IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 November 2023, accepted 7 December 2023, date of publication 12 December 2023,
date of current version 19 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3341587

APPLIED RESEARCH

Drought Forecasting: Application of Ensemble and Advanced Machine Learning Approaches

GEETABAI S. HUKKERI[1], SUJAY RAGHAVENDRA NAGANNA[2], DAYANANDA PRUTHVIRAJA[3], (Senior Member, IEEE), NAGARAJ BHAT[4,5], AND R. H. GOUDAR[6]

[1]Department of Computer Science and Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[2]Department of Civil Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[3]Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[4]Department of Computer Science and Engineering, Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Udupi 574115, India
[5]Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, AL 35487, USA
[6]Department of Computer Science and Engineering, Visvesvaraya Technological University, Belagavi, Karnataka 590018, India

Corresponding authors: Sujay Raghavendra Naganna (sujay.n@manipal.edu) and Dayananda Pruthviraja (dayananda.p@manipal.edu)

This work was supported by the Manipal Academy of Higher Education (Open Access Funding).

ABSTRACT Depending on the severity and spatial-temporal variability, droughts can have a wide range of impacts such as crop failure, water shortages, and food insecurity. Accurate and timely forecasting is necessary to mitigate the hazards of extreme weather events, such as droughts, brought on by climate change. A district like Chitradurga in India, which typically receives around 450-600 mm of annual rainfall, will require advanced drought mitigation strategies and plans before the onset of the drought. This research focuses on 1-step lead time forecasting of meteorological drought episodes making use of the 6-month Standardised Precipitation Index (SPI-6) as indicator. The fine resolution rainfall data ($0.25° \times 0.25°$) obtained from the Indian Meteorological Department was used to derive the 6-month SPI data of 23 grid stations. The 1-step lead time SPI-6 time series was forecast considering the antecedent SPI-6 time series data as model input. The Mutual Information was used to determine the most relevant input features for drought forecasting. The standard Artificial Neural Network, an advanced machine learning framework - Multivariate Adaptive Regression Splines, and the ensemble learning-based *CatBoost* Regression and Gradient Tree Boosting paradigms were employed to forecast drought episodes. Error and efficiency metrics were employed for performance evaluation of the simulated models. The multivariate adaptive regression splines and gradient tree boosting forecasts had slightly higher accuracy and lower error rates than the artificial neural network model, which suggests that they may be more reliable for drought forecasting. The root mean square error and normalized Nash-Sutcliffe efficiency ranges of the multivariate adaptive regression splines model (during test phase) were 0.37–0.54 and 0.78–0.87, respectively. The thematic maps that were created using spatial interpolation of model forecasts from all the stations also confirmed that the district as a whole experienced drought in April 2019.

INDEX TERMS CatBoost regression, droughts, ensemble learning, gradient tree boosting, multivariate adaptive regression splines, mutual information gain, standardized precipitation index.

I. INTRODUCTION

Drought is a highly persistent natural disaster that can have effects ranging from subtle to severe over a range of timescales in a specific region. Persistent precipitation deficit is the key factor contributing to different forms of droughts [1], [2]. Region-specific factors such as topography, climate change, temperature, rainfall distribution, etc. determine both drought frequency and severity [3]. People and the environment can be severely affected by droughts. They can lead to water shortages, crop failures, wildfires, and the spread of disease. Droughts can also have a significant

The associate editor coordinating the review of this manuscript and approving it for publication was Mouquan Shen.

economic impact, as they can lead to job losses and damage to infrastructure [4].

The functional classification of droughts is based on rainfall deficits that accumulate over different timescales. These droughts can be permanent, seasonal, or contingent [2], [5]. Agricultural, meteorological, hydrological, and socio-economic droughts have compounding effects based on their complexity, severity, and impact [6]. Meteorological drought is classified based on the rainfall deficiency compared to the long-term average. As per the Indian Meteorological Department, following are the classifications: slight drought: rainfall deficiency of 25% or less; moderate drought: rainfall deficiency of 26% to 50%; and severe drought: rainfall deficiency of more than 50% [7]. Meteorological drought is usually a measure of the dryness of the atmosphere [8].

The intensity of a drought is measured by its magnitude, and its severity is determined by its duration. In general, droughts that last for longer periods tend to be more severe than those that occur over shorter time scales. This is due to the fact that prolonged droughts have more time to impact the availability of water, vegetation, and other environmental factors [9], [10]. The relationship between drought time scale and magnitude is not always readily apparent [9]. Shorter-term droughts can occasionally be just as severe as those that last for a longer period of time. This is due to the fact that a drought's timing can be just as crucial as its duration. For instance, a drought that strikes during a critical time of year, such as the growing season, can be more severe than a drought that occurs during a less critical time of year [11]. The different drought time scales can also have a different impact on different environments. For example, droughts that occur in arid environments tend to be more severe than droughts that occur in humid environments. This is because arid environments have less water available to begin with, so even a small reduction in precipitation can have a significant impact. Droughts can also have a significant impact on natural ecosystems, leading to changes in vegetation, wildlife populations, and soil health [12].

Spatio-temporal forecasting of drought can have a profound impact on the way we prepare for, monitor, and respond to droughts [13]. By providing advanced warning of drought, we can take steps to mitigate its effects, such as conserving water, planting drought-resistant crops, and creating firebreaks. Spatio-temporal forecasting can also help us to compare the severity of previous droughts with the existing ones, which can help us to better understand the risks of drought and to develop more effective mitigation strategies [14]. Arid and semi-arid regions are more likely to be affected by drought because they receive less precipitation than other regions. This means that they are more vulnerable to the effects of drought, such as water shortages, crop failures, and wildfires [15]. There exists a number of methods that can be used to forecast droughts. Some of the most common methods include: statistical methods [16], [17], [18], numerical weather prediction (NWP) models [19], and machine learning methods [20]. The apt method for forecasting drought will vary depending on the region and the specific application. Spatio-temporal forecasting is a valuable tool that can help us to better understand and respond to drought. The Standardized Precipitation Index (SPI) is a versatile and widely used index for quantifying meteorological drought. It is based on the long-term precipitation record for any particular location, and can be calculated for different time periods. This makes it a valuable tool for assessing the severity and impact of droughts, as well as for forecasting the likelihood of them in the future [21]. The choice of Standardized Precipitation Index (SPI) time scale depends on the specific application and the type of drought impact being assessed. For the purpose of evaluating medium-term drought conditions and their effects on hydrological and agricultural systems, time scales–such as six months–work better. Hence, 6-month SPI (SPI-6) was chosen to better reflect the cumulative effects of precipitation deficits on soil moisture, streamflow, and baseflow levels. The 6-month SPI is relatively insensitive to short-term precipitation variability. Gridded and satellite indices can be valuable tools for drought monitoring, but their accuracy and reliability can be affected by various factors, including data scarcity, spatial and temporal resolution, interpolation methods, model biases, and uncertainties in input data [22], [23]. Despite these uncertainties, gridded SPI data can be a valuable tool for drought monitoring.

A. RELATED LITERATURE

Drought characterization, analysis, and forecasting have been the subject of many meta-analysis and systematic review papers. The review panellists have addressed a wide range of issues, including the challenges of drought analysis and forecasting [24], [25], [26], the current state of scientific knowledge and interdisciplinary analyses of drought events [27], [28], drought effects on environmental systems [29], multiple drought-related variables and multivariate drought indices [30], [31], resilience and coping with transboundary drought risks [32], [33].

In recent years, there has been a growing interest in using machine learning (ML) to improve the forecast accuracy of the SPI or other drought indices. Here we present a literature review of some of the recent studies related to drought forecasting using machine learning. A recent systematic review by Gyaneshwar et al. [34] sheds insight on the different types of deep learning (DL) models that have been used for drought prediction, as well as the challenges and limitations of those models. Further, an organized and inclusive survey presented by Prodhan et al. [35] emphasize on the basic concepts and the development of ML and DL algorithms for efficient and effective drought forecasting task. Recent research by Basak et al. [36] found that the Prophet model was more accurate and robust than support vector regression (SVR) and multiple linear regression (MLR) models for short-term drought forecasting using SPI. Likewise in another study by Elbeltagi et al. [37], the Random Subspace (RSS) model was hybridized with

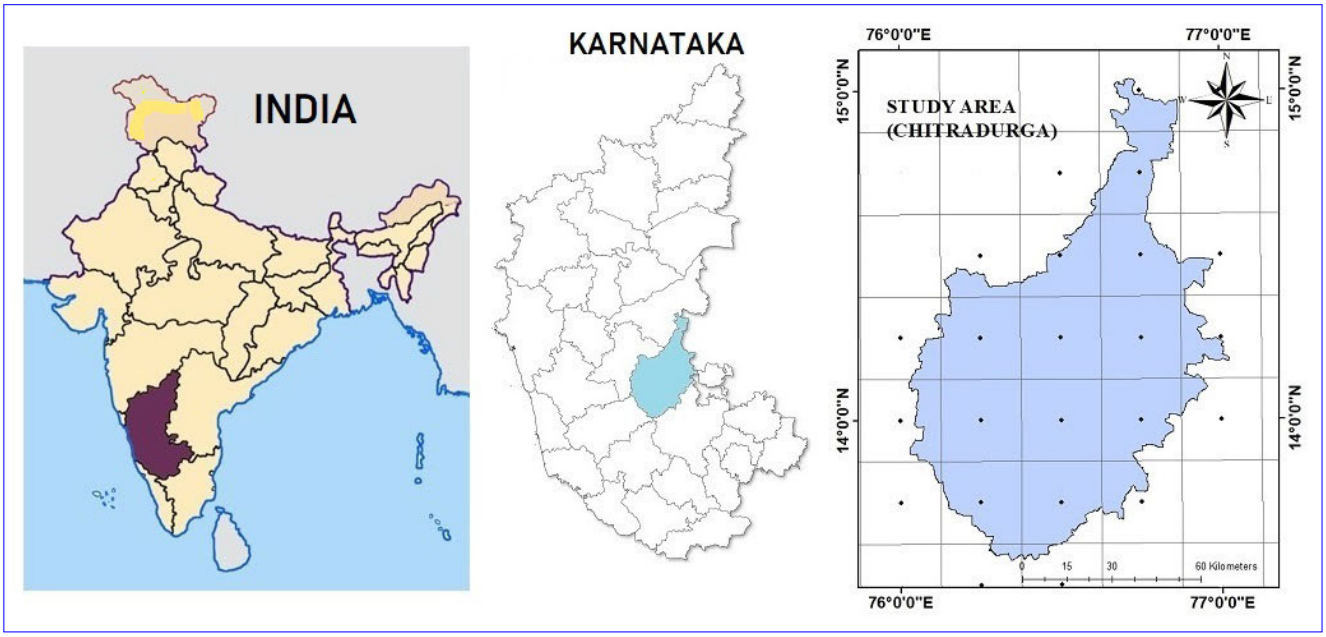

FIGURE 1. Chitradurga District of Karnataka, India.

M5 Pruning tree (M5P), Random Tree (RT) and Random Forest (RF) paradigms to estimate SPI. The RSS-M5P model provided the most accurate SPI forecasts. The M5P algorithm improved the performance of the RSS model. The standard ML models such as support vector machine (SVM), random forest, additive regression, bagging, and random subspace models were implemented for forecasting SPI for different time-scales (3, 6, 9, and 12) of the Wadi Mina basin, Algeria by Achite et al. [38]. The research by Gul et al. [39] found that the Extreme Gradient Boosting (XgBoost) model outperformed the Adaptive Boosting (AdaBoost), and Gradient Boosting (GradBoost) models in forecasting SPI-3, SPI-6 and SPI-12 timeseries of the Aegean region in Türkiye. The hybrid ML models can be more interpretable than standard ML models; considering this Mohammadi [10] applied artificial neural network with firefly algorithm (ANN-FA) to model SPI-3, SPI-6, SPI-18, and SPI-24 of Lima meteorological station in Peru and obtained promising forecasts with higher accuracy. The FA is a metaheuristic algorithm that was used to optimize the parameters of the ANN. To summarize the application of machine learning (ML) for drought forecasting is advantageous in terms of accuracy, robustness, interpretability, and scalability. In general, ML models serve as valuable tools for drought modeling. They help to improve the accuracy of drought forecasts, simplify the handling of drought data, and provide insights into the factors that contribute to droughts.

TABLE 1. Descriptive statistics of SPI-6 dataset derived from IMD precipitation data (Station 1 - 12).

Station	Dataset	MAX	MIN	MEAN	STD	SKEW
Station 1 15°0'N, 76°45'E	Full	2.99	-3.00	0.0	1.0	-0.0835
	Train	1.76	-3.00	0.0	1.0	-0.2200
	Test	2.99	-3.00	0.0	1.0	0.0240
Station 2 14°45'N, 76°30'E	Full	2.15	-3.34	0.0	1.0	-0.4832
	Train	1.73	-3.34	0.0	1.0	-0.5828
	Test	2.15	-1.69	0.0	1.0	-0.1945
Station 3 14°45'N, 76°45'E	Full	2.40	-2.74	0.0	1.0	-0.2296
	Train	1.97	-2.51	0.0	1.0	-0.3283
	Test	2.40	-2.74	0.0	1.0	-0.2517
Station 4 14°30'N, 76°15'E	Full	2.62	-3.14	0.0	1.0	-0.1832
	Train	2.23	-3.14	0.0	1.0	-0.3126
	Test	2.62	-2.21	0.0	1.0	0.0474
Station 5 14°30'N, 76°30'E	Full	2.06	-3.86	0.0	1.0	-0.4732
	Train	2.02	-3.86	0.0	1.0	-0.5163
	Test	2.06	-2.16	0.0	1.0	-0.2713
Station 6 14°30'N, 76°45'E	Full	2.19	-2.95	0.0	1.0	-0.2453
	Train	2.19	-2.95	0.0	1.0	-0.2456
	Test	2.04	-2.16	0.0	1.0	-0.2021
Station 7 14°30'N, 77°0'E	Full	2.22	-3.42	0.0	1.0	-0.3665
	Train	2.15	-3.42	0.0	1.0	-0.4542
	Test	2.22	-2.59	0.0	1.0	-0.2461
Station 8 14°15'N, 76°0'E	Full	2.33	-4.11	0.0	1.0	-0.3927
	Train	2.33	-4.11	0.0	1.0	-0.4662
	Test	1.51	-2.43	0.0	1.0	-0.2191
Station 9 14°15'N, 76°15'E	Full	2.14	-2.72	0.0	1.0	-0.3313
	Train	2.06	-2.72	0.0	1.0	-0.3858
	Test	2.14	-2.26	0.0	1.0	-0.2011
Station 10 14°15'N, 76°30'E	Full	2.33	-2.72	0.0	1.0	-0.2154
	Train	2.00	-2.72	0.0	1.0	-0.3204
	Test	2.33	-1.95	0.0	1.0	-0.0835
Station 11 14°15'N, 76°45'E	Full	2.09	-2.41	0.0	1.0	-0.2254
	Train	2.09	-2.41	0.0	1.0	-0.2287
	Test	1.92	-2.15	0.0	1.0	-0.2257
Station 12 14°15'N, 77°0'E	Full	2.07	-3.34	0.0	1.0	-0.5234
	Train	2.07	-3.34	0.0	1.0	-0.5634
	Test	2.03	-3.10	0.0	1.0	-0.4380

Note: MIN - Minimum; MAX - Maximum; STD - Standard Deviation; SKEW - Skewness

B. STUDY AREA

The study area chosen was Chitradurga, a district of Karnataka state in the southern part of India. It is located at an elevation of 732 meters above sea level, and has a total area of 8,437 square kilometers covering about 4.4% of the total area of the Karnataka state. The annual mean temperature in Chitradurga is 25.4 °C, and the average annual precipitation is 655 mm [40]. The major soil types in the

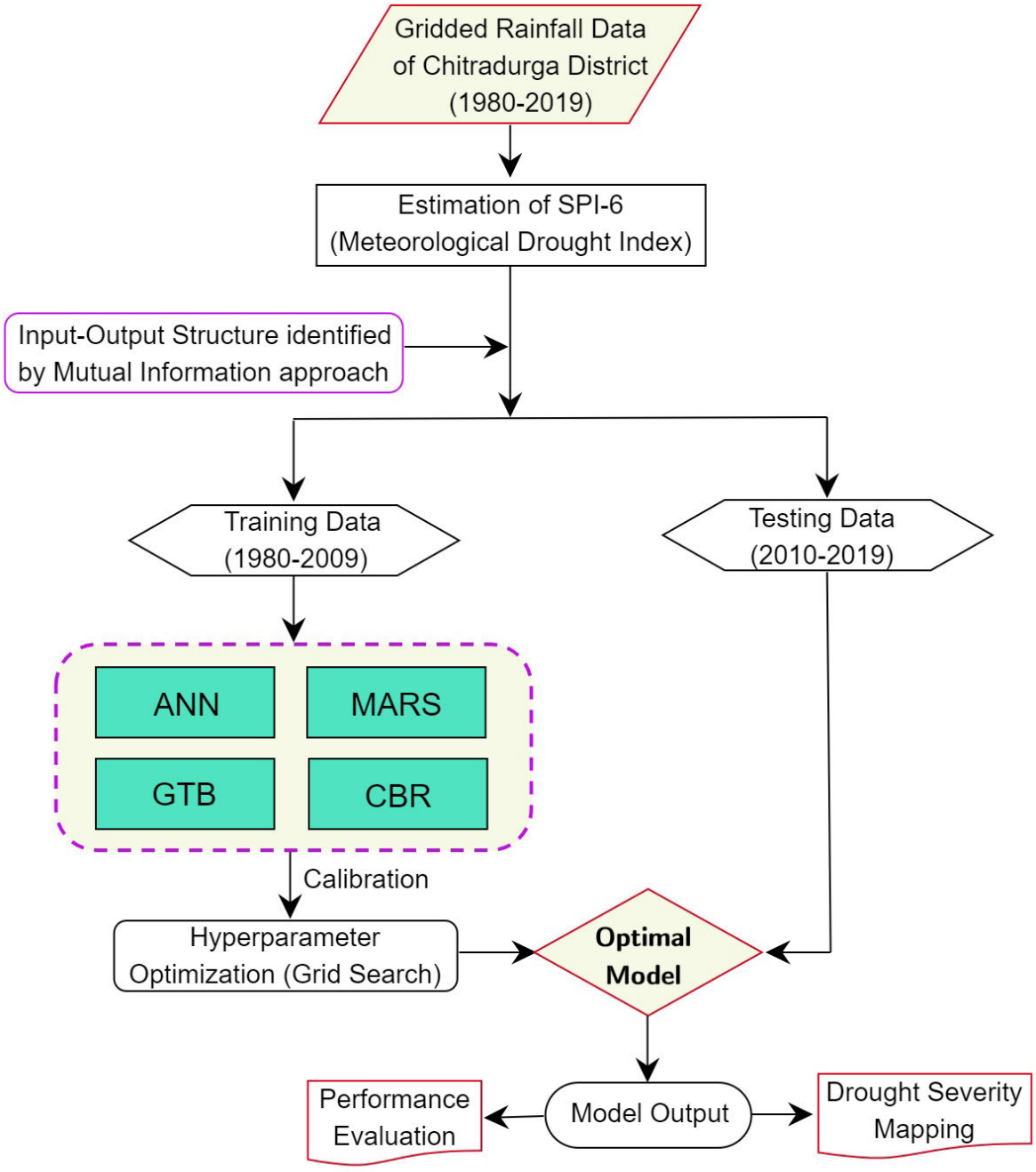

FIGURE 2. Methodology flowchart.

district are deep and shallow black soil, mixed red and black soil, red loamy soil, and sandy soil. The land usage in Chitradurga includes a forest area of 737.17 square km and a net sown area of 4,698.37 square km. The district comes under the Krishna river basin and sub-basins of Vedavathi, Janagahalli, Swarnamukhi, and Yakanahalli Nala [41]. According to Koppen and Geiger's classification, the climate of Chitradurga is classified as 'BSh' ('B' denotes dry; 'S' denotes semi-arid or steppe; and 'h' denotes hot climate) [42]. The monthly rainfall data of the study area was obtained from the Indian Meteorological Department (IMD) for the period from January 1980 to December 2019. The SPI was determined for 23 grid locations within the study area using the precipitation data from IMD. The location details of all the grid locations are given in Figure 1.

C. MOTIVATION AND OBJECTIVES

In India, 68% of the country is drought-prone, with 35% of that area receiving between 750 and 1125 mm of rainfall annually, and 33% receiving less than 750 mm annually [43].

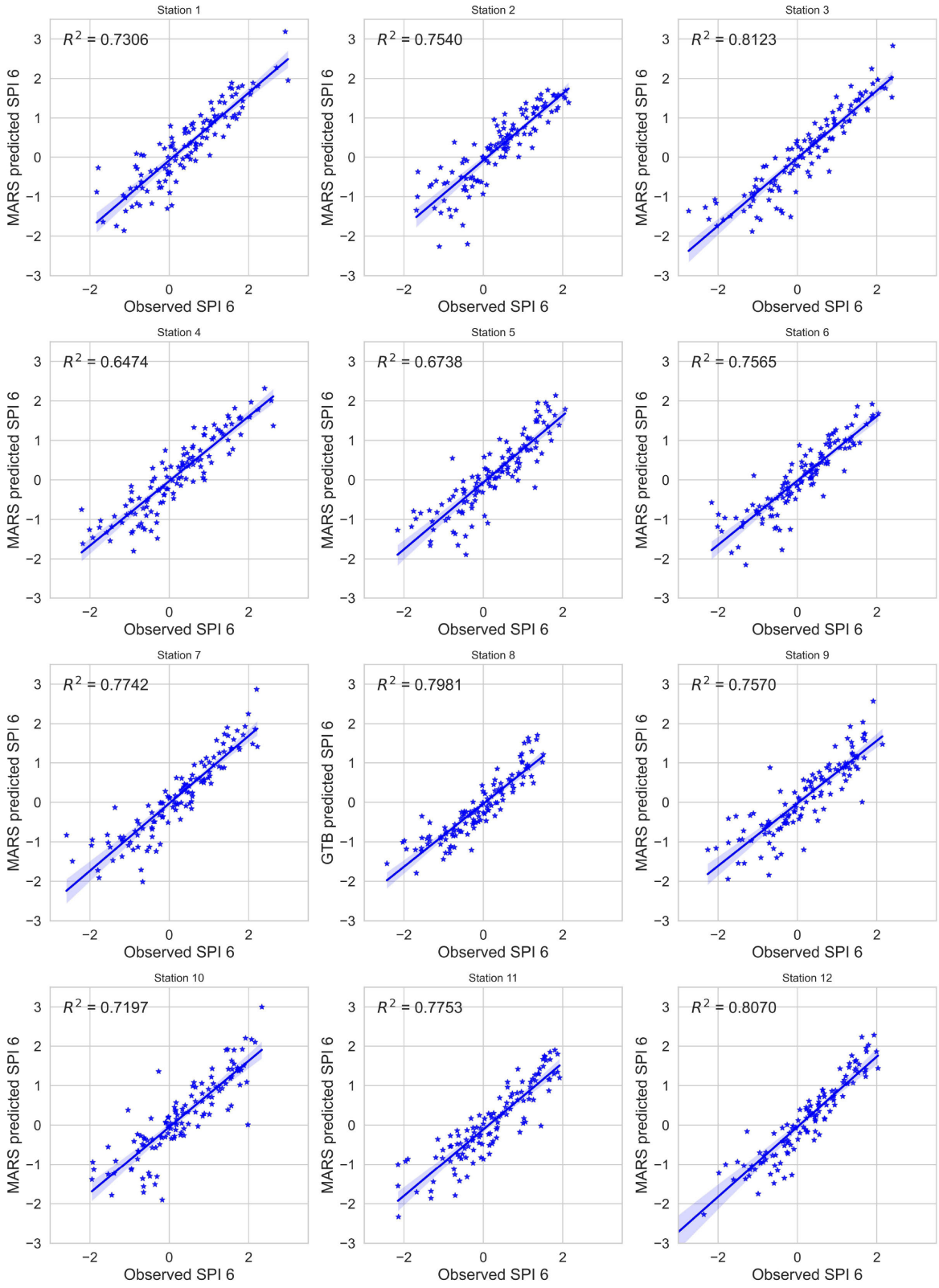

FIGURE 3. Scatter plots of best performing models used for SPI-6 forecasting (Stations 1-12) (Test phase).

Chitradurga is a drought-prone district in Karnataka, India. It receives an average annual precipitation of 655 mm, which is well below the state average of 747 mm. Additionally, the district experiences temperatures as high as 44°C during the summer season with low humidity, which makes it even more susceptible to drought stress. Drought analysis and

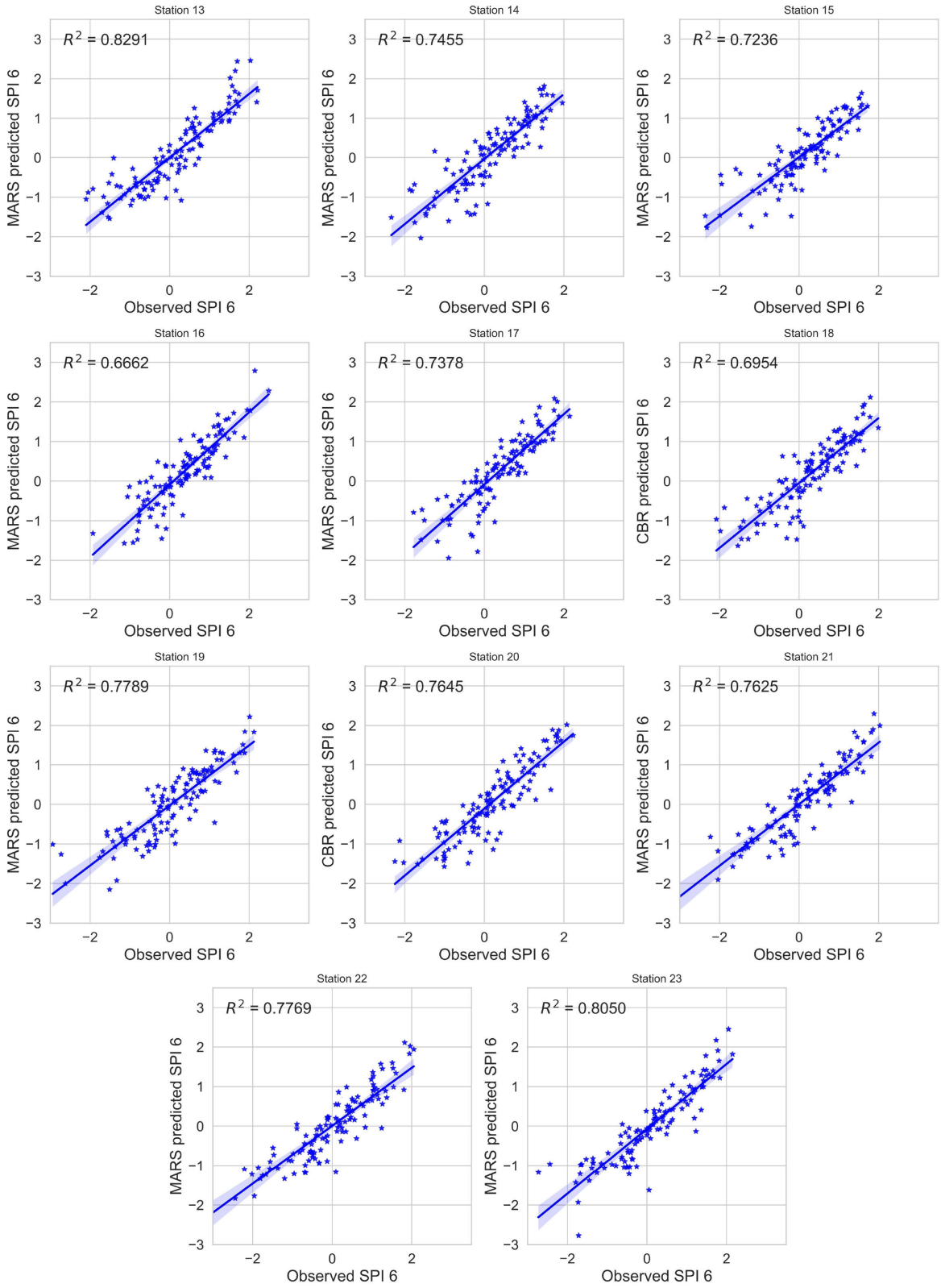

FIGURE 4. Scatter plots of best performing models used for SPI-6 forecasting (Stations 13-23) (Test phase).

forecasting aids to delineate the areas that are likely to experience drought event. The forecast information could be used to develop and hone drought mitigation and adaptation strategies. The SPI-6 index, which is often used to foresee reduced streamflow, water shortages, and crop failures, was considered in the current research. The objective of

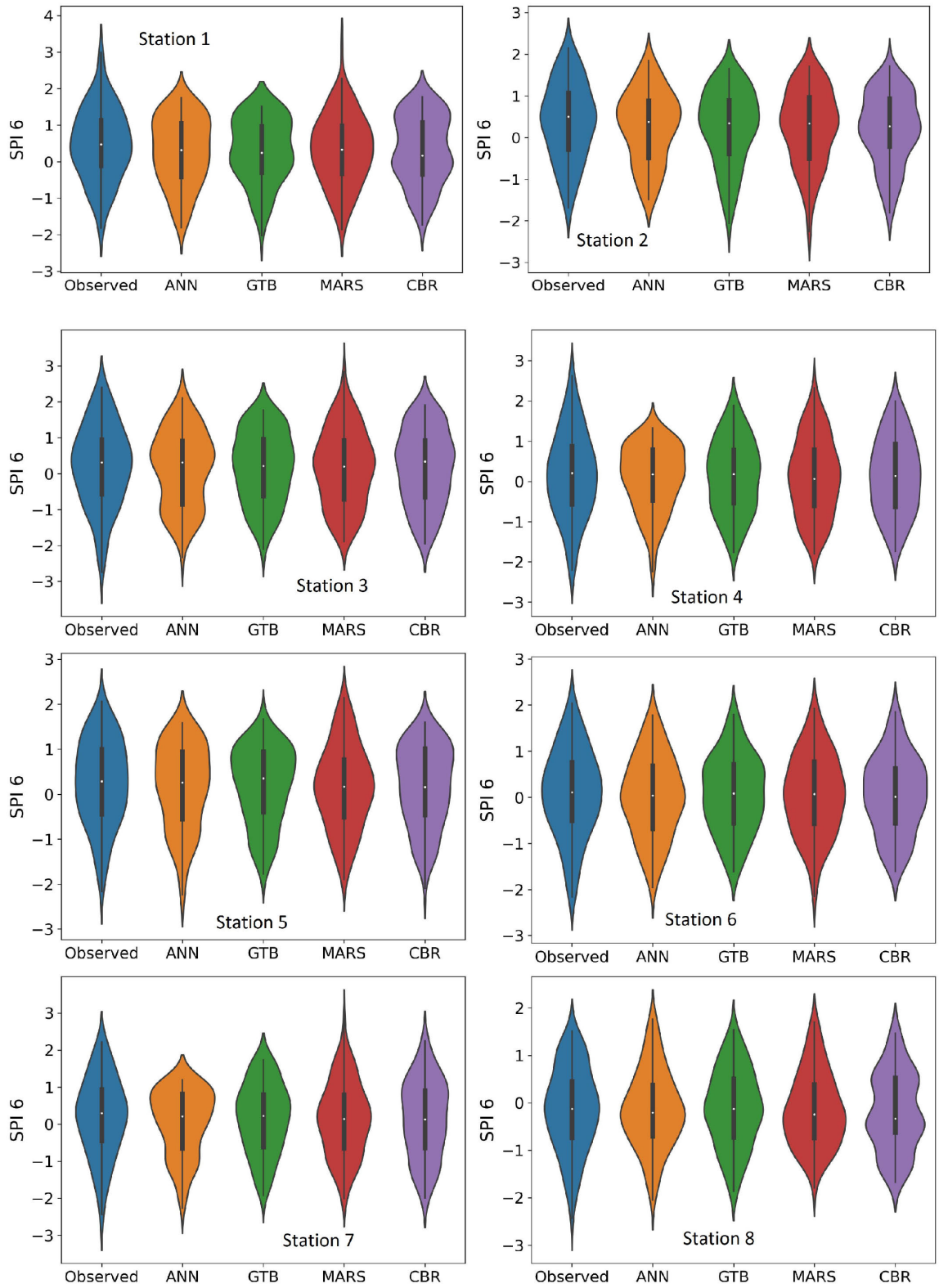

FIGURE 5. Violin plots of ML models used for SPI-6 forecasting (Stations 1-8) (Test Phase).

the study was to quantitatively forecast the one-step lead time meteorological drought (SPI-6) using Artificial Neural Network, Gradient Tree Boosting, Multivariate Adaptive Regression Splines and CatBoost Regressor paradigms, compare the effectiveness of these models, and then endorse the best model based on their performance.

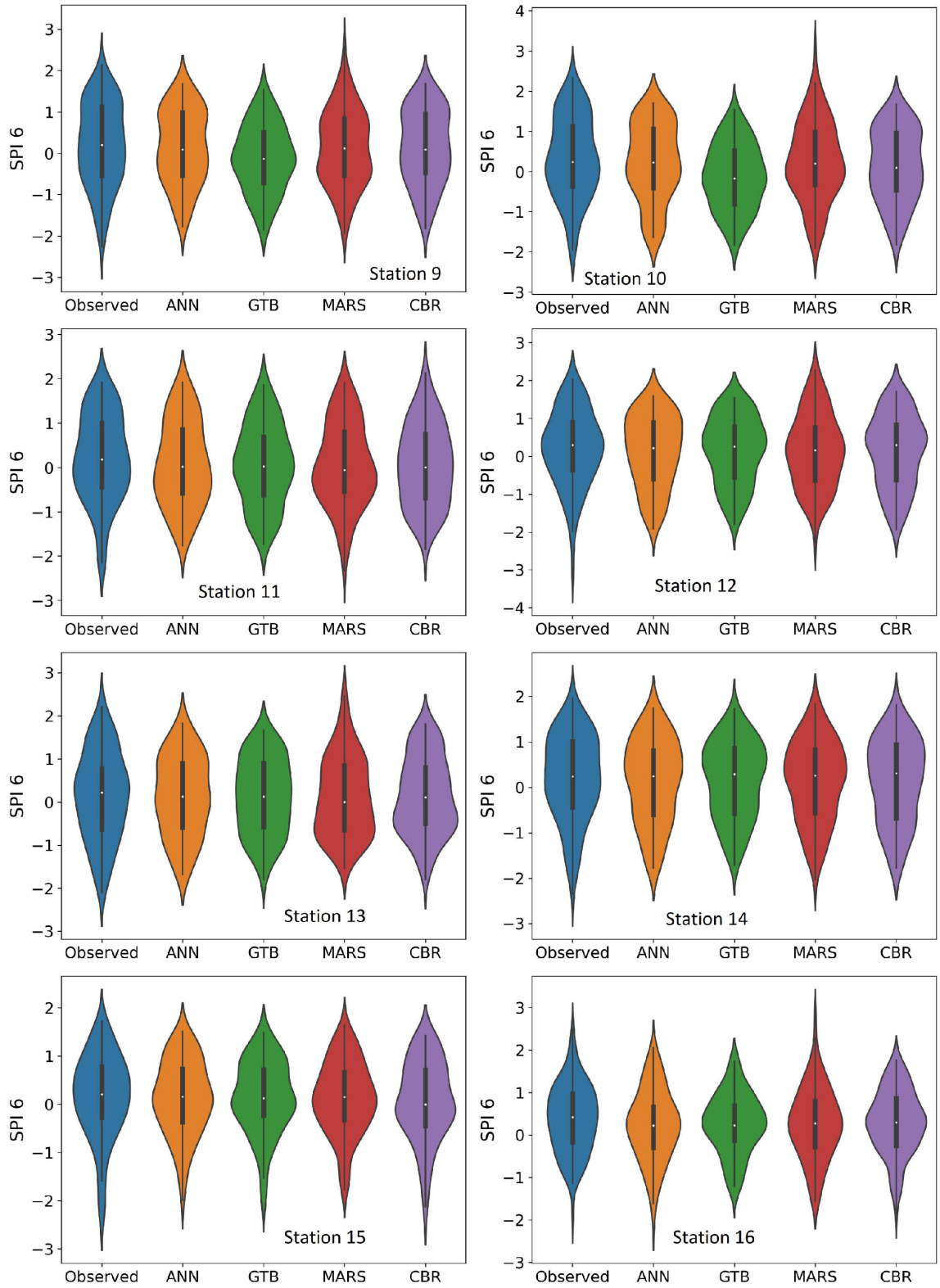

FIGURE 6. Violin plots of ML models used for SPI-6 forecasting (Stations 9-16) (Test Phase).

The selection of multivariate adaptive regression splines (MARS), CatBoost regression (CBR), and gradient tree boosting (GTB) paradigms was based on their ability to handle complex non-linear relationships, their robustness to noise and outliers, and their ability to produce accurate predictions. Additionally, CBR and GTB are both relatively

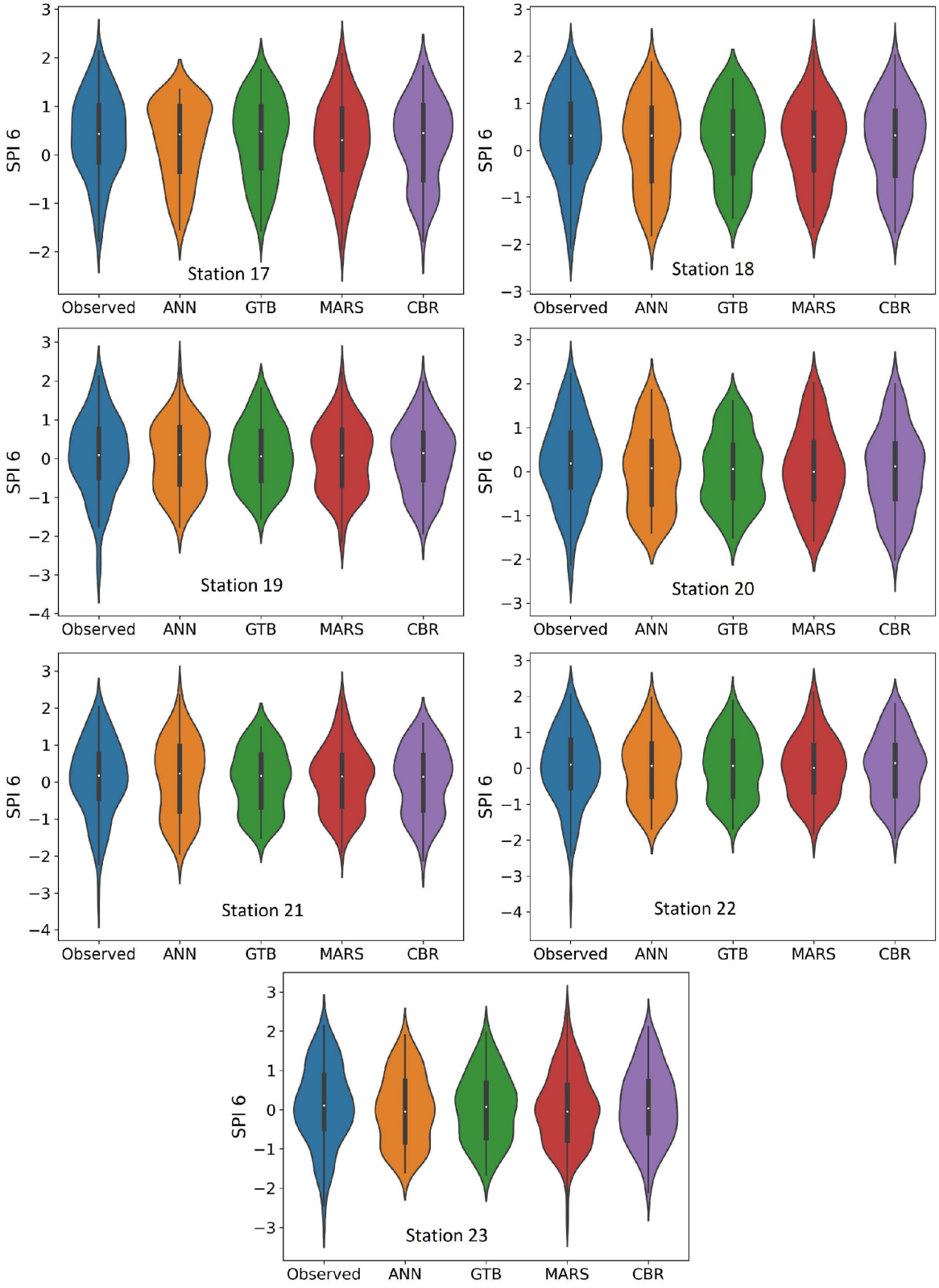

FIGURE 7. Violin plots of ML models used for SPI-6 forecasting (Stations 17-23) (Test Phase).

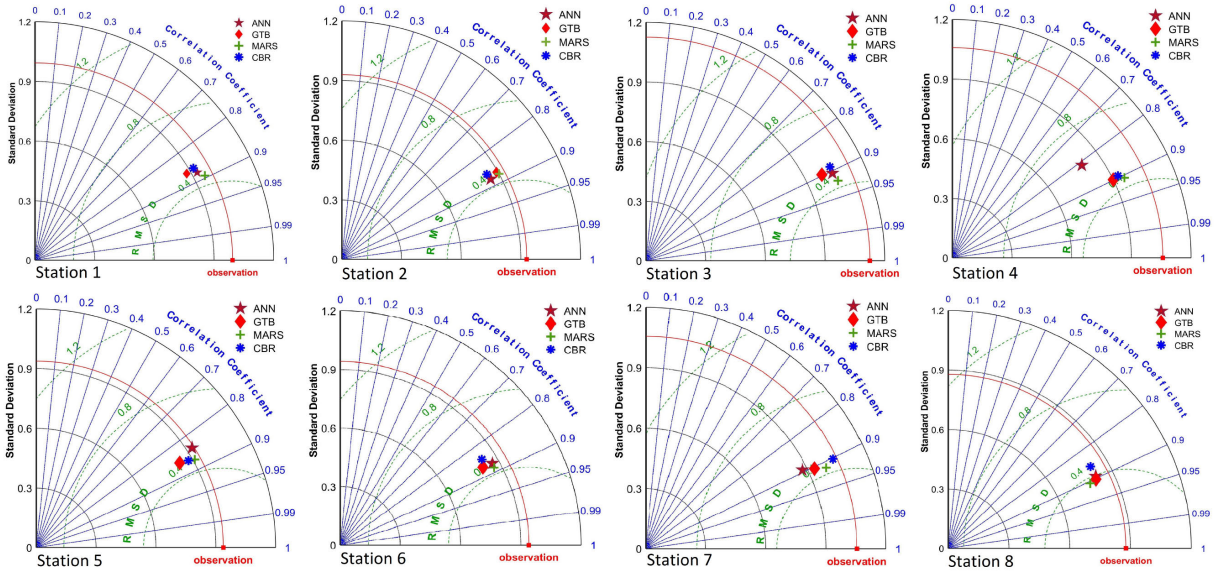

FIGURE 8. Taylor diagram for comparative evaluation of ML models used for SPI-6 forecasting (Stations 1-8) (Test Phase).

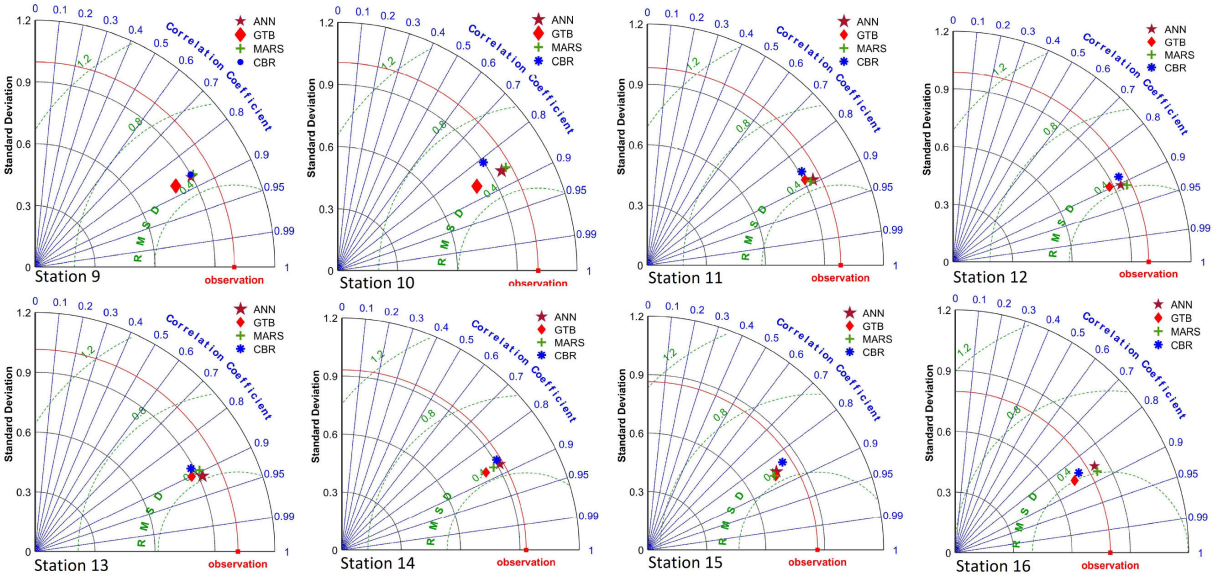

FIGURE 9. Taylor diagram for comparative evaluation of ML models used for SPI-6 forecasting (Stations 9-16) (Test Phase).

fast and efficient. These algorithms are well-suited for non-linear prediction tasks and have been shown to be effective in a variety of other studies. As a benchmark model, a standard artificial neural network was utilized to assess how well the ensemble models performed.

II. THEORETICAL OVERVIEW

A. ARTIFICIAL NEURAL NETWORKS (ANN)

The fundamental principle underlying Artificial Neural Networks is inspired (in part) by the way a person's neural system analyzes data as well as information in order to learn and grasp knowledge. A neural network is a system of interconnected processing units, called neurons, that work together to arrive at optimal solutions. The input layer is where the raw data is fed into the network, the hidden layers process the data, performs all sorts of computation using complex non-linear functions, and the output layer produces the final results. The activation functions determine whether or not a neuron should be activated, and backpropagation wouldn't be feasible without them. Activation functions provide the gradients and error that are needed to update the weights and biases of the network, and they also allow the network to re-learn and update during training. Through a chain rule mechanism, the backpropagation algorithm trains the neural network more efficiently by minimizing the loss function using the gradient descent optimization method. A feedforward backpropagation network architecture was implemented in the current study. Refer to [44] and [45] for more details on ANN, its various architectures, the model tuning process, and optimisation.

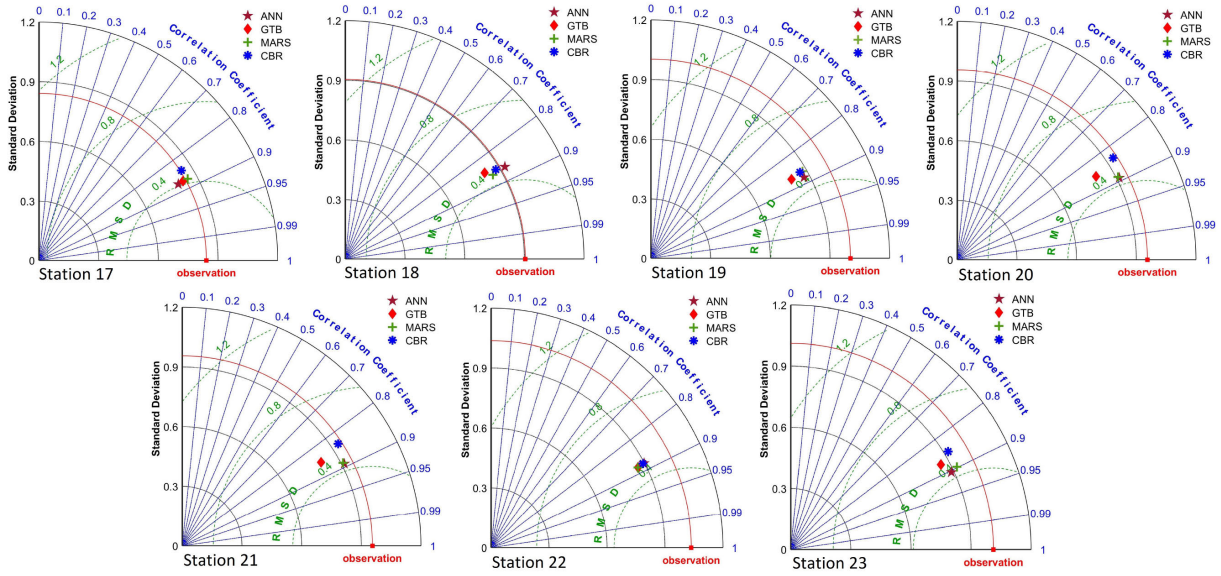

FIGURE 10. Taylor diagram for comparative evaluation of ML models used for SPI-6 forecasting (Stations 17-23) (Test Phase).

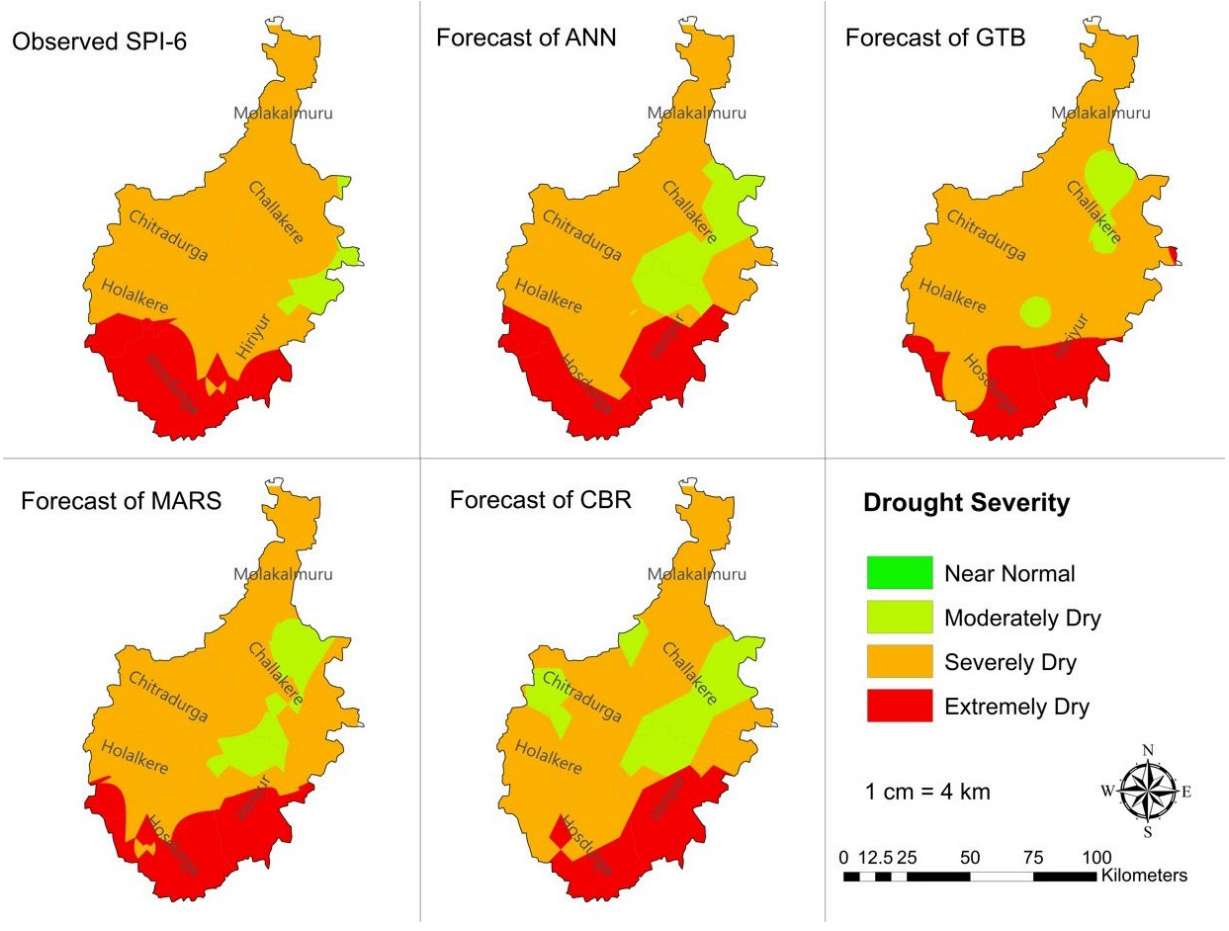

FIGURE 11. The drought severity maps.

B. GRADIENT TREE BOOSTING (GTB)

Gradient Tree Boosting is an ensemble learning paradigm that builds a final model from a set of individual models. Usually, these individual models are decision trees, which are weak learners. The prediction capability of these weak models often leads to overfitting or underfitting issues, but in collective/ensemble form produce a robust model [46]. To govern the contribution of each tree to the ensemble,

TABLE 2. Descriptive statistics of SPI-6 dataset derived from IMD precipitation data (Station 13 - 23).

Station	Dataset	MAX	MIN	MEAN	STD	SKEW
Station 13 14°0'N, 76°0'E	Full	2.21	-2.87	0.0	1.0	-0.3167
	Train	2.11	-2.87	0.0	1.0	-0.4086
	Test	2.21	-2.10	0.0	1.0	-0.0895
Station 14 14°0'N, 76°15'E	Full	2.05	-2.99	0.0	1.0	-0.3978
	Train	2.05	-2.99	0.0	1.0	-0.3696
	Test	1.96	-2.34	0.0	1.0	-0.4280
Station 15 14°0'N, 76°30'E	Full	1.85	-2.39	0.0	1.0	-0.6809
	Train	1.85	-2.39	0.0	1.0	-0.6597
	Test	1.72	-2.36	0.0	1.0	-0.6763
Station 16 14°0'N, 76°45'E	Full	2.49	-4.12	0.0	1.0	-0.8382
	Train	2.27	-4.12	0.0	1.0	-0.8879
	Test	2.49	-1.93	0.0	1.0	-0.0029
Station 17 14°0'N, 77°0'E	Full	2.14	-3.87	0.0	1.0	-0.4748
	Train	1.98	-3.87	0.0	1.0	-0.4392
	Test	2.14	-1.78	0.0	1.0	-0.3146
Station 18 13°45'N, 76°0'E	Full	2.15	-2.71	0.0	1.0	-0.4626
	Train	2.15	-2.71	0.0	1.0	-0.4351
	Test	1.99	-2.08	0.0	1.0	-0.4395
Station 19 13°45'N, 76°15'E	Full	2.12	-3.02	0.0	1.0	-0.3629
	Train	2.00	-3.02	0.0	1.0	-0.3437
	Test	2.12	-2.94	0.0	1.0	-0.4262
Station 20 13°45'N, 76°30'E	Full	2.23	-3.14	0.0	1.0	-0.2858
	Train	2.10	-3.14	0.0	1.0	-0.3156
	Test	2.23	-2.25	0.0	1.0	-0.1352
Station 21 13°45'N, 76°45'E	Full	2.13	-3.16	0.0	1.0	-0.3718
	Train	2.13	-2.90	0.0	1.0	-0.3518
	Test	2.03	-3.16	0.0	1.0	-0.4419
Station 22 13°30'N, 76°15'E	Full	2.05	-3.64	0.0	1.0	-0.3642
	Train	2.05	-3.44	0.0	1.0	-0.3176
	Test	2.05	-3.64	0.0	1.0	-0.4916
Station 23 13°30'N, 76°30'E	Full	2.22	-2.73	0.0	1.0	-0.3062
	Train	2.22	-2.72	0.0	1.0	-0.3275
	Test	2.15	-2.73	0.0	1.0	-0.2469

Note: MIN - Minimum; MAX - Maximum; STD - Standard Deviation; SKEW - Skewness

TABLE 3. Input-output (I/O) structure identified by mutual information criterion.

Stations	I/O structure
1-10, 12-19, 21 and 22	$(SPI)_{t-1} + (SPI)_t \mapsto (SPI)_{t+1}$
11, 20, 23	$(SPI)_{t-5} + (SPI)_{t-1} + (SPI)_t \mapsto (SPI)_{t+1}$

a learning rate is considered. This learning rate scales the contribution of each tree, promoting stable and accurate learning. Gradient Descent is employed to optimize the ensemble's loss function, which typically measures the difference between actual target values and ensemble predictions. The boosting process continues until a predefined number of trees (boosting rounds) are built or until the model's performance on a validation set no longer improves. GTB often utilize regularization techniques to prevent overfitting issues. These methods, such as feature subsampling and tree depth constraints, enhance the model's generalization capabilities. The ability of GTB to handle complex, non-linear features and robustness against overfitting have made it a favored choice among data scientists and machine learning practitioners. The authors advise referring to [47] and [48] for comprehensive information regarding GTB.

C. MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS)

MARS is a non-parametric regression technique capable of modeling the relationship between a dependent variable and multiple independent variables. It was introduced by an American statistician, Jerome H. Friedman, where the model is built using piecewise linear models, also known as basis functions, obtained by iteratively searching through all possible potential knots and across all variable interactions [49]. The knot positions are automatically selected using an adaptive regression method. The way the basis functions are derived is significant in the MARS algorithm. There are two phases here: first, the generation phase or forward pass, followed by the pruning phase or backward pass. In the forward pass, MARS searches for the best basis functions to add to the model. It does this by performing a forward stepwise search, greedily selecting the basis function that reduces the model's error the most. After the forward pass, MARS enters the pruning phase, where it iteratively removes basis functions from the model to reduce overfitting. The output of each basis function is weighted by a coefficient, and is a combinations of hinge functions. Basis functions are produced by a combination of hinge functions and the output of each basis function is often weighted by a coefficient. By summing the weighted output of all of the basis functions, the final output of the MARS model satisfying the generalised cross-validation criterion is obtained. In essence, the MARS model is an ensemble of linear functions. For further detailed information on MARS algorithm, refer to [50] and [51].

D. CATBOOST REGRESSION (CBR)

CatBoost or Categorical Boosting is based on ensemble learning strategy that uses gradient boosting and decision trees to generate regression models. In CBR the features are handled directly without being encoded. CatBoost also has an array of extra functionalities such as the provisions for early stopping and cross-validation, a variety of loss function algorithms that users can select from, the potential to manage missing data, and supports both L1 and L2 regularization strategies to avoid overfitting. Prompting leaf-wise growth, CatBoost builds trees in order to lower the loss function at each and every leaf itself and thereby enhance the precision of the model. CatBoost, in general, is a sophisticated and versatile regression technique that applies a cross-validation method internally to choose the optimal hyperparameters for the model. Refer to the following literature [52] and [53] for more details on the CatBoost algorithm.

III. METHODOLOGY

A. MODEL DEVELOPMENT

The Standardized Precipitation Index (SPI) developed by McKee et al. [54] is widely used to detect and characterize meteorological drought. It can be used to represent drought conditions over a wide range of timescales (3, 6, 12, 24, and 48 months), making it suitable for both short- and long-term droughts. A long-term precipitation record at the chosen station is fitted to a probability distribution (such as the gamma distribution) in order to produce the SPI (Standard Precipitation Index), which is then transformed into a normal distribution with mean SPI as zero and unit

TABLE 4. Model parameter settings used for calibration of the developed models (Station 1 - 12).

Models	Model Parameters	*Station 1*	*Station 2*	*Station 3*	*Station 4*	*Station 5*	*Station 6*	*Station 7*	*Station 8*	*Station 9*	*Station 10*	*Station 11*	*Station 12*
ANN	No. of HL neurons	3	4	14	10	13	10	5	5	3	10	8	3
	Hidden layer AF	tansig	logsig	tansig	tansig	tansig	logsig	logsig	logsig	logsig	logsig	tansig	logsig
	Output layer AF	tansig	purelin	purelin	tansig	tansig	purelin	purelin	purelin	tansig	tansig	purelin	tansig
	Training function	trainlm	trainlm	trainlm	trainoss	trainrp	trainlm	trainrp	trainrp	trainlm	trainlm	trainlm	trainlm
	Epocs	200	400	500	500	200	200	200	200	200	200	200	200
GTB	Loss function	huber	huber	huber	huber	huber	huber	huber	huber	huber	huber	huber	huber
	learning_rate	0.08	0.1	0.15	0.17	0.1	0.2	0.2	0.2	0.2	0.3	0.3	0.09
	n_estimators	66	66	40	30	50	20	20	30	30	20	23	40
	max_depth	3	5	4	3	3	3	3	3	3	3	3	3
	min_samples_split	10	6	4	2	4	2	2	2	3	2	2	2
MARS	max_degree	2	3	3	2	8	5	8	5	5	5	3	5
	penalty	1	4	5	5	8	6	8	9	11	10	12	10
	minspan_alpha	0.05	0.01	0.01	0.1	0.1	0.5	0.4	0.09	0.2	0.9	0.5	1
	endspan_alpha	0.05	0.02	0.01	0.1	0.1	1.5	5	5	5	1.5	0.5	1
	endspan	5	8	5	5	5	5	8	5	5	7	2	5
CBR	Iterations	5	8	10	13	10	7	10	5	7	7	7	8
	learning_rate	1	1	1	1	1	1	1	1	1	1	1	1
	depth	7	6	5	5	5	5	7	5	5	5	5	5
	l2_leaf_reg	0.95	1.78	1.85	1.85	4.1	1.8	1	5	1.9	1.9	1.5	1.6

Note: HL - Hidden Layer; AF - Activation Function.

TABLE 5. Model parameter settings used for calibration of the developed models (Station 13 - 23).

Models	Model Parameters	*Station 13*	*Station 14*	*Station 15*	*Station 16*	*Station 17*	*Station 18*	*Station 19*	*Station 20*	*Station 21*	*Station 22*	*Station 23*
ANN	No. of HL neurons	3	3	5	12	3	15	3	3	15	3	5
	Hidden layer AF	logsig	logsig	tansig	logsig	logsig	logsig	logsig	tansig	tansig	tansig	tansig
	Output layer AF	tansig	tansig	purelin	tansig	purelin	purelin	purelin	purelin	purelin	purelin	tansig
	Training function	trainlm	trainlm	trainlm	trainlm	trainrp	trainlm	trainlm	trainlm	trainlm	trainlm	trainlm
	Epocs	200	200	200	200	200	200	200	200	500	200	200
GTB	Loss function	huber	huber	huber	huber	huber	huber	huber	huber	huber	huber	huber
	learning_rate	0.09	0.1	0.09	0.05	0.11	0.12	0.2	0.2	0.09	0.09	0.2
	n_estimators	40	30	50	50	60	20	20	23	40	100	19
	max_depth	4	4	3	4	3	3	3	2	3	3	3
	min_samples_split	2	3	2	2	3	5	2	3	3	3	2
MARS	max_degree	5	5	5	5	5	5	5	5	3	3	3
	penalty	10	11	10	10	12	9	6	12	8	5	10
	minspan_alpha	0.5	0.5	1	0.7	0.9	0.5	0.5	0.2	0.4	0.3	0.5
	endspan_alpha	0.5	0.5	1	0.7	0.9	0.5	1.5	0.2	0.4	0.3	0.5
	endspan	5	5	5	5	5	5	5	8	5	3	5
CBR	Iterations	11	8	10	10	7	12	19	10	7	8	13
	learning_rate	1	1	1	1	1	1	1	1	1	1	1
	depth	3	5	5	5	5	3	3	3	5	4	3
	l2_leaf_reg	1.9	1.9	1.5	1.8	0.9	2.1	1.9	1.5	1.9	2	2.2

Note: HL - Hidden Layer; AF - Activation Function.

standard deviation. Droughts can be detected and classified using the SPI. Generally speaking, a drought is indicated by an SPI value of -1.5 or less, with larger negative values suggesting more severe drought conditions. The SPI can also be used to compare drought conditions in various regions and track the onset of a drought over time.

In the present study, SPI-6 was calculated from monthly rainfall data of 1980-2019 (40 years) procured from IMD, India. Further, the SPI-6 data was divided into two parts, namely the training data (1980-2009) and testing data (2010-2019). The descriptive statistics of SPI-6 timeseries of 23 grid stations within the study area are presented in Tables 1 and 2. Descriptive statistics is a versatile tool used to summarize and describe the main features of a dataset, such as the central tendency and dispersion. Mutual Information (MI) approach was employed to choose the best lag intervals of SPI-6 timeseries so as to serve as inputs for forecasting 1-step lead-time SPI-6 time series. The mutual information between the original time series and the lagged version of the time series (that accounts for temporal dependence) is utilized in the selection of input features for building models. The input-output structure of the models developed for the SPI-6 data of 23 grid stations are presented in Table 3. The input-output structure of an ML model is an important factor to consider when designing and applying the model. The input data should be relevant to the task that the model is being used for, and map the output data. The model hyperparameters of ANN, GTB, MARS and CBR were optimized based on automated grid search approach. Figure 2 illustrates the multiple steps followed in this research methodology.

B. PERFORMANCE EVALUATION

The degree of agreement between the forecasted and actual SPI-6 data was assessed using error and efficiency metrics. Mean Absolute Error (MAE) is a non-parametric measure, computed as the average of the absolute differences between the forecasted values and the actual values. Similarly, Root Mean Square Error (RMSE) is a measure of the average squared difference between forecasted values and actual values. RMSE is also more sensitive to outliers than MAE, making it more susceptible to being influenced by a

TABLE 6. Model performance evaluation using statistical metrics (Station 1 - 12).

Station	Models	TRAIN				TEST			
		MAE	RMSE	NNSE	WI	MAE	RMSE	NNSE	WI
1	ANN	0.321	0.446	0.817	0.934	0.402	0.516	0.788	0.925
	GTB	0.259	0.383	0.859	0.951	0.399	0.523	0.783	0.918
	MARS	0.335	0.465	0.804	0.927	0.376	0.472	0.816	0.939
	CBR	0.321	0.437	0.823	0.937	0.408	0.529	0.779	0.920
2	ANN	0.338	0.468	0.819	0.932	0.370	0.461	0.803	0.930
	GTB	0.157	0.244	0.943	0.984	0.367	0.485	0.786	0.925
	MARS	0.320	0.454	0.827	0.939	0.362	0.475	0.793	0.929
	CBR	0.311	0.421	0.848	0.949	0.402	0.507	0.771	0.914
3	ANN	0.278	0.385	0.858	0.952	0.379	0.488	0.842	0.947
	GTB	0.177	0.270	0.925	0.978	0.377	0.499	0.836	0.942
	MARS	0.306	0.418	0.837	0.943	0.324	0.439	0.868	0.958
	CBR	0.263	0.349	0.880	0.962	0.397	0.518	0.825	0.940
4	ANN	0.422	0.572	0.743	0.885	0.470	0.628	0.739	0.876
	GTB	0.260	0.395	0.859	0.951	0.362	0.469	0.836	0.940
	MARS	0.350	0.484	0.802	0.925	0.349	0.452	0.846	0.947
	CBR	0.295	0.390	0.862	0.955	0.367	0.474	0.833	0.941
5	ANN	0.403	0.534	0.781	0.919	0.403	0.536	0.754	0.912
	GTB	0.289	0.420	0.852	0.947	0.366	0.486	0.789	0.921
	MARS	0.382	0.523	0.788	0.917	0.369	0.477	0.794	0.930
	CBR	0.324	0.436	0.843	0.946	0.369	0.481	0.792	0.926
6	ANN	0.363	0.474	0.824	0.937	0.347	0.464	0.804	0.930
	GTB	0.293	0.398	0.869	0.955	0.349	0.461	0.806	0.927
	MARS	0.361	0.476	0.823	0.937	0.320	0.436	0.823	0.939
	CBR	0.341	0.448	0.840	0.946	0.383	0.500	0.780	0.915
7	ANN	0.322	0.451	0.822	0.934	0.383	0.502	0.816	0.930
	GTB	0.235	0.361	0.878	0.959	0.339	0.453	0.845	0.946
	MARS	0.316	0.446	0.825	0.937	0.316	0.436	0.855	0.953
	CBR	0.245	0.332	0.895	0.968	0.370	0.466	0.837	0.948
8	ANN	0.303	0.447	0.845	0.947	0.291	0.394	0.832	0.943
	GTB	0.215	0.338	0.905	0.970	0.286	0.380	0.842	0.947
	MARS	0.316	0.453	0.842	0.946	0.290	0.373	0.847	0.947
	CBR	0.326	0.443	0.847	0.948	0.363	0.454	0.789	0.924
9	ANN	0.361	0.486	0.808	0.929	0.360	0.492	0.805	0.929
	GTB	0.556	0.713	0.663	0.853	0.480	0.586	0.744	0.893
	MARS	0.373	0.518	0.788	0.917	0.373	0.501	0.799	0.927
	CBR	0.351	0.471	0.818	0.935	0.370	0.504	0.797	0.926
10	ANN	0.333	0.447	0.828	0.939	0.389	0.532	0.781	0.922
	GTB	0.577	0.725	0.647	0.847	0.576	0.703	0.672	0.853
	MARS	0.347	0.463	0.818	0.934	0.397	0.536	0.778	0.923
	CBR	0.535	0.674	0.680	0.857	0.483	0.627	0.720	0.885
11	ANN	0.298	0.402	0.861	0.954	0.350	0.466	0.817	0.939
	GTB	0.218	0.309	0.913	0.973	0.373	0.488	0.802	0.931
	MARS	0.306	0.421	0.849	0.949	0.354	0.468	0.815	0.938
	CBR	0.283	0.376	0.876	0.961	0.425	0.528	0.776	0.918
12	ANN	0.280	0.401	0.861	0.955	0.307	0.433	0.838	0.947
	GTB	0.233	0.351	0.890	0.963	0.305	0.444	0.832	0.941
	MARS	0.305	0.428	0.844	0.947	0.290	0.421	0.846	0.952
	CBR	0.279	0.387	0.869	0.957	0.331	0.477	0.810	0.936

TABLE 7. Model performance evaluation using statistical metrics (Station 13 - 23).

Station	Models	TRAIN				TEST			
		MAE	RMSE	NNSE	WI	MAE	RMSE	NNSE	WI
13	ANN	0.319	0.455	0.828	0.938	0.311	0.420	0.854	0.951
	GTB	0.219	0.331	0.901	0.968	0.328	0.443	0.840	0.942
	MARS	0.357	0.489	0.806	0.927	0.350	0.453	0.834	0.943
	CBR	0.315	0.437	0.839	0.945	0.363	0.479	0.818	0.933
14	ANN	0.374	0.494	0.811	0.930	0.356	0.470	0.797	0.931
	GTB	0.267	0.388	0.874	0.956	0.343	0.453	0.809	0.930
	MARS	0.356	0.493	0.811	0.930	0.348	0.464	0.801	0.930
	CBR	0.350	0.469	0.826	0.938	0.395	0.495	0.780	0.923
15	ANN	0.346	0.496	0.799	0.923	0.344	0.454	0.783	0.917
	GTB	0.284	0.416	0.850	0.946	0.313	0.436	0.797	0.923
	MARS	0.353	0.497	0.798	0.923	0.319	0.441	0.793	0.921
	CBR	0.296	0.423	0.845	0.947	0.377	0.493	0.755	0.909
16	ANN	0.368	0.534	0.794	0.921	0.360	0.461	0.750	0.916
	GTB	0.290	0.462	0.838	0.936	0.329	0.420	0.783	0.918
	MARS	0.372	0.541	0.789	0.918	0.333	0.426	0.778	0.928
	CBR	0.348	0.474	0.830	0.940	0.344	0.435	0.771	0.915
17	ANN	0.355	0.491	0.812	0.929	0.332	0.431	0.792	0.927
	GTB	0.244	0.369	0.885	0.962	0.314	0.423	0.799	0.932
	MARS	0.350	0.493	0.811	0.930	0.336	0.442	0.784	0.928
	CBR	0.324	0.431	0.849	0.950	0.360	0.486	0.750	0.911
18	ANN	0.374	0.502	0.807	0.927	0.375	0.499	0.767	0.922
	GTB	0.311	0.455	0.836	0.935	0.366	0.486	0.776	0.916
	MARS	0.369	0.515	0.799	0.923	0.352	0.466	0.791	0.926
	CBR	0.351	0.486	0.817	0.934	0.355	0.482	0.779	0.923
19	ANN	0.343	0.470	0.822	0.934	0.352	0.472	0.819	0.934
	GTB	0.214	0.316	0.910	0.972	0.373	0.497	0.803	0.921
	MARS	0.367	0.503	0.801	0.924	0.385	0.504	0.799	0.925
	CBR	0.286	0.396	0.866	0.956	0.371	0.504	0.798	0.924
20	ANN	0.334	0.459	0.830	0.939	0.348	0.464	0.809	0.936
	GTB	0.299	0.414	0.857	0.949	0.348	0.464	0.809	0.936
	MARS	0.351	0.472	0.822	0.936	0.351	0.465	0.809	0.935
	CBR	0.317	0.406	0.862	0.955	0.430	0.574	0.735	0.904
21	ANN	0.352	0.475	0.818	0.939	0.345	0.489	0.808	0.938
	GTB	0.287	0.413	0.856	0.948	0.351	0.485	0.811	0.926
	MARS	0.362	0.494	0.807	0.927	0.326	0.456	0.829	0.938
	CBR	0.353	0.460	0.828	0.940	0.358	0.501	0.801	0.927
22	ANN	0.339	0.466	0.820	0.934	0.354	0.490	0.818	0.933
	GTB	0.247	0.344	0.893	0.966	0.354	0.488	0.818	0.930
	MARS	0.351	0.471	0.817	0.933	0.357	0.488	0.819	0.930
	CBR	0.346	0.448	0.831	0.942	0.365	0.490	0.817	0.932
23	ANN	0.318	0.426	0.848	0.946	0.331	0.447	0.837	0.943
	GTB	0.241	0.331	0.902	0.968	0.357	0.496	0.806	0.925
	MARS	0.329	0.441	0.839	0.944	0.334	0.454	0.832	0.942
	CBR	0.277	0.366	0.883	0.963	0.376	0.532	0.783	0.918

few extreme values. Normalized Nash–Sutcliffe Efficiency (NNSE) is a dimensionless metric of the accuracy of hydrological models. It is also a more sensitive measure of accuracy than the coefficient of determination (R^2), making it more likely to detect large errors. The Willmott Index (WI) can be used to compare the agreement between two time series, regardless of their scales. A Willmott index of 1 indicates a perfect fit between the simulated and observed values.

- Mean Absolute Error (MAE)

$$MAE = \sum_{i=1}^{n} \frac{|y_i - x_i|}{N} \tag{1}$$

- Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{N}} \tag{2}$$

- Normalized Nash–Sutcliffe Efficiency (NNSE)

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] \tag{3}$$

$$NNSE = \frac{1}{2 - NSE} \tag{4}$$

- Willmott Index (WI)

$$WI = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(|y_i - \bar{x}| + |x_i - \bar{x}|)^2} \tag{5}$$

where, x_i represnts the observed value, y_i represents the model estimated value, $\bar{x}$ represents the mean of the observed data, and N is the number of data points.

IV. RESULTS AND DISCUSSION

The study was conducted to forecast the one-step-ahead Standardized Precipitation Index (SPI-6) of 23 rainfall gauging stations in the Chitradurga district of India. Four machine learning models, namely ANN, GTB, MARS and CBR were developed to generate the forecast. The performance of the models was assessed using various performance indices, such as MAE, RMSE, NNSE, and WI. The model parameters were arrived at using an automated grid search approach that systematically scans across a range of potential values. The parameter settings of the calibrated models are presented in Tables 4 and 5.

The results of one-step ahead SPI-6 forecasts using four ML models revealed that the ensemble learning based models showed superior performance than the typical ANN considering all evaluation indices. Tables 6 and 7 presents the performance statistics for comparative evaluation of all the four ML models with reference to stations 1-23. Considering all evaluation metrics, the MARS model outperformed GTB, CBR, and ANN models. For instance, referring to station 4 testing results, the improvement in the MAE, RMSE, NNSE and WI of the MARS as compared to the ANN was by 25.53%, 27.11%, 14.86% and 7.95%, respectively. However, the corresponding statistics were 2.77%, 2.27%, 1.19% and 1.06%, respectively, when the MARS and GTB forecasts were compared. The MARS performed relatively superior than the GTB. The MARS model is also an ensemble learning model that combines multiple linear regression splines to improve forecast accuracy. Looking at the testing results of stations 11, 20 and 23, the additional input variable ($SPI_{(t-5)}$) that was considered during the calibration of the ANN model for these stations may have helped to reduce the noise in the data and improve the performance of the ANN model. As a result, the ANN model performed relatively similar to that of MARS for stations 11, 20 and 23. The MARS model produced the best SPI-6 forecasts among the four implemented ML models, with the least error and higher efficiency indices. The MARS model had the highest NNSE values for 17 out of the 23 stations. The SPI-6 forecasts of stations 8 and 12 were significantly accurate with the least error criteria (MAE=0.29) compared to the other 21 stations. With the exception of ANN model, the performance of the GTB, MARS, and CBR paradigms was undeniably almost similar for the all the stations.

The scatter plots presented in Figures 3 and 4 portray the degree of association between the forecasted and observed SPI-6 data. It could be observed that the points in the scatter plots of the MARS and CBR models are more tightly clustered around the line of perfect fit compared to the other models. Having a coefficient of determination (R^2) between 0.7 and 0.9 indicates that the variables (observed and forecasted) exhibit the highest degree of association. The test phase results were considered for plotting.

The violin plots, which are relatively easy to interpret, prove to be a good choice for representing data and communicate to a wider audience. A violin plot offers the benefits of a box plot and a kernel density plot. The Figures 5, 6 and 7 portray the violin plots that represent the results of SPI-6 forecasts from the ANN, GTB, MARS and CBR models with respect to test phase. The violin plots also aid in visualising and comparing the distributions of observed and forecated SPI-6 data. Based on skew, symmetry, and variability attributes, the violin plots of the observed SPI-6 data and the MARS model estimates relatively mimic each other. Referring to both tabular results and violin plots, it could be inferred that the GTB model outperformed MARS for station 8, and the CBR model forecasts were relatively superior to those of MARS for stations 18 and 20.

The Taylor diagram [55] is a graphical tool employed for comparing the performance of different models (in test phase) considering three statistical metrics, namely the correlation coefficient, the centered root-mean-square and the standard deviation. The Taylor diagrams presented in Figures 8, 9 and 10 show that chronologically the MARS, GTB and CBR models produced relatively accurate forecasts of the SPI-6 timeseries for all 23 stations within the Chitradurga district, India. From the viewpoint of Taylor diagrams, the performance of the models varied from station to station.

For example, the MARS model had the best performance for stations 1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 21, 22 and 23. The GTB model had the best performance for station 8. The CBR model had the best performance for stations 7, 18, 20. The performance of the MARS, GTB and CBR models was relatively consistent across several stations.

Here, we compare the performance of the proposed models with other models that were identified in the literature. For the SPI-6 forecast in the Aegean region of Türkiye, the XgBoost model developed by Gul et al. [39] achieved WI = 0.901. Similarly, Basak et al. [36] conducted drought research focusing on semi-arid regions of western India. According to their findings, the test phase RMSE range of the Prophet model for the SPI-6 forecast was 0.29–0.63. However, according to the results obtained in the current case study, for station 3, the SPI-6 forecasts from the MARS model achieved WI = 0.96, and in most stations, WI > 0.94 was attained. The RMSE and NNSE ranges of the MARS model (during test phase) were 0.37–0.54 and 0.78–0.87, respectively. Shahdad and Saber [56] investigated the efficiency of the standalone Reduced Error Pruning Tree (REPT) model and its integration with Bagging (BA), Additive Regression (AR), Dagging (DA), and Random Committee (RC) for modeling SPI-6 in the Karkheh watershed in Iran. They found that hybrid algorithms improved the modeling capabilities of the standalone REPT algorithm. The DA-REPT model outperformed the other hybrid models with the lowest RMSE of 0.387. Overall, the hybrid ML modelling approach is a promising approach for SPI-6 forecasting. High accuracy can be attained while being computationally efficient.

The thematic maps were developed to depict the spatial extent of drought severity in the Chitradurga district. The maps were developed for the observed SPI-6 and model forecasted SPI-6 data. A spatial interpolation technique called Ordinary Kriging was used to generate the thematic maps, considering the data from all the 23 grid locations. The drought severity map derived for April 2019 is presented in Figure 11. The maps generated from MARS and GTB model forecasts better emulate the observed SPI-6 map.

V. CONCLUSION

In a semi-arid climatic area like Chitradurga, droughts have an devastating impact on living beings. So, the attempt of alerting the local authorities to mitigate or counteract the consequences is of vital importance. Hence, the objective was to determine the most effective machine learning model for forecasting meteorological drought. SPI was chosen due to its very limited data requirement, that is precipitation, and its wide temporal flexibility. Model complexity and computational effort were reduced by using mutual information to choose model input features. Among the implemented models, MARS was found to have best performance in forecasting the one-step lead-time drought using SPI-6 timeseries. The GTB and CBR models too showed satisfactory values of NNSE with lower RMSE values as compared to MARS in most of the grid locations. It is clear from the thematic maps that, the entire district suffered drought in the summer of 2019 (April 2019). In the drought severity analysis using SPI-6 timeseries, the prediction maps from the MARS and GTB models converged with the observed SPI-6 map. The study only considered one-step lead-time forecasting. It would be interesting to see how the MARS model performs for longer lead times. The limitation of parameter optimization in ML models could be overcome by employing meta-heuristic techniques. To enhance the task of drought forecasting, the parameter optimization of ML models is feasible through the use of either evolution-based, swarm-based, or physics-based approaches. Furthermore, the use of SPI data in conjunction with other meteorological variables as input could considerably improve modelling performance.

AUTHOR CONTRIBUTIONS

Conceptualization, Sujay Raghavendra Naganna and Nagaraj Bhat; methodology, Sujay Raghavendra Naganna; software, Sujay Raghavendra Naganna and Nagaraj Bhat; validation, Dayananda Pruthviraja and R. H. Goudar; formal analysis, Geetabai S. Hukkeri; investigation, Geetabai S. Hukkeri; resources, Sujay Raghavendra Naganna and Nagaraj Bhat; data curation, Sujay Raghavendra Naganna and Nagaraj Bhat; writing—original draft, Geetabai S. Hukkeri, Dayananda Pruthviraja, and Sujay Raghavendra Naganna writing—review and editing, Sujay Raghavendra Naganna; visualization, Geetabai S. Hukkeri and Nagaraj Bhat; supervision, Dayananda Pruthviraja and R. H. Goudar. All authors have read and agreed to the published version of the manuscript.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

"During the preparation of this work, the authors used Google Bard in order to draft some sentences in the introduction and method sections. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication."

DATA AVAILABILITY

"Data cannot be shared openly but is available on request from authors."

ACKNOWLEDGMENT

The authors would like to thank the Indian Meteorological Department, Government of India, for providing the necessary data required for research.

CONFLICTS OF INTEREST

"The authors have no conflicts of interest to declare that are relevant to the content of this article."

REFERENCES

[1] A. K. Mishra and V. P. Singh, "A review of drought concepts," *J. Hydrol.*, vol. 391, nos. 1–2, pp. 202–216, Sep. 2010.

[2] M. Allaby and R. Garratt, *Droughts* (Dangerous Weather). New York, NY, USA: Facts On File, 2014.

[3] J. Spinoni, G. Naumann, H. Carrao, P. Barbosa, and J. Vogt, "World drought frequency, duration, and severity for 1951–2010," *Int. J. Climatol.*, vol. 34, no. 8, pp. 2792–2804, Jun. 2014.

[4] Y. Ding, M. J. Hayes, and M. Widhalm, "Measuring economic impacts of drought: A review and discussion," *Disaster Prevention Manage., Int. J.*, vol. 20, no. 4, pp. 434–446, Aug. 2011.

[5] A. Kikon, B. M. Dodamani, S. D. Barma, and S. R. Naganna, "ANFIS-based soft computing models for forecasting effective drought index over an arid region of India," *AQUA-Water Infrastruct., Ecosyst. Soc.*, vol. 72, no. 6, pp. 930–946, Jun. 2023.

[6] D. A. Wilhite and M. H. Glantz, "Understanding the drought phenomenon: The role of definitions," *Water Int.*, vol. 10, no. 3, pp. 111–120, 1985.

[7] M. P. Shewale and S. Kumar, "Climatological features of drought incidences in India," Meteorol. Monograph, Climatol., Dept. India Meteorol., Pune, India, Tech. Rep. 21/2005, 2005. [Online]. Available: https://www.imdpune.gov.in/Reports/drought.pdf

[8] W. C. Palmer, *Meteorological Drought*, vol. 30. Washington, DC, USA: U.S. Department of Commerce, Weather Bureau, 1965.

[9] S. Stefanidis, D. Rossiou, and N. Proutsos, "Drought severity and trends in a Mediterranean oak forest," *Hydrology*, vol. 10, no. 8, p. 167, Aug. 2023.

[10] B. Mohammadi, "Modeling various drought time scales via a merged artificial neural network with a firefly algorithm," *Hydrology*, vol. 10, no. 3, p. 58, Feb. 2023.

[11] S. Bahmani, S. R. Naganna, M. A. Ghorbani, M. Shahabi, E. Asadi, and S. Shahid, "Geographically weighted regression hybridized with Kriging model for delineation of drought-prone areas," *Environ. Model. Assessment*, vol. 26, no. 5, pp. 803–821, Oct. 2021.

[12] P. Das, S. R. Naganna, P. C. Deka, and J. Pushparaj, "Hybrid wavelet packet machine learning approaches for drought modeling," *Environ. Earth Sci.*, vol. 79, no. 10, p. 221, May 2020.

[13] A. K. Mishra and V. P. Singh, "Drought modeling—A review," *J. Hydrol.*, vol. 403, nos. 1–2, pp. 157–175, 2011.

[14] D. A. Wilhite, M. D. Svoboda, and M. J. Hayes, "Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness," *Water Resour. Manage.*, vol. 21, pp. 763–774, May 2007.

[15] S. Jha and R. Srivastava, "Impact of drought on vegetation carbon storage in arid and semi-arid regions," *Remote Sens. Appl., Soc. Environ.*, vol. 11, pp. 22–29, Aug. 2018.

[16] M. L. Parry, T. R. Carter, and N. T. Konijn, *The Climatology of Droughts and Drought Prediction*. Dordrecht, The Netherlands: Springer, 1988, pp. 305–323.

[17] B. G. Hunt, "The simulation and prediction of drought," *Vegetatio*, vol. 91, nos. 1–2, pp. 89–103, Jan. 1991.

[18] A. K. Mishra and V. R. Desai, "Drought forecasting using stochastic models," *Stochastic Environ. Res. Risk Assessment*, vol. 19, no. 5, pp. 326–339, Nov. 2005.

[19] J. S. Oguntoyinbo, "Drought prediction," *Clim. Change*, vol. 9, no. 1, pp. 79–90, Aug. 1986.

[20] A. AghaKouchak, B. Pan, O. Mazdiyasni, M. Sadegh, S. Jiwa, W. Zhang, C. A. Love, S. Madadgar, S. M. Papalexiou, S. J. Davis, K. Hsu, and S. Sorooshian, "Status and prospects for drought forecasting: Opportunities in artificial intelligence and hybrid physical–statistical forecasting," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 380, no. 2238, Dec. 2022, Art. no. 20210288.

[21] N. B. Guttman, "Comparing the Palmer drought index and the standardized precipitation index," *J. Amer. Water Resour. Assoc.*, vol. 34, no. 1, pp. 113–121, Feb. 1998.

[22] X. Bai, W. Shen, X. Wu, and P. Wang, "Applicability of long-term satellite-based precipitation products for drought indices considering global warming," *J. Environ. Manage.*, vol. 255, Feb. 2020, Art. no. 109846.

[23] G. Naumann, P. Barbosa, H. Carrao, A. Singleton, and J. Vogt, "Monitoring drought conditions and their uncertainties in Africa using TRMM data," *J. Appl. Meteorol. Climatol.*, vol. 51, no. 10, pp. 1867–1874, 2012.

[24] U. S. Panu and T. C. Sharma, "Challenges in drought research: Some perspectives and future directions," *Hydrol. Sci. J.*, vol. 47, no. s1, pp. S19–S30, Aug. 2002.

[25] A. Zargar, R. Sadiq, B. Naser, and F. I. Khan, "A review of drought indices," *Environ. Rev.*, vol. 19, pp. 333–349, Jul. 2011.

[26] J. M. Ndayiragije and F. Li, "Effectiveness of drought indices in the assessment of different types of droughts, managing and mitigating their effects," *Climate*, vol. 10, no. 9, p. 125, Aug. 2022.

[27] G. Kallis, "Droughts," *Annu. Rev. Environ. Resour.*, vol. 33, pp. 85–118, Nov. 2008. [Online]. Available: http://dx.doi.org/10.1146/annurev.environ.33.081307.123117

[28] A. F. Van Loon, "Hydrological drought explained," *Wiley Interdiscip. Rev., Water*, vol. 2, no. 4, pp. 359–392, Jul. 2015.

[29] S. M. Vicente-Serrano, S. M. Quiring, M. Peña-Gallardo, S. Yuan, and F. Domínguez-Castro, "A review of environmental droughts: Increased risk under global warming?" *Earth-Sci. Rev.*, vol. 201, Feb. 2020, Art. no. 102953.

[30] Z. Hao and V. P. Singh, "Drought characterization from a multivariate perspective: A review," *J. Hydrol.*, vol. 527, pp. 668–678, Aug. 2015.

[31] S. Bachmair, K. Stahl, K. Collins, J. Hannaford, M. Acreman, M. Svoboda, C. Knutson, K. H. Smith, N. Wall, B. Fuchs, N. D. Crossman, and I. C. Overton, "Drought indicators revisited: The need for a wider consideration of environment and society," *Wiley Interdiscip. Rev., Water*, vol. 3, no. 4, pp. 516–536, Jul. 2016.

[32] G. Gebremeskel Haile, Q. Tang, S. Sun, Z. Huang, X. Zhang, and X. Liu, "Droughts in east africa: Causes, impacts and resilience," *Earth-Sci. Rev.*, vol. 193, pp. 146–161, Jun. 2019.

[33] S. V. R. K. Prabhakar, "Implications of regional droughts and transboundary drought risks on drought monitoring and early warning: A review," *Climate*, vol. 10, no. 9, p. 124, Aug. 2022.

[34] A. Gyaneshwar, A. Mishra, U. Chadha, P. M. D. R. Vincent, V. Rajinikanth, G. P. Ganapathy, and K. Srinivasan, "A contemporary review on deep learning models for drought prediction," *Sustainability*, vol. 15, no. 7, p. 6160, Apr. 2023.

[35] F. A. Prodhan, J. Zhang, S. S. Hasan, T. P. P. Sharma, and H. P. Mohana, "A review of machine learning methods for drought hazard monitoring and forecasting: Current research trends, challenges, and future research directions," *Environ. Model. Softw.*, vol. 149, Mar. 2022, Art. no. 105327.

[36] A. Basak, A. T. M. S. Rahman, J. Das, T. Hosono, and O. Kisi, "Drought forecasting using the prophet model in a semi-arid climate region of western India," *Hydrol. Sci. J.*, vol. 67, no. 9, pp. 1397–1417, Jul. 2022.

[37] A. Elbeltagi, M. Kumar, N. L. Kushwaha, C. B. Pande, P. Ditthakit, D. K. Vishwakarma, and A. Subeesh, "Drought indicator analysis and forecasting using data driven models: Case study in jaisalmer, India," *Stochastic Environ. Res. Risk Assessment*, vol. 37, no. 1, pp. 113–131, Jan. 2023.

[38] M. Achite, N. Elshaboury, M. Jehanzaib, D. Vishwakarma, Q. Pham, D. Anh, E. Abdelkader, and A. Elbeltagi, "Performance of machine learning techniques for meteorological drought forecasting in the wadi Mina basin, Algeria," *Water*, vol. 15, no. 4, p. 765, Feb. 2023.

[39] E. Gul, E. Staiou, M. J. S. Safari, and B. Vaheddoost, "Enhancing meteorological drought modeling accuracy using hybrid boost regression models: A case study from the Aegean region, Türkiye," *Sustainability*, vol. 15, no. 15, p. 11568, Jul. 2023.

[40] "Soils and land use of Chitradurga District, Karnataka," Nat. Bur. Soil Surv. Land Use Planning (NBSSLUP), Nagpur, India, Tech. Rep., 2022. [Online]. Available: https://nbsslup.icar.gov.in/wp-content/uploads/2021/Publications/District_Reports/Chitradurga.pdf

[41] "Ground water information booklet, Chitradurga district, Karnataka state," Central Ground Water Board (CGWB), Bangalore, India, Tech. Rep., 2013. [Online]. Available: https://cgwb.gov.in/District_Profile/karnataka/2012/CHITRADURGA-2012.pdf

[42] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, "Present and future Köppen-Geiger climate classification maps at 1-km resolution," *Sci. Data*, vol. 5, no. 1, pp. 1–12, 2018.

[43] "Manual for drought management," Dept. Agricult. Cooperation, NIDM, Minsitry Agricult., GOI, New Delhi, India, Tech. Rep., 2009. [Online]. Available: https://agricoop.nic.in/sites/default/files/DroughtManual.pdf

[44] N. Karayiannis and A. Venetsanopoulos, *Artificial Neural Networks: Learning Algorithms, Performance Evaluation, and Applications* (The Springer International Series in Engineering and Computer Science). NY: Springer, 2010.

[45] K. Du and M. Swamy, *Neural Networks and Statistical Learning*. London, U.K.: Springer, 2019.

[46] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Ann. Statist.*, vol. 29, no. 5, Oct. 2001.
[47] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers Neurorobotics*, vol. 7, p. 21, Dec. 2013.
[48] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 3149–3157. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
[49] J. H. Friedman, "Multivariate adaptive regression splines," *Ann. Statist.*, vol. 19, no. 1, pp. 1–67, Mar. 1991.
[50] J. H. Friedman and C. B. Roosen, "An introduction to multivariate adaptive regression splines," *Stat. Methods Med. Res.*, vol. 4, no. 3, pp. 197–217, 1995.
[51] T. Hastie, R. Tibshirani, and J. Friedman, "Additive models, trees, and related methods," in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009, pp. 295–336.
[52] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: An interdisciplinary review," *J. Big Data*, vol. 7, no. 1, p. 94, Dec. 2020.
[53] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Red Hook, NY, USA: Curran Associates, 2018, pp. 6639–6649. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf
[54] T. McKee, N. Doesken, and J. Kleist, "The relationship of drought frequency and duration to time scales," in *Proc. 8th Conf. Appl. Climatol.* Anaheim, CA, USA: American Meteorological Society, pp. 179–184.
[55] K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," *J. Geophys. Res., Atmos.*, vol. 106, no. D7, pp. 7183–7192, Apr. 2001.
[56] M. Shahdad and B. Saber, "Drought forecasting using new advanced ensemble-based models of reduced error pruning tree," *Acta Geophysica*, vol. 70, no. 2, pp. 697–712, Apr. 2022.

GEETABAI S. HUKKERI received the integrated B.E. and Ph.D. degrees in computer science and engineering from Visvesvaraya Technological University, Belagavi, India. Since 2023, she has been an Assistant Professor with the Computer Science and Engineering Department, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India. She is the author of two books, four journal articles, and more than five conference publications. Her research interests include computer vision, information retrieval, big data, machine learning, and deep learning and its applications.

SUJAY RAGHAVENDRA NAGANNA received the B.E. degree in civil engineering from Visvesvaraya Technological University, Belagavi, India, in 2011, and the M.Tech. and Ph.D. degrees in water resources engineering from the National Institute of Technology Karnataka, Surathkal, India, in 2014 and 2019, respectively. He is currently an Assistant Professor with the Department of Civil Engineering, Manipal Institute of Technology, Bengaluru, Karnataka, India. He has authored several research articles in various international journals and conferences. His principal research interests include estimation of streambed hydraulic properties, stream-aquifer interaction modeling, statistical and geostatistical analysis, advanced concrete technology, the built environment, and sustainable design of architectural concrete.

DAYANANDA PRUTHVIRAJA (Senior Member, IEEE) received the B.E., M.Tech., and Ph.D. degrees in computer science and engineering from Visvesvaraya Technological University, Belagavi, India. Since 2023, he has been a Professor and the Head of the Information Technology Department, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India. In previous assignment, he was a Professor and the Head of the Department of Information Science and Engineering, JSSATE, Bengaluru. He has published many articles in national and international journals in the field of image processing and retrieval. He has few research grants and conducted consultancy works. His research interests include image processing and information retrieval, machine learning, and deep learning and its applications.

NAGARAJ BHAT received the B.E., M.Tech., and Ph.D. degrees in computer science and engineering from Visvesvaraya Technological University, Belagavi, India. He was a Visiting Scientist with the Department of Civil, Construction, and Environmental Engineering, The University of Alabama, Tuscaloosa, USA, from October 2021 to October 2023. Currently, he is a Professor with the Computer Science and Engineering Department, Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal. He is the author of 25 journal articles and more than 20 conference publications. His research interests include weather informatics, remote sensing, and GIS applications for climate data analysis. His expertise lies in developing tools for climate data analysis, integrating numerical weather prediction (NWP) models with GIS platforms, and utilizing deep learning algorithms for multi-disciplinary research.

R. H. GOUDAR received the B.E. degree in computer science and engineering, the M.Tech. degree in geo-informatics, and the Ph.D. degree in computer science and engineering from Visvesvaraya Technological University, Belagavi, India. He is currently an Associate Professor with the Department of Computer Network Engineering, Visvesvaraya Technological University. He has 19 years of teaching experience at professional institutes across India. He was a Faculty Member with the International Institute of Information Technology, Pune, for four years, and the Indian National Satellite Master Control Facility, Hassan, India. He has published over 145 papers in international journals, book chapters, and conferences of high reputation. His research interests include semantic web, cloud, big data, machine learning, and deep learning and its applications. He has received various awards like the Outstanding Faculty Award, the Research Performance Award, the Young Research Scientist Award from VGST Karnataka, and the Eminent Engineer Award from the Honorable Chief Minister of Karnataka.