Drought Prediction and Validation for Desert Region using Machine Learning Methods

Azmat Raja, Gopikrishnan T
Department of Civil Engineering
National Institute of Technology, Patna, India

***Abstract*—Drought prediction serves as an early warning to the effective management of water resources to avoid the drought impact. The drought prediction is carried out for arid, semi-arid, sub-humid, and humid climate types in the desert region. The drought is predicted using Standardized precipitation evapotranspiration index (SPEI). The application of machine learning methods such as artificial neural network (ANN), K-Nearest Neighbors (KNN), and Deep Neural Network (DNN) for the drought prediction suitability is analyzed. The SPEI is predicted using the aforesaid machine learning methods with inputs used to calculate SPEI. The predictions are assessed using statistical indicators. The coefficient of determination of ANN, KNN, and DNN are 0.93, 0.83, and 0.91 respectively. The mean square error of ANN, KNN, and DNN are 0.065, 0.512, and 0.52 respectively. The mean absolute error of ANN, KNN, and DNN are 0.001, 0.512, and 0.01 respectively. Based on results of statistical indicator and validations it is found that DNN is suitable to predict drought in all the four types of desert region.**

Keywords—Drought; SPEI; machine learning; water resources; prediction

I. Introduction

Drought is a recurring natural phenomenon characterized by prolonged water crisis driven by below-normal precipitation over a considerable length of time ranging from months to years [11]. Because of its complex nature and extensive occurrence, it is difficult to describe drought and identify its characteristics [16]. For sustainable agricultural activity and water resource management, accurate drought prediction and management are critical. The duration, frequency, intensity, and geographic distribution of rainfall, as well as the water needs of humans, animals, crops, and the region's vegetative cover, assess the severity of the drought. Drought is divided into four categories based on its effects: meteorological drought, agricultural drought, hydrological drought, and socioeconomic drought [8] [25]. Drought can reduce food production and have an impact on a community's socioeconomic viability [32].

Agricultural drought analysis and forecasting are more important than other types of drought in India since it is an agrarian country with 68 percent of the population dependent on agriculture. Due to a lack of water, anomalous rainfall, and harsh climatic conditions, the dry region of western India is at risk of severe droughts. The state of Rajasthan in India comprises of desert region which is classified as arid, semi-arid, sub-humid, and humid. Rajasthan is vulnerable to drought due to harsh climate. Drought is unavoidable, but it can be minimized by careful planning and preparation. Effective drought responses do not come from simply understanding the situation. While resources should be invested in enhancing the accuracy of drought monitoring and early warning systems, equal or more emphasis should be placed on upgrading drought governance structures to ensure a more effective division of responsibilities between national and local governments. As a result, numerous studies have been conducted in order to provide an objective and quantitative assessment of drought severity. Drought indices, which are proxies based on climate data and assumed to appropriately characterize the degree of drought hazard, are commonly used to quantify the effects of drought.

Drought indices have been developed in large numbers and are widely used in drought evaluation, monitoring, and forecasting. The standardized precipitation evapotranspiration index (SPEI) [36] is a developed form of standardized precipitation index (SPI) [27] by taking monthly climatic water balance. The SPEI's multi-scalar properties allow it to distinguish between different drought types and impacts, which is a significant advantage over the most generally used drought indices, which assess the effect of potential evapotranspiration on drought severity [5].

Machine learning is the subset of artificial intelligence. Machine learning (ML) algorithms are a set of commands that allow systems to learn and improve from prior data without requiring complex programming. ML techniques have been used to implement prediction or forecasting of drought. These algorithms work by simulating a model from input datasets known test sets, and then using the model findings to forecast, predict, or make various types of judgments in various application domains. K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and Deep Neural Network (DNN) machine learning techniques are extensively used in prediction of drought [1]-[4], [7], [9]-[10], [12]-[15], [17]-[24], [28]-[31], [33]-[35].

KNN is also called Lazy Learner. It does not require any training. It stores the training dataset and only uses it to make real-time predictions. As a result, the KNN method is much faster than other algorithms that require training. KNN is a simple algorithm to use. KNN can be implemented with only two parameters: the value of K and the distance function. ANN have a number of advantages, including the ability to detect complex nonlinear relationships between dependent and independent variables, the ability to recognize all potential relationships between predictor factors [10]. ANN is suitable for both large and small data sets. Deep learning is significantly responsible for the current boom in artificial intelligence (AI) usage. DNN has the biggest advantage in that it learns high-level features from data in a gradual manner. DNN gives the high quality results over other traditional methods. Overfitting is a major problem in neural networks which can be easily handled in DNN using dropout layer.

Various studies have used machine learning approaches to predict the drought using SPEI. For drought estimation, we take into account all climatic parameters that influence evapotranspiration. All of the activation functions are used to determine the best activation function for drought prediction. In this study we use SPEI drought indices to measure drought in western Rajasthan from 1979 to 2013. Drought prediction and selection of acceptable methodologies for drought measures in dry regions utilizing ANN, KNN, and Deep Neural Network machine learning algorithms.

II. Materials and Methods

The computation process of the drought prediction using both SPEI drought indices and machine learning techniques is shown in work flow diagram (Fig. 1). The working procedure start from the data collection and end at accuracy assessments of machine learning models.

A. Study Area

Rajasthan is located in north-western India between latitudes of 230 N and 300 N and longitudes of 690 E and 780 E. Fig. 2 depicts the study area in western Rajasthan, which consists of 12 districts in the state of Rajasthan. Due to insufficient rainfall and vast desert, the 12 districts are at risk of drought. Jaisalmer and Bikaner are in the arid region; Ganganagar, Hanumangarh, and Jodhpur are in the semi-arid region; Jalor, Jhunjhunu, Pali, and Sikar are in the sub-humid region; and Barmer district is in the humid region.

B. Data Collection

Daily and monthly weather data for all 12 districts in western Rajasthan for 24 weather stations were retrieved from global weather data (https://globalweather.tamu.edu/) from 1979 to 2013. Maximum and minimum temperatures, precipitation, wind speed, humidity, and solar radiation are all included in the weather data.

C. Standardized Precipitation Evapotranspiration Index (SPEI)

Vicente et al. [36] created SPEI, which is derived from SPI and uses both precipitation and potential evapotranspiration (PET) as input parameters to calculate the monthly climatic water balance.

$$D = P - PET \tag{1}$$

Where D is a simple aggregation of water shortages or excess over a period of time, P is the precipitation in mm, and PET is potential evapotranspiration in mm. The PET is computed using penman- Monteith (PM) equation using meteorological data [5-6].

The SPEI value can be calculated using the standardized values of F(x) [36].

$$SPEI = W - \frac{C_0 + C_1 W + C_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \tag{2}$$

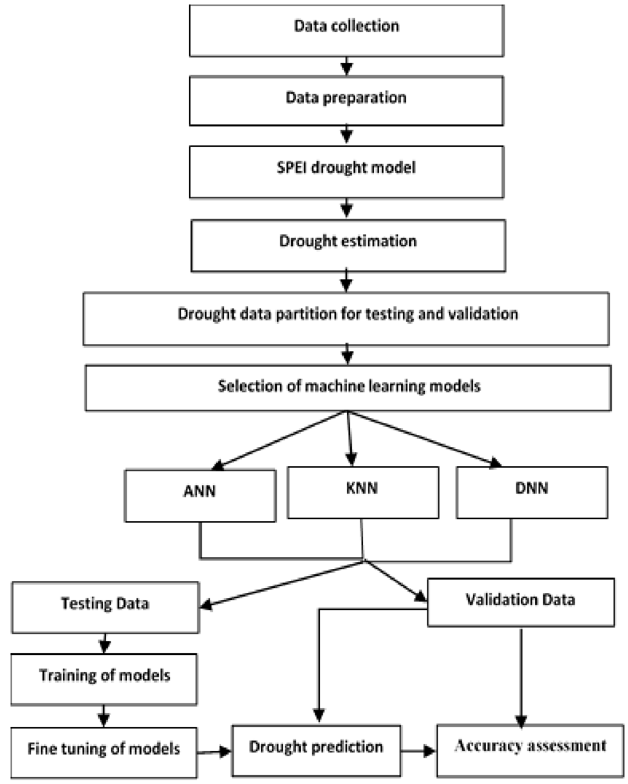

Fig. 1. The Workflow Diagram.

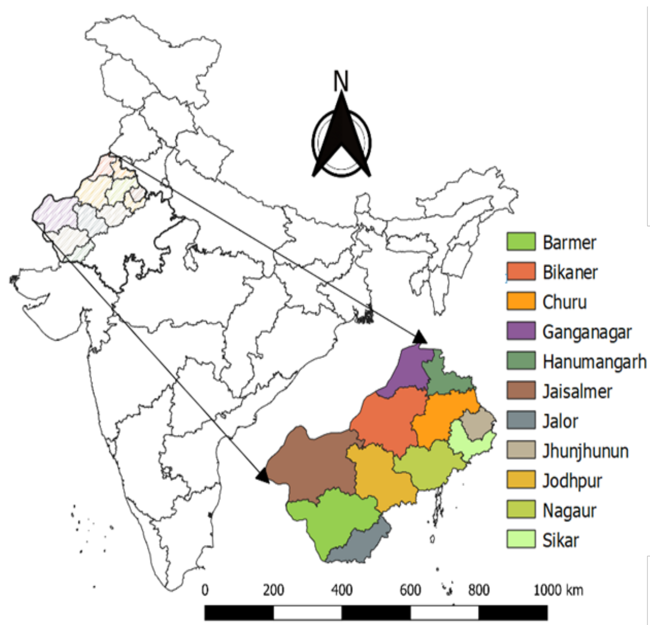

Fig. 2. The Map of the Study Area.

Where, $W = \sqrt{-2lnp}$ for P ≤ 0.5 and P is the probability of exceeding a determined D value, P =1-F(x). If P > 0.5, then P is replaced by 1 - P and the sign of the resultant SPEI is reversed. The constants are C_0= 2.515517, C_1 = 0.802853, C_2= 0.010328, d_1 = 1.432788, d_2 = 0.189269 and d_3= 0.001308.

According to the classifications of [27] [5] the SPEI can be classified mainly in five classes as shown in Table I.

TABLE I. SPEI DROUGHT CATEGORY CLASSIFICATION

Drought class	SPEI
No-Drought	Greater than -0.5
Mild	-0.5 to -0.99
Moderate	-1 to -1.49
Severe	-1.50 to -1.99
Extreme	Less than -2

D. Machine Learning Methods

Artificial neural network (ANN): McCulloch and Pitts [26] created the technique in the 1940s, and it has since evolved alongside developments in calibrating methodologies. One of the benefits of the ANN method is that the simulation is not required to completely characterize the intermediate relationships [10], [12], [21]. An ANN consists of a series of input layers, hidden layers, and output layers, each with its unique weightage as shown in Fig. 3. In this study we used 7 inputs variables and two hidden layers (first hidden layer consists eight nodes and second hidden layer consists five nodes) to compute the SPEI at output node.

1) K-Nearest Neighbors (KNN): The K-Nearest Neighbors (KNN) algorithm is a supervised machine learning algorithm that can be used to address classification and regression problems. The KNN is successfully applied in drought forecasting globally [13]-[14] [17] [28]. The basic logic behind the KNN is computing the Euclidean distance, the other alternative distance can be used are Manhattan distance, Hamming Distance, and Minkowski distance. The outcome of a KNN regression is the mean of the k closest data points. We choose odd numbers as k as a rule of thumb. KNN is a lazy learning model in which only runtime computations are performed. KNN has advantage over ANN as neural networks need large training data. The importance of meteorological data and PET on drought prediction is shown in Fig. 4.

2) Deep Neural Network (DNN): Deep learning has risen in popularity in recent years as a result of its superiority in prediction when compared to traditional machine learning approaches. Deep learning takes a lot of data because the network learns on its own. Traditional machine learning is just a set of algorithms for parsing and learning from data. The DNN has been successfully used for prediction in the past with positive results [7], [18], [21], [23], [33]-[34]. We used Keras and tensor flow package in R which is based on deep learning. The high-level interface and the automatic differentiation feature of TensorFlow make simple to implement the algorithm in an efficient manner [24]. The process of computation involved in prediction is shown in Fig. 5.

Relu, Selu, Tanh, and Sigmoid are the most common activation functions utilized in DNN. As stated in Table II, we used all four activation functions as well as their combinations. The DNN has a high prospect of overfitting the model in most cases. We used 20% of the training data for validation, as indicated in Fig. 6, to avoid overfitting the model.

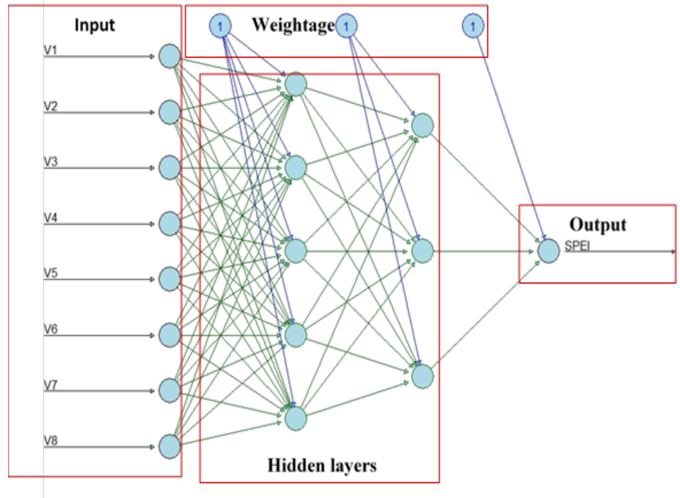

Fig. 3. The Components of ANN.

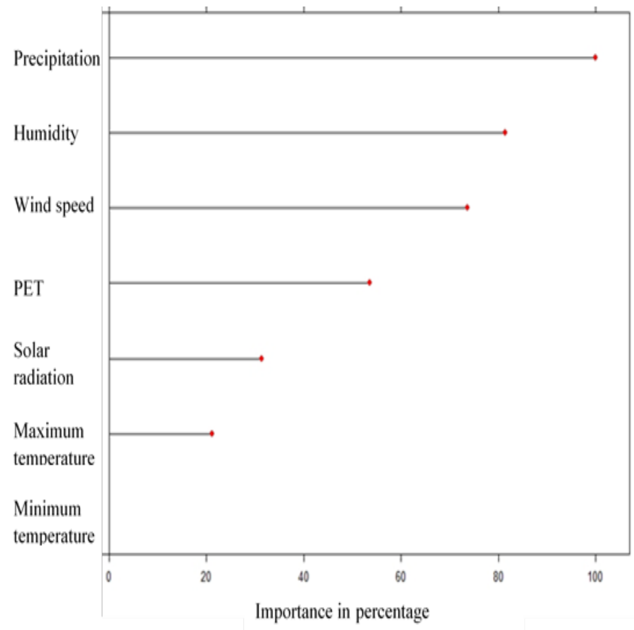

Fig. 4. The Importance of Meteorological Parameters on Drought Prediction using KNN.

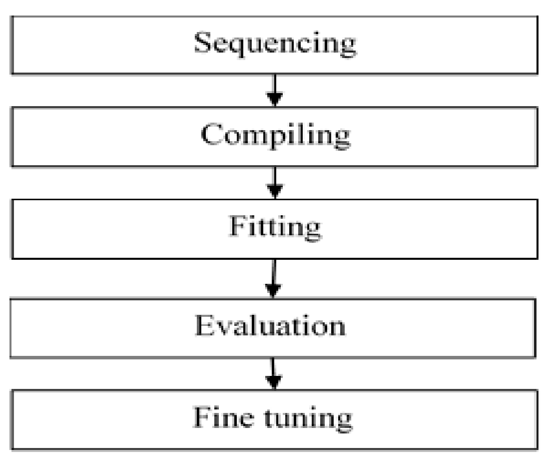

Fig. 5. Computation Process of Drought.

TABLE II. ACTIVATION FUNCTION USED IN DNN MODELLING

Activation function	Abbreviation
Relu without fine-tuning	M
Relu	M1
Selu	M2
Tanh	M3
Sigmoid	M4
Combination of Relu, Selu, and Tanh	M5
Combination of Relu, Selu, Sigmoid, and Tanh	M6

3) Performance assessment: The mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) are three typical performance indicators used to assess the models performance.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Actual - Predicted| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Actual - Predicted)^2} \quad (4)$$

$$R^2 = 1 - \frac{Squared\ sum\ error\ of\ regrssion\ line}{Squared\ sum\ error\ of\ mean\ line} \quad (5)$$

Where, N is the number of samples.

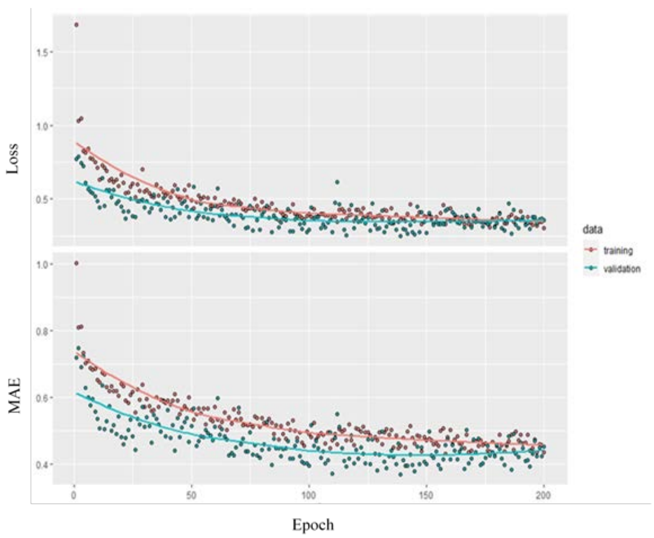

Fig. 6. Validation Spilt to avoid Overfitting of the Model.

III. RESULT AND DISCUSION

A. Temporal Variation of SPEI

The temporal variation of SPEI from 1979 to 2013 for different regions of western Rajasthan is shown in Fig. 7. The variation of drought pattern is varying with climatic condition. Each region observed severe to extreme drought in the year 1987, 1999, 2002, 2006, and 2009. The arid and semi-arid region showed more vulnerable to drought than the humid and sub-humid region Due to their location in the driest section of the Rajasthan, the districts of Jaisalmer and Bikaner experienced severe drought whereas other districts had moderate drought. Drought is less of a concern in Barmer, but it is more of a problem in Jaisalmer.

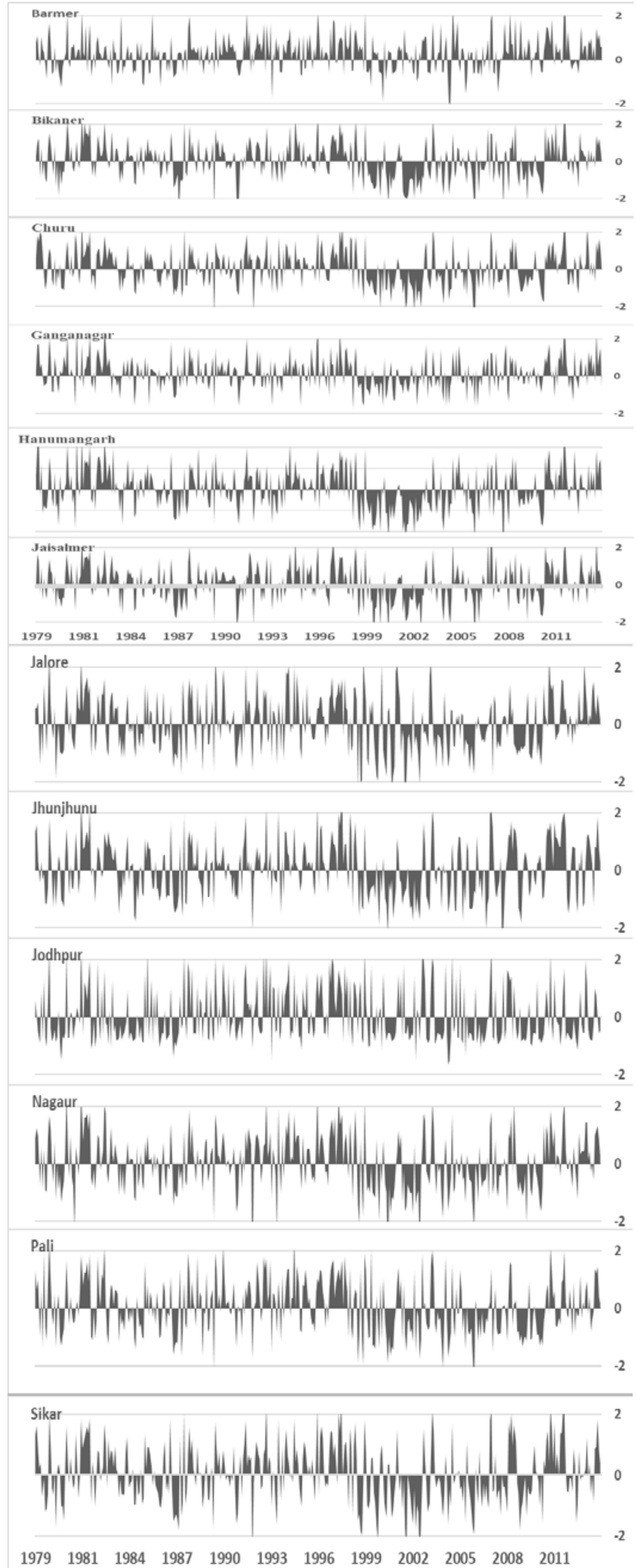

Fig. 7. Temporal Variations of SPEI over the Years 1979-2013.

B. Validation of ANN Drought

The variation of drought predicted by ANN for four different meteorological conditions is shown in Fig. 8. With slightly high or low SPEI values, the predicted values showed a similar pattern of variation as the test target. All of the regions showed good results, although the humid zone produced slightly better results. We used normalized data because there is always a problem with convergence with ANN when dealing with large data sets.

C. Validation of KNN Drought

The variation SPEI values predicted by KNN is shown Fig. 9. All the region showed similar kind of variation. The ANN showed slight better variation than KNN with target SPEI values. For sub-humid and humid locations, the K-NN based SPEI values best matched lower test target SPEI values, indicating that K-NN is better suited for drought prediction in humid regions.

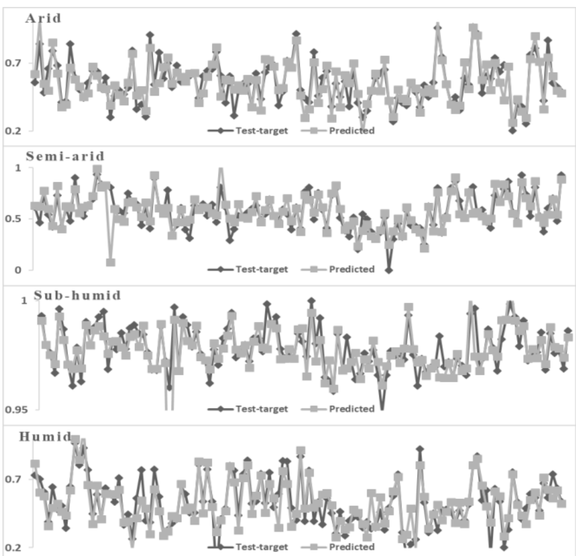

Fig. 8. Comparison of ANN Predicted and Test Target SPEI.

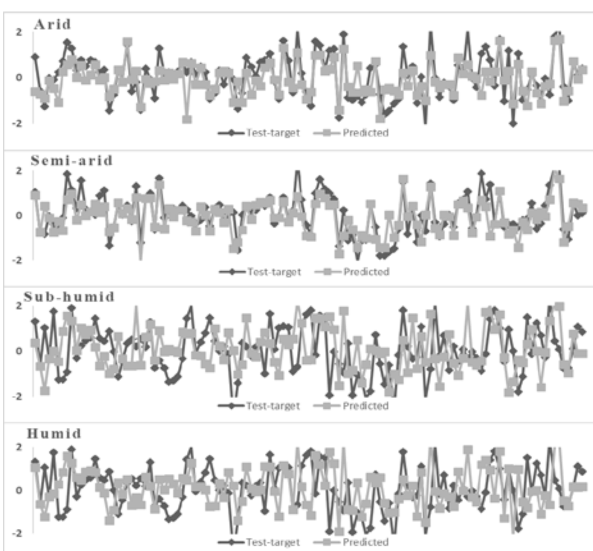

Fig. 9. Comparison of KNN Predicted and Test Target SPEI.

D. Validation of DNN Drought

The variation SPEI values predicted by DNN using a fine-tuned relu activation functions shown Fig. 10. The variation of SPEI values predicted by the DNN showed similar pattern of variation as target value with almost same SPEI values. The results of DNN is better than the both ANN and KNN for all the regions. The predicted values do not exceed the target values because the DNN has overfitting control. All the activation functions and their combinations showed good results.

E. Performance Assesment

Three different statistical indicators are used to evaluate the drought prediction models' performance: coefficient of determination, mean square error, and mean absolute error. The performance statistics of the ANN, KNN, and DNN is shown in Fig. 11, Fig. 12, and Fig. 13 respectively for arid, semi-arid, sub-humid, and humid regions. After fine-tuning, both ANN and DNN showed considerable improvements in statistical parameters, however KNN did not show any substantial improvement. In ANN and DNN, the coefficient determination is higher, whereas in KNN, it is slightly lower but still satisfactory. The highest coefficient of determination of ANN, KNN, and DNN are 0.93, 0.83, and 0.91 in semi-arid, sub-humid, and semi-arid region respectively. All the three models resulted in low MSE and very low MAE values. All of the activation functions in DNN worked well, however Relu and the combination of activation functions worked very well. The fine-tuning of the model improves the results, but it can also lead to overfitting, therefore when fine-tuning is applied we need to examine the pattern and similarity of predicted values with target values. The activation function Relu after fine-tuning showed highest agreement in arid and humid region with R^2 values 0.87 and 0.89 respectively. The highest agreement observed in semi-arid and sub-humid region by fine-tuned Selu with R^2 values 0.91 and 0.87 respectively.

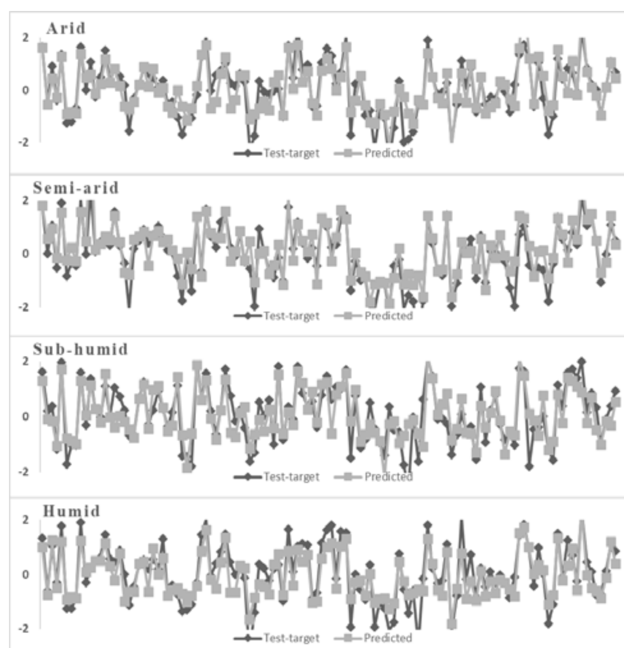

Fig. 10. Comparison of DNN Predicted and Test Target SPEI.

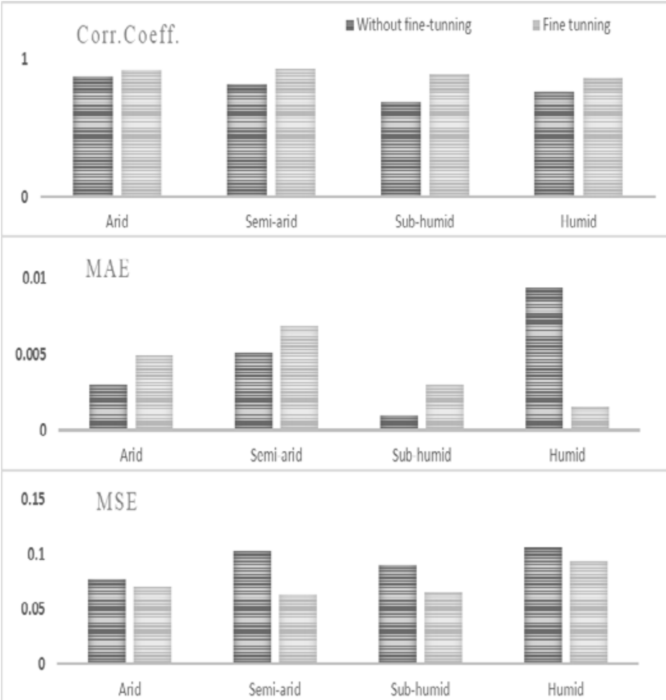

Fig. 11. Statistical Indices of ANN for different Climatic Condtions.

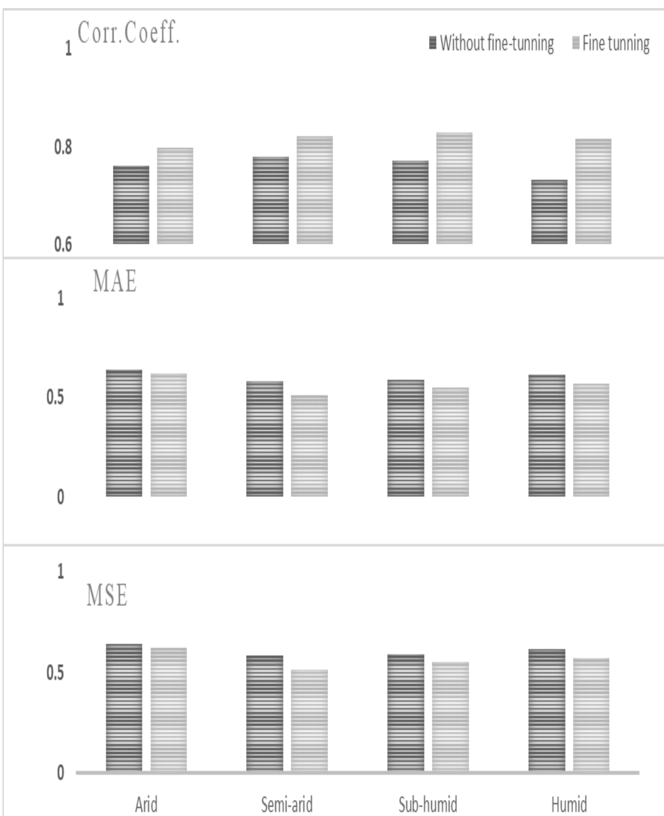

Fig. 12. Statistical Indices of KNN for different Climatic Condtions.

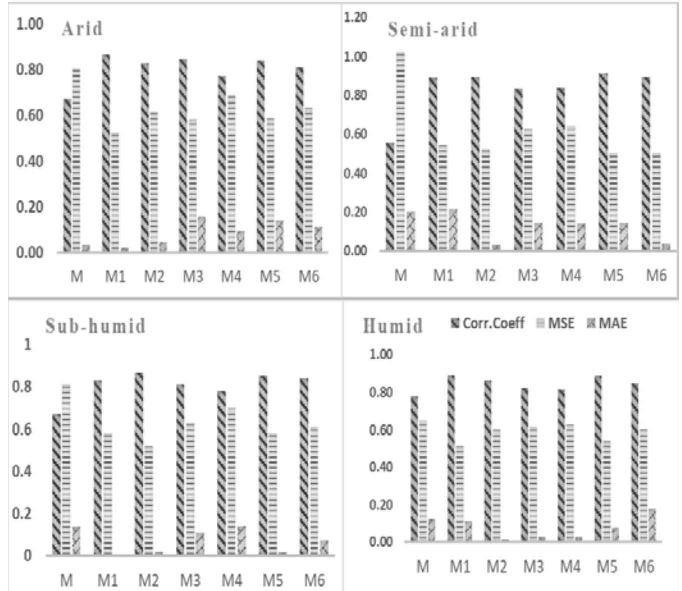

Fig. 13. Statistical Indices of DNN for different Climatic Condtions.

IV. Conclusion

The ANN, K-NN, and DNN are used to predict the drought in arid, semi-arid, sub humid, and humid conditions for the years 1979 to 2013. The KNN is one of the oldest machine learning algorithms but it is still in use for comparison to other machine learning methods. The DNN performed better than the both ANN and KNN for predicting drought in all the climate conditions. The ANN predicted better results than the KNN. It is also observed that both the ANN and KNN showed very similar results to the SPEI prediction in humid region. In all climatic conditions, DNN and ANN have larger coefficients of determination, whereas KNN has a comparatively lower coefficient of determination. The activation Relu showed best results compared to other activation function. Based on the results DNN can be used for drought prediction of any climatic condition with large data sets.

References

[1] A. Canziani, A. Paszke, and E. Culurciello, “An analysis of deep neural network models for practical applications,” arXiv preprint arXiv: 1605.07678, 2016.

[2] A. Dikshit, B. Pradhan, and A. M. Alamri, “Temporal hydrological drought index forecasting for New South Wales, Australia using machine learning approaches,” Atmosphere (Basel), vol. 11, no. 6, p. 585, 2020.

[3] A. Dikshit, B. Pradhan, and A. Huete, “An improved SPEI drought forecasting approach using the long short-term memory neural network,” J Environ Manage, vol. 283, p. 111979, 2021.

[4] A. Mokhtar et al., “Estimation of SPEI Meteorological Drought Using Machine Learning Algorithms,” IEEE Access, vol. 9, pp. 65503–65523, 2021.

[5] A. Raja and T. Gopikrishnan, “Drought Analysis Using the Standardized Precipitation Evapotranspiration Index (SPEI) at Different Time Scales in an Arid Region”, Eng. Technol. Appl. Sci. Res., vol. 12, no. 4, pp. 9034–9037, Aug. 2022.

[6] Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56.

[7] Belayneh, A., & Adamowski, J. (2013). Drought forecasting using new machine learning methods. Journal of Water and Land Development, 18(9), 3–12. https://doi.org/10.2478/jwld-2013.

[8] Boken, V. K., Cracknell, A. P., & Heathcote, R. L. (2005). Monitoring and predicting agricultural drought: a global study. Oxford University Press.

[9] Bourquin, J., Schmidli, H., van Hoogevest, P., & Leuenberger, H. (1998). Comparison of artificial neural networks (ANN) with classical modelling techniques using different experimental designs and data from a galenical study on a solid dosage form. In European Journal of Pharmaceutical Sciences (Vol. 6).

[10] Cao, W., Wang, X., Ming, Z., & Gao, J. (2018). A review on neural networks with random weights. Neurocomputing, 275, 278–287.

[11] Dai, A. (2011). Drought under global warming: a review. Wiley Interdisciplinary Reviews: Climate Change, 2(1), 45–65.

[12] Deo, R. C., & Şahin, M. (2015). Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research, 153, 512–525. https://doi.org/10.1016/j.atmosres.2014.10.016.

[13] Fadaei-Kermani, E., Barani, G. A., & Ghaeini-Hessaroeyeh, M. (2017). Drought monitoring and prediction using K-nearest neighbor algorithm. Journal of AI and Data Mining, 5(2), 319–325.

[14] Fathabadi, A., Gholami, H., Salajeghe, A., Azanivand, H., & Khosravi, H. (2009). Drought forecasting using neural network and stochastic models. Advances in Natural and Applied Sciences, 3(2), 137–147.

[15] Fung, K. F., Huang, Y. F., Koo, C. H., & Mirzaei, M. (2020). Improved SVR machine learning models for agricultural drought prediction at downstream of Langat River Basin, Malaysia. Journal of Water and Climate Change, 11(4), 1383–1398.

[16] Homdee, T., Pongput, K., & Kanae, S. (2016). A comparative performance analysis of three standardized climatic drought indices in the Chi River basin, Thailand. Agriculture and Natural Resources, 50(3), 211–219.

[17] Jiang, Z., Rashid, M. M., Johnson, F., & Sharma, A. (2021). A wavelet-based tool to modulate variance in predictors: an application to predicting drought anomalies. Environmental Modelling & Software, 135, 104907.

[18] K. Sundararajan et al., “A contemporary review on drought modeling using machine learning approaches,” CMES-Computer Modeling in Engineering and Sciences, vol. 128, no. 2, pp. 447–487, 2021.

[19] Kanfar, R., Shaikh, O., Yousefzadeh, M., & Mukerji, T. (2020). Real-time well log prediction from drilling data using deep learning. International Petroleum Technology Conference.

[20] Kaur, A., & Sood, S. K. (2020). Deep learning based drought assessment and prediction framework. Ecological Informatics, 57, 101067.

[21] N. Khan, D. A. Sachindra, S. Shahid, K. Ahmed, M. S. Shiru, and N. Nawaz, “Prediction of droughts over Pakistan using machine learning algorithms,” Advances in Water Resources, vol. 139, p. 103562, 2020.

[22] Kim, N., Na, S.-I., Park, C.-W., Huh, M., Oh, J., Ha, K.-J., Cho, J., & Lee, Y.-W. (2020). An artificial intelligence approach to prediction of corn yields under extreme weather conditions using satellite and meteorological data. Applied Sciences, 10(11), 3785.

[23] Kim, T.-W., & Valdés, J. B. (2003). Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. Journal of Hydrologic Engineering, 8(6), 319–328.

[24] Li, G., Hari, S. K. S., Sullivan, M., Tsai, T., Pattabiraman, K., Emer, J., & Keckler, S. W. (2017). Understanding error propagation in deep learning neural network (DNN) accelerators and applications. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 1–12.

[25] Liu, M., & Grana, D. (2019). Accelerating geostatistical seismic inversion using TensorFlow: A heterogeneous distributed deep learning framework. Computers & Geosciences, 124, 37–45.

[26] Lloyd-Hughes, B., & Saunders, M. A. (2002). A drought climatology for Europe. International Journal of Climatology: A Journal of the Royal Meteorological Society, 22(13), 1571–1592.

[27] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4), 115–133.

[28] McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. Proceedings of the 8th Conference on Applied Climatology, 17(22), 179–183.

[29] O. Gupta and R. Raskar, “Distributed learning of deep neural network over multiple agents,” Journal of Network and Computer Applications, vol. 116, pp. 1–8, 2018.

[30] O.-J. Jang, H.-T. Moon, and Y.-I. Moon, “Drought Forecasting for Decision Makers Using Water Balance Analysis and Deep Neural Network,” Water (Basel), vol. 14, no. 12, p. 1922, 2022.

[31] Park, S., Im, J., Han, D., & Rhee, J. (2020). Short-term forecasting of satellite-based drought indices using their temporal patterns and numerical model output. Remote Sensing, 12(21), 1–21. https://doi.org/10.3390/rs12213499.

[32] Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., Zhou, L., Liu, H., Ma, Y., & Ding, Y. (2010). The impacts of climate change on water resources and agriculture in China. Nature, 467(7311), 43–51.

[33] Poornima, S., & Pushpalatha, M. (2019). Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. Soft Computing, 23(18), 8399–8412. https://doi.org/10.1007/s00500-019-04120-1.

[34] S. Shamshirband et al., “Predicting standardized streamflow index for hydrological drought using machine learning models,” Engineering Applications of Computational Fluid Mechanics, vol. 14, no. 1, pp. 339–350, 2020.

[35] Qian, Y., Fan, Y., Hu, W., & Soong, F. K. (2014). On the training aspects of deep neural network (DNN) for parametric TTS synthesis. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3829–3833.

[36] Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. Journal of Climate, 23(7), 1696–1718.