



# Comparison of LSTM network, neural network and support vector regression coupled with wavelet decomposition for drought forecasting in the western area of the DPRK

Yong-Sik Ham<sup>1</sup> · Kyong-Bok Sonu<sup>1</sup> · Un-Sim Paek<sup>1</sup> · Kum-Chol Om<sup>1</sup> · Sang-Il Jong<sup>1</sup> · Kum-Ryong Jo<sup>1</sup>

Received: 20 July 2022 / Accepted: 11 December 2022 / Published online: 23 December 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

Drought forecasting is very important in reducing the drought damage and optimizing water resources. This paper focuses on confirming the advantage of wavelet long short-term memory network (WLSTMN) through comparison with wavelet artificial neural network (WANN) and wavelet support vector regression (WSVR) for drought forecasting in the west area of the Democratic People's Republic of Korea. The standardized precipitation index with 6 and 12-month timescales (SPI-6 and SPI-12) was used in this study. In order to increase the number of training samples for the development of data-driven models, SPIs were calculated at ten days' intervals and input data was lagged combinations of time series that decomposed using Haar wavelet mother function at 1–10 decomposition levels. The performances of the three models with several decomposition levels and lags at 1-month lead time were estimated with determination coefficient ( $R^2$ ), Lin's concordance correlation coefficient (LCCC), root-mean-square error (RMSE) and mean absolute error (MAE). Area-averaged performance measures of optimal models show that  $R^2$ , LCCC, RMSE and MAE of WLSTMN for SPI-6 were 0.709, 0.806, 0.572 and 0.427, respectively, better than those of other models. And  $R^2$ , LCCC, RMSE and MAE of WLSTMN for SPI-12 were 0.919, 0.950, 0.296 and 0.190, respectively. It has a better performance compared to the other models. Consequently, WLSTMN model for drought indices with two time-scales outperformed traditional WANN and WSVR, which have smaller  $R^2$  and LCCC, larger RMSE and MAE.

**Keywords** Standardized precipitation index · Drought prediction · Long short-term memory · WANN · WSVR · DPRK

---

✉ Yong-Sik Ham  
ys.ham1221@ryongnamsan.edu.kp

<sup>1</sup> Meteorology Institute, Faculty of Global Environmental Science, Kim Il Sung University, Pyongyang, Democratic People's Republic of Korea

## 1 Introduction

Drought is the natural hazard defined as a lack in precipitation compared to the average condition, and it causes great economic losses, which accounts for an overwhelming proportion of global loss caused by natural disasters (Keshavarz et al. 2013). Especially, drought will be death-dealing in the areas where mean precipitation observations at monthly or seasonal timescales were a little (Mishra and Singh 2010; Husak et al. 2013). Since the Korean peninsula is located at the monsoonal region whose climate varies with season, its precipitation depends on season and geographical position and this region is frequently affected by drought and flood. In the Democratic People's Republic of Korea (DPRK), drought mostly occurs in spring and autumn, which causes the damage in the fields of agriculture and water power. In particular, the growth of young crops suffers from spring drought. For example, during the springs of 2001, 2017 and 2019 most areas of the DPRK were affected from severe droughts. Especially, in the rice-cultivating areas of the North and South Hwanghae Provinces and North and South Phyongan Provinces rice seedlings were affected. Also, autumn drought may affect the production of hydroelectric power stations during winter characterized by relatively less precipitation. Therefore, improving the drought monitoring and its prediction method is essential in minimizing the damages in the fields of agriculture and power production in the DPRK.

Damage caused by drought occurs with time lags, so that drought is generally estimated with timescales and its persistence. It is very important to establish an early drought warning system and take a rational help to be prepared for the coming damage (Belayneh et al. 2012; Mouatadid et al. 2018). Drought warning system is accomplished by applying effective forecasting models using several drought indices.

Several drought indices have been suggested for quantitative estimation of lack of precipitation, such as standardized precipitation index (SPI) (McKee et al. 1993), standardized precipitation evapotranspiration index (SPEI) (Vicente-Serrano et al. 2010), Palmer drought severity index (PDSI) (Palmer 1965), drought severity index (DSI) (Nalbantis and Tsakiris 2009) and so on. SPI is used as the most universal drought index because not only calculation of SPI is very simple and the value of SPI is the standardized value but also it is possible to analyze multi-timescale characteristic of drought (WMO 2012). Therefore, SPI is selected as multi-timescale drought index for assessment of dry or wet condition in this study.

Simplistic approaches like the autoregressive integrated moving average (ARIMA) model as well as more complex nonlinear approaches using support vector regression (SVR) and artificial neural network (ANN) models can be applied to drought forecasting (Rezaeianzadeh et al. 2016; Belayneh et al. 2012; Zhang et al. 2017; Zhang et al. 2019; Fathabadi et al. 2009; Mokhtarzad et al. 2017; Deo et al. 2017a; Djerbouai and Souag 2016). Although ARIMA model is the most common model for time series forecasting, it has the limitations that is not suitable for hydro-meteorological time series forecasting with non-stationary and nonlinearity (Zhang et al. 2017). Some recent studies show that SVR and ANN models can reflect nonlinear relation between input and output more accurately than ARIMA model in drought forecast. The performances of SVR and ANN models based on machine learning technique vary with applied fields. Deo et al. (2017) used multivariate adaptive regression, least square support vector machine and M5 Tree model for forecasting SPI in eastern Australia. The results by Khan and Coulibaly (2006) showed that an SVR model for lake water levels forecasting outperformed ANN models. SVR model for SPI drought forecasting at Khorasan province of Iran was more effective

than ANN model (Mokhtarzad et al. 2017). While SVR model for SPEI-12 forecasting in the Sanjiang Plain of China didn't outperform ARIMA model, which was the best model of used three models (Zhang et al. 2019). Although some of the applications of ANN model showed that ANN model didn't outperform ARIMA and SVR models, the ability of ANN is still high because of model's structure that well describes nonlinearity relation between input and output (ASCE 2000).

The ability of machine learning method to forecast non-stationary drought time series is limited (Belayneh et al. 2014). A solution to overcome this limitation is to apply data pre-processing methods such as wavelet transform and empirical mode decomposition. Wavelet transform is a useful data pre-processing tool that gives successful decomposition of original data, and it is effective in forecasting nonlinear and non-stationary time series (Renaud et al. 2005; Murtagh et al. 2004). It can be implemented on multi-resolution levels for capturing useful information, and the utilization of decomposed data instead of original data can improve the ability of a machine learning model. Belayneh et al. (2016) compared the bootstrap SVR, the bootstrap ANN, the boosting SVR, the boosting ANN, the wavelet bootstrap SVR, the wavelet bootstrap ANN, the wavelet boosting ANN and wavelet boosting SVR models. The results showed that wavelet boosting machine learning methods outperformed other methods. Djerbouai and Souag (2016) designed WANN models coupled wavelet decomposition on several resolution levels using several mother wavelets from db1 to db17, and compared with ARIMA. Synthetic analysis of previous researches for forecasting SPI drought emphasizes the utility of wavelet analysis (Anshuka et al. 2019). Ali et al. (2019) used the multivariate empirical mode decomposition (MEMD) formalized by Huang et al. (1998) as a data pre-processing tool to predict SPI drought. Pre-processing of 12 synoptic-scale climate indices using MEMD explicitly improved the ability of the kernel ridge regression algorithms used for SPI forecasting. Although data pre-processing methods can make some improvement about the ability of models, SVR and ANN models have some disadvantages because of the limitation of itself (Tu 1996; Sapankevych and Sankar 2009).

Deep learning has been made a big stride in artificial intelligence technique in recent years. Augmentation of training data and improvements of training algorithms of deep layer models increase the usability of the artificial intelligence technique (Nikhil 2017; Grover et al. 2015). Shi et al. (2015) developed the convolutional long short-term memory (ConvLSTM) layers for precipitation nowcasting. Veillette et al. (2018) represented the usability of deep learning by consisting ConvLSTM for the generation of radar image and verifying the effectiveness. Du et al. (2018) estimated the ability of deep layer neural network for precipitation forecasting with training data from 200 to 2000. Results showed the deep belief network based on multi-layer restricted Boltzmann machines can overcome the shortcomings of traditional forecasting methods. Although deep belief network for precipitation forecasting has been suggested, it has not been applied to drought forecasting.

The purpose of this paper is to investigate the ability of LSTM network coupled with a wavelet decomposition on multi-resolution levels for SPI drought forecasting in the west area of the DPRK. In this study, Haar mother wavelet function was used as mother wavelet, and wavelet-based LSTM network (WLSTMN) was compared with WANN and WSVR models. WANN and WSVR models for SPI drought forecasting have been developed in some other studies, but they haven't been investigated in the DPRK yet (Anshuka et al. 2019). Also, traditional SVR and ANN models didn't outperform wavelet-based models, therefore, in this study only wavelet-based models have been investigated. This study forecasted SPI-6 and SPI-12 (SPI with 6 and 12-month timescales) which are factors of medium and long-term drought conditions, and the performances of models at

different decomposition levels for 1-month lead time were estimated by normal performance measures.

The rest of the paper is organized as follows. Section 2 presents study area, used data and the SPI calculation method, and provides a brief description of the machine learning models coupled with wavelet decomposition and their performance measures. Section 3 presents the results from three models, and the detailed discussion and conclusion are described in Sect. 4.

## 2 Materials and methodology

### 2.1 Study area and data

Study area encompasses the west area of the DPRK including 6 spots (Pyongyang, Kanggye, Sinuiju, Phyongsong, Sariwon and Haeju) (Fig. 1), which are the main weather stations of Pyongyang city, Jagang Province, North and South Phyongan Provinces, North and South Hwanghae Provinces. North and South Phyongan Provinces, North and South Hwanghae Provinces are the main granaries of the DPRK. The geographical locations and statistical characteristic values of precipitation for six weather stations are presented in Table 1. Study area is located at the East Asian monsoon region, so precipitation varies terribly during an annual cycle, but is mostly concentrated around the rainy season from June to September and drought mainly occurred from March to May.

The daily precipitation data from 6 weather stations covering a period of 1960–2020 are used, and these data were taken from the State Hydro-Meteorological Administration (SHMA) of the DPRK. The meteorological observations in study area started at different time: Pyongyang (1907), Sinuiju (1931), Haeju (1945), Kanggye (1952) and Sariwon (1954). Missing data is 5% of daily precipitation data before 1960 and few after 1960. The maximum daily precipitation, 411.5 mm, was observed in Sinuiju and the maximum and minimum monthly mean precipitation days were 14.1 in July in Haeju and 2.5 in January in Sinuiju, respectively. These data have been exactly corrected and quality-controlled by using contemporaneous data from neighboring spots.

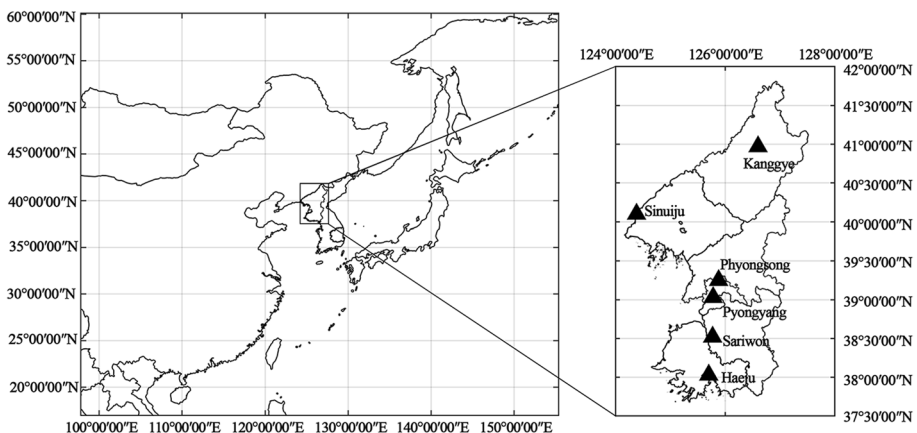


Fig. 1 Study area

**Table 1** Geographical locations and statistical characteristic values of precipitation for six meteorological stations

Station	Site Code	Latitude (N)	Longitude (E)	Max annual precipitation (mm)	Min annual precipitation (mm)	Mean annual precipitation (mm)	Standard deviation (mm)
Pyongyang	47.058	39.03	125.78	1712.2	504.9	967.8	283.9
Kanggye	47.020	40.97	126.6	1345.4	543.4	858.0	210.6
Sinuiju	47.035	40.1	124.38	1627.4	647.2	1017.3	258.1
Phyongsong	99.103	39.25	125.88	2005.0	614.5	1080.2	321.5
Sariwon	47.065	38.52	125.77	1743.1	496.2	873.2	282.4
Haeju	47.069	38.03	125.7	2164.2	616.9	1100.2	346.0

## 2.2 Standardized precipitation index (SPI)

SPI suggested by McKee et al. (1993) and McKee et al. (1995) is the meteorological drought index, which represents the degree of deviation from the mean precipitation. The main advantage of the SPI index is that the calculation of SPI is uncomplicated because it only depends on precipitation records (Logan et al. 2010) and it is possible to analyze the strength and duration of drought on multi-timescales (Tsakiris and Vangelis 2004; Mishra and Desai 2006; Mishra et al. 2007). The detailed calculation method of SPI is presented in Thom (1958), Edwards and McKee (1997) Abramowitz and Stegun (1965) and Zhang et al (2017).

Drought classification by SPI values is shown in Table 2 (WMO 2012).

In this study, drought indices forecasted by the machine learning models are the SPI-6 and SPI-12 series, which were derived at 10 days' intervals. SPI series based on monthly precipitation include 12 elements for a year. However, according to the main purpose of this study to elucidate the ability of deep learning we derived SPI with 10 days' intervals instead of monthly SPI so that the length of training samples could be extended. Calculation of SPIs was implemented at the 10th, 20th and the last days of every month. For example, SPI-6 of January 10, 1961 is calculated by the cumulated precipitation from July 11, 1960 to January 10 and SPI-12 by the cumulated precipitation from January 11, 1960 to January 10, 1961. In case of leap year, the precipitation of February 29 is added to that of February 28. In this way, SPIs are produced by the cumulated precipitation of 6 or 12 months moving the interval to every 10 or 11 days. The length of SPI time series is  $(2020-1960)*3*12 = 2160$ .

## 2.3 Wavelet decomposition

Wavelet transform is a useful mathematical tool to analyze nonstationary time series. Wavelet analysis can be applied to reveal aspects of breakdown points and discontinuities, and compress or de-noise a signal (Kim and Valdes 2003). In addition, it can be used for Prediction of time series (Renaud et al. 2005; Murtagh et al. 2004). Wavelet transform can be implemented according to two algorithms; the first method is the continuous wavelet transform (CWT) and the latter is the discrete wavelet transform (DWT). CWT at time  $t$  for a time series  $f(t)$  is defined as following Eq. (1) (Nason and Von Sachs 1999),

**Table 2** Drought classification of SPI value

SPI value	Drought class
$\geq 2.00$	Extremely wet
1.50–1.99	Very wet
1.00–1.49	Moderately wet
–0.99 ~ 0.99	Near normal
–1.0 ~ –1.49	Moderately dry
–1.5 ~ –1.99	Severely dry
$\leq -2.00$	Extremely dry

$$W_{a,b}(t) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt \tag{1}$$

where  $a$  is the scale parameter,  $b$  is the position parameter, and  $\psi$  is the mother wavelet function (Kim and Valdes 2003).

CWT inheres complexity and requires tedious computation time, so DWT is often used to forecast time series as follows (Cannas et al. 2006):

$$W(j, k)_t = 2^{-j/2} \sum \psi(2^{-j/2} - k) f(t) dt \tag{2}$$

where integers  $j$  and  $k$  are known as the scale and position parameters, which control the scale and translation respectively.

Disadvantage of DWT for forecasting application is that the usual DWT is based on a decimated one. This algorithm leads to decrease of decomposed time series; therefore, non-redundant DWT can't be applied to problems related to forecasting. Another disadvantage of DWT is that DWT uses future data values. Clearly, this becomes a difficulty for prediction problems that must be given attention to the boundary. This problem has been sufficiently discussed in Renaud et al. (2005), Murtagh et al. (2004), Adamowski and Sun (2010) and Belayneh et al. (2014). In order to overcome these disadvantages of DWT, nondecimated or redundant version, known as the a' trous algorithm has been suggested by Mallat (1998). The smoother versions  $c_p$  and detail coefficients  $d_p$  of original series  $x(t)$  at decomposition level  $p$  are defined at different decomposition levels as given by Eqs. (3)–(5)

$$c_0(t) = f(t) \tag{3}$$

$$c_{j+1}(t) = \sum_{l=-\infty}^{+\infty} h(l) c_j(t + 2^{j-1} l) \tag{4}$$

$$d_j(t) = c_{j-1}(t) - c_j(t) \tag{5}$$

where  $c_0(t)$  is the original signal,  $h$  is the low pass filter;  $j$  is the decomposition level;  $l$  is the shift frequency.

### 2.4 Wavelet support vector regression (WSVR)

Support vector machines (SVM) can be applied to classification and regression problems (Gao et al. 2001). Since the initial purpose of this study is to forecast quantitative SPI drought, the SVM for regression problem known as SVR was used. SVR is a supervised learning model that is based on the structural risk minimization principle (Vapnik 1995). In other words, the goal of SVR model is to minimize generalization error, assuming the linear relation between a set of predictors  $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N\}$  and targets  $\{y_1, y_2, \dots, y_N\}$ . If the relationship between predictor and target is not linear, predictors can be mapped to high dimensional space. The formula for the function between predictor and target as follows:

$$f(x) = w^T \cdot \varphi(x) + b \tag{6}$$

where  $w$  and  $b$  are coefficient vector and scalar that have to be estimated from predictor and target data, respectively, and  $\varphi(x)$  defines the nonlinear mapping of  $x$ . Minimization of  $\epsilon$  loss SVR proposed by Vapnik (1995) is defined as Eq. (7). This optimization problem is converted to the quadratic programming problem as following Eq. (11) by utilizing Lagrange multiplier method, and then support vectors refer to predictor vectors corresponded to the positive Lagrange multipliers

$$\min_w R(f) = \frac{1}{2}w^T w + C \sum (\xi_i + \xi_i^*) \tag{7}$$

subject to

$$\begin{aligned} \forall n; y_n - (w^T \cdot \varphi(x_n) + b) &\leq \epsilon + \xi_n \\ \forall n; (w^T \cdot \varphi(x_n) + b) - y_n &\leq \epsilon + \xi_n^* \\ \forall n; \xi_i &\geq 0, \xi_i^* \geq 0 \end{aligned}$$

where  $C$  is the regularization parameter;  $\xi_i$  and  $\xi_i^*$  are nonnegative slack variables that represent the deviation of the prediction and observation, respectively.

$$\max \text{imize} - \frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i \cdot x_j) - \epsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \tag{8}$$

subject to  $\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C]$ . where  $\alpha_i, \alpha_i^*$  are the Lagrange multipliers,  $K(x_i \cdot x_j)$  is called kernel function defined as scalar product in the feature space, regression function for new  $x$  is given by Eq. (9).

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i \cdot x) + b \tag{9}$$

Used kernel function is the radial basis function (RBF), which is nonlinear function as follows:

$$K(x, y) = \exp\left(-\frac{x - y}{2\gamma^2}\right) \tag{10}$$

In this study, predictors used to predict the SPI series are the wavelet-based decomposed series at different decomposition levels and the models were trained using fivefold cross-validation. The optimal model was decided through training and validation stage. Validation data correspond to the set of 2012–2020.

### 2.5 Wavelet artificial neural network (WANN)

ANN models are machine learning models that adhere to the empirical risk minimization principle as opposed to the structural risk minimization principle used by traditional SVR (Vapnik 1995). The advantage of ANN is that it automatically creates relationships between input (predictors) and target without analysis of variables (Hydrology 2000). The other advantage of ANN is that it controls noises contained in training data for itself by training. Recently, with the development of statistical programs, the application of ANN is very easy and many studies for drought forecasting show that ANN models improved



the ability of drought forecasting (Fathabadi et al. 2009; Belayneh et al. 2012; Zhang et al. 2017; Mokhtarzad et al. 2017; Djerbouai and Souag 2016). Multi-layer perceptron (MLP) feed-forward network has been used extensively to forecast SPI drought series and more detailed architecture can be found in Mishra and Desai (2006) and Belayneh et al. (2012).

This study used MLP feed forward network with a hidden layer, the input and output layer, which was trained according to the Levenberg–Marquardt backpropagation algorithm. The decomposed series was used as input data of the ANN models and the number of neurons of hidden layer was decided by trial and error method. Sampled data was partitioned into 3 sets of training, validation and testing.

## 2.6 Wavelet long short-term memory network (WLSTMN)

In recent years, deep learning has achieved rapid development of the artificial intelligence fields such as image classification and time series prediction. Long short-term memory network (LSTMN) can avoid vanishing gradient problem of traditional recurrent neural network (RNN) by adding a way to carry information across long time steps as the extended formula of RNN (Hochreiter and Schmidhuber 1997). Input layer of LSTM network is composed of a sequence input (SI) layer, which is the core component of LSTMN. Time series data is inputted into the LSTM layer by a SI layer. LSTM layer contains LSTM units composed of an input gate, a forget gate, a cell with a self-recurrent connection, and an output gate, which can remove or add successfully information. The updating algorithm of an LSTM unit and the role of four components were introduced by Hochreiter and Schmidhuber (1997) and Felix et al. (2014). The stochastic gradient descent (SGD) (Murphy 2012), root-mean-square propagation (RMSProp) (Hinton 2012) and adaptive moments (Adam) (Kingma and Ba 2014) optimizers can be applied to update the LSTMN parameters for minimizing the loss function.

In general, the increase of the number of hidden layers may lead to overfitting (Srivastava et al. 2014). Overfitting can happen in every supervised learning problem, but in case the model is trained with insufficient sampled data, it can happen severely. Reducing the size of the model is the way to avoid overfitting. Another way is to add weight regularization is known as L1 and L2 regularization (Neumaier et al. 1998) and dropout layer (Srivastava et al. 2014). In addition, cross-validation (Loughrey et al. 2005) and early stopping (Kohavi 1995) can be adopted to overcome overfitting.

LSTMN used to forecast SPI series in this study has a SI layer, a LSTM layer with 100 neurons, two fully connected (FL) layers with 50 and 1 neurons, respectively, a dropout layer with 50 neurons, a regression output layer with 1 neuron and the structure of model is shown in Fig. 2. The dimension of SI layer depends on wavelet decomposition level and dropout ratio in a dropout layer varies from 0.1 to 0.5. Sampled data was partitioned as in WANN model. All the WSVR, WANN and WDNN models were developed with the MATLAB (R.2021a) software (Deep Learning Toolbox, Statistics and Machine Learning Toolbox and Wavelet Toolbox).

## 2.7 Performance measures

The performance of developed models was estimated by the coefficient of determination ( $R^2$ ), Lin's concordance correlation coefficient (LCCC) (Lin 1989), root-mean-squared error (RMSE) and mean absolute error (MAE) expressed as Eq. (11)–(14):

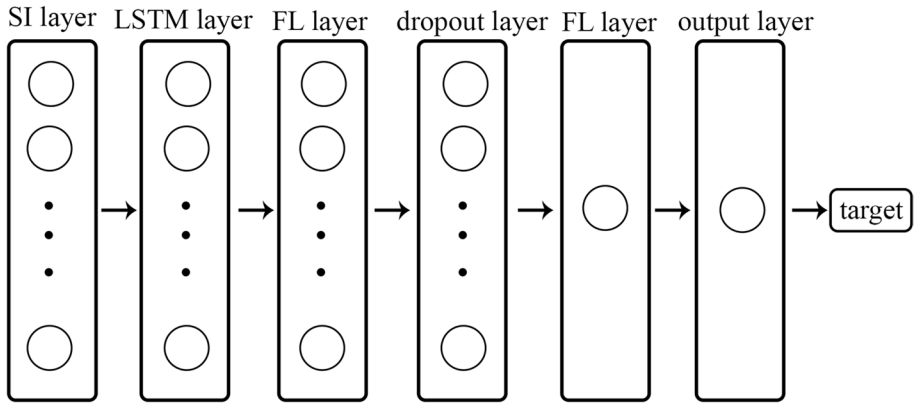


Fig. 2 The structure of WLSTMN to forecast SPI drought

$$R^2 = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \tag{11}$$

$$LCCC = \frac{2\rho\sigma_o\sigma_p}{\sigma_o^2 + \sigma_p^2 + (\mu_o - \mu_p)^2} \tag{12}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \tag{13}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \tag{14}$$

where  $y_i$  is the observed SPI,  $\hat{y}_i$  is the predicted value,  $\rho$  is the Pearson correlation coefficient between the observed and predicted SPI,  $\sigma_o$  and  $\sigma_p$  are the corresponding variances of the observed and predicted SPI,  $\mu_o$  and  $\mu_p$  are the means for the observed and predicted SPI.

The coefficient of determination is the value of 0 to 1, and the highest value of  $R^2$  represents the best model. The LCCC indicates the degree to which pairs of the observed and predicted SPI fall on the 45° line through the origin. The smaller the values of RMSE and MAE, the better the performance of developed model.

### 3 Results

In this study, WSVR, WANN and WLSTMN models to predict 1 month-ahead SPI drought were developed for six stations in the west area of the DPRK (Pyongyang, Kanggye, Sinuiju, Phyongsong, Sariwon and Haeju). Predictions were generated using the above-mentioned three models for SPI-6 and SPI-12. The mother wavelet used to pre-process data was

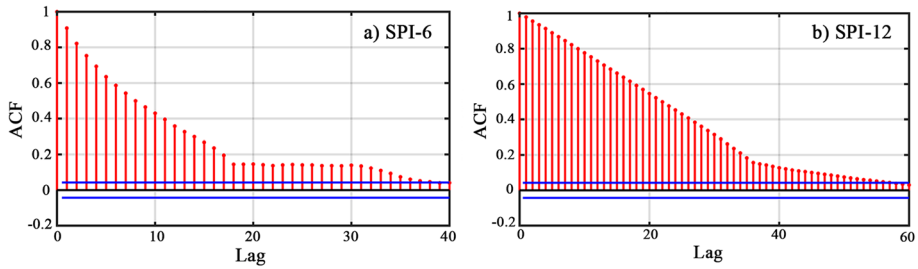
the Haar wavelet and SPI time series was decomposed at 1–10 levels. For WSVR models, 85% of the data was used to train and validate, and 15% of the data was used to test. All the WANN and WLSTMN models were trained using 70% of the data, while 15% of the data was used to validate models and the final 15% of the data was used to test models. Then, the testing period corresponds to the period of 2012–2020. The predicted SPI time series by the optimal models determined through optimization on training and validation stages were compared with the observed SPI time series for the testing period. Three performance measures and optimal decomposition level (ODL) for SPI-6 and SPI-12 are presented in Table 3. As shown in Table 3, all the models exhibited better results for SPI-12 compared to forecasts of SPI-6. In addition, the results showed the optimal models for the study area were the WLSTMN models which had the highest  $R^2$  and LCCC, the smallest RMSE and MAE for most stations and all the timescales. The detailed results by six models at 6 stations will be mentioned in following subsections.

### 3.1 WSVR models

As mentioned earlier, wavelet decomposed series with time lag was used as the inputs of WSVR models. The limitation of lag with significant correlation was determined by the auto-correlation function (ACF). Figure 3 is ACF figures for SPI-6 and SPI-12 at Pyongyang station, respectively (other stations were omitted). The number of inputs were chosen to have between 1 and 16 lags for SPI-6, while the values from 1 and 34 lags were tested for SPI-12. The number of lag and decomposition level with the optimal performance selected

**Table 3** The performance measures and ODL for SPI-6 and SPI-12 during the period of testing

Model	Station	SPI-6					SPI-12				
		ODL	$R^2$	LCCC	RMSE	MAE	ODL	$R^2$	LCCC	RMSE	MAE
WSVR	Pyongyang	3	0.680	0.799	0.554	0.401	5	0.903	0.928	0.298	0.200
	Kanggye	3	0.625	0.728	0.614	0.444	5	0.837	0.877	0.391	0.262
	Sinuiju	3	0.685	0.813	0.613	0.461	5	0.903	0.944	0.331	0.216
	Phyongsong	3	0.705	0.820	0.602	0.427	5	0.905	0.944	0.323	0.190
	Sariwon	3	0.660	0.796	0.581	0.418	5	0.932	0.944	0.258	0.152
	Haeju	3	0.701	0.820	0.662	0.456	5	0.932	0.944	0.328	0.200
WANN	Pyongyang	5	0.694	0.786	0.541	0.409	6	0.905	0.946	0.289	0.190
	Kanggye	3	0.641	0.720	0.598	0.430	6	0.876	0.929	0.332	0.199
	Sinuiju	2	0.687	0.772	0.607	0.472	5	0.895	0.938	0.340	0.216
	Phyongsong	6	0.693	0.740	0.624	0.449	7	0.891	0.935	0.341	0.240
	Sariwon	5	0.680	0.773	0.573	0.425	6	0.930	0.958	0.263	0.159
	Haeju	5	0.719	0.806	0.638	0.467	5	0.918	0.937	0.365	0.256
WLSTMN	Pyongyang	3	0.722	0.838	0.522	0.390	3	0.914	0.947	0.287	0.184
	Kanggye	4	0.665	0.752	0.577	0.425	5	0.906	0.951	0.288	0.175
	Sinuiju	2	0.705	0.799	0.586	0.449	5	0.908	0.948	0.318	0.203
	Phyongsong	3	0.724	0.822	0.575	0.422	5	0.916	0.955	0.296	0.179
	Sariwon	4	0.706	0.796	0.544	0.418	5	0.936	0.947	0.260	0.176
	Haeju	4	0.728	0.825	0.631	0.455	5	0.935	0.955	0.326	0.224



**Fig. 3** ACF for SPI-6 and SPI-12 time series for the Pyongyang station

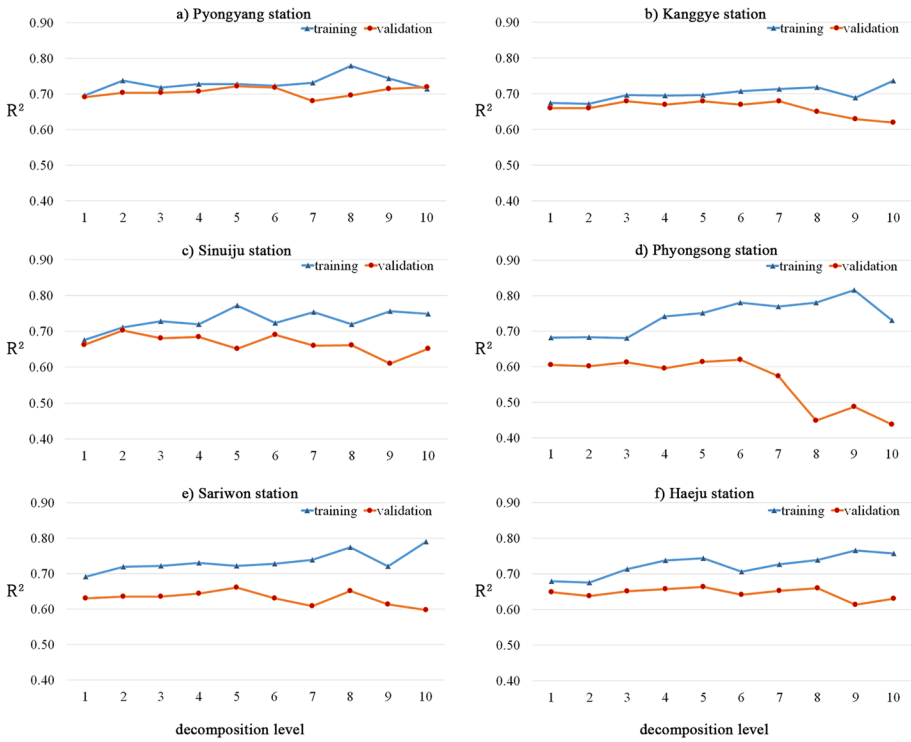
using a trial and error procedure were 12 and 3 for SPI-6 at all stations respectively, while the values were 15 and 5 for SPI-12 at all stations, respectively.

The best WSVR model for SPI-6 was proposed from Phyongsong, where the optimal model had  $R^2$  of 0.705, LCCC of 0.820, RMSE of 0.602, and MAE of 0.427. The best SVR model for SPI-12 had  $R^2$  of 0.932, LCCC of 0.944, with RMSE and MAE values of 0.258 and 0.152 in Sariwon, respectively. For SPI-6, the smallest  $R^2$  and LCCC were 0.625 and 0.728 in Kanggye, while the largest RMSE values 0.662 in Haeju. For SPI-12, the worst performance measures are  $R^2$  of 0.837, LCCC of 0.877, RMSE of 0.391, and MAE of 0.262 in Kanggye (Table 3).

### 3.2 WANN models

The values of  $R^2$  measured by WANN models at 1–10 levels for six stations during the training and validation period are shown in Figs. 4 and 5, in which the blue line shows comparison for the training period and the red line shows the validation period. The decomposition level with the highest  $R^2$  for the validation period of SPI-6 was not unique for all the six stations and the WANN models with more than 6 decomposition level had lower generalization capacity (Fig. 4a, b, c, d, e and f). Especially, as shown in Fig. 4d, the value of  $R^2$  for SPI-6 at Phyongsong station had a sharper decreasing trend with increasing of decomposition level (more than 6) than other stations. For SPI-12 forecasts, increasing of decomposition level (more than 7) depressed the performance measure during the validation period as opposed to the training period (Fig. 5). According to the comparison of the performance criterion at different decomposition levels, the optimal decomposition level of wavelet-based decomposition for SPI-6 forecasting was 5 in Pyongyang, Sariwon and Haeju, and 2, 3, 6 in Sinuiju, Kanggye and Phyongsong, respectively, while the values for SPI-12 were 6 in Pyongyang, Kanggye and Sariwon, 5 in Sinuiju and Haeju, and 7 in Phyongsong. In an optimal decomposition level, the differences between  $R^2$  of training and validation data do not exceed 0.02 in Pyongyang (Fig. 5a) Sinuiju (Fig. 5c), Sariwon (Fig. 5e), Haeju (Fig. 5f), but 0.04 and 0.07 in Kanggye and Phyongsong, respectively, which is larger than other stations. These results represent that the generalization capacity of the WANN models for Kanggye and Phyongsong stations is lower than the generalization capacity of models for all the other stations.

For SPI-6, the  $R^2$  and LCCC values of WANN models during testing period were greater than 0.64 and 0.72 respectively, with the highest values of 0.719 and 0.806 in Haeju, and the lowest RMSE and MAE values were 0.541 and 0.409 in Pyongyang, respectively (Table 3). For SPI-12, the  $R^2$  and LCCC values of the forecasted results were greater



**Fig. 4** The coefficient of determination measured by the WANN models at different decomposition levels for SPI-6 during the training and validation period

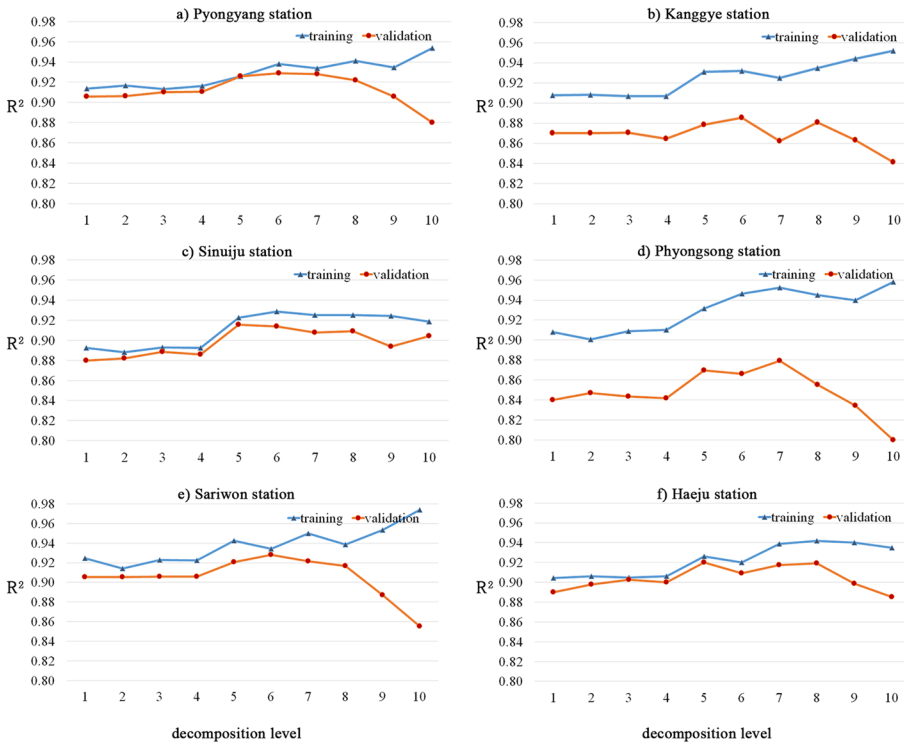
than 0.87 and 0.92 respectively, with the highest value of 0.930 and 0.958, and the lowest RMSE and MAE values of 0.263 and 0.159, respectively in Sariwon.

### 3.3 WLSTMN models

The results by the WLSTMN models vary with the maximum number of training epochs. The performance measures of all the WLSTMN models trained with more than 100 epochs were extremely high for training data, while low for validation data. The optimal maximum numbers of training epochs determined by the trial and error method ranged from 30 to 50. Also, Here, Adam optimizer was used as the solver of training network.

Figures 6 and 7 show the  $R^2$  values of WLSTMN models with 1–10 levels for SPI-6 and SPI-12, respectively. Figure 6 shows that the WLSTMN models with more than 5-level have the low value of  $R^2$  for the validation data of SPI-6, and the optimal decomposition level for six stations ranges from 2 to 4. The optimum values were 3 in Pyongyang and Phyongsong, 2 in Sinuiju and 4 in Kanggye, Sariwon and Haeju for SPI-6. The results in Fig. 7 indicate that the  $R^2$  values for the validation data of SPI-12 exhibit the same conformation as indicated by variation of  $R^2$  values for SPI-6 with decomposition levels, and the optimal decomposition level for SPI-12 is 3 in Pyongyang and 5 in all the other stations.

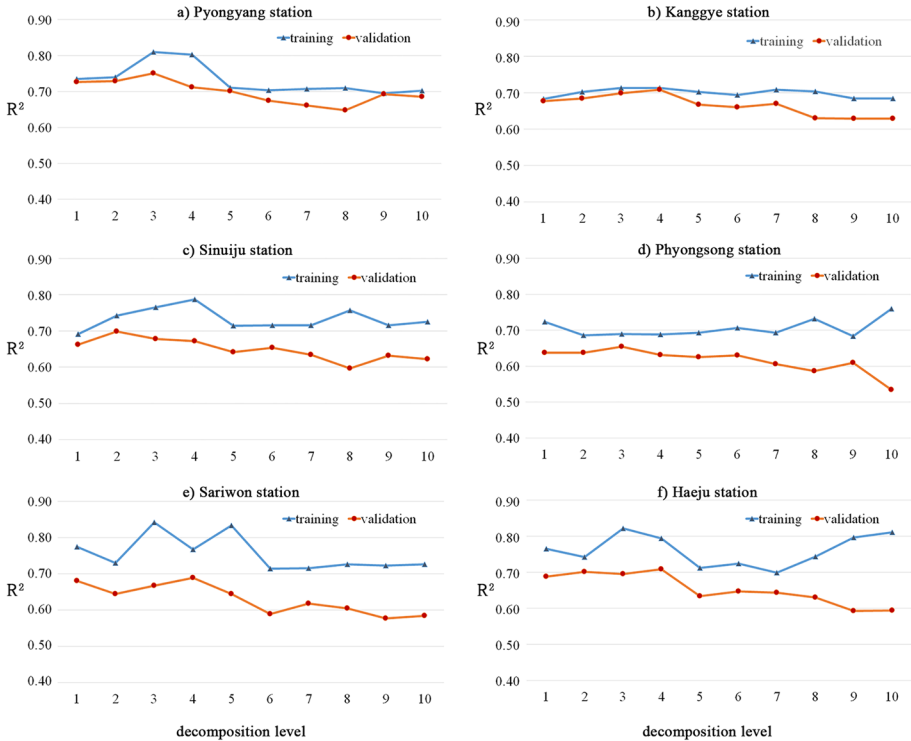
For the validation data of SPI-6, the  $R^2$  values of the best model (WLSTMN) were greater than 0.65, with the highest value of 0.75 in Pyongyang and the lowest value in



**Fig. 5** The coefficient of determination measured by the WANN models at different decomposition levels for SPI-12 during the training and validation period

Phyongsong (Fig. 6), while, for the validation of SPI-12, the  $R^2$  values were greater than 0.88, with the highest value of 0.93 in Sariwon and the lowest value in Phyongsong (Fig. 7). And the  $R^2$  value in Phyongsong was lower than the other stations (Fig. 7d). During the testing period of SPI-6, the  $R^2$  and LCCC values were greater than 0.66 and 0.75, respectively, with the highest  $R^2$  of 0.728 (Haeju) and LCCC of 0.838 (Pyongyang) and the lowest RMSE and MAE values were 0.522 and 0.390, respectively in Pyongyang, (Table 3). During the testing period of SPI-12 the  $R^2$  and LCCC values were greater than 0.90 and 0.94, respectively, with the highest value of 0.936 and 0.955 (Sariwon and Haeju), and the lowest RMSE and MAE values were 0.260 and 0.175 (Sariwon and Kanggye), respectively.

As shown in Table 3, the best forecasting results of SPI-6 in Pyongyang were obtained by WLSTMN model, in which the values of  $R^2$ , LCCC, RMSE, and MAE were 0.722, 0.838, 0.522, and 0.390, respectively. Meanwhile, the values of the best performance measures ( $R^2$ , LCCC, RMSE, MAE) of SPI-12 for WLSTMN were (0.914, 0.947, 0.287, 0.184) in Pyongyang. Namely, the best results of SPI-12 in Pyongyang were also obtained by WLSTMN model. For SPI-6, LCCC of WSVR in Sinuiju was greater than that of WLSTMN, but its other measures ( $R^2$ , RMSE and MAE) were worse than WLSTMN. Also, for SPI-12 in Sariwon, LCCC, RMSE and MAE of WANN were better than WLSTMN. In other stations, all measures of WLSTMN were the best as

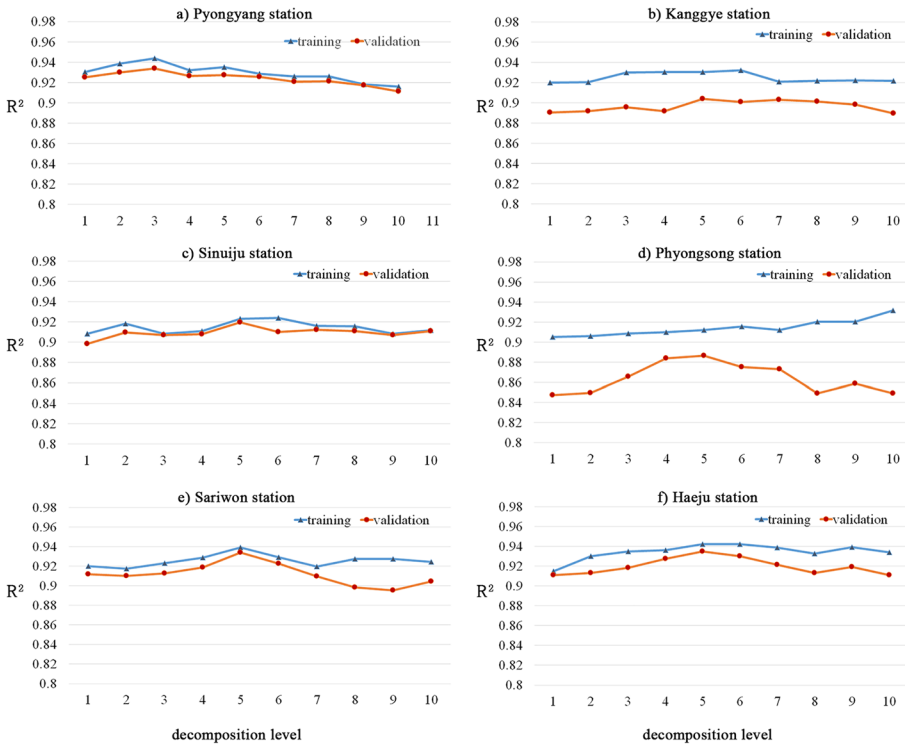


**Fig. 6** The coefficient of determination measured by the WLSTMN models at different decomposition levels for SPI-6 during the training and validation period

compared to WANN and WSVR. The above results show that WLSTMN for SPI-12 has the best performance in all stations except for WANN in Sariwon.

The scatter plots of SPI observations and values predicted by WSVR, WANN and WLSTMN models at the six stations for the testing period are shown in Figs. 8 and 9. These figures indicate most datapoints around the trend line and several points below or over this line under a certain level of underestimation or overestimation. In particular, datapoints are closer to the trend line for SPI-12 than for SPI-6. Figures 8 and 9 show the improvement of results predicted by WLSTMN compared to WSVR and WANN. For SPI-6, the differences of  $R^2$  and LCCC values between WLSTMN and other models are particularly large in Pyongyang and the scatter plot of WLSTMN (Fig. 8a) indicates the data points closer to the trend line than other models. These improvements are visible in other stations with difference of a certain degree. Especially the locations of data points around extreme value ( $SPI < -1.5$ ) were clearly improved. (Figs. 8 and 9).

Figure 10 shows the average performance measures at six stations for the testing period. The WLSTMN model provides the better performance than the WSVR and WANN models, with the highest  $R^2$  of 0.709 and LCCC of 0.806, the lowest RMSE of 0.572 and the lowest MAE of 0.427 for SPI-6 (Fig. 10a), and the highest  $R^2$  of 0.919 and LCCC of 0.950, the lowest RMSE of 0.296 and the lowest MAE of 0.190 for SPI-12 (Fig. 10b). The performances of SPI-6 for WSVR and WANN models were similar, with  $R^2$  values of 0.676 and 0.686, LCCC values of 0.795 and 0.766, respectively, and RMSE of 0.604 and 0.597,



**Fig. 7** The coefficient of determination measured by the WLSTMN models at different decomposition levels for SPI-12 during the training and validation period

MAE values of 0.434 and 0.442, respectively. All the performance measures of SPI-12 were similar for WSVR and WANN models. According to the results shown in Table 3 and Figs. 8, 9 and 10, it can be concluded that the best SPI forecasting model is the WLSTMN model for both of SPI-6 and SPI-12 time series.

### 4 Discussion and conclusion

The results from this study represent that the WLSTMN model outperformed the accuracy of WSVR and WANN models. The use of the LSTM unit allowed the SPI prediction models to be more reliable. Although deep learning requires enormous sampled data, the WLSTMN model developed in this study outperforms the other models for all the SPIs and all the stations. Due to the limitation of sample size, it cannot achieve a remarkable improvement of performance measures. However, deep learning with LSTM unit for forecasting time series can clearly overcome the shortcoming of traditional SVR and ANN models (Du et al. 2018; Zhao et al. 2019).

SVR on the structural risk minimization principle may outperform ANN on the empirical one. However, with the introduction of wavelet decomposition of input data, ANN can make a comparable prediction of SPI time series with SVR. WANN model used in this



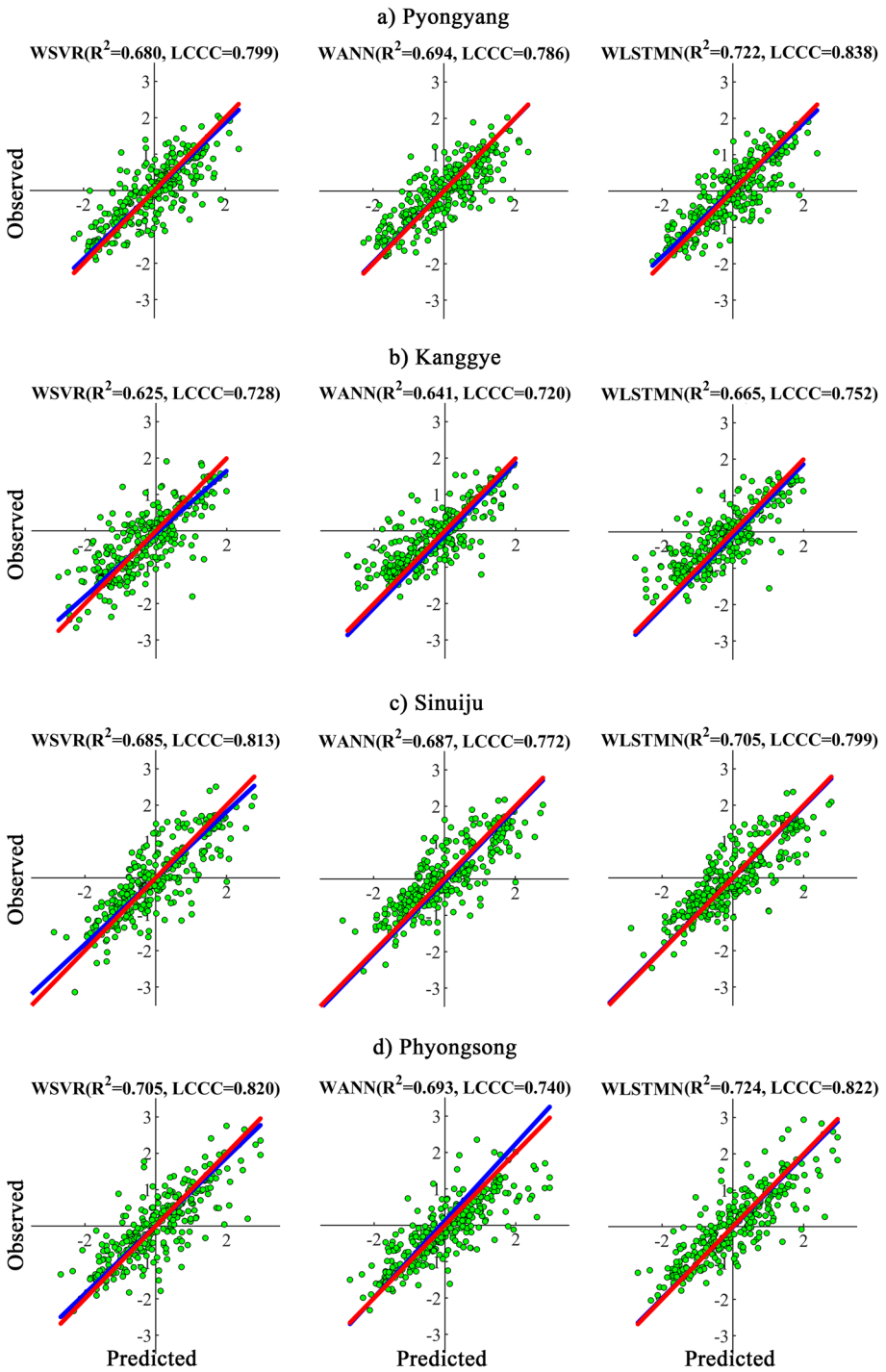
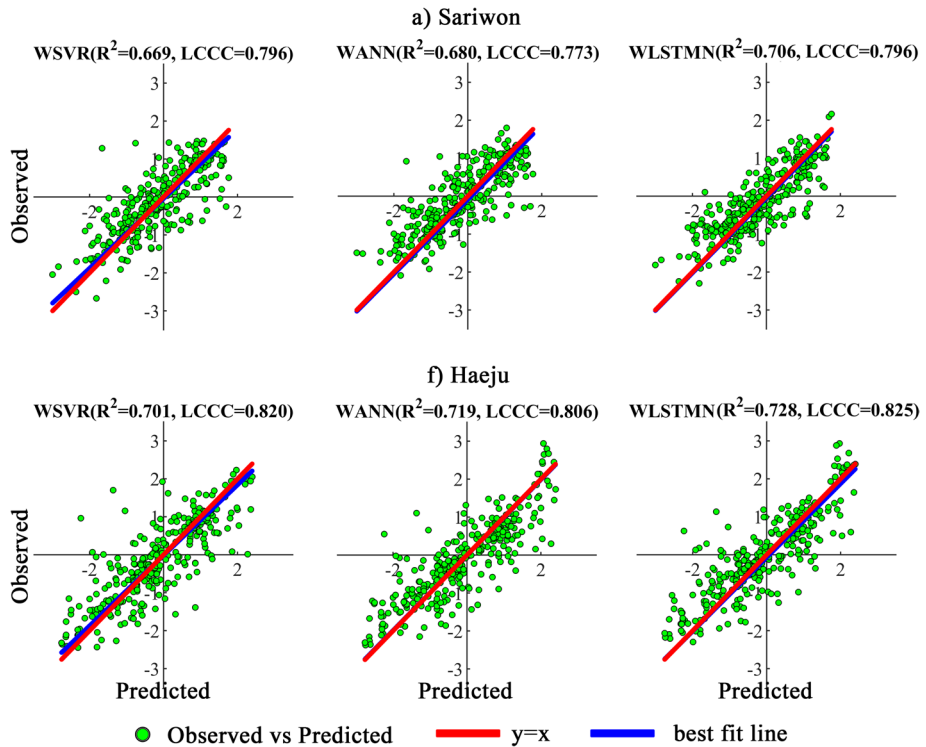


Fig. 8 Scatter plots of the three models at six stations (SPI-6)



**Fig. 8** (continued)

study outperformed slightly WSVR (Fig. 10). This result is consistent with previous studies (Belayneh et al. 2016).

It is known that drought prediction performance in the Korean Peninsula is lower than other regions. In particular, the smaller the temporal scale is, the larger the prediction error is (Anshuka et al. 2019). The result in this study shows that the performance of drought prediction for SPI-12 is better than for SPI-6. As shown in Figs. 8, 9 and Table 3, performance measures for all models are available.

Although the performance of models coupled wavelet decomposition is good, it depends on the decomposition level (Djrbouai and Souag 2016). The optimal decomposition level is determined differently with region, temporal scale and model. The reasonable decomposition level is decisive in improving the performance of SPI time series prediction. Overlarge decomposition level may lead to the increase of performance for the training data, but decrease for validation data. In case the model with many parameters has limited training data, overfitting is inevitable. The limitation of WANN models in SPI time series prediction is represented in a bit large differences between training data and validation data performances in Phyongsong (Fig. 5d) and Kanggye (Fig. 5b) for SPI-12. This shows that the generalization performance of WANN is low in SPI prediction in over-mentioned stations. In order to check whether this result is the production of random separation of training sample or not, we modified the training and validation data from the whole data repeatedly, but the difference between their performances did not reduce. This implies WANN for Kanggye and Phyongsong did not represent better the characteristics of time

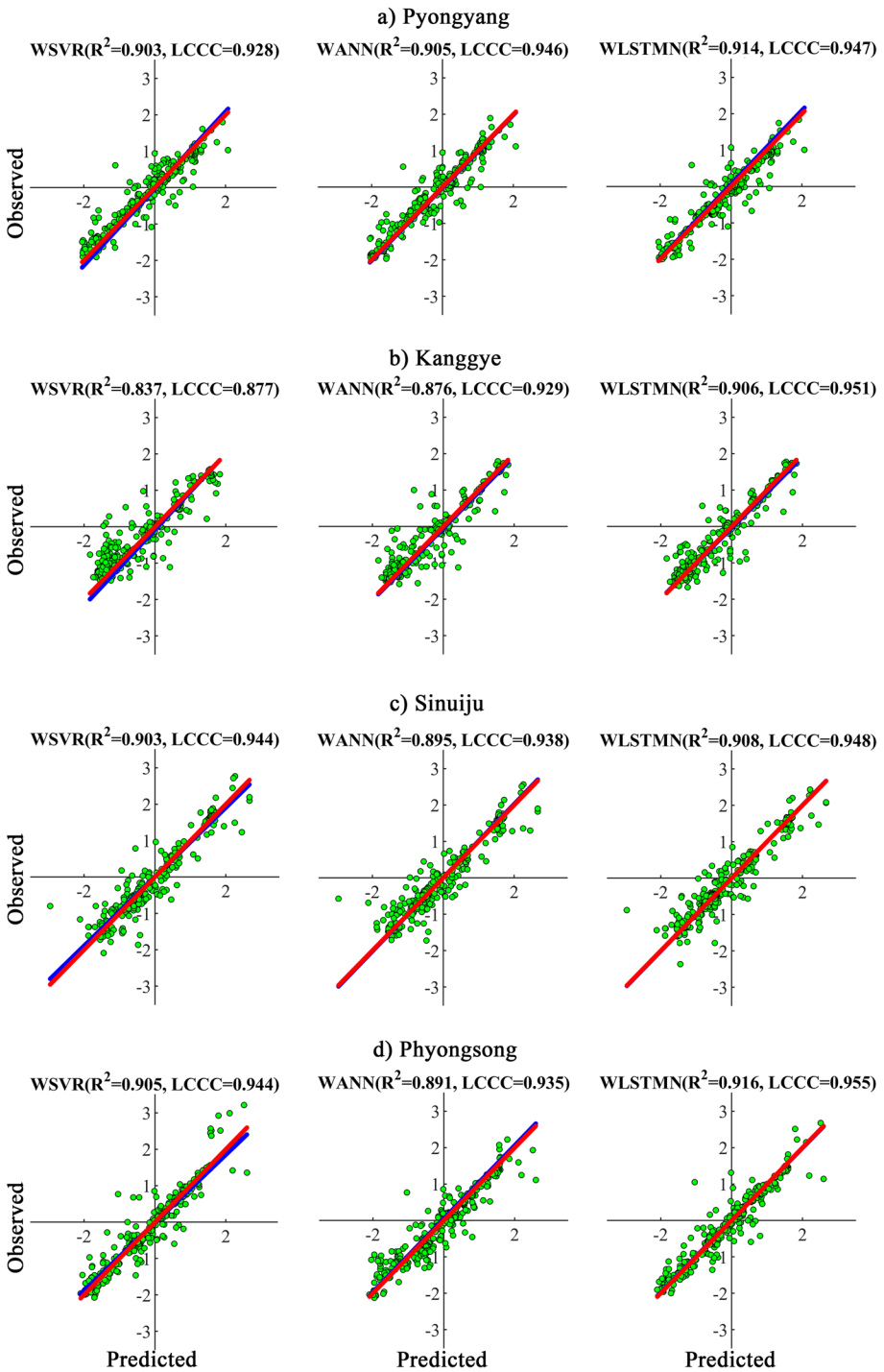


Fig. 9 Scatter plots of the three models at six stations (SPI-12)

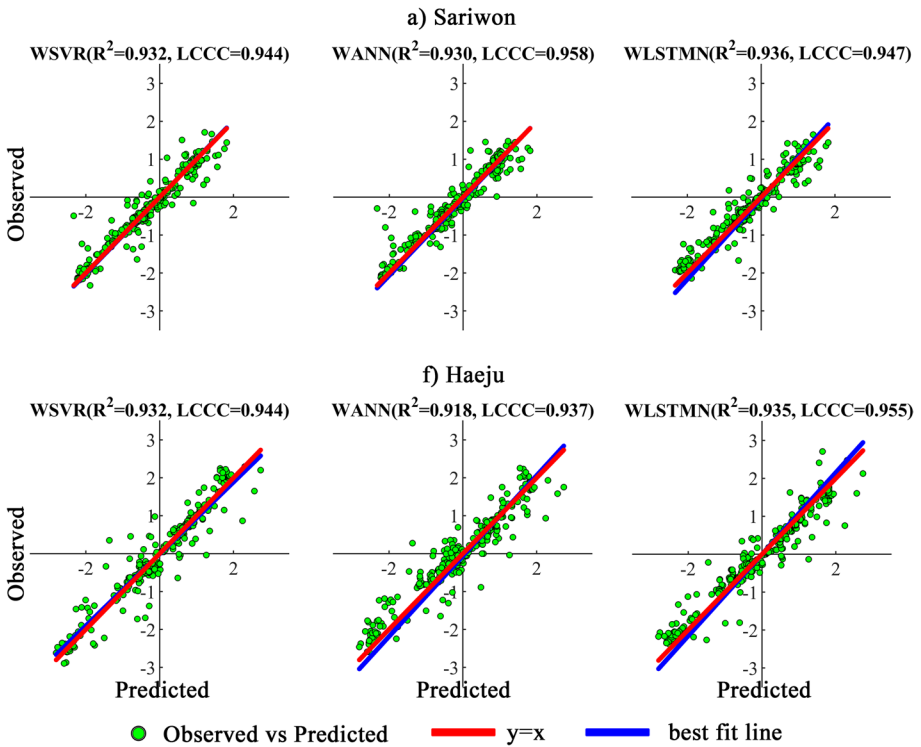


Fig. 9 (continued)

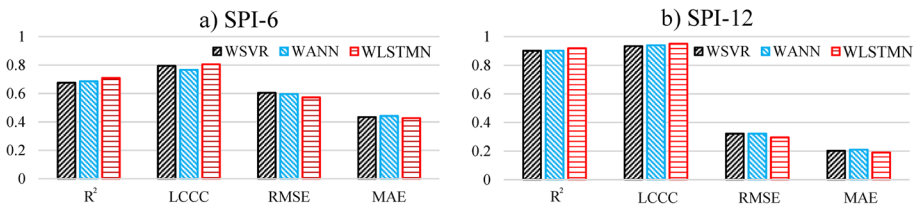


Fig. 10 The averages of performance measures for the testing period

series compared to the models for other stations. But the difference between  $R^2$  values of the training and validation data has a maximum of 0.07, which is not bad as compared to the result of (Zhang et al. 2019; Deo et al. 2017b, a). For SPI-6, the difference is 0.15 in Phyongsong (Fig. 4d), larger than other stations. This may result from the largest decomposition level of 6 in Phyongsong, hence the number of neurons of input layer more than other stations. The performance of model could be improved through the increased number of hidden layer and intrusion of dropout layer (Du et al. 2018; Grover et al. 2015).

This limitation of WANN was overcome in some degree through the application of WLSTM (Figs. 6d, 7b and d). The differences between performance measures of the training and validation data in Kanggye and Phyongsong were reduced in WLSTM. The addition of LSTM layer lead to the solution of the vanishing-gradient problem and the safe transfer of the characteristics of time series at the former time step to the next steps.

Furthermore, the overestimation or underestimation around extreme values by WSVR and WANN was improved in WLSTMN (Figs. 8 and 9). Figure 11 shows the observed and predicted values of three models for the severely or extremely dry events with SPI less than  $-1.5$  for SPI-6 and SPI-12 in Pyongyang (other stations were not shown). In some cases, WLSTMN has a greater error than WSVR or WANN, but for most of drought events its predicted values are closer to the observed values than other models. With a note that the prediction around extreme values is essential in drought forecasting, it could be conclusive that WLSTMN is more superior.

Figure 12 shows the confusion matrix of drought obtained by the quantitative predicted values of SPI-6 and SPI-12 in three models. As shown in the confusion matrices of WSVR and WANN models, they were not effective for extreme droughts (Fig. 12a, b). For the testing data, two extreme droughts at 6-month time scale occurred, but WSVR and WANN predicted ‘moderately dry’ events while WLSTMN predicted ‘severely dry’ events for these two events. Also, during the testing period, 31 severe droughts were observed, exactly predicted numbers of WSVR, WANN and WLSTMN are 10, 11 and 13, respectively. For the moderate droughts, the exact predicted number of WLSTMN is greater than other models. But WLSTMN overestimated the moderate droughts into ‘severely dry’, whose number is greater than other models. For SPI-12, six extreme droughts occurred during the testing period, WLSTMN predicted them into ‘severely dry’ and others misclassified two extreme droughts into ‘moderately dry’. Generally, the classification accuracy was the highest in WLSTMN for extreme and severe droughts, but similar in all models for moderate droughts. For SPI-6 and SPI-12, the misclassified drought classes did not exceed two classes for all the events and dry events were not predicted into ‘wet’. This implies that the models developed in this study could be applied to not only drought forecasting but its classification prediction. However, prior to the application to the classification prediction, the comparison of performances of WSVR, WANN and WLSTMN for the classification could be required through further research.

As shown in Fig. 10, area-averaged performance measures of WLSTMN are better than those of WSVR and WANN. Also,  $R^2$  and LCCC values of these three models for SPI-6 are larger than those for SPI-12 (Fig. 8 and 9). Thus, it can be pointed out that the models developed in this study predict the SPI-12 series with moderate change better than SPI-6 series with sharp change. These results are consistent with Zhang et al. 2017 and Anshuka et al. 2019. RMSE and MAE of WLSTMN were less than those of other models. RMSE and MAE of all the developed models are larger than those of data-driven models

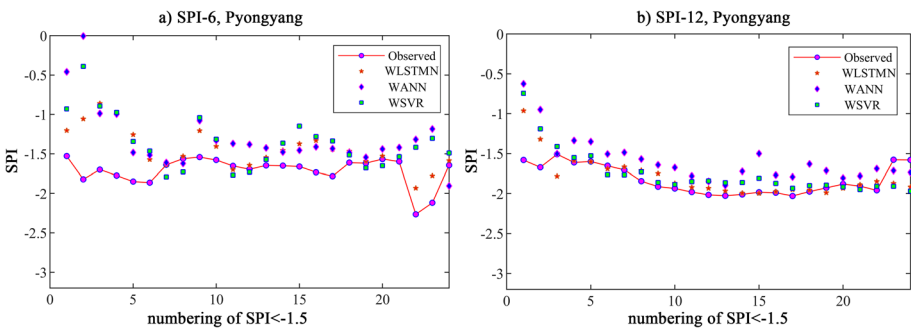
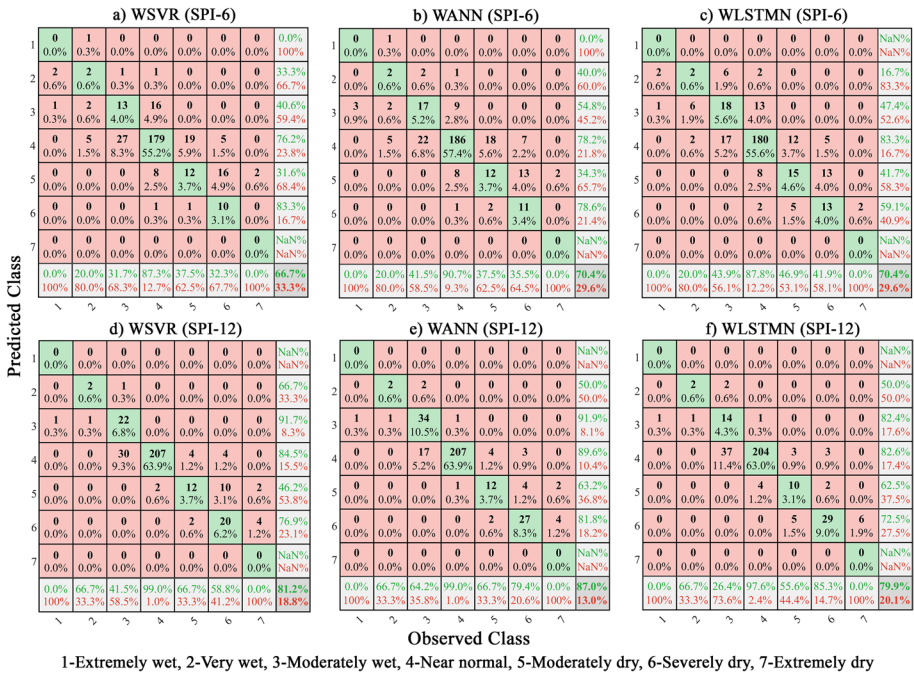


Fig. 11 The extreme value predictions of three models for the testing data in Pyongyang



**Fig. 12** The confusion matrix of SPI-6 and SPI-12 for the testing data in Pyongyang

suggested by Belayneh et al. (2016) and Ali et al. (2019). This is because drought forecasting scores depend on the area and time scale as mentioned in Anshuka et al. (2019). With a regard to the low performance of SPI drought forecasting in the Korean Peninsula, the values of RMSE and MAE obtained by this study are not bad. In particular,  $R^2$  and LCCC showed the better accuracy. The numbers of neurons of LSTM layer, FL and dropout layer were fixed 100, 50 and 50, respectively, and the training data was not enough. In addition, wavelet mother function was restricted to Haar. Due to these limitations, WLSTMN model may not outperform significantly WSVR and WANN.

In this study, WSVR, WANN and WLSTMN models based on coupling machine learning technique with Haar wavelet decomposition were proposed to forecast SPI drought series during the period of 1961–2020 in the west area of the DPRK. WLSTMN with the best performance is operational in the drought forecasting of the study area. Through the comparison of the prediction performances of all the models, the following conclusions were derived.

First, WLSTMN models provided better prediction results compared to WANN and WSVR models. The WLSTMN model had an advantage compared to the WANN and WSVR models for drought predictions in the west area of the DPRK, with respect to the  $R^2$ , LCCC, RMSE, MAE values. Next, the comparisons of area-averaged performances of the WANN and WSVR models showed that the WANN model exhibited higher  $R^2$  and lower LCCC during the testing period, while their RMSE and MAE values were approximately the same. Then, the results of all machine learning models for SPI-12 showed better prediction results than SPI-6 for all the stations.

Although wavelet analysis was an effective pre-processing tool, incorporating models with overlage decomposition levels didn't avoid overfitting. This study presents the

drought predictions of WLSTMN, WANN and WSVR models using Haar wavelet at 1-month lead time. Further studies should be required to estimate the performance of machine learning models coupled with different Daubechies wavelets (Djrbouai and Souag 2016), and especially WLSTMN with different neurons of hidden layers should be explored. Also, these new models should be applied in any other stations with different characteristics and at longer lead times.

**Acknowledgements** We acknowledge the State Hydro-Meteorological Administration (SHMA) of the DPRK and the data providers, Prof. Dr. C.-M. Song and S.-H. Kim.

**Funding** The authors have not disclosed any funding.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Abramowitz M, Stegun IA (1965) Handbook of mathematical functions. Dover publications, New York
- Ali M, Deo RC, Maraseni T, Downs NJ (2019) Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. *J Hydrol* 576:164–184
- Anshuka A, Ogtrop FF, Vervoort RW (2019) Drought forecasting through statistical models using standardised precipitation index: a systematic review and meta-regression analysis. *Nat Hazards*. <https://doi.org/10.1007/s11069-019-03665-6>
- Belayneh A, Adamowski J (2012) Standard precipitation Index drought forecasting using neural networks, wavelet neural networks, and support vector regression. *Appl Comput Intell Soft Comput*. <https://doi.org/10.1155/2012/794061>
- Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *J Hydrol* 508:418–429
- Belayneh A, Adamowski J, Khalil B, Quilty J (2016) Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmos Res* 172:37–47
- Cannas B, Fanni A, Sias G, Tronci S, Zedda MK (2006) River flow forecasting using neural networks and wavelet analysis. In: Proceedings of the European geosciences union
- Deo RC, Kisi O, Singh VP (2017) Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. *Atmos Res* 184:149–175
- Djrbouai S, Souag-Gamane D (2016) Drought forecasting using neural networks, wavelet neural networks, and stochastic models: case of the Algerois Basin in North Algeria. *Water Resour Manag* 30:2445–2464
- Du JL, Liu YY, Liu ZJ (2018) Study of precipitation forecast based on deep belief networks. *Algorithms* 11:1132. <https://doi.org/10.3390/a11090132>
- Edwards DC, McKee TB (1997) Characteristics of 20th century drought in the United States at multiple time scales. Colorado State University, Fort Collins. Climatology Report No. 97–2, CO
- Fathabadi A, Gholami H, Salajeghe A, Azanivand H, Khosravi H (2009) Drought forecasting using neural network and stochastic models. *Adv Nat Appl Sci* 3(2):137–146
- Felix AG, Schmidhuber JA, Cummins FA (2014) Learning to forget: continual prediction with LSTM. *Neural Comput* 12(10):2451–2471
- Gao JB, Gunn SR, Harris J, Brown M (2001) A probabilistic framework for SVM regression and error bar estimation. *Mach Learn* 46:71–89
- Govindaraju R (2000) ASCE Task committee on application of artificial neural networks in hydrology (2000) Artificial neural networks in hydrology. I. Preliminary concepts. *J Hydrol Eng* 5(2):124–137
- Grover A, Kapoor A, Horvitz E (2015) A deep hybrid model for weather forecasting. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, KDD 15, pp 379–386, New York, NY, USA, ACM

- Hinton G (2012) Neural networks for machine learning. Coursera Video Lect 264:2146–2153
- Hochreiter S, Schmidhuber J (1997) Long short—term memory. *Neural Comput* 9(8):1735–1780
- Huang NE et al (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond A Math Phys Eng Sci* 454(1971):903–995
- Husak GJ, Funk CC, Michaelsen J, Magadzire T, Goldsberry KP (2013) Developing seasonal rain-fall scenarios for food security early warning. *Theoret Appl Climatol*. <https://doi.org/10.1007/s00704-013-0838-8>
- Keshavarz M, Karami E, Vanclay F (2013) The social experience of drought in rural Iran. *J L Use Policy* 30:120–129
- Khan MS, Coulibaly P (2006) Application of support vector machine in lake water level prediction. *J Hydrol Eng* 11(3):199–205
- Kim T, Valdes JB (2003) Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *J Hydrol Eng* 8:319–328
- Kingma D, Ba J (2014) Adam: a method for stochastic optimization. preprint <http://arxiv.org/abs/1412.6980>
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *Int Joint Conf Artif Intell* 14:1137–1143
- Lin LJ (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45:255–268
- Logan KE, Brunsell NA, Jones AR, Feddema JJ (2010) Assessing spatiotemporal variability of drought in the US central plains. *J Arid Environ* 74:247–255
- Loughrey J, Cunningham P (2005) Using early stopping to reduce overfitting in wrapper-based feature weighting. Trinity College Dublin Department of Computer Science, TCD-CS-2005-41, pp 12
- Mallat SG (1998) A wavelet tour of signal processing. Academic, San Diego, p 577
- McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In: Paper presented at 8th conference on applied climatology. American Meteorological Society, Anaheim, CA
- McKee TB, Doesken NJ, Kleist J (1995) Drought monitoring with multiple time scales. In: 9th Conference on applied climatology, American meteorological society, Boston, pp 233–236
- Mishra AK, Desai VR (2006) Drought forecasting using feed-forward recursive neural network. *Ecol Model* 198(1–2):127–138
- Mishra AK, Singh VP (2010) A review of drought concepts. *J Hydrol* 391(1):202–216
- Mishra AK, Desai VR, Singh VP (2007) Drought forecasting using a hybrid stochastic and neural network model. *J Hydrol Eng* 12:626–638
- Mokhtarzad M, Eskandari F, Vanjani NJ, Arabasadi A (2017) Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environ Earth Sci* 76:729. <https://doi.org/10.1007/s12665-017-7064-0>
- Mouatadid S, Raj N, Deo RC, Adamowski JF (2018) Input selection and data-driven model performance optimization for predicting standardized precipitation and evaporation index in a drought-prone region. *Atmos Res*. <https://doi.org/10.1016/j.atmosres.2018.05.012>
- Murphy KP (2012) Machine learning: a probabilistic perspective. The MIT Press, Cambridge, Massachusetts
- Murtagh F, Starck JL, Renuad O (2004) On neuro-wavelet modeling. *Decis Support Syst* 37(4):475–484
- Nalbantis I, Tsakiris G (2009) Assessment of hydrological droughts revisited. *Water Resour Manag* 23:881–897
- Nason GP, Von Sachs R (1999) Wavelets in time-series analysis. *philosophical transactions of the royal society a: mathematical. Phys Eng Sci* 357:2511–2526
- Neumaier A (1998) Solving ill-conditioned and singular linear systems: a tutorial on regularization. *Siam Rev* 40(3):636–666
- Nikhil B (2017) Fundamentals of deep learning, O'Reilly
- Palmer WC (1965) Meteorological drought. US Weather Bureau, Washington
- Renaud O, Starck J, Murtagh F (2005) Wavelet-based combined signal filtering and prediction. *IEEE Trans Syst Man Cyber Part B Cybern* 35(6):1241–1251
- Rezaeianzadeh M, Stein A, Cox JP (2016) Drought forecasting using Markov chain model and artificial neural networks. *Water Resour Manag* 30:2245–2259
- Sapankevych NI, Sankar R (2009) Time series prediction using support vector machines: a survey. *IEEE Comput Intell Mag* 4:24–38
- Shi XJ, Chen ZR, Wang H, Yeung DY, Wong WK, Woo WC (2015) Convolutional lstm network: a machine learning approach for precipitation nowcasting. In: *Advances in neural information processing systems*, pp 802–810
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15:1929–1958



- Thom HCS (1958) A note on gamma distribution. *Mon Weather Rev* 86:117–122
- Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 49:1225–1231. [https://doi.org/10.1016/S0895-4356\(96\)00002-9](https://doi.org/10.1016/S0895-4356(96)00002-9)
- Vapnik V (1995) *The nature of statistical learning theory*. Springer, New York, NY, USA
- Veillette MS, Hassey EP, Mattioli CJ, Iskenderian H, Lamey MP (2018) Creating synthetic radar imagery using convolutional neural networks. *J Atmos Oceanic Tech* 35(12):2323–2338
- Vicente-Serrano SM, Begueria S, Lopez-Moreno JI (2010) A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J Clim* 23:1696–1718. <https://doi.org/10.1175/2009JCLI2909.1>
- World Meteorological Organization (WMO) (2012) *Standardized precipitation index user guide*. world meteorological organization, available at: [http://library.wmo.int/pmb\\_ged/wmo\\_8\\_en-2012.pdf](http://library.wmo.int/pmb_ged/wmo_8_en-2012.pdf)
- Zhang YH, Li WW, Chen QH, Pu X, Xiang L (2017) Multi-models for SPI drought forecasting in the north of Haihe river basin. *China Stoch Environ Res Risk Assess* 31:2471–2481. <https://doi.org/10.1007/s00477-017-1437-5>
- Zhang YH, Yang HR, Cui HJ, Chen QH (2019) Comparison of the ability of ARIMA, WNN and SVM models for drought forecasting in the Sanjiang plain, China. *Nat Resour Res*. <https://doi.org/10.1007/s11053-019-09512-6>
- Zhao JF, Mao X, Chen LJ (2019) Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomed Signal Process Control* 47:312–323

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.