWT-LSTM, with *LMI* = 0.821, *ENI*0.823, *WI* − 0.824 for Melbourne Airport, and *LMI* = 0.823, *ENI* = 0.817, *WI* = 0.811 for Melbourne Botanical Gardens, and *LMI* = 0.824, *ENI* = 0.818, *WI* = 0.817 for Preston Reservoir. Results by LSTM and ARIMA in Tables 3 and 4 demonstrated that the results obtained by ORDWT-ARIMA were better than those by ORWDT-LSTM.

4.2. Monthly rainfall prediction based on the hybrid ORDWT-ARIMA-LSTM model

The efficiency of the proposed hybrid ORDWT-ARIMA-LSTM model was evaluated using Melbourne Airport, Melbourne Botanical Gardens, and Preston Reservoir stations. From Table 7, it becomes evident that the proposed decomposition ORDWT model gained more accuracy when

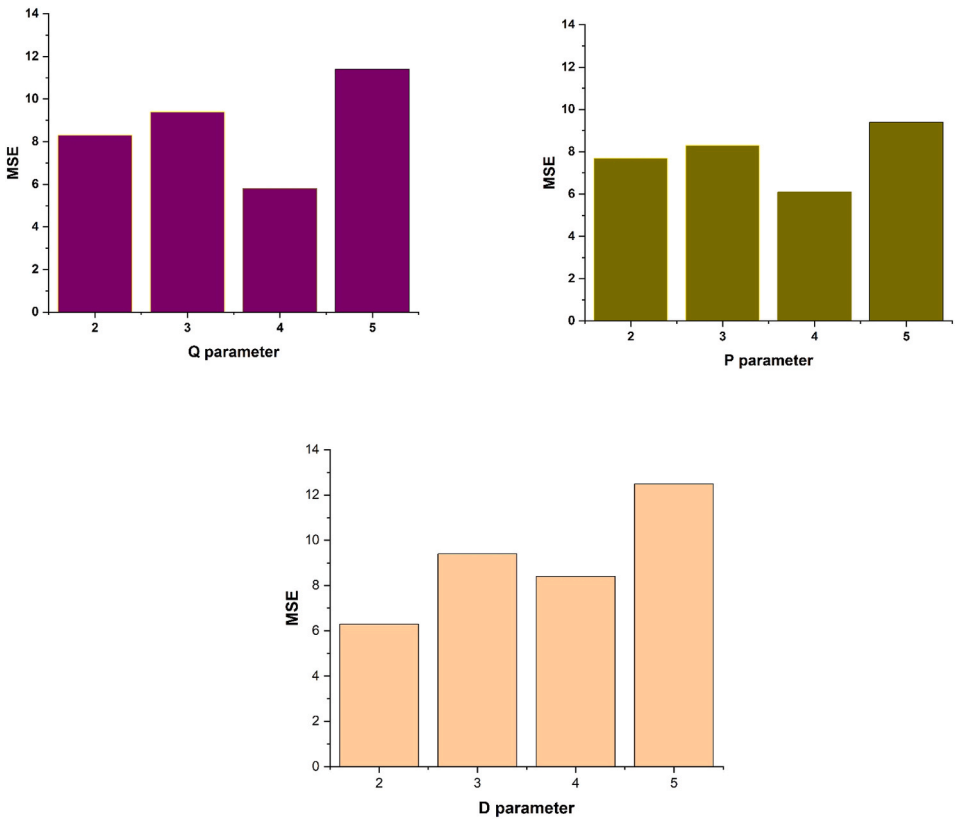

Fig. 10. The selection of ARIMA parameters.

Table 5
The performance of ORDWT-ARIMA, EMD-ARIMA, and DWT-ARIMA models in each station.

Station	Model	RMSE	MAE	RSE	LMI	ENI	WI
Melbourne Airport	ORDWT-ARIMA	**7.59**	**6.51**	**0.735**	**0.851**	**0.842**	**0.851**
	EMD-ARIMA	9.43	8.32	0.843	0.702	0.713	0.714
	DWT-ARIMA	10.32	9.23	0.932	0.681	0.662	0.683
Melbourne Botanical Gardens	ORDWT-ARIMA	**7.75**	**6.31**	**0.762**	**0.832**	**0.831**	**0.842**
	EMD-ARIMA	9.78	7.91	0.856	0.711	0.721	0.723
	DWT-ARIMA	10.76	8.16	0.956	0.661	0.665	0.671
Preston Reservoir	ORDWT-ARIMA	**7.93**	**6.45**	**0.792**	**0.851**	**0.850**	**0.863**
	EMD-ARIMA	10.10	8.17	0.892	0.717	0.701	0.711
	DWT-ARIMA	11.21	9.13	0.965	0.673	0.670	0.664

Table 6
The performance of ORDWT-LSTM, EMD-LSTM, and DWT-LSTM models in each station.

Station	Model	RMSE	MAE	RSE	*LMI*	*ENI*	*WI*
Melbourne Airport	ORDWT-LSTM	**8.81**	**6.21**	**0.801**	**0.821**	**0.823**	**0.824**
	EMD-LSTM	11.23	10.92	0.9541	0.693	0.721	0.724
	DWT-LSTM	13.12	12.43	0.961	0.631	0.622	0.615
Melbourne Botanical Gardens	ORDWT-LSTM	**8.96**	**6.76**	**0.832**	**0.823**	**0.817**	**0.811**
	EMD-LSTM	11.75	10.87	0.978	0.704	0.717	0.719
	DWT-LSTM	14.42	12.78	0.981	0.621	0.617	0.607
Preston Reservoir	ORDWT-LSTM	**9.12**	**7.61**	**0.893**	**0.824**	**0.817**	**0.818**
	EMD-LSTM	12.31	11.71	9.891	0.710	0.717	0.719
	DWT-LSTM	14.78	12.95	0.991	0.613	0.612	0.617

integrated with ARIMA-LSTM compared with the results in Tables 4 and 5 for ORDWT-ARIMA and ORDWT-LSTM. The proposed ORDWT-ARIMA-LSTM model scored the lowest MAE, RMSE, RSE errors and the highest LMI, ENI, and WI compared to EMD- ARIMA-LSTM, DWT-ARIMA-LSTM. For Melbourne Airport, the proposed model ORDWT-ARIMA-LSTM gained RMSE = 2.91, MAE = 1.93, RSE = 0.215, *LMI* = 0.98, *ENI* = 0.97, *WI* = 0.99. For Melbourne Botanical Garden, these metrics were RMSE = 3.84, MAE = 2.65, RSE = 0.287, *LMI* = 097, *ENI* = 0.96, *WI* = 097, while the proposed model ORDWT-ARIMA-LSTM gained RMSE = 3.94, MAE = 2.87, RSE = 0.310, *LMI* = 09.8, *ENI* = 0.97, *WI* = 0.97 for Preston Reservoir. It is to be worth noted that the DWT-ARIMA-LSTM obtained the lowest prediction results compared to ORDWT-ARIMA-LSTM and EMD-ARIMA-LSTM. The EMD-ARIMA-LSTM recorded the second-highest prediction results for three stations.

Table 7
Rainfall prediction using ORDWT-ARIMA-LSTMA and other hybrid comparing models.

Station	Model	RMSE	MAE	RSE	*LMI*	*ENI*	*WI*
Melbourne Airport	ORDWT-ARIMA-LSTM	**2.91**	**1.93**	**0.215**	**0.980**	**0.970**	**0.990**
	EMD-ARIMA-LSTM	6.54	5.98	0.851	0.831	0.832	0.834
	DWT-ARIMA-LSTM	7.98	6.31	0.861	0.762	0.763	0.755
Melbourne Botanical Gardens	ORDWT-ARIMA-LSTM	**3.84**	**2.65**	**0.287**	**0.971**	**0.962**	**0.971**
	EMD-ARIMA-LSTM	8.12	7.81	0.842	0.831	0.834	0.824
	DWT-ARIMA-LSTM	9.02	8.71	0.912	0.732	0.722	0.715
Preston Reservoir	ORDWT-ARIMA-LSTM	**3.94**	**2.87**	**0.310**	**0.982**	**0.971**	**0.973**
	EMD-ARIMA-LSTM	7.42	6.21	0.841	0.852	0.841	0.841
	DWT-ARIMA-LSTM	8.71	7.95	0.891	0.761	0.751	0.742

Error differences across sites align with their hydro-climatic regimes. Preston Reservoir (94 m) shows higher intra-annual variability and a heavier high-rainfall tail than Melbourne Airport (29 m), likely reflecting slight orographic uplift and local land water, urban influences. The larger share of high-intensity months at Preston drives a higher RMSE there, as residuals grow with intensity. An error intensity analysis confirms that most excess error at Preston occurs during peak rainfall months. These findings indicated that adding static physiographic covariates (elevation, orographic indices, distance-to-coast, urban fraction) and exposure corrections (e.g., wind-induced under catch) is a promising avenue to narrow site gaps in future work.

From the results, we can observe that The ORDWT provides an overcomplete, and shift-invariant representation compared to EMD, and DWT, allowing it to capture both fine-scale fluctuations and long-term dependencies in the rainfall data more effectively. EMD suffers from mode mixing and instability across stations, and DWT, which was limited by fixed dyadic scales, however, ORDWT adaptively decomposed the time series data into smoother and more stable subcomponents. This improved decomposition quality enables the hybrid model ARIMA-LSTM to learn temporal features more efficiently, leading to consistent improvements in prediction accuracy across multiple stations.

4.3. Benchmarking against comparing models

Table 8 reports the results of Melbourne Airport, Melbourne Botanical, and Preston Reservoir stations in terms of RSME, MAE, KGE, and R. It becomes evident that the incorporation of ORDWT in all prediction models ARIMA, SVR, GRU, BPNN, BIL-LSTM, BIL-GRU presented substantial improvement over the EMD and DWT. The ORDWT-ARIMA-LSTM showed an improvement over ORDWT-SVR, ORDWT-GRU, ORDWT -BILSTM, ORDWT-BIGRU, and ORDWT-BPNN with an increase in KGE, R, by 10, and 9 percent and reduction in RMSE, and MAE by 4.21 %, 4.81 %, respectively, showing the enhanced prediction ability enabled by ARIMA-LSTM architecture within the ORDWT technique. For Melbourne Airport, the proposed model ORDWT-ARIMA-LSTM gained KGE = 0.98, R = 0.98; Melbourne Botanical Garden, KGE = 0.97, R = 0.97, and Preston Reservoir, KGE = 0.97, R = 0.98.

Similarly, ORDWT-BiLSTM outperforms ORDWT-GRU, ORDWT-SVR, ORDWT-BPNN, and ORDWT-SVR, showing noticeable reductions in RMSE and MAE highlighting the efficiency of bidirectional LSTM in improving rainfall prediction accuracy over other models. Additionally, the combination of prediction model ARIMA-LSTM with both EMD and DWT showed notable increae in rainfall prediction compared to GRU, SVR, BILSTM, BIGRU, and BPNN. EMD-ARIMA-LSTM achieves the second lowest results in RMSE and MAE compared to EMD-BILSTM, EMD-SVR, EMD-BIGRU, and EMD-BPNN. The results highlight the collaborative advantages of integrating ARIMA-LSTM with EMD. In contrast, the DWT improved when it was integrated with BI-LSTM. Table 6 indicated that the proposed ORDWT-ARIMA-LSTM performed better than the baseline models in estimating rainfall prediction based on R, KGE MAE and RMSE criteria.

The boxplot in Fig. 11 demonstrates the absolute prediction error between the predicted and observed rainfall of each model. The results showed that the ORDWT-ARIMA-LSTM model outperformed all methods in achieving the lowest absolute error rates. In addition, the proposed ARIMA-LSTM prediction model gained the lowest prediction rates with DWT and EMD as compared to hybrid version of SVR, GRU, BPNN, BIL-LSTM, and BIL-GRU models. Overall, the ORDWT-ARIMA-LSTM models performed better for all three stations to predict monthly rainfall.

The scatter plots are used to analyse the proposed model ORDWT-ARIMA-LSTM comprehensive validation and comparison against other models. Fig. 12 show the scatter plots of the proposed ORDWT based

Table 8
Performance evaluation based on different prediction models.

	Melbourne Airport				Melbourne Botanical				Preston Reservoir			
Model	RMSE	MAE	KGE	R	RMSE	MAE	KGE	R	RMSE	MAE	KGE	R
EMD-ARIMA-LSTM	6.54	5.98	0.832	0.831	8.12	7.81	0.834	0.831	7.42	6.21	0.841	0.832
EMD-SVR	6.87	7.18	0.761	0.771	8.87	8.12	0.751	0.760	8.89	8.65	0.740	0.740
EMD-GRU	7.37	6.16	0.811	0.812	8.49	7.89	0.800	0.800	7.32	6.76	0.810	0.810
EMD-BPNN	9.21	8.46	0.771	0.771	9.47	8.11	0.750	0.760	8.42	8.16	0.780	0.770
EMD-BILSTM	6.33	5.76	0.822	0.821	8.46	7.88	0.810	0.810	6.49	6.89	0.800	0.800
EMD-BI-GRU	6.56	5.79	0.813	0.811	8.34	7.73	0.800	0.800	6.75	6.91	0.800	0.800
DWT-ARIMA-LSTM	7.98	6.31	0.763	0.750	9.02	8.71	0.722	0.712	8.71	7.95	0.751	0.742
DWT-SVR	13.22	12.81	0.592	0.580	14.21	14.83	0.550	0.560	14.76	12.97	0.580	0.570
DWT-GRU	11.62	10.66	0.631	0.620	11.77	10.68	0.620	0.690	11.83	10.98	0.600	0.610
DWT-BPNN	10.32	9.53	0.651	0.640	11.34	10.92	0.620	0.620	11.11	10.21	0.630	0.630
DWT-BILSTM	8.22	6.76	0.741	0.730	8.54	6.82	0.710	0.720	8.65	6.92	0.720	0.710
DWT-BI-GRU	9.27	7.16	0.710	0.710	9.59	7.77	0.700	0.700	9.76	7.82	0.710	0.710
ORWDT-SVR	4.83	3.89	0.850	0.840	4.86	3.90	0.830	0.820	4.88	3.92	0.840	0.830
ORWDT-GRU	5.65	4.97	0.810	0.800	5.67	4.98	0.800	0.800	5.72	4.99	0.810	0.810
ORDWT-BPNN	3.78	3.14	0.850	0.840	3.79	3.18	0.840	0.830	3.81	3.33	0.820	0.820
ORDWT-BILSTM	3.54	2.74	0.890	0.900	3.58	2.77	0.880	0.880	3.59	2.78	0.900	0.890
ORDWT-BI-GRU	4.57	4.14	0.810	0.800	4.76	4.81	0.790	0.800	4.60	4.16	0.811	0.821
ORDWT-ARIMA-LSTM	**2.91**	**1.93**	**0.981**	**0.981**	**3.84**	**2.65**	**0.971**	**0.970**	**3.94**	**2.87**	**0.971**	**0.981**

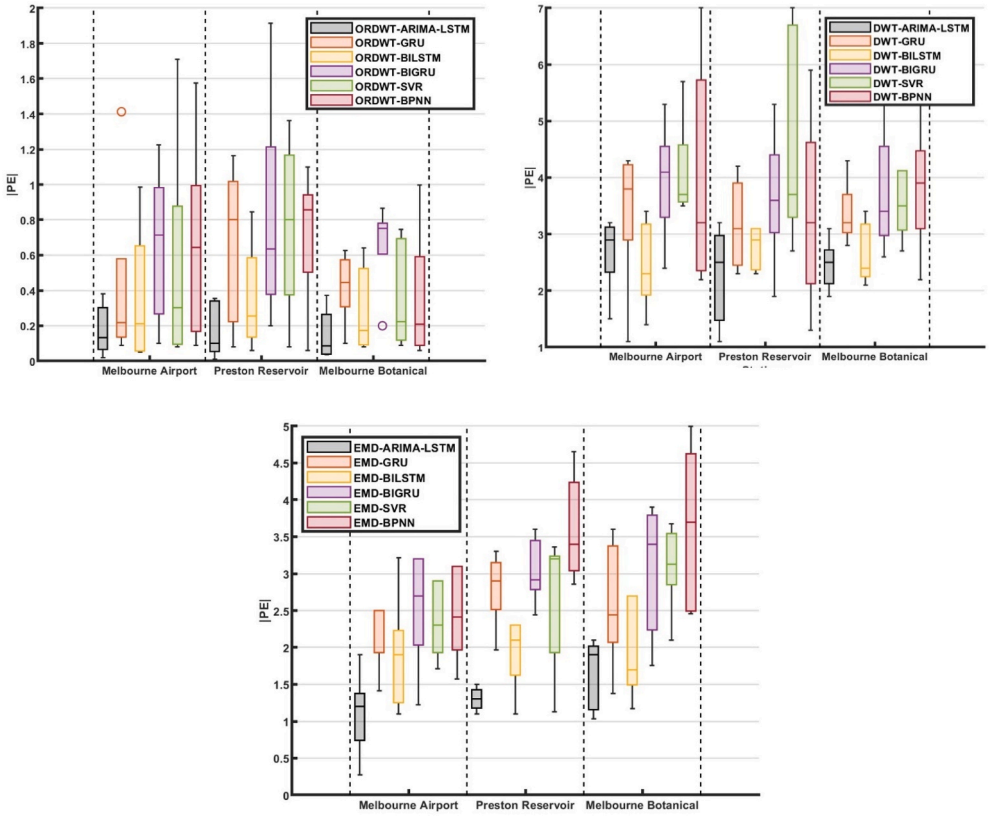

Fig. 11. Boxplot of the absolute prediction error |PE| generated by the ORDWT-ARIMA-LSTM *vs.* ORDWT-GRU, ORDWT-BILSTM, ORDWT-BIGRU, ORDWT-SVR, ORDWT-BPNN, and hybrid version DWT and EMD models for each station.

model along with DWT, and EMD version of the models for each station. This experiment was designed to examine the generalization and consistency of the proposed ORDWT-ARIMA-LSTM model on these stations. The comparison results highlighted the strengths and weaknesses of all DWT and EMD models against the ORDWT, advancing the clarity and reliability of the ORDWT-ARIMA-LSTM for rainfall prediction. The results in Table 7, Figs. 11 and 12 demonstrated a clear advantage for ORDW-ARIMA-LSTM using several metrics as compared to other models for monthly rainfall prediction. The ORDWT-ARIMA-LSTM outperformed all hybrid models, reflecting superior accuracy and reliability in extracting rainfall patterns. The scatter plots proved the robustness of the ORDWT-ARIMA-LSTM model, highlighting its ability to perform rainfall prediction tasks compared to DWT and EMD methods (see Fig. 13).

Additionally, the proposed ORDWT-ARIMA-LSTM compared against individual models SVR, GRU, BPNN, BI-LSTM, and BI-GRU, as shown in Table 9. The comparison results revealed a substantial improvement in computational efficiency. Specifically, the ORDWT-ARIMA-LSTM model achieved a remarkable improvement in the predictions by demonstrating the advantages of ORDWT in revealing the patterns of rainfall. The proposed ORDWT decomposition model significantly enhanced the prediction results of ARIMA-LSTM, making ORDWT-ARIMA-LSTM accurate and more suitable for real-time prediction and resource-constrained applications.

4.4. Monthly rainfall levels categorisation

The proposed model was examined to categorized monthly rainfall into three categories {Light, Heavy, and Extreme-heavy}. The training sample was used to fit all preprocessing steps to train model parameters, the validation set was employed only for hyperparameter selection and early stopping, while the testing set held out for final evaluation once. No observations from validation or test were used during training. All transforms and statistics were assessed on the training block only and then applied forward to validation and test. During training or validation, we did not shuffle sequences.

To create three non-overlapping tiers per station and month-of-year mmm, we computed percentile thresholds from the training climatology: Light: $y < P75$, Heavy : $P75 \le y < P95$, Extreme-heavy: $y \ge P95$. The numeric ranges are as follows *Light* : $[0.8, P75)$ mm·month^{-1}, *Heavy*: $[P75, P95)$ mm·month^{-1}, *Extreme heavy*: $[P95, 216]$ mm·month^{-1}. To evaluate the performance of the proposed model, four metrics POD (Probability of Detection), ETS (Equitable Threat Score), CRA (Correct Rate/Accuracy), and FAR (False Alarm Ratio) were employed. POD ranges from 0 to 1, where 1.0 is perfect and values near 0 mean most events were missed. ETS ranges from about 0.33 to 1. Where 1.0 is perfect; 0 means no skill beyond chance; negative values indicate worse than random. For rare events, 0.3–0.6 is typically considered respectable skill. CRA ranges from 0 to 1. Where 1.0 is perfect overall correctness; values near 0 mean predictions are mostly wrong. FAR ranges from 0 to 1. Where 0 is perfect (no false alarms); values near 1 mean almost every "event" prediction did not occur while lower is better.

We computed one-*vs*-rest for each class. Splits are strictly chronological; thresholds were estimated on training only. As shown in Fig. 13, the proposed model showed a high reliability for Light months across three stations Melbourne Airport, Melbourne Botanical Gardens, Preston Reservoir. at Melbourne Airport, the proposed model achieved an average of POD ≈0.93–0.83, CRA ≈0.90–0.85. The proposed model detected Heavy rain months with solid scores (POD ≈ 0.90–0.86, ETS ≈ 0.0.21–0.0.3, FAR ≈ 0.51–0.54) indicating useful discrimination at the operational 1-month lead. As expected for rare events, Extreme-heavy months exhibit low performance with scores (POD ≈ 0.82–0.73, ETS ≈ 0.32–0.58, FAR ≈ 0.53–0.62) but remain informative and not over-

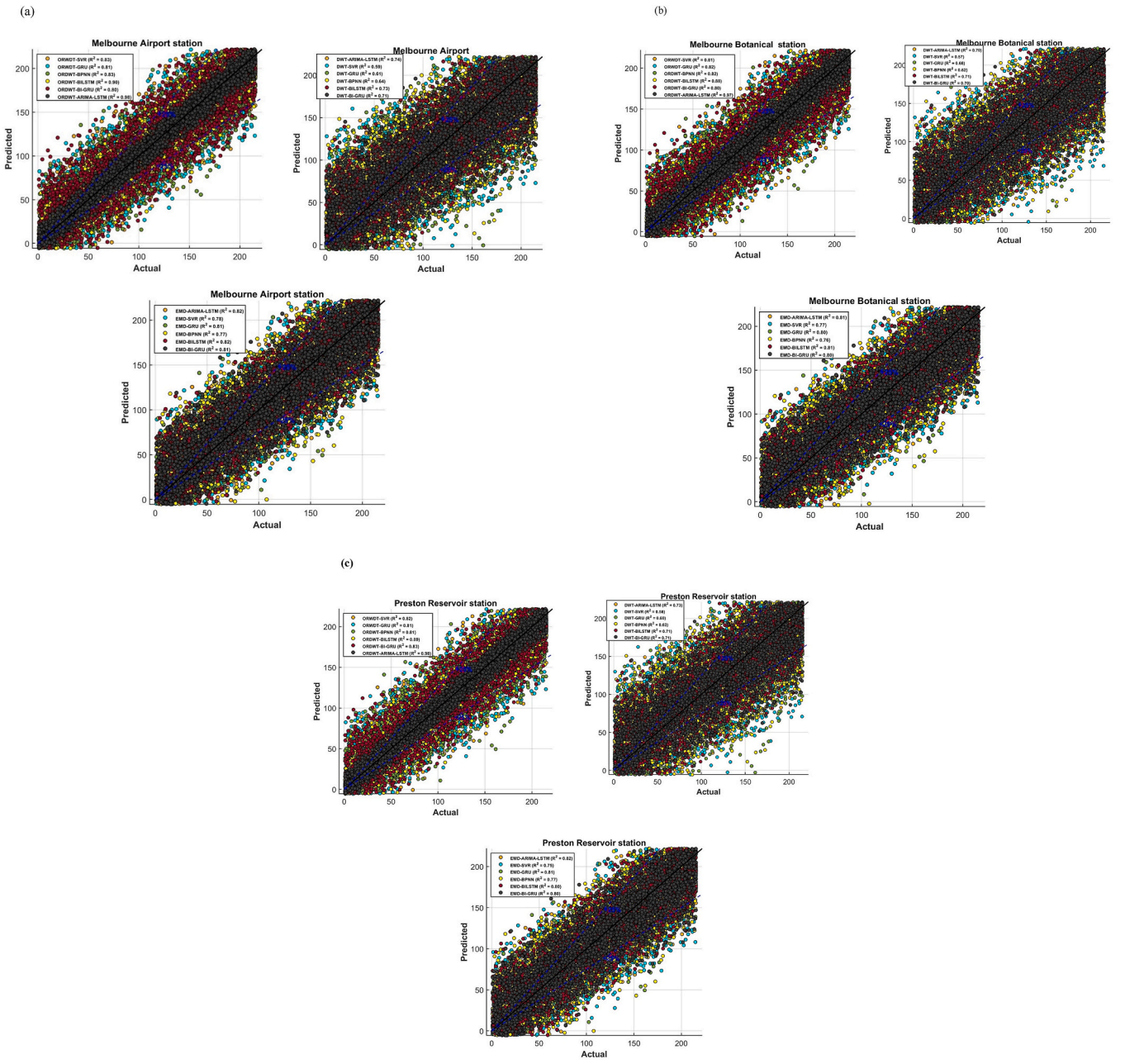

Fig. 12. Scatter plots of the proposed ORDWT-ARIMA-LSTM and oother hybrid version of the models for (a) Melbourne Airport; (b) Melbourne Botanical Garden; and (c) Preston Reservoir stations.

triggered. Melbourne Botanical Gardens scores were marginally higher, whereas Preston Reservoir was slightly lower across levels, consistent with its higher inter-monthly variability and heavier wet-month tail. Overall, these results showed that beyond low continuous errors, the proposed model meaningfully separates intensity levels at 1-month lead.

4.5. Ablation study: systematic evaluation of model complexity

To analyse the complexity of the proposed model, we compare the time complexity of each model component, as shown in Table 10. Suppose m represents the length of the time series, p is the LSTM hidden layer dimension, p and q refer the ARIMA model parameters, and J is the number of ORDWT decomposition levels. The LSTM model required a time complexity of $O\left(n \cdot h^2\right)$, since it executed sequential operations through prediction process.

The proposed hybrid ARIMA–LSTM added an extra time $O(p+q)$ corresponding to the autoregressive and moving-average operations which increased the computational load slightly while improving temporal representation. The proposed ORDWT–ARIMA–LSTM adds the decomposition step of ORDWT with complexity, $O(J \cdot n)$ due to the multi-level decomposition. However, this cost remains linear with respect to sequence length and did not dominate the training phase. Overall, the ORDWT–ARIMA–LSTM obtained a favourable trade-off between accuracy and computational time. The additional complexity time produced by the ORDWT led to significantly improvement in predictive accuracy and feature robustness while maintaining practical efficiency for both CPU and GPU environments.

5. Further Discussion, limitations, and future work

The results indicated that the proposed ORSWT-ARIMA-LSTM model

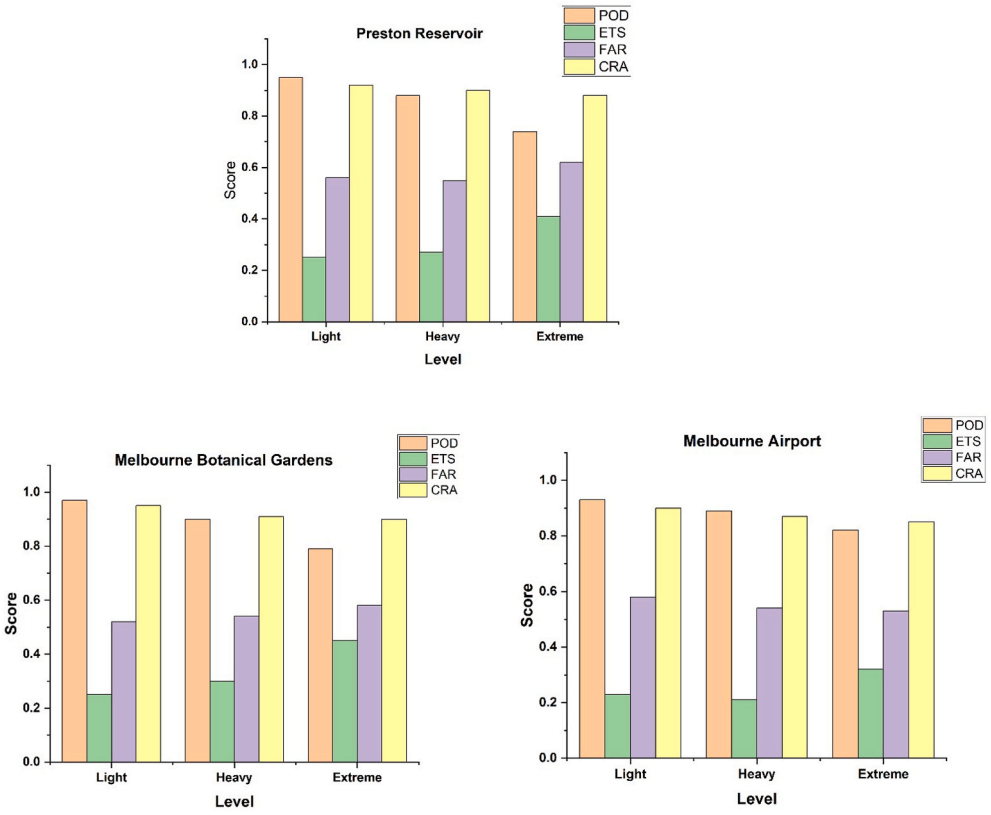

Fig. 13. Prediction events of 1-month-lead. Bars show POD, ETS, FAR, and CRA for Light (<P75), Heavy (P75–P95), and Extreme (≥P95) months. Thresholds are computed per site and month from the training climatology.

Table 9
Performance evaluation of the ORDWT-ARIMA-LSTM against individual models.

	Melbourne Airport							
Model	RMSE	MAE	KGE	R	LMI	ENI	WI	APB
SVR	16.4	14.4	71.5	72.3	71.2	71.7	72.3	7.2
GRU	15,4	14.6	0.71	0.70	0.71	0.69	0.68	8.4
BPNN	17.3	15.3	0.67	0.66	0.68	0.66	0.65	8.6
BILSTM	10.3	9.9	0.75	0.77	0.776	0.75	0.76	7.1
BI-GRU	11.2	‘10.7	0.74	0.73	0.74	0.72	0.72	7.3
ORDWT-ARIMA-LSTM	**2.91**	**1.93**	**0.98**	**0.98**	**0.98**	**0.97**	**0.99**	**1.2**
	Melbourne Botanical Garden							
SVR	17.2	15.1	70.2	71.6	70.6	0.71	0.71	7.1
GRU	15,8	14.9	0.70	0.69	0.69	0.68	0.69	8.3
BPNN	17.6	15.7	0.66	0.65	0.66	0.65	0.66	8.7
BILSTM	10.7	10.3	0.74	0.76	0.76	0.76	0.75	7.3
BI-GRU	11.8	‘10.9	0.73	0.72	0.73	0.73	0.71	7.4
ORDWT-ARIMA-LSTM	**3.84**	**2.65**	**0.97**	**0.97**	**0.97**	**0.96**	**0.97**	**1.3**
	Preston Reservoir							
SVR	16.4	14.4	71.5	72.3	71.2	71.7	72.3	6.8
GRU	15,4	14.6	0.71	0.70	0.71	0.69	0.68	7.7
BPNN	17.3	15.3	0.67	0.66	0.68	0.66	0.65	7.9
BILSTM	10.3	9.9	0.75	0.77	0.776	0.75	0.76	6.8
BI-GRU	11.2	‘10.7	0.74	0.73	0.74	0.72	0.72	6.8
ORDWT-ARIMA-LSTM	**3.94**	**2.87**	**0.97**	**0.98**	**0.98**	**0.97**	**0.97**	**1.1**

Table 10
Analysis of the proposed model time complexity.

Model	Time complexity
LSTM	$O(m.\beta^2)$
ARIMA-LSTM	$O(p+q+m.\beta^2)$
ORDWT-ARIMA-LSTM	$O(J.n+p+q+m.\beta^2)$

significantly outperforms its competing counterparts, DWT, and EMD in monthly rainfall predictions for Melbourne Airport, Melbourne Botanical Garden, and Preston Reservoir stations. The proposed model can support decision-makers in achieving highly accurate rainfall predictions which can be utilise in the environmental, hydrological and agricultural sector to improve resource management and optimization. The analysis of the obtained results reveals several important findings.

Table 11
Average and standard deviation of the proposed model performance over 10 independent runs with different random seeds.

Model	RMSE (mean ± Std)	MAE (mean ± Std)	R (mean ± Std)
LSTM	7.98 ± 0.14	6.21 ± 0.023	0.83 ± 0.011
ARIMA	7.27 ± 0.13	6.51 ± 0.011	0.85 ± 0.021
The proposed hybrid ARIMA-LSTM	8.77 ± 0.10	6.21 ± 0.015	0.95 ± 0.018

Table 12
Average and standard deviation of the proposed model performance over 10 independent runs with different random seed.

Metrics	ORDWT-BI-GRU (A)	ORDWT-ARIMA-LSTM (B)	Diff (A-B)	95 % CI of Diff	DM stat	P-vale
MSE	5.87 ± 0.04	2.11 ± 0.01	−3.76	[−5.80, −1.72]	−2.29	0.0019
RMSE	2.42 ± 0.02	1.45 ± 0.001	−0.97	[−1.55, −0.38]	−2.29	0.0015

- Rainfall at a point reflects many interacting processes but their footprints appear in the local time series as seasonality, multi-scale variance, and regime shifts. The proposed model was designed to make those footprints learnable. In our model, the ORDWT produces shift-invariant multi-scale sub bands that separate the annual cycle, intra-seasonal swings, and bursty events. The ARIMA was used to capture the linear, persistent component within each sub band, while LSTM extracts the nonlinear residual dynamics associated with regime changes and extremes. Because all transforms and parameters were fit only on past data, the model learns statistical surrogates of the physical drivers without look-ahead. When exogenous information is available, the same architecture naturally extends to include static covariates and dynamic predictors. In short, the prediction model systematically exploits the multi-scale structure and persistence those processes imprint on the monthly rainfall record, and it can incorporate targeted physics-based covariates when provided.
- We conducted ten repeated experiments using different random initialization seeds to validate the consistency of the obtained results and ensure robust metric selection. Because the study involves time series prediction, the temporal order and data split were kept fixed across all runs to protect the sequential structure of the data. Only the initialization of model parameters was varied in each repetition. The reported results in Table 11 represented the average performance across these ten runs, with the standard deviations included to indicate statistical variability. This procedure improved the robustness, reproducibility, and reliability of the findings while maintaining the integrity of the temporal dataset.
- Fig. 14 shows the Taylor diagram of the proposed model for all three stations. The Taylor diagram visually compares the performance of predictive models by computing the Standard Deviation (*SD*) on the polar axis and the angular location of the inverse cosine of the pearson correlation (R) on the radial axis. It highlights the best model that aligns with actual data during testing by combining several metrics to identify the optimal fit. The results in Fig. 11 shows that, across three stations, the ORDWT-ARIMA-LSTM achieved the highest *r* value, which indicates its estimation was the closest to the actual data. In contrast, the other decomposition-based hybrid models, ORDWT-BI-GRU and ORDWT-GRU, showed lower prediction performance compared to the proposed ARIMA-LSTM model. These findings highlighted the superior predictive ability of the ORDWT-ARIMA-LSTM model in accurately capturing the complex patterns in rainfall data, making it the most successful predictive model among the others.
- In this experiment, the Combined Performance Index (*CPI*, %) was employed to evaluate the predictive capability of the proposed ORDWT-AARIMA-LSTM model. The *CPI* is computed as a weighted sum of multiple measures that incorporate both dispersion and distribution function similarity information:

$$CPI = \left(KSI + OVER + \frac{2RMSE}{4} \right)$$

where *KSI* refers to the Kolmogorov-Smirnov test integral, and *OVER* represents the relative frequency of exceedance. A lower *CPI* value refers to better performance. Fig. 15 shows *CPI* values for each model across three stations. The results highlight that the proposed ORDWT-ARIMA-LSTM model achieved the lowest *CPI* values at all three stations. It scored approximately 1.8 %, 1.7 %, and 2.1 % for Melbourne Airport,

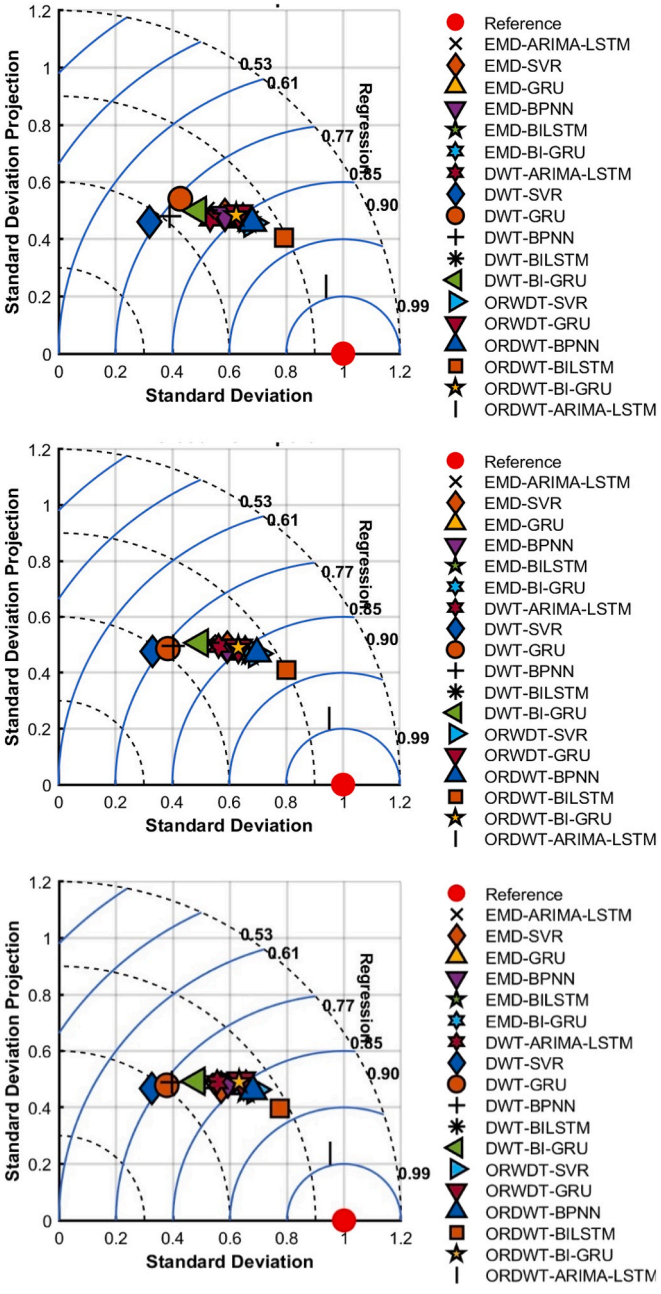

Fig. 14. Taylor diagram of the proposed ORDWT-ARIMA-LSTM against other comparing models for Melbourne Airport, Melbourne Botanical Garden, and Presten Reservoir stations.

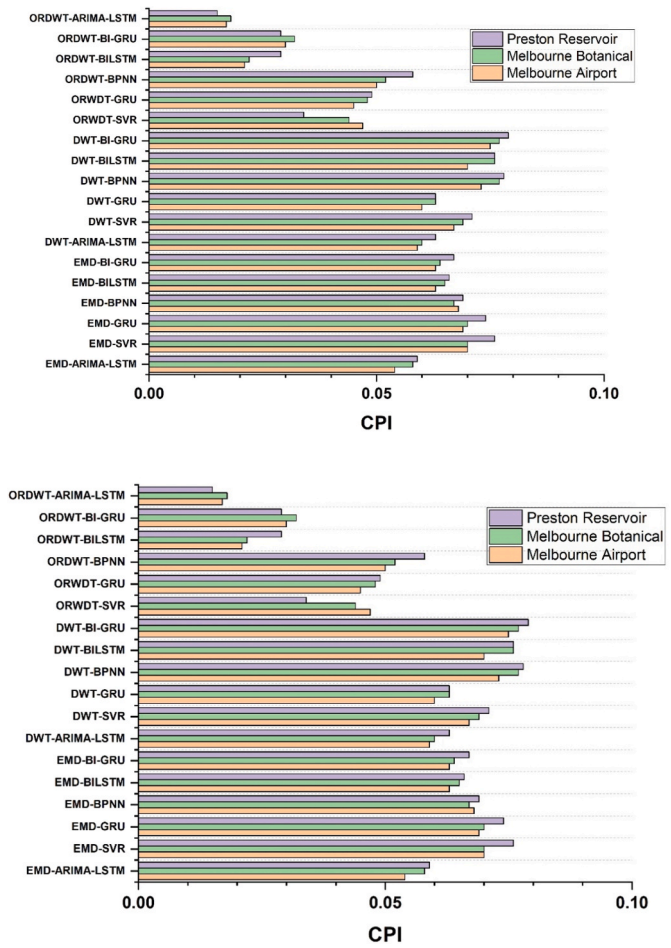

Fig. 15. Performance evaluation of each model using the Combined Performance Index (CPI).

Melbourne Botanical, and Preston Reservoir, respectively. Similarly, all prediction models scored the lowest CPI when integrated with the ORDWT. These results indicate that the ORDWT model outperforms DWT, and EMD. In addition, the hybrid ORDWT-ARIMA-LSTM outperformed the benchmark DWT, and EMD-based prediction models (EMD-GRU, EMD-BI-GRU, EMD-BILSTM, EMD-BPNN, EMD-SVR, DWT-SVR, DWT-BPNN, DWT-GRU, DWT-BI-GRU, DWT-BILSTM), This makes the proposed ORDWT-ARIMA-LSTM more suitable for monthly rainfall predictions.

- Although the traditional models such as SVR, GRU, BPNN, BI-LSTM, and BI-GRU showed promising results, these models fail to extract the linear and nonlinear features of rainfall data. They cannot extract temporal features, resulting in lower performance than the proposed ARIMA-LSTM model. Integrating LSTM and ARIMA reduced prediction errors, significantly improving prediction as evidenced by the assessment metrics.
- The proposed ORDWT decomposition markedly improves the prediction rate. The ORDWT effectively decomposed the rainfall data and captured relevant characteristics in the rainfall data, resulting in more precise predictions. The results showed that the ORDWT-ARIMA-LSTM consistently outperformed other hybrid version of EMD and DWT across various metrics, including MAE, MAPE, and RMSE. This consistent superiority emphasizes that the ORDWT was a robust data decomposition technique, offering absolute reliability and accuracy in extracting rainfall data patterns. Our findings strongly demonstrated that ORDWT-based prediction models were more effective for rainfall prediction than DWT and EMD.
- The proposed ORDWT-ARIMA-LSTM hybrid model demonstrated excellent performance compared to individual models based on by the assessment metrics. Combining the strengths of the two architectures, ARIMA and LSTM, this improvement was obtained. The ARIMA effectively captures nonlinear patterns, and the LSTM captures linear patterns. Their combination with ORDWT created a strength that enables the hybrid ORDWT-ARIMA-LSTM model to reveal hidden patterns, yielding more reliable and accurate rainfall predictions.
- The ORDWT-ARIMA-LSTM and ORDWT-BI-GRU models were compared using statistical testes. The results in Table 12 showed that the proposed ORDWT-ARIMA-LSTM achieved significantly lower errors than ORDWT-BI-GRU on the same test months. For MSE, the mean difference between them was -3.76 with a 95 % confidence interval CI $[-5.80, -1.72]$. Based on RMSE, the statistical differences were -0.97 with 95 % confidence intervals (CI) $[-1.55, -0.38]$. In both cases the confidence intervals were strictly negative and the Diebold–Mariano (DM) statistic was -2.29 with $p < 0.01$–0.02, confirming that the error reductions were statistically significant under a prediction comparison test. The statistical evaluation indicated that the proposed model ORDWT-ARIMA-LSTM delivered more accurate monthly rainfall totals with lower absolute error. Our findings confirmed that the ORDWT-ARIMA-LSTM was strong predictor for the monthly rainfall prediction task.
- Accurate monthly rainfall prediction enables timely, sector-specific actions. In irrigated agriculture, forecasting a light month supports earlier irrigation scheduling, while a heavy and extreme month justifies delaying field operations, and reducing yield risk and wasted inputs. Water managers can also align releases and allocations with outlooks. A forecast dry month can help in cautious releases and drought allocations, while a wet month enables refill targets and flood-space management. Public works can bring forward maintenance to cut overflow incidents. Energy and insurance benefit from constant hydropower scheduling and more precise risk pricing when heavy months are flagged in advance. Overall, the proposed model provides decision-relevant guidance for water and drought management, helping experts select appropriate policies and preparedness measures.

6. Conclusions

This paper proposed a hybrid modelling approach for monthly rainfall prediction, integrating ORDWT with ARIMA and LSTM models to design ORDWT-ARIMA-LSTM. The ORDWT was employed to decompose the nonlinear and non-stationary rainfall data into predictable sub-bands (i.e., coefficients). The ARIMA-LSTM model was designed to capture essential linear and non-linear features for predicting monthly rainfall. The proposed ORDWT-ARIMA-LSTM model was tested in three stations Melbourne Airport, Melbourne Botanical, and Preston Reservoir to predict the monthly rainfall. The findings showed that classic decompositions methods such as DWT and EMD struggled to effectively analyse nonlinear and non-stationary data. The proposed ORDWT-ARIMA-LSTM method outperformed DWT and EMD techniques to show the potential applicability for rainfall predictions. The scope of the proposed ORDWT-ARIMA-LSTM model can extend for other applications areas such as renewable and sustainable energy, environmental and agriculture sectors to offer decision makers the ability to make accurate and timely decisions.

CRediT authorship contribution statement

Mohammed Diykh: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Mumtaz Ali:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Conceptualization. **Aitazaz Ahsan Farooque:** Writing – review & editing, Supervision,

Investigation, Funding acquisition, Conceptualization. **Anwar Ali Aldhafeeri:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Mehdi Jamei:** Writing – review & editing, Visualization, Methodology, Investigation, Conceptualization. **Abdulhaleem Labban:** Writing – review & editing, Methodology, Investigation, Conceptualization.

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.

Acknowledgements

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Project No. KFU254413]. The authors are thankful to the Bureau of Meteorology, Australia for providing the rainfall data for the required stations.

Data availability

Data will be made available on request.

References

Abbot, J., Marohasy, J., 2014. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. Atmos. Res. 138, 166–178.
Abbot, J., Marohasy, J., 2018. Forecasting of medium-term rainfall using artificial. Engineering and Mathematical Topics in Rainfall: 33.
Acworth, I., Bernardi, T., Andersen, M.S., Rau, G.C., 2024. Hydrological complexity and climate implications in Australia's arid zone: a decade of high-resolution rainfall observations. J. Hydrol.: Reg. Stud. 51, 101643.
Aderyani, F.R., Mousavi, S.J., Jafari, F., 2022. Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN. J. Hydrol. 614, 128463.
Al-Saadi, Y.R., Tapou, M.S., Badi, A.A., Abdulla, S., Diykh, M., 2022. Developing smart self orienting solar tracker for mobile PV power generation systems. IEEe Access 10, 79090–79099.
Anand, P., Rastogi, R., Chandra, S., 2020. A class of new support vector regression models. Appl. Soft Comput. 94, 106446.
Awad, M., Khanna, R., Awad, M., Khanna, R., 2015. Support vector regression. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, pp. 67–80.
Bagirov, A.M., Mahmood, A., Barton, A., 2017. Prediction of monthly rainfall in Victoria, Australia: clusterwise linear regression approach. Atmos. Res. 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.
Baig, F., Ali, L., Faiz, M.A., Chen, H., Sherif, M., 2024. How accurate are the machine learning models in improving monthly rainfall prediction in hyper arid environment? J. Hydrol. 633, 131040.
Basha, C.Z., Bhavana, N., Bhavya, P., Sowmya, V., 2020. Rainfall prediction using machine learning & deep learning techniques. In: 2020 International Conference on Electronics and Sustainable Communication Systems. ICESC), IEEE.
Bayram, I., Selesnick, I.W., 2007. Design of orthonormal and overcomplete wavelet transforms based on rational sampling factors. Wavelet App. Indust. Process. V SPIE.
Bayram, I., Selesnick, I.W., 2009. Frequency-domain design of overcomplete rational-dilation wavelet transforms. IEEE Trans. Signal Process. 57 (8), 2957–2972.
Bian, C., He, H., Yang, S., Huang, T., 2020. State-of-charge sequence estimation of lithium-ion battery based on bidirectional long short-term memory encoder-decoder architecture. J. Power Sources 449, 227558.
Buscema, M., 1998. Back propagation neural networks. Subst. Use Misuse 33 (2), 233–270.
Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE). Geosci. Model Dev. Discuss. (GMDD) 7 (1), 1525–1534.
Chen, B., Zhang, Z., Sun, C., Li, B., Zi, Y., He, Z., 2012. Fault feature extraction of gearbox by using overcomplete rational dilation discrete wavelet transform on signals measured from vibration sensors. Mech. Syst. Signal Process. 33, 275–298.
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 Sens. 12 (19), 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.
Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
Cilimkovic, M., 2015. Neural networks and back propagation algorithm. Instit. Technol. Blanchardstown Blanchardstown Road North Dublin 15 (1), 18.
Cramer, S., Kampouridis, M., Freitas, A.A., 2018. Decomposition genetic programming: An extensive evaluation on rainfall prediction in the context of weather derivatives. Applied Soft Computing 70, 208–224.
Cui, X., Wang, Z., Pei, R., 2023. A VMD-MSMA-LSTM-ARIMA model for precipitation prediction. Hydrological Sciences Journal 68 (6), 810–839.
Devanand, A., Evans, J., Pitman, A., Pal, S., Gochis, D., Sampson, K., 2024. The influence of lateral flow on land surface fluxes in southeast Australia varies with model resolution. EGUsphere 2024, 1–28.
De Vos, N., Rientjes, T., 2005. Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation. Hydrol. Earth Syst. Sci. 9 (1/2), 111–126.
Di Bucchianico, A., 2008. Coefficient of determination (R 2). Encyclopedia of Statistics in Quality and Reliability.
Diykh, M., Ali, M., Jamei, M., Abdulla, S., Uddin, M.P., Farooque, A.A., Labban, A.H., Alabdally, H., 2024. Empirical curvelet transform based deep DenseNet model to predict NDVI using RGB drone imagery data. Computers and Electronics in Agriculture 221, 108964.
Diykh, M., Ali, M., Jamei, M., Prasad, R., Labban, A.H., Abdulla, S., Singh, N., Farooque, A.A., 2025. Multi-step ahead significant wave height forecasting using global and local view graph representation based on GRU model. Ocean Engineering 340, 122314.
El-Shafie, A., Noureldin, A., Taha, M., Hussain, A., Mukhlisin, M., 2012. Dynamic versus static neural network model for rainfall forecasting at Klang River Basin, Malaysia. Hydrol. Earth Syst. Sci. 16 (4), 1151–1169.
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. Sci. Total Environ. 854, 158760.
Fan, D., Sun, H., Yao, J., Zhang, K., Yan, X., Sun, Z., 2021. Well production forecasting based on ARIMA-LSTM model considering manual operations. Energy 220, 119708.
Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91.
George, J., Letha, J., Jairaj, P.G., 2016. Daily rainfall prediction using generalized linear bivariate model–a case study. Procedia Technology 24, 31–38.
Gupta, S., Gupta, A., 2019. Dealing with noise problem in machine learning data-sets: a systematic review. Procedia Comput. Sci. 161, 466–474.
Jamei, M., Ali, M., Malik, A., Karbasi, M., Rai, P., Yaseen, Z.M., 2023. Development of a TVF-EMD-based multi-decomposition technique integrated with Encoder-Decoder-Bidirectional-LSTM for monthly rainfall forecasting. Journal of Hydrology 617, 129105.
Jaseena, K., Kovoor, B.C., 2021. Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks. Energy Convers. Manag. 234, 113944.
Kamble, A., Ghare, P.H., Kumar, V., 2023. Optimized rational dilation wavelet transform for automatic imagined speech recognition. IEEE Trans. Instrum. Meas. 72, 1–10.
Lam, K.F., Mui, H., Yuen, H., 2001. A note on minimizing absolute percentage error in combined forecasts. Comput. Oper. Res. 28 (11), 1141–1147.
Legates, D.R., McCabe, G.J., 2013. A refined index of model performance: a rejoinder. Int. J. Climatol. 33 (4), 1053–1056.
McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash–Sutcliffe efficiency index. J. Hydrol. Eng. 11 (6), 597–602.
Nayak, M.A., Ghosh, S., 2013. Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier. Theor. Appl. Climatol. 114, 583–603.
Othman, G., Zeebaree, D.Q., 2020. The applications of discrete wavelet transform in image processing: a review. J. Soft Comput. Data Mining 1 (2), 31–43.
Patel, M.K., Darji, M., Patel, B., 2025. Application of Liquid State Machine for Accurate Rainfall Prediction. Procedia Computer Science 258, 1554–1564.
Rana, R., 2016. Gated recurrent unit (GRU) for emotion classification from noisy speech. arXiv preprint arXiv:1612.07778.
Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45 (11), 2673–2681.
Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys. Nonlinear Phenom. 404, 132306.
Shin, J., Yoon, S., Park, N.S., Kim, Y., 2025. Development of water quality prediction model using LTSF-Linear and complete ensemble empirical mode decomposition. Desalination and Water Treatment 101254.
Staudemeyer, R.C., Morris, E.R., 2019. Understanding LSTM–a tutorial into long short-term memory recurrent neural networks. arXiv preprint arXiv:1909.09586.
Tao, L., He, X., Li, J., Yang, D., 2021. A multiscale long short-term memory model with attention mechanism for improving monthly precipitation prediction. J. Hydrol. 602, 126815.
Torres, M.E., Colominas, M.A., Schlotthauer, G., Flandrin, P., 2011. A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
Trivedi, D., Pattnaik, S., Sharma, O., 2025. *Reducing the real-time heavy rainfall forecast error associated with Monsoon depressions and deep depressions over the Indian region using deep learning* (No. Copernicus Meetings.
Wang, Y., Skerry-Ryan, R., Stanton, D., Wu, Y., Weiss, R.J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Z., Bengio, S., 2017. Tacotron: towards end-to-end speech synthesis. arXiv preprint arXiv:1703.10135.
Wang, Y., Stanton, D., Zhang, Y., Ryan, R.-S., Battenberg, E., Shor, J., Xiao, Y., Jia, Y., Ren, F., Saurous, R.A., 2018. Style tokens: unsupervised style modeling, control and

transfer in end-to-end speech synthesis. In: International Conference on Machine Learning. PMLR.
Waqas, M., Humphries, U.W., Wangwongchai, A., Dechpichai, P., Ahmad, S., 2023. Potential of artificial intelligence-based techniques for rainfall forecasting in Thailand: a comprehensive review. Water 15 (16), 2979.
Webb, C., 2010. Bureau of meteorology reference evapotranspiration calculations. Clim. Serv. Centre Queensland Reg. Office Bureau Meteorol.
Willmott, C.J., Robeson, S.M., Matsuura, K., 2012. A refined index of model performance. Int. J. Climatol. 32 (13), 2088–2094.
Yildirim, Ö., 2018. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Comput. Biol. Med. 96, 189–202.
Yu, Y., Si, X., Hu, C., Zhang, J., 2019. A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput. 31 (7), 1235–1270.
Zeiler, A., Faltermeier, R., Keck, I.R., Tomé, A.M., Puntonet, C.G., Lang, E.W., 2010. Empirical mode decomposition-an introduction. In: The 2010 International Joint Conference on Neural Networks (IJCNN), IEEE.
Zhang, J., Liu, H., Bai, W., Li, X., 2024. A hybrid approach of wavelet transform, ARIMA and LSTM model for the share price index futures forecasting. N. Am. J. Econ. Finance 69, 102022.