



Machine learning–based prediction of agricultural drought using global climatic indices for the Palakkad district in India

Saranya Das K.¹ · N. R. Chithra¹

Received: 30 May 2023 / Accepted: 9 February 2024 / Published online: 19 February 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2024

Abstract

Agricultural drought refers to soil moisture deficit, which causes adverse effects on the crop production and economy of a nation. This work compared the capability of artificial neural network (ANN) and support vector machine (SVM) algorithm in predicting agricultural drought in the Palakkad district of Kerala, India. Also, the influence of various global climatic indices on soil moisture stress in the study area is assessed. Two models were developed to investigate the impact of global climatic indices. Model 1 considered only local meteorological variables as predictors, and model 2 included global climatic indices along with meteorological variables. The results showed that ENSO has commendable influence on the early prediction of agricultural drought in Palakkad and are more evident at higher lead times (2 to 4 months). For the first model of ANN and SVM, the R^2 values at a 4-month lead range from 0.56 to 0.76 and 0.62 to 0.77, respectively. Similarly, for model 2, the R^2 varies from 0.61 to 0.77 and 0.75 to 0.82 for ANN and SVM models, respectively. Further, the results indicated that the SVM model shows clear advancement in prediction over ANN especially at higher lead times, even though both show a comparable performance at 1-month lead time. The study provided useful information regarding the potential predictors of agricultural drought in the study area and suggest suitable models for the early prediction. This will support the decision makers in drought prevention and water resource management.

1 Introduction

Drought is a destructive natural disaster, which propagates very slowly with its impacts remaining, even after the drought event (Pendergrass et al. 2020). Among the other natural hazards, drought causes great damage and a massive loss to the agricultural sector and hence economic stability of a nation (Zargar et al. 2011; Li et al. 2020). Drought is mainly of four types: meteorological drought, agricultural drought, hydrological drought, and socio-economic drought (Mishra & Singh 2010). Among these, agricultural drought has higher importance as it directly affects food security in India (Mishra et al. 2014). The water demand has increased due to the increasing rate of population and development (Sang et al. 2023). The water stress developed as a result of a drought event retards the crop root growth, which decreases

agricultural productivity. Extensive studies were reported around the world, related to the assessment and monitoring of agricultural drought. But only a limited number of studies were reported on agricultural drought prediction.

Early prediction of agricultural drought is most important to develop suitable mitigation measures and to reduce the damage caused by the disaster. In the beginning years of agricultural drought prediction, researchers considered only the meteorological variables or drought indices as predictors (Dutta et al. 2013; Maity et al. 2016). Fung et al. (2020) attempted to predict the Standardized Precipitation Evapotranspiration Index (SPEI) at various timescales using the SVR model and fuzzy-support vector regression (F-SVR) and boosted-support vector regression (BS-SVR) models. They observed that the accuracy of prediction increases with increasing timescale of the SPEIs. Prediction of agricultural drought by integrating machine learning methods and remotely sensed data was done by Feng et al. (2019). SPEI-3 was calculated using various climate and vegetation indices obtained from remote sensing data. They suggested that the proposed approach will be suitable for vegetated regions with limited in situ data availability.

✉ Saranya Das K.
saranya_p210117ce@nitc.ac.in; saranyadask@gmail.com
N. R. Chithra
chithranr@nitc.ac.in

¹ Department of Civil Engineering, National Institute of Technology Calicut, Kozhikode, Kerala, India

The adoption of global climatic indices as predictors in drought prediction is a significant milestone in this area. These indices highly influence the precipitation and temperature pattern of a region at different seasons (Wu et al. 2022). Chiew et al. (1998) studied the teleconnections between the El Niño Southern Oscillation (ENSO) and the hydroclimate of Australia using empirical methods. They found that there is a significant correlation between rainfall and streamflow with the Southern Oscillation Index (SOI) in most parts of Australia. Schepen et al. (2011) found that lagged ENSO events can be used as a potential predictor of seasonal rainfall in Australia. These findings encouraged the researchers to include large-scale climatic indices for predicting drought with a longer lead time (Dikshit et al. 2021). Zhang et al. (2021) predicted Standardized Soil Moisture Index (SSI) using Standardized Precipitation Index (SPI) and ENSO events using the meta-Gaussian model. They found that the prediction performance was highly region specific, and the model was better performed in the spring season than summer season of China. Tian et al. (2018) executed a study to predict agricultural drought in China using SVM. They developed seven different models by including ENSO and western Pacific subtropical high (WPSH) to predict SPEI-6. It was mentioned that the incorporation of climatic indices improved the prediction accuracy up to 3 months of lead time.

A machine learning (ML)-based approach is very popular to exhibit the complex nonlinear relation between independent and dependent variables. Singh and Borah (2013) suggested that time series analysis and ANN can be used to resolve the prediction problems of Indian summer monsoon rainfall. ANN was used successfully for the prediction of short- and long-lead-time meteorological drought (Erol Keskin et al. 2011; Le et al. 2016) and agricultural drought (Manatsa et al. 2008; Marj & Meijerink 2011; Shukla et al. 2011). Similarly, SVM is highly efficient to handle noisy data and has been used in many prediction works (Jehanzaib et al. 2021). Mokhtarzad et al. (2017) executed a study to forecast drought using SPI by ANN, adaptive neuro-fuzzy inference system (ANFIS), and SVM. They found that all three models hold low error and are suitable for prediction studies. They also highlighted the advantage of the SVM model over the other two models.

In India, former studies including global climatic indices are more limited to the prediction of Indian monsoon (Mooley & Parthasarathy 1983; Kirtman & Shukla 2000; Krishnamurthi 2000). Shukla et al. (2011) attempted to forecast Indian Summer Monsoon Rainfall Index (ISMRI) considering the ENSO indices. They reported that ISMRI shows a good correlation with Niño 3, Niño 3.4, and Niño 4 indices at various lag periods, and hence, these indices can accurately predict ISMRI. Shukla and Paolino (1983)

demonstrated that the antecedent Southern Oscillation Index (SOI) was a useful predictor of Indian monsoon rainfall.

The Kerala state in India is experiencing intermittent drought events in the past few decades. One of the reasons for this was the decreasing trend of southwest monsoon in Kerala. In Kerala, 2016 was identified as one of the driest years. According to the report by the India meteorological department, Kerala faces a 32% deficit from summer monsoon and 62% from post-monsoon during 2016. Abhilash et al. (2019) tried to identify the reason for this extreme dry period in Kerala. They investigated the relationship between Pacific teleconnection and Kerala rainfall. They found that stronger El Niño events negatively influence the summer monsoon in Kerala, whereas there was a positive relationship with post-monsoon rainfall. They also reported that the inter-annual variability and influence of teleconnection on Kerala Summer Monsoon Rainfall (KSMR) and ISMR are different.

The recurrent water stress adversely affects cultivation in Kerala, as more than 70% of agricultural fields depend on rainfall alone. The prolonged period of deficiency in rainfall and high rate of evaporation lead to the reduction in soil moisture content. In the past few years, the agriculture sector in the state of Kerala in India is facing challenges due to uncertainties caused by climate change. Economic Review 2021 by the Kerala state planning board reveals that the gross state value added (GSVA) from the agriculture sector shows a decline from 12.37% in 2013–2014 to 9.44% in 2020–2021. If the agricultural drought is predicted accurately for higher lead time, the disaster can be mitigated effectively. But studies on agricultural drought prediction are limited in Kerala. Similarly, the evaluation of direct impact of local meteorological variable on soil moisture deficiency is also limited.

This study aims (a) to identify the most suitable predictors from local meteorological variables and global climatic indices for the prediction of agricultural drought in the Palakkad district of Kerala, (b) to study the effect of global climatic indices on the soil moisture deficit in the study area, and (c) to suggest a suitable model for the early prediction of agricultural drought incorporating the advantages of ANN and SVM. Palakkad is the largest district in the State of Kerala and is known as the Granary of Kerala. According to the agricultural statistics (2018–19) of Kerala, Palakkad is in the first position for the production of rice. Midland and highland regions of Palakkad are under threat of lack of moisture content, and such area having low moisture content had been tremendously increased from 2001 to 2021 in Palakkad (Viswambharan, et al. 2022).

The remaining part of the paper includes four sections. Section 2 provides the study area, data used, and methodology for developing the models; results and discussions were

summarized in Sect. 3; and the conclusion of the work is given in Sect. 4.

2 Materials and methods

2.1 Study area

Palakkad is located within a latitude ranging from 10°20' to 11°14' North and a longitude from 76°02' to 76°54' East (Viswambharan, et al. 2022). It is situated at the foot of the Western Ghats and stands as a northern gateway to Kerala (Fig. 1). Palakkad is called as the Rice Bowl of Kerala.

The total district area is 4480 km² and the net cultivated area is 2840 km², 64% of the entire district area. Based on the physical features of the area, the Palakkad district is divided into Midland and Highland (District Census handbook-2011).

Palakkad has a tropical and dry climate. The major portion of rainfall (71%) is received during the southwest monsoon season, and the remaining is contributed from the northeast monsoon. The annual rainfall varies from 1883 to 3267 mm. The maximum temperature and minimum temperature ranges from 28.1 to 37.4 °C and from 22.2 to 25.3 °C, respectively. Humidity is high during morning hours throughout the year, and it is higher during the monsoon period. The direction of wind is predominantly from the west and the wind speed is high during August (13.6 kmph) (Central Ground Water Board Kerala Region, 2013).

2.2 Data

This study included daily rainfall data with 0.25° resolution, daily maximum temperature, and daily minimum temperature data with 1° resolution from six grid points on various parts of the study area. Data were obtained for a period of 1971 to 2020 from the India Meteorological Department.

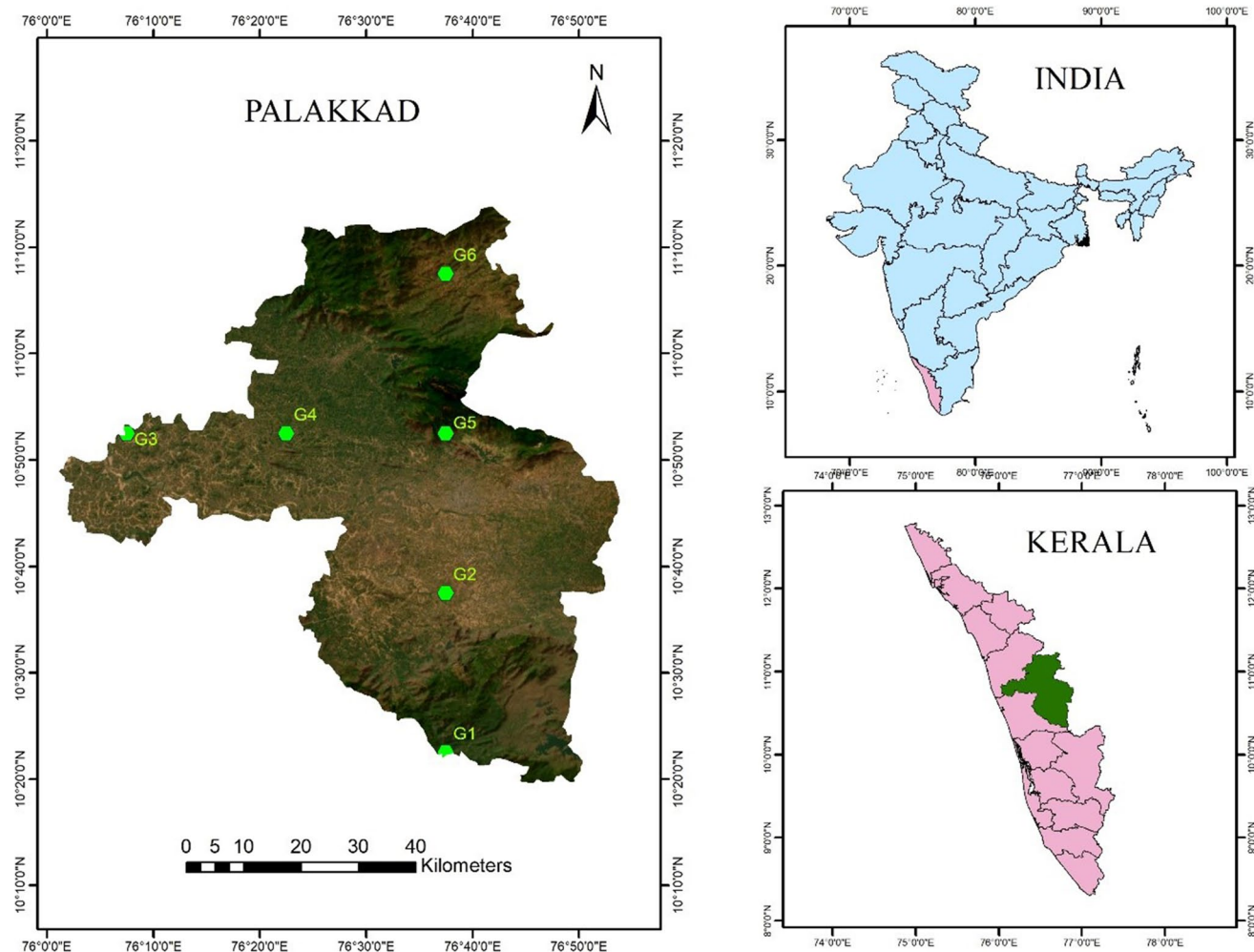


Fig. 1 Location of study area

The details of grid points along with location map of the study area are presented in Fig. 1.

The temperature data was interpolated using bilinear interpolation to match with rainfall data. Monthly time series values of root zone soil moisture, specific humidity, and wind speed were acquired from NASA Global Land Data Assimilation System (GLDAS) for the same period.

Climate index (CI) is a numerical value used to describe the state and change in the climate system. This study considered various types of climatic indices to evaluate their influence on the soil moisture variability in the study area. The indices include (1) indices based on sea surface temperature anomalies like the Atlantic Multi-decadal Oscillation (AMO), Indian Ocean Dipole (IOD), Pacific Decadal Oscillation (PDO), ENSO indices (Nino 1 + 2, Nino 3,

Nino 3.4, and Nino 4) and (2) indices based on sea pressure anomalies such as Arctic Oscillation (AO), North Atlantic Oscillation (NAO), Southern Oscillation Index (SOI), and Pacific–North American teleconnection pattern (PNA). The monthly time series values of these climatic indices were downloaded from <http://psl.noaa.gov> and <http://www.bom.gov.au> for the duration of 1971 to 2020.

2.3 Methodology

The overall procedure of the study is depicted in Fig. 2. The suitable predictors for soil moisture on the study area were identified based on correlation analysis (Pearson correlation and Spearman's correlation) performed between meteorological variables and global climatic indices with soil

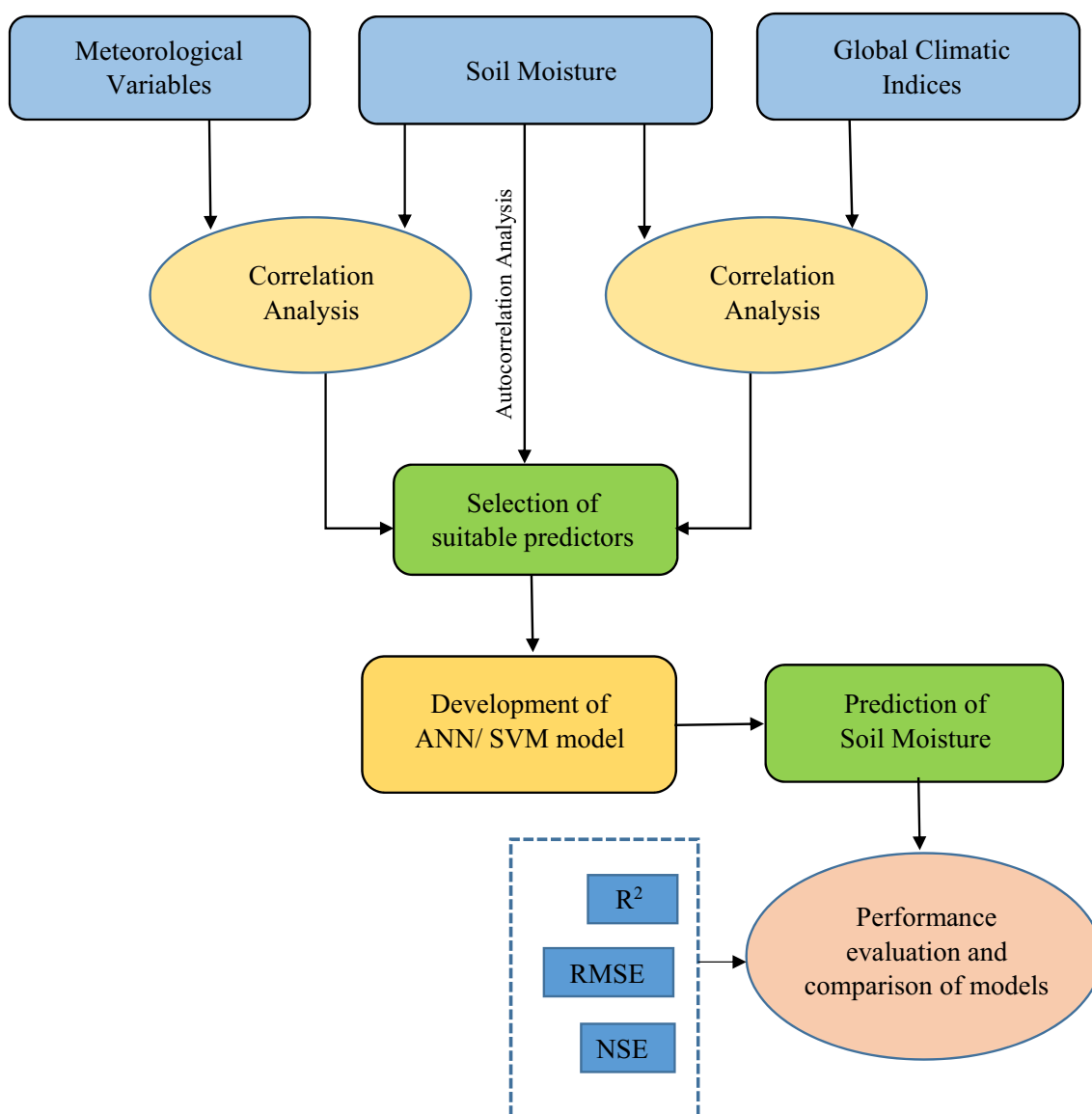


Fig. 2 Methodological flow-chart of the study

moisture value. The selected predictors were used as the inputs in various models by ANN and SVM. By comparing the observed data and predicted data from different models, the best model for accurate prediction of future soil moisture was identified.

2.3.1 Selection of suitable predictors

The selection of relevant predictors having high influence on the predictand is the most important step in a prediction study. Those variables which show strong correlation with soil moisture (as an indicator of agricultural drought) are selected as potential predictors. Autocorrelation analysis of soil moisture data was performed to identify the degree of correlation existing between soil moisture at two successive time intervals. Autocorrelation function (ACF) values range from -1 to 1 . The temporal lag at which maximum correlation obtained is considered in the prediction. Correlation analysis of meteorological variables and climatic indices at different lag periods with soil moisture was also made to identify the optimum lag at which these variables have a significant correlation with soil moisture.

2.3.2 Artificial neural network (ANN) modelling

ANN is a soft computation technique; its structure and operation resemble the functioning of the human brain. The method was developed by McCulloch and Pitts in 1943. ANN is capable of developing a relationship between independent and dependent variables which are respectively given as the input and output of a network (Morid et al. 2007).

ANN consists of many artificial neurons which are linked together in various layers according to a specific network architecture. The main three layers of an ANN architecture are (1) input layer: the input data is given into the network through the input neurons present in this layer. (2) Hidden layer: the processing of input data (weight assign, activation function) is taken place in this intermediate layer. A neural network may have zero or more hidden layers. As the number of hidden layers increases, the complexity of the network increases. (3) Output layer: outputs from the network are produced in this layer. Each neuron in one layer will be linked to all the neurons in the succeeding layer. More details on ANN can be found in Fausett (2006).

2.3.3 Support vector machine (SVM)

Support vector machines (SVM) are a supervised learning algorithm used for both classification and regression (Zhu et al. 2021) and were introduced by Vapnik in 1995. The basic concept of SVM is statistical learning theory and the structural risk minimization principle (Jehanzaib et al. 2021). The main aim of an SVM is to allow good generalization by reducing the risk.

SVM establishes a relationship between independent and dependent variables by combining these through kernel functions. This allows to predict the value of a dependent variable corresponding to new independent variables (Zhang et al. 2020). The equation for the same can be represented as:

$$y = f(x) + \varepsilon = w \cdot \varnothing(x) + b \quad (1)$$

where x is the independent variable, y is the dependent variable, w is the weight vector, b is a constant, and $\varnothing(x)$ is the non-linear mapping function.

Kernel functions are used to transform the training data into the required form to achieve better regression. Linear, polynomial, Gaussian, exponential, sigmoid, and radial basis function are some of the commonly used kernel functions.

2.3.4 Model evaluation measures

The performance of the model was valued by R^2 , RMSE, NSE, and MAPE. The R^2 value ranges from 0 to 1. As R^2 approaches 1, the prediction accuracy increases. A lower RMSE value indicates a better prediction (Luo et al. 2022). The range of NSE varies from $-\infty$ to 1. As the value increases, accuracy of the prediction also increases. A model having the NSE value 1 implies a perfect prediction. A negative NSE value represents unreliable prediction (Zhang et al. 2021). MAPE is the percentage equivalent of mean absolute error (MAE), and it measures the average magnitude of error produced by a model. The equations to calculate R^2 , RMSE, NSE, and MAPE are given below:

$$R^2 = \left[\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right]^2 \quad (2)$$

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

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - O_i}{O_i} \right| \quad (5)$$

where O is observed soil moisture, \bar{O} is the average of observed soil moisture, P is the predicted soil moisture, \bar{P}

is the average of predicted soil moisture, and n is the total number of observations.

3 Results and discussion

3.1 Autocorrelation of soil moisture

Autocorrelation analysis determines the linear dependence existing within a series. Figure 3 depicts the results of autocorrelation analysis of soil moisture data for temporal lags of 1–3 months at the six grid points. At all grid points, the ACF values give the highest positive value at 1-month lag, and the value decreases as the lag increases. At a lag of 3 months, the value of ACF changes its direction. Hence, antecedent soil moisture at 1-month lag was considered as one of the predictors.

3.2 Correlation between soil moisture and meteorological variables

The correlation between various local meteorological variables and soil moisture at lag periods of 1 to 6 months is computed by Pearson correlation coefficient and Spearman correlation coefficient. The correlation coefficient values show slight changes among the two methods, but the lag time at which each local meteorological variables shows strong correlation with soil moisture is similar in case of both analysis. The results of Pearson correlation analysis at various grid points are presented in Fig. 4. As rainfall and soil moisture are varying in the side direction, the temporal lag at which both show the highest positive correlation was considered in the prediction. At all grid points, rainfall shows a significant correlation at 1 month (Pearson correlation coefficient varies from 0.59 to 0.77 among six grid points) and 2 months (0.44 to 0.69) lag,

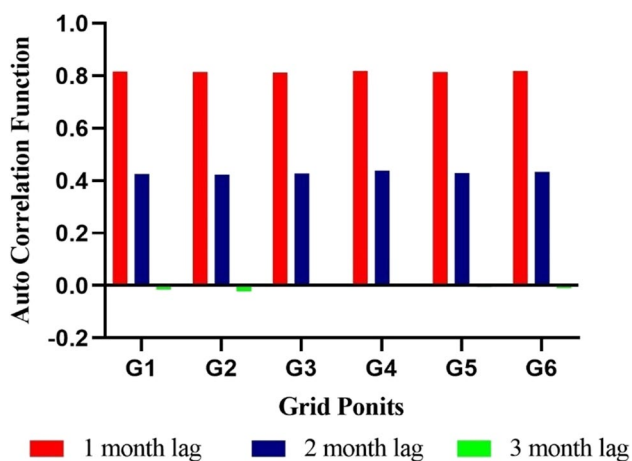


Fig. 3 Autocorrelation results at G1, G2, G3, G4, G5, and G6

and hence, rainfall values at these 2 lag times were considered as predictors. In case of analysis between rainfall and soil moisture, the study by Zhang et al. (2021) also reported similar results. They found that soil moisture and antecedent precipitation has high correlation at 1-month lag, and it decreases as the lag time increases. Lag times were selected for other variables also, based on the values of correlation coefficient (r). At all grid points, specific humidity shows a strong correlation at a lag of 2 months with r value ranging from 0.77 to 0.81. The lag period selected for various local meteorological variables and corresponding correlation coefficients at G1 is tabulated in Table 1.

3.3 Correlation between soil moisture and global climatic indices

To choose the relevant predictors, among global climatic indices which have a greater influence on agricultural drought in the study area, correlation analysis was performed between soil moisture and various global climatic indices like AMO, AO, IOD, NAO, Nino 1+2, Nino 3, Nino 3.4, Nino 4, PDO, PNA, and SOI at time lags of 1–12 months. Those indices which show a higher r value were selected as predictors, and the time lag at which these higher correlations were obtained was also noted.

Figure 5 depicts the result of correlation analysis at first grid point. Among the 11 indices considered, only the ENSO events, Nino 1+2, Nino 3, and Nino 3.4, give a significant correlation with soil moisture, and hence, other indices were excluded from the predictor's list. Similar results were obtained at other grid points also. Among the three indices selected, Nino 1+2 shows the strongest relationships with soil moisture at 6-month lag with an r value of 0.81 at all grid points. The selected time lag for each index and the corresponding correlation coefficient values at first grid point are presented in Table 2.

3.4 Development of prediction models

This study developed two ANN models and two SVM models to predict soil moisture. The first model considered only local meteorological variables as predictors and the second model included selected global climatic indices along with meteorological variables. The details of the two models developed are:

Model 1:

$$SM_{(t+n)} = f(SM_{t-n}, R_{t-n}, R_{t-n-1}, Tmax_{t-n}, Tmin_{t-n-1}, Tmin_{t-n-2}, H_{t-n-1}, W_{t-n-1})$$

Model 2:

$$SM_{(t+n)} = f(SM_{t-n}, R_{t-n}, R_{t-n-1}, Tmax_{t-n}, Tmin_{t-n-1}, Tmin_{t-n-2}, H_{t-n-1}, W_{t-n-1}, Nino1 + 2_{t-n-5}, Nino1 + 2_{t-n-6}, Nino3_{t-n-4}, Nino3.4_{t-n-3})$$

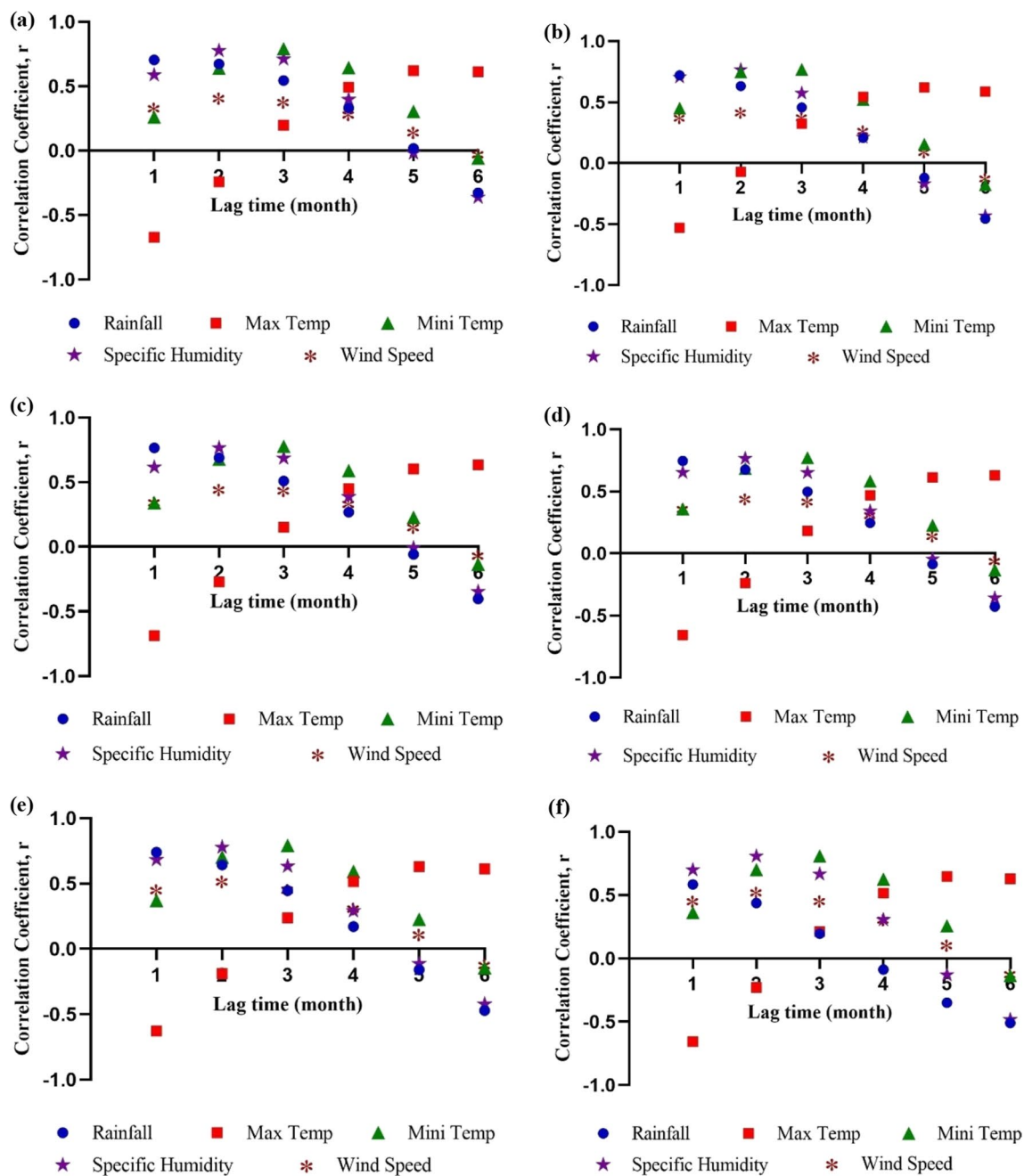


Fig. 4 Correlation coefficient between soil moisture and meteorological variables at the grid points a G1, b G2, c G3, d G4, e G5, and f G6

where SM is soil moisture, R is rainfall, Tmax is maximum temperature, Tmin is minimum temperature, H is specific humidity, W is wind speed and n is the lead time of prediction. For example, to predict the SM for a month t with 1-month lead using model 1, the SM data of the previous month was the first predictor. Rainfall values of the previous

2 months were the other two predictors. Similarly Tmax, Tmin, H , and W data with corresponding lag period can be selected accordingly.

1. *ANN modelling*: Levenberg–Marquardt back-propagation algorithm was used in this study, for training and

Table 1 Selected lag time and corresponding correlation coefficients of local meteorological variables at G1

Variables	Lag time selected (month)	Correlation coefficients	
		Pearson	Spearman's
Rainfall	1	0.704	0.822
	2	0.672	0.811
Maximum temperature	1	-0.672	-0.638
Minimum temperature	2	0.639	0.617
Specific humidity	3	0.790	0.790
	2	0.776	0.706
Wind speed	2	0.412	0.313

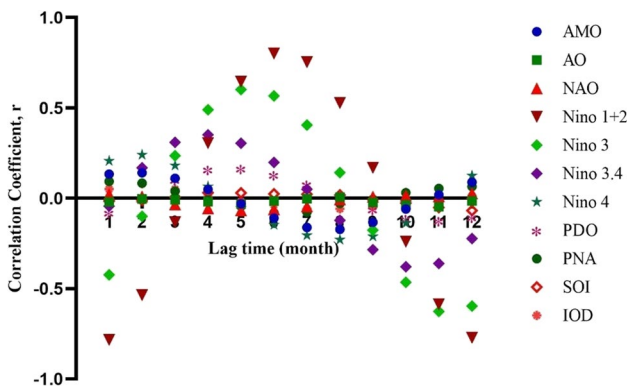


Fig. 5 Correlation coefficient between soil moisture and climatic indices at G1

Table 2 Selected lag time and corresponding correlation coefficient of climatic indices at G1

Climatic index	Time lag selected (month)	Correlation coefficient	
		Pearson	Spearman
Nino 1 + 2	6	0.801	0.803
	7	0.753	0.735
Nino 3	5	0.602	0.621
Nino 3.4	4	0.351	0.367

testing the data. This algorithm was based on Newton’s method and is specially designed for reducing the residual sum of squares (Lv et al. 2018). The data from 1971 to 2010 was used for training, and data from 2011 to 2020 was used for testing the network. The number of inputs for the first model was eight, and for the second model, it was 12. The network includes one hidden layer, and the number of neurons was fixed by testing different architectures with varying numbers of neurons. By

comparing the performance of these trials, the number of neurons was selected as four.

2. *SVM modelling*: Out of the total data, 80% was used for training and 20% was used for testing the models. Thus the first 40 years of data (1971–2010) were used for training, and the remaining data (2011–2020) was used for testing the model. Training of data was done with different types of kernel functions: linear, quadratic, fine Gaussian, medium Gaussian, and coarse Gaussian. Among the trained models, one with the least RMSE value (medium Gaussian) was selected for the prediction of testing events. Fivefold cross-validation was used to avoid the problem of overfitting. Hence, during the cross-validation, input data will be divided into 5 equal subsets. The validation process will be repeated 5 times, with each of the 5 subsets used as the validation data, and the other 4 as training data, alternatively. The final performance will be the average of validation performances in 5 subsets.

All the models were trained and tested for different lead periods from 1 to 4 months. The R^2 , RMSE, and NSE values of ANN prediction are presented in Table 3. For a 1-month lead time, the R^2 for the first and second models at various grid points varies from 0.87 to 0.93 and 0.91 to 0.95, respectively. This indicates that both models are capable of achieving good results at this lead time. Even though the global climate indices are not able to perform well at this short lead time, they produce a significant improvement in the prediction accuracy at longer lead times. As the lead time increases, the RMSE values of model 1 show a drastic increase compared to model 2, and the first model fails to acquire agreeable results. But the second model performs better than the first model by reducing RMSE up to 15% at 4-month lead time.

Table 4 describes the prediction performance of SVM models at different lead times. Similar to the ANN models, both the SVM models show very close values of R^2 and NSE at 1-month lead time. At longer lead times, the second model performs better than the first model. The second SVM model was able to achieve NSE values greater than 0.5 at all the grid points at 3- and 4-month lead times. As mentioned by Moriasi et al. (2007), models with an NSE value greater than 0.5 have a satisfactory performance. These results highlighted that meteorological variables such as rainfall, temperature, humidity, and wind speed are the most effective factors in improving the accuracy of prediction at short lead times (Mokhtarzad et al. 2017). But at a longer lead time, local meteorological variables do not provide adequate predictive information, whereas the ENSO events were more efficient to provide accurate

Table 3 Prediction performance of ANN models at different lead time

Grid point		Model 1				Model 2			
		Lead 1	Lead 2	Lead 3	Lead 4	Lead 1	Lead 2	Lead 3	Lead 4
G1	R^2	0.93	0.84	0.77	0.76	0.95	0.85	0.80	0.78
	RMSE	39.22	74.43	82.77	86.51	36.40	68.38	77.84	81.64
	NSE	0.92	0.73	0.66	0.63	0.93	0.77	0.70	0.67
	MAPE	8.39	17.10	18.41	19.97	7.65	15.16	17.45	18.07
G2	R^2	0.87	0.62	0.60	0.57	0.91	0.72	0.70	0.70
	RMSE	44.20	86.08	89.31	93.55	38.47	72.86	75.82	76.40
	NSE	0.81	0.28	0.23	0.15	0.86	0.49	0.44	0.43
	MAPE	16.99	36.76	39.89	42.18	12.94	30.06	32.72	33.10
G3	R^2	0.91	0.75	0.68	0.67	0.92	0.79	0.76	0.77
	RMSE	36.21	75.64	84.05	86.63	32.86	63.83	71.54	72.45
	NSE	0.88	0.46	0.33	0.29	0.90	0.62	0.52	0.50
	MAPE	6.00	13.56	15.20	16.00	5.43	10.46	12.12	13.21
G4	R^2	0.91	0.71	0.66	0.63	0.93	0.73	0.73	0.74
	RMSE	41.49	89.10	92.71	94.22	38.62	73.40	84.56	86.91
	NSE	0.86	0.36	0.31	0.28	0.88	0.56	0.42	0.39
	MAPE	8.20	18.89	19.36	20.57	8.08	14.97	17.73	18.52
G5	R^2	0.90	0.70	0.63	0.62	0.92	0.79	0.68	0.65
	RMSE	50.03	86.96	94.91	108.69	46.79	80.22	90.11	94.07
	NSE	0.82	0.47	0.36	0.17	0.85	0.55	0.43	0.38
	MAPE	10.97	19.88	22.45	26.92	11.19	18.76	20.55	20.63
G6	R^2	0.89	0.68	0.61	0.56	0.92	0.74	0.64	0.59
	RMSE	43.91	76.97	82.77	90.08	36.42	62.19	76.21	81.79
	NSE	0.85	0.55	0.48	0.38	0.90	0.70	0.56	0.45
	MAPE	10.12	17.49	18.95	22.47	8.37	14.42	16.16	19.27

predictive information of soil moisture. These findings illustrate that in the study region, where strong ENSO impacts are present, the second model will be the best choice for the early prediction of agricultural drought. Also, the results show that the performance of SVM models is better than ANN models.

Figure 6 presents the variation between observed and predicted values of soil moisture from ANN and SVM models at the fourth grid point for different lead times (1 month to 4 months).

At the 1-month lead time, not much variation is found between the graphs. As lead time increases, the second models from ANN and SVM are showing more agreement with the observed values compared to the first model. This clearly defines the influence of global climatic indices on soil moisture. Major ENSO events were recorded in the years 1972–1973, 1982–1983, 1997–1998, and 2015–2016 (Mike Carlowicz and Stephanie Schollaert Uz 2017). According to Kothawale et al. (2010), ENSO events are responsible for 60% of droughts all over India. Thomas and Prasannakumar

(2016) found out that ENSO events are the major reason for the irregular distribution of annual rainfall in Kerala which results in a short-term meteorological drought in various parts of the state. This work also achieves a similar result, where the ENSO events are found more responsible for the drought event in the Palakkad district. The second model, developed in this study by incorporating global climatic indices as predictors, was able to demonstrate the impact of these large-scale drivers on soil moisture deficiency in the study area and to achieve good prediction results.

While comparing the prediction abilities of ANN and SVM, the two models yielded good results in a short lead period. Additionally, SVM models outperformed ANN models, especially at longer lead periods. This may be because SVM models are functioning under the structural risk minimization principle, and hence, these models are not sensitive to local minima or maxima (Zhang et al. 2020). The outcome of this work goes in line with earlier studies (Borji et al. 2016).

Table 4 Prediction performance of SVM models at different lead time

Grid point		Model 1				Model 2			
		Lead 1	Lead 2	Lead 3	Lead 4	Lead 1	Lead 2	Lead 3	Lead 4
G1	R^2	0.95	0.86	0.80	0.77	0.94	0.84	0.82	0.82
	RMSE	36.28	61.39	76.35	81.34	40.39	65.10	74.16	75.11
	NSE	0.94	0.81	0.71	0.67	0.92	0.79	0.73	0.72
	MAPE	7.94	13.59	17.38	19.16	8.62	14.11	17.04	17.63
G2	R^2	0.86	0.74	0.61	0.61	0.88	0.80	0.77	0.78
	RMSE	47.28	73.21	86.66	87.16	43.30	65.39	70.97	65.01
	NSE	0.78	0.48	0.27	0.26	0.82	0.59	0.51	0.59
	MAPE	19.33	30.96	38.15	41.09	17.57	27.92	31.31	29.9
G3	R^2	0.94	0.85	0.78	0.77	0.92	0.83	0.81	0.81
	RMSE	33.81	51.93	65.64	72.17	34.54	52.49	62.20	65.87
	NSE	0.89	0.75	0.59	0.51	0.89	0.74	0.63	0.59
	MAPE	5.95	9.17	12.16	13.61	5.83	8.84	11.30	12.41
G4	R^2	0.91	0.82	0.74	0.75	0.91	0.83	0.80	0.80
	RMSE	38.61	63.09	81.49	87.18	39.38	63.15	71.86	80.96
	NSE	0.88	0.68	0.46	0.39	0.87	0.68	0.58	0.47
	MAPE	8.03	13.06	17.67	19.62	7.93	13.01	15.31	18.03
G5	R^2	0.91	0.79	0.71	0.68	0.92	0.82	0.78	0.78
	RMSE	40.55	70.93	84.31	92.83	41.55	65.69	73.42	77.35
	NSE	0.88	0.64	0.50	0.39	0.88	0.70	0.62	0.58
	MAPE	9.06	16.59	19.76	23.27	9.21	15.03	17.21	19.21
G6	R^2	0.91	0.76	0.65	0.62	0.91	0.79	0.75	0.75
	RMSE	37.94	62.05	73.51	76.21	37.79	57.58	63.53	63.16
	NSE	0.89	0.71	0.59	0.56	0.89	0.75	0.69	0.69
	MAPE	8.15	14.06	17.04	18.01	8.2	12.78	14.69	15.14

A seasonal analysis was conducted to evaluate the performance of models in different seasons. Four seasons were considered for this evaluation: dry season (Jan, Feb, Mar), pre-monsoon season (Apr, May), south-west monsoon season (Jun, Jul, Aug, Sept), and post-monsoon season (Oct, Nov, Dec). The results indicated that, the performance of the models varies among the seasons. The ANN and SVM models give better performance during the south-west monsoon season (with an average R^2 value of 0.7) than the other seasons. The main reason for the good prediction performance during south-west monsoon may be due to the high correlation of rainfall with soil moisture in the study area.

The findings of this study will be beneficial for early agricultural drought prediction and warning, and formulating proactive agricultural drought mitigation strategies. The performance of the models especially at higher lead times can be improved by incorporating deep learning techniques and thereby the predictions can be further refined.

4 Conclusions

This study identifies relevant meteorological and global climatic indices that influence the agricultural drought in the Palakkad district and develops models using ANN and SVM for the prediction of root zone soil moisture as an indicator of agricultural drought, up to 4 months of lead time. Various global climatic indices like AMO, AO, IOD, NAO, Nino 1 + 2, Nino 3, Nino 3.4, Nino 4, PDO, PNA, and SOI were considered in this study, to assess their influence on agricultural drought in Palakkad. The correlation analysis was performed to select the best predictors from the climatic indices, and based on the correlation coefficient values, the optimum lag at which each of the predictors strongly correlated with soil moisture was chosen. A similar analysis was done between soil moisture and local meteorological variables: rainfall, maximum temperature, minimum temperature, specific humidity, and wind speed. Two ANN models and two SVM models were

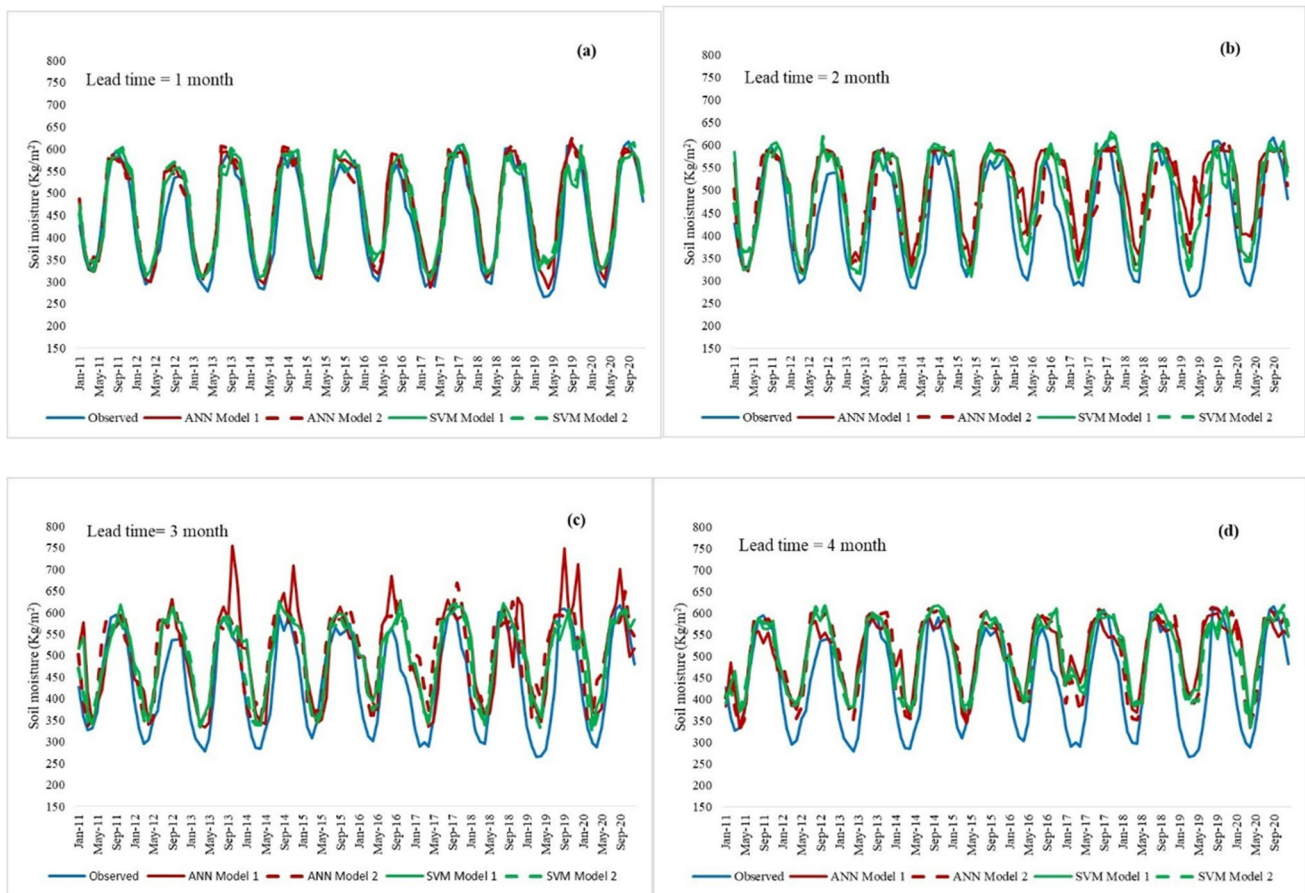


Fig. 6 Variation of observed and predicted soil moisture values at G4 for a 1-month, b 2-month, c 3-month, and d 4-month lead time

trained between selected predictors and soil moisture. For model 1, meteorological variables were the only predictors, whereas model 2 used both meteorological variables as well as global climatic indices as predictors. The prediction results of the models were compared employing R^2 , RMSE, NSE, and MAPE values. The major findings of the study are given below:

- The autocorrelation analysis of soil moisture shows a strong positive correlation at a lag of 1 month. Antecedent rainfall shows better correlation results at temporal lags of 1 and 2 months. The specific humidity and wind speed also have considerable influence on the soil moisture at 2-month lag period.
- ENSO events, Nino 1 + 2, Nino 3, and Nino 3.4, show a good correlation with soil moisture at a longer lag time (6 to 7 months).
- ANN models and SVM models were found to be effective for the short-term (1 month) prediction of agricultural drought.
- In the case of second model, SVM was able to achieve almost the same accuracy in 3-month and 4-month lead times. But the accuracy of the ANN model declines at higher lead time.
- The performance of the models was improved by the addition of global climatic indices as predictors in both ANN and SVM models from 2-month lead time onwards.

The findings of this study will aid the early warning of agricultural drought and devise drought mitigation plans.

Author contribution Saranya Das K.: conceptualization, methodology, data collection, analysis, writing—original draft preparation. N. R. Chithra: supervision, writing—review and editing.

Funding No funding was received for conducting this study.

Data availability The gridded rainfall and temperature data were obtained from the India Meteorological Department (available at <https://www.imdpune.gov.in>). The monthly time series values of root zone soil moisture, specific humidity, and wind speed were acquired from NASA GLDAS (available at <https://disc.gsfc.nasa.gov>). The monthly values of various climatic indices were downloaded from <http://psl.noaa.gov> and <http://www.bom.gov.au>.

Declarations

Competing interests The authors declare no competing interests.

References

- Abhilash S, Krishnakumar EK, Vijaykumar P, Sahai AK, Chakrapani B, Gopinath G (2019) Changing Characteristics of Droughts over Kerala, India : Inter-Annual Variability and Trend. *Asia-Pac J Atmos Sci* 55:1–17
- Borji M, Malekian A, Salajegheh A, Ghadimi M (2016) Multi-time-scale analysis of hydrological drought forecasting using support vector regression (SVR) and artificial neural networks (ANN). *Arab J Geosci* 9:725. <https://doi.org/10.1007/s12517-016-2750-x>
- Chiew FHS, Piechota TC, Dracup JA, McMahon TA (1998) El Niño / Southern Oscillation and Australian rainfall, stream-flow and drought : Links and potential for forecasting. *J Hydrol* 204:138–149
- Dikshit A, Pradhan B, Alamri A M (2021) Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. *Sci Total Environ* 755. <https://doi.org/10.1016/j.scitotenv.2020.142638>
- Dutta D, Kundu A, Patel NR (2013) Predicting agricultural drought in eastern Rajasthan of India using NDVI and standardized precipitation index. *Geocarto Int* 28(3):192–209. <https://doi.org/10.1080/10106049.2012.679975>
- Erol Keskin M, Terzi Ö, Dilek Taylan E, Küçükyaman D (2011) Meteorological drought analysis using artificial neural networks. *Sci Res Essays* 6(21):4469–4477. <https://doi.org/10.5897/sre10.1022>
- Faussett LV (2006) *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*. Pearson Education, India
- Feng P, Wang B, Liu DL, Yu Q (2019) Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agric Syst* 173:303–316. <https://doi.org/10.1016/j.agsy.2019.03.015>
- Fung KF, Huang YF, Koo CH, Mirzaei M (2020) Improved SVR machine learning models for agricultural drought prediction at downstream of langat river basin, Malaysia. *J Water Clim Change* 11(4):1383–1398. <https://doi.org/10.2166/wcc.2019.295>
- Jehanzaib M, Bilal Idrees M, Kim D, Kim TW (2021) Comprehensive Evaluation of Machine Learning Techniques for Hydrological Drought Forecasting. *J Irrig Drain Eng* 147(7):1–11. [https://doi.org/10.1061/\(asce\)ir.1943-4774.0001575](https://doi.org/10.1061/(asce)ir.1943-4774.0001575)
- Kirtman BP, Shukla J (2000) Influence of the Indian Summer Monsoon on ENSO. *Q J R Meteorol Soc* 126:213–239
- Kothawale DR, Munot AA, Kumar KK (2010) Surface air temperature variability over India during 1901–2007, and its association with ENSO. *Climate Res* 42(2):89–104. <https://doi.org/10.3354/cr00857>
- Krishnamurthi G (2000) Indian Monsoon-ENSO relationship on interdecadal timescale. *American Meteorological Society* 579–595
- Le MH, Perez GC, Solomatine D, Nguyen LB (2016) Meteorological Drought Forecasting Based on Climate Signals Using Artificial Neural Network - A Case Study in Khanhhoa Province Vietnam. *Proc Eng* 154:1169–1175. <https://doi.org/10.1016/j.proeng.2016.07.528>
- Li J, Wang Z, Wu X, Xu CY, Guo S, Chen X (2020) Toward monitoring short-term droughts using a novel daily scale, standardized antecedent precipitation evapotranspiration index. *J Hydrometeorol* 21(5):891–908. <https://doi.org/10.1175/JHM-D-19-0298.1>
- Luo Z, Wang H, Li S (2022) Prediction of International Roughness Index Based on Stacking Fusion Model. *Sustainability* 14:6949. <https://doi.org/10.3390/su14126949>
- Lv C, Xing Y, Zhang J, Na X, Li Y, Liu T (2018) Levenberg-Marquardt Backpropagation Training of Multilayer Neural Networks for State Estimation of A Safety Critical Cyber-Physical System. *IEEE Trans Industr Inf* 14(8):3436–3446. <https://doi.org/10.1109/TII.2017.2777460>
- Maity R, Suman M, Verma NK (2016) Drought prediction using a wavelet based approach to model the temporal consequences of different types of droughts. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2016.05.042>
- Manatsa D, Chingombe W, Matarira CH (2008) The impact of the positive Indian Ocean dipole on Zimbabwe droughts. *Int J Climatol* 2029:2011–2029. <https://doi.org/10.1002/joc>
- Marj AF, Meijerink AMJ (2011) Agricultural drought forecasting using satellite images, climate indices and artificial neural network. *Int J Remote Sens* 32(24):9707–9719. <https://doi.org/10.1080/01431161.2011.575896>
- Mike Carlowicz and Stephanie Schollaert Uz (2017) El Niño. NASA earth observatory. <https://earthobservatory.nasa.gov/features/EINiño>. Accessed 14 February 2017
- Mishra AK, Singh VP (2010) A review of drought concepts. *J Hydrol* 391(1–2):202–216. <https://doi.org/10.1016/j.jhydrol.2010.07.012>
- Mishra V, Shah R, Thrasher B (2014) Soil moisture droughts under the retrospective and projected climate in India. *J Hydrometeorol* 15(6):2267–2292. <https://doi.org/10.1175/JHM-D-13-0177.1>
- Mokhtarzad M, Eskandari F, Jamshidi Vanjani N, Arabasadi A (2017) Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environ Earth Sci* 76(21):1–10. <https://doi.org/10.1007/s12665-017-7064-0>
- Mooley DA, Parthasarathy B (1983) Indian summer monsoon and El Niño. *Pure Appl Geophys* 121(2):339–352. <https://doi.org/10.1007/BF02590143>
- Moriasi DN, Arnold JG, Liew M, Bingner RL, Harmel RD, Veith TL (2007) Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans ASABE* 50:885–900
- Morid S, Smakhtin V, Bagherzadeh K (2007) Drought forecasting using artificial neural networks and time series of drought indices. *Int J Climatol* 27(15):2103–2111. <https://doi.org/10.1002/joc.1498>
- Pendergrass AG, Meehl GA, Pulwarty R, Hobbins M, Hoell A, Agha-Kouchak A, Bonfils CJW, Gallant AJE, Hoerling M, Hoffmann D, Kaatz L, Lehner F, Llewellyn D, Mote P, Neale RB, Overpeck JT, Sheffield A, Stahl K, Svoboda M, Woodhouse CA (2020) Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nat Clim Chang* 10(3):191–199. <https://doi.org/10.1038/s41558-020-0709-0>
- Sang L, Zhu G, Xu Y, Sun Z, Zhang Z, Tong H (2023) Effects of Agricultural Large-And Medium-Sized Reservoirs on Hydrologic Processes in the Arid Shiyang River Basin. *Northwest China Water Resour Res* 59(2):1–15. <https://doi.org/10.1029/2022WR033519>

- Schepen A, Wang QJ, Robertson D (2011) Evidence for Using Lagged Climate Indices to Forecast Australian Seasonal Rainfall. *J Clim* 25(4):1230–1246. <https://doi.org/10.1175/jcli-d-11-00156.1>
- Shukla J, Paolino DA (1983) The Southern Oscillation and Long-Range Forecasting of the Summer Monsoon Rainfall over India. *Mon Weather Rev* 111(9):1830–1837. [https://doi.org/10.1175/1520-0493\(1983\)111](https://doi.org/10.1175/1520-0493(1983)111)
- Shukla RP, Tripathi KC, Pandey AC, Das IML (2011) Prediction of Indian summer monsoon rainfall using Niño indices: A neural network approach. *Atmos Res* 102(1–2):99–109. <https://doi.org/10.1016/j.atmosres.2011.06.013>
- Singh P, Borah B (2013) Indian summer monsoon rainfall prediction using artificial neural network. *Stoch Env Res Risk Assess* 27(7):1585–1599. <https://doi.org/10.1007/s00477-013-0695-0>
- Thomas J, Prasannakumar V (2016) Temporal analysis of rainfall (1871–2012) and drought characteristics over a tropical monsoon-dominated State (Kerala) of India. *J Hydrol* 534:266–280. <https://doi.org/10.1016/j.jhydrol.2016.01.013>
- Tian Y, Xu YP, Wang G (2018) Agricultural drought prediction using climate indices based on Support Vector Regression in Xiangjiang River basin. *Sci Total Environ* 622–623:710–720. <https://doi.org/10.1016/j.scitotenv.2017.12.025>
- Viswambharan S, Kumaramkandath IT, Tali JA (2022) A geospatial approach in monitoring the variations on surface soil moisture and vegetation water content: a case study of Palakkad District, Kerala, India. *Environ Earth Sci* 81(20):1–15. <https://doi.org/10.1007/s12665-022-10611-6>
- Wu X, Liu P, Lai C (2022) Long-range precipitation forecast based on multipole and preceding fluctuations of sea surface temperature. *Int J Climatol* 42:8024–8039. <https://doi.org/10.1002/joc.7690>
- Zargar A, Sadiq R, Naser B, Khan FI (2011) A review of drought indices. *Environ Rev* 19(1):333–349. <https://doi.org/10.1139/a11-013>
- Zhang Y, Yang H, Cui H, Chen Q (2020) Comparison of the Ability of ARIMA, WNN and SVM Models for Drought Forecasting in the Sanjiang Plain China. *Nat Resour Res* 29(2):1447–1464. <https://doi.org/10.1007/s11053-019-09512-6>
- Zhang Y, Hao Z, Feng S, Zhang X, Xu Y, Hao F (2021) Agricultural drought prediction in China based on drought propagation and large-scale drivers. *Agric Water Manag* 255:107028. <https://doi.org/10.1016/j.agwat.2021.107028>
- Zhu Q, Luo Y, Zhou D, Xu YP, Wang G, Tian Y (2021) Drought prediction using in situ and remote sensing products with SVM over the Xiang River Basin China. *Nat Haz* 105(2):2161–2185. <https://doi.org/10.1007/s11069-020-04394-x>

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.