IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 1 June 2024, accepted 7 July 2024, date of publication 11 July 2024, date of current version 23 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3426614

RESEARCH ARTICLE

Improving Meteorological Drought Prediction in Tamil Nadu Through Weighted Dataset Construction and Multi-Objective Optimization

KARPAGAM SUNDARARAJAN[1], KATHIRAVAN SRINIVASAN[2], (Senior Member, IEEE), AND JAYAKUMAR KALIAPPAN[2]

[1]School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore 632014, India
[2]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore 632014, India

Corresponding author: Kathiravan Srinivasan (kathiravan.srinivasan@vit.ac.in)

ABSTRACT Droughts typically develop gradually, and early prediction is crucial for the government to formulate effective mitigation plans. Our approach does not involve predicting specific drought index values. Instead, we forecast whether a particular year will experience drought. Insufficient investigation has been carried out regarding variations in additional climatic indicators like shortwave radiation, wind speed, sea level, and pollution in the context of droughts in the state of Tamil Nadu, India. In the study period taken from 1995 to 2020, only three years (2002, 2009, and 2017) experienced drought occurrences, resulting in an imbalanced dataset. To enhance the classification performance of this imbalanced dataset, a weighted dataset is constructed using a feature weighting approach known as the Single Objective Scorer (SOS) based Multi-objective PSO(MPSO) in conjunction with the Gradient Boosting Classifier. The proposed model facilitates objective-based multi-population formation and neighborhood learning. Precision and recall are crucial metrics, particularly in measuring imbalanced dataset classification performance. The application of multi-objective optimization techniques helps to strike a suitable balance between precision and recall. In addition to the Standardized Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI), 14 climatic indicators based on land, atmosphere, and sea are utilized. By employing the weighted dataset created with SOS-based MPSO, a significant improvement in recall value of 0.81 is achieved. Based on the weights assigned to the features, it is identified that the Mean Sea Level of the Arabian Sea and CO_2 are significant indicators for predicting meteorological drought. The Explainable AI techniques SHAP and LIME are employed for interpreting the drought prediction model, providing insights into its workings.

INDEX TERMS Particle swarm optimization algorithm, multi-population, climatic indicators, imbalanced dataset, pollution, mean sea level.

I. INTRODUCTION

This section provides an introduction to various concepts utilized in this work, including feature weighting, the neighborhood learning strategy of the Particle Swarm Optimization (PSO) algorithm, the precision-recall trade-off, multi-objective optimization, drought indices, and climatic indicators.

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

A. FEATURE WEIGHTING

Feature weighting refers to the determination of the significance of features in the classification process [1]. The magnitude of a feature indicates its level of influence on classification performance, whether it is high or low. Various methods exist for feature weighting, including those based on the Pearson correlation coefficient, Fisher coefficient, Information theory, and decision tree ranking. In addition, the Intelligent Minkowski k-means (imwk-means) approach is employed to weigh features for selection purposes. When multiple weights are assigned to a single feature, either the maximum or the mean value is typically chosen. Feature weighting methods can be categorized according to the learning approach, techniques utilized, and the presence of feedback [2].

B. STANDARD PSO ALGORITHM

The standard Particle Swarm Optimization (PSO) algorithm was originally introduced by Kennedy and Eberhart in 1995 [3]. Since then, numerous researchers have made advancements to the algorithm, making it widely applicable across various fields [4], [5]. The formulas for updating particle positions and velocities are presented in Equations 1 and 2.

$$\mathbf{X}(\mathbf{t}+\mathbf{1}) = \mathbf{X}(\mathbf{t}) + \mathbf{V}(\mathbf{t}+\mathbf{1}) \tag{1}$$

$$\begin{aligned} \mathrm{V}(\mathrm{t}+1) = w_i * \mathrm{V}(\mathrm{t}) + \mathrm{c1} * \mathrm{rand}() * (\mathrm{Xpbest} - \mathrm{X}(\mathrm{t}) \\ + c2 * \mathrm{rand}() * (\mathrm{Xgbest} - \mathrm{X}(\mathrm{t})) \end{aligned} \tag{2}$$

w_i Inertia weight
V(t) Particle's velocity at the time 't'
X(t) Position of the particle at the time 't'
c1 Personal learning factor
c2 Neighborhood learning factor
Rand Random number distributed between 0 and 1 uniformly
Xpbest Particle's best position
Xgbest Global best position

A linearly decreasing inertia weight is utilized in this approach. The value of the inertia weight is determined according to the current iteration and is represented by Equation 3.

$$w_i = \mathrm{wmax} - \left(\frac{\mathrm{wmax} - \mathrm{wmin}}{\mathrm{max\_iteration}}\right) * I \tag{3}$$

wmax = 0.9, wmin = 0.2, w_i = weight at iteration 'i ', max_iteration = Maximum Iteration

The PSO algorithm has experienced improvements in various aspects, including population initialization, neighborhood learning, parameter tuning, multi-population strategies, and learning methodologies [6].

C. MULTI-OBJECTIVE OPTIMIZATION

Evolutionary algorithms have proven to be successful in solving a wide range of multi-objective optimization problems due to their ability to generate diverse populations. Various applications of multi-objective optimization using PSO include reservoir operation for flood control [7], reactive power optimization in distribution network systems to minimize power loss and voltage deviation [8], declustering seismic catalogs into mainshocks, aftershocks, and foreshocks using the Chimps multi-objective optimization algorithm [9], determining reservoir operation policies for a three-reservoir hydropower system in different time periods in Iran [10], and fault location in distributed networks [11]. Several methodologies are employed in the search for multi-objective optimized solutions, such as grid dominance ranking, grid clustering in grid space [12], dynamic neighborhood learning, offspring competitive learning, and reference point mechanisms [13], among others. Particle ranking with multi-objective optimization serves as a key tool for feature weighting [14].

D. MULTI-POPULATION

In the implementation of multi-population techniques, various parameters need to be determined. These include deciding whether a fixed or variable subpopulation count is required, determining the communication interval and policy, establishing the connection topology between subpopulations, defining the search area size, determining whether overlapping is needed, and specifying whether the search strategies of subpopulations should be uniform or different [15]. The utilization of multi-population techniques has shown greater success in addressing combinatorial optimization, multi-objective optimization, and large-scale optimization problems.

E. CLIMATE INDICATORS AND DROUGHT INDICES

The Global Climate Observing System (GCOS) is jointly sponsored by the World Meteorological Organization (WMO), the United Nations Environment Programme (UNEP), the Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific, and Cultural Organization (IOC-UNESCO), and the International Science Council (ISC). Given the increasing complexity of studying the global climate, GCOS, in collaboration with WMO, has identified seven climatic indicators that are particularly effective for climate research, as depicted in Figure 1.

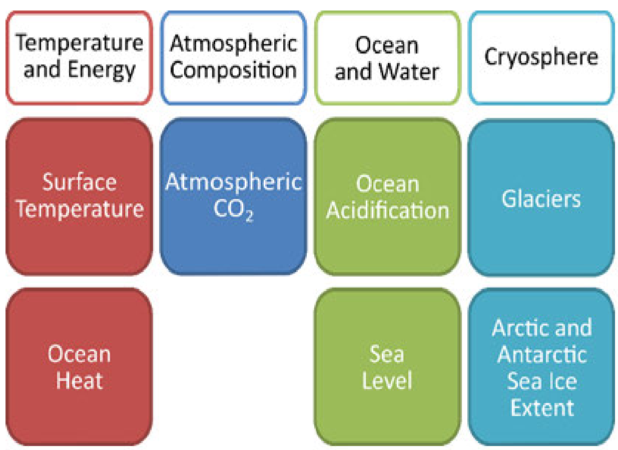

FIGURE 1. Climatic Indicators given by GCOS (Courtesy:https://gcos.wmo.int/en/global-climate indicators).

The global climatic indicators identified by GCOS are divided into four categories: Temperature and Energy, Atmospheric Composition, Ocean and Water, and Cryosphere [16] (gcos.wmo.int). However, in the specific study area of Tamil Nadu, the cryosphere category is not applicable since it lacks glaciers and sea ice. Under the Temperature and Energy category, the indicators considered are minimum, maximum, and mean land temperatures. From the Atmospheric Composition category, the focus is on the CO2 levels. Finally, for the indicators from the Ocean and Water category, there is no open-source data discovered for the ocean acidification parameter for the supplied geographical area and the selected study period. And regarding the sea level, a relative measure of the mean sea level changes of the Arabian Sea, Bay of Bengal, and Indian Ocean are taken. The Essential Climate Variables (ECV) list [17] (public.wmo.int) provided by WMO consists of physical, chemical, or biological variables, or a set of related variables, that significantly contribute to the understanding of Earth's climate. These variables are further grouped and listed under the Atmosphere, Land, and Ocean categories. Precipitation, wind speed, and shortwave radiation are examples of variables falling under the Atmosphere category.

Water vapour is a very good measure for determining atmospheric temperature and precipitation. Water vapour significantly affects the climate system's dynamic and radiative properties. Vapour pressure describes the partial pressure of water vapour in the atmosphere [18]. Dew point temperature refers to the atmospheric temperature lowered to the point of saturation [19]. The changes occurring in the sea also affect the climate of the earth. Various factors bring changes to the ocean volume of the world, resulting in global uniform mean sea level changes [20].

The two most commonly used indicators worldwide for detecting and characterizing meteorological droughts are SPI and SPEI [20]. SPI indicates the precipitation conditions for a specific period within a long time series. It uses precipitation data alone and can characterize both wetness and dryness. SPEI is an extension to the SPI that takes potential evapotranspiration into account. It measures normalized anomalies in precipitation minus potential evapotranspiration. SPI and SPEI are calculated at various timescales and represented as SPI1, SPI6, SPEI3, SPEI6, and so on. SPEI1 is the index determined over a 1-month period, SPEI3 over a 3-month period, and SPEI6 over a 6-month period.

F. STUDY AREA

The state of Tamil Nadu, India, receives roughly 945 mm (37.2 in) of rainfall annually, of which 48% comes from the northeast monsoon and 32% from the southwest monsoon. Because the state's water resources are entirely dependent on rainfall, monsoon failures result in severe drought and acute water scarcity. The seven agroclimatic zones of Tamil Nadu are as follows: heavy rainfall, high altitude hilly, west, south, northeast, and Kaveri Delta (the most fertile agricultural zone). The elevation map [21] of the state is given in Figure 2.

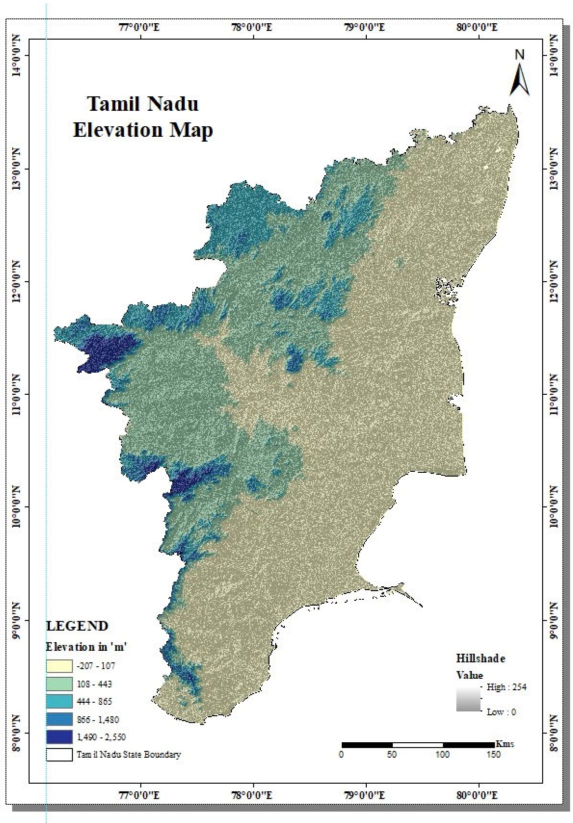

FIGURE 2. **Tamil nadu elevation map.**

G. PRECISION AND RECALL TUG OF WAR

Precision refers to the proportion of correctly classified positive instances out of the total instances classified as positive, regardless of whether they are actually correct or incorrect. On the other hand, recall specifically measures the number of positive instances that are correctly classified. The formulas for precision and recall are provided below in Equation 4 and Equation 5.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{4}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{5}$$

The classification threshold plays a crucial role in determining whether an instance belongs to the positive or negative class. Depending on the value of the classification threshold, the number of false positives or false negatives can increase or decrease. An increase in false positives leads to a decrease in precision, while an increase in false negatives results in a decrease in recall [22].

H. MACHINE LEARNING APPROACHES USED

1) DECISION TREE

The decision tree is a supervised machine learning algorithm used to solve classification as well as regression problems. The internal nodes or decision nodes hold the attribute values, and new branches originate based on its value. The leaf nodes give the final outcome.

A decision tree algorithm operates by recursively choosing the attribute for each internal node that holds the maximum information gain. This step is repeated until the tree reaches a maximum depth or minimum number of samples in a leaf node.

2) RANDOM FOREST

Random Forest is also a supervised machine learning algorithm used in solving classification and regression task. Construct decision tree for each random subsets of training data. The final outcome is the based on the majority voting of the decision tree outcomes [23].

3) GRADIENT BOOSTING

Gradient boosting is one type of ensemble approach; it first builds a model on the training dataset and then the second model is built to rectify the errors in the first model. The base model prediction was done initially by taking the average of the outcomes. The loss function commonly used for regression in Mean Square Error (MSE) and for classification is cross entropy. Our target is to minimize the loss function. In the second step, the residuals are calculated which is the difference between the observed value and the predicted value. In the third step, decision tree is built to predict these residuals. Last step, is to iterate over the third step [23].

In Gradient Tree Boosting, the employment of fixed-size decision trees as base learners is a prevalent approach. This technique is distinguished by its capacity to enhance the accuracy and efficiency of predictions. It achieves this by integrating multiple decision trees into a cohesive, unified model.

I. MOTIVATION AND CONTRIBUTION OF THIS WORK

Many real-world problems involve imbalanced datasets, such as spam prediction, disease diagnosis, and natural disaster prediction (e.g., drought, earthquake, landslide databases). Recognizing the significance of these imbalanced datasets, our motivation was to enhance classification performance while achieving a balance between precision and recall. Our research focuses on studying various climatic indicators in meteorological drought occurrences, driven by the health risks faced by livestock, plants, and humans due to drought. We also aim to improve multi-objective optimization using a multiswarm approach and refine the learning strategy.

The increase in PM2.5 affects the metrological parameters. The increase in CO2 increases the global warming and hence affects the precipitation and evapotranspiration [24]. Naumann [25] studied the relation between the duration of drought and the increase in global warming. It is seen that global mean drought length will be 2.0 months per degree Celsius and accelerating quickly to reach 4.2 months per degree Celsius when global warming approaches 3°C. Dai [26] in his work concludes that there will be increased frequency of drought in the next 30–90 years over many land areas resulting from either decreased precipitation and/or increased evaporation.

The contributions of this work include:

1. Development of a novel Single Objective Scorer (SOS) -based MPSO algorithm, which creates multiple swarms. Each swarm group excels in a specific objective, and learning is facilitated among these swarms.
2. Creation of a weighted dataset using the weights obtained from the SOS-based MPSO in combination with the Gradient Boosting classifier for meteorological drought forecasting.
3. Investigation of the influence of climatic indicators and drought indices on meteorological drought occurrence prediction in the state of Tamil Nadu using Explainable AI techniques.

II. LITERATURE SURVEY

The problem statement for the proposed work was identified through a survey conducted in areas such as feature weighting, multi-objective optimization, and the development of PSO variants.

A. FEATURE WEIGHTING

A feature-weighted Naïve Bayes model was designed by incorporating feature weights into the Naïve Bayesian formula [25]. These feature weights are calculated based on the correlation between features and classes, as well as the inter-correlation between features. To ensure their relevance, the weights are normalized using a sigmoidal function, bringing them within the range of 0 to 1. Jiang introduced a class-specific attribute-weighted Naïve Bayesian approach [27]. Wrapper-based techniques were employed to determine the attribute weights, and two gradient-based feature weighting techniques were also proposed by Jiang. In another study, Jiang incorporated feature weights into the conditional probability estimation and referred to it as the deep feature weighting approach [28]. Correlation-based measures were utilized to calculate the feature weights. This method was subsequently applied to text classification tasks [29].

B. PSO IMPROVEMENTS

Cui designed two archive mechanisms aimed at improving the convergence and divergence processes. The convergence archive population focuses on achieving Pareto dominance [30], while the diversion archive population aims to enhance population diversity. The global leader is selected from these two archives, and flight parameters are adjusted adaptively. Building on this work, Xia et al. [31] devised the Expanded PSO algorithm, inspired by human learning from multiple exemplars and forgetting ability.

In a related study, Wei et al. [32] applied distinct learning strategies to each sub-swarm. Particle behavior is influenced by Adaptive Learning Exemplars (ALE), which are dynamically selected exemplars, and the adaptive population size (APS). However, this approach is time-consuming, making it less suitable for simple unimodal functions.

De Campos et al. [33] explored two parallel PSO techniques, namely Pareto dominance and decomposition, to enhance communication between sub-swarms. The Pareto dominance approach selects dominant solutions for multi-objective problems and identifies the best particles. In the decomposition approach, fitness evaluation is performed for each sub-problem in multi-objective optimization to find the best particle. Both strategies were tested under asynchronous and synchronous communication models.

Building upon these studies, Li et al. [34] proposed the multi-population cooperative particle swarm optimization (MPCPSO) algorithm, which incorporates two learning strategies: the dynamic segment-based mean learning strategy (DSMLS) for exemplar selection and coevolution of populations, and the multidimensional comprehensive learning strategy (MDCLS) for convergence. However, this algorithm encounters challenges in finding the global optimum for complex functions. To address this, a novel mutation operator was introduced to increase diversity, although further improvements are still needed.

Ye et al. [35] classified particles into two categories: communication particles and ordinary particles. Communication particles are utilized for exploitation, using local best solutions, while ordinary particles aid in exploration by considering ‘m’ local best solutions within the subswarms. The algorithm employs a dynamic searching process based on probability values. Evaluations conducted on 10-Dimension and 30-Dimension problems revealed that convergence is slowed down.

Sun et al. [36] introduced two neighborhood selection strategies: all-dimension-neighborhood (ADS) and randomly selected neighbors (RSN). RSN facilitates exploration and is primarily used in earlier stages, while ADS promotes exploitation and is employed in later stages, employing shrinking and random expansion operators.

In the context of feature selection, Kilic et al. [37] proposed a novel multi-population-based PSO algorithm. During population initialization, two population categories are created: one with random initialization and the other with Relief-based measures. The Relief-based measure assigns relevance values between -1 and 1 to each feature, while the random initialization assigns a value between 0 and 1 to each feature. To convert from continuous space to binary space, transfer functions are required.

Wang et al. [38] addressed the time-consuming nature of calculating diversity levels in the population by proposing the diversity-enhanced PSO with neighborhood search (DNSPSO). This approach introduces trial particles into every particle to enhance diversity. Two search strategies, namely Local Neighbourhood Search (LNS) and Global Neighbourhood Search (GNS), are employed. LNS creates new particles based on position values from the ‘K’ nearest neighbors, while GNS generates particles by combining two randomly selected particles in the swarm to facilitate exploration.

Li et al. [39] developed a multi-population approach consisting of an elite population and a shoddy population based on the fitness of solutions. Notably, the Dynamic Opposition-Based Learning strategy allows for a comeback after a stagnated search. This is achieved by monitoring fitness improvement at each iteration and modifying the particle update formula if no improvement occurs consecutively for five iterations, thus avoiding learning from the previous best solution.

Zhang et al. [40] proposed a Dynamic Neighborhood Learning strategy and an offspring competition mechanism. The neighborhood selection is done randomly, and a cross-mutation operator is used for breeding. Through experiments conducted on 11 multimodal functions, the modified particle swarm optimization approach demonstrated improved efficiency.

Yazd et al. [41] applied KNN for selecting the days which are similar to the days of our interest from historical record. Totally 3 variables precipitation, minimum temperature and maximum temperature are taken and choice of data selection is from 4 stations.

C. DROUGHT PREDICTION WITH MACHINE LEARNING

Nabipour et al. [42] forecasted hydrological drought, which is essential for water resource management. They compared the forecasting performance of the standard Artificial Neural Network (ANN) with a hybridized ANN that incorporates nature-inspired optimization algorithms, namely the Salp Swarm Algorithm (SSA), Grasshopper Optimization Algorithm (GOA), Particle Swarm Optimization (PSO), and Biogeography-based optimization (BBO). The PSO algorithm demonstrated the best forecasting performance.

Dikshit et al. [43] conducted research on spatiotemporal drought forecasting using the Standard Precipitation Evaporation Index (SPEI) and climatic indicators such as rainfall, cloud cover, potential evapotranspiration, vapor pressure, and temperature (maximum, minimum, and mean). They employed the Random Forest Regressor for index value prediction and Random Forest classifier for drought class classification. The results indicate that the model performs well in predicting SPEI1 and SPEI3, with potential evapotranspiration (PET) serving as a prominent indicator in the forecasting process.

Danandeh Mehr et al. [44] proposed a fuzzy random forest model to predict SPEI in ungauged catchment areas. The model utilizes global drought information from multiple satellite images and meteorological data. Although the model was tested only for a one-month lead time of SPEI6, forecasting at higher lead times is necessary for effective drought mitigation planning.

Ali et al. [45] conducted research on monthly SPI predictions for Pakistan using a novel drought prediction framework called the Committee Extreme Learning Machine (Comm-ELM) model. This model is based on the committee particle swarm optimization-adaptive neuro-fuzzy inference system (Comm-PSO-ANFIS) and committee multiple linear regression (Comm-MLR) models.

Behifar et al. [46] conducted research on 13 satellite-based indexes. The best metrics for determining the standardized precipitation index (SPI) with a three-month time scale were found to be the indexes based on actual evapotranspiration, precipitation, and soil moisture. Additionally, the drought map was created using two additional ideal measures, the precipitation condition index (PCI) and the evapotranspiration condition index (ETCI), for prediction purposes using RandomForest.

Nematchoua et al. [47] evaluated the performance of six machine learning algorithms in predicting daily global solar radiation and air temperature in 27 cities located across 27 countries. Among the six algorithms tested (Decision Trees (DT), Linear model (LM), Random Forest (RF), Support Vector Machine (SVM), Deep Learning (DL), and Gradient Boosted Trees (GBT)), the performance of Deep Learning (DL) was outstanding. The input variables used for the prediction included wind speed, daily air temperature, solar radiation, and relative humidity recorded in these cities.

D. RESEARCH CHALLENGES AND LIMITATIONS

a) Bioinspired optimization algorithms have been utilized to tune the hyperparameters of various classifiers, including ANN, ANFIS [40], [45], and others. However, there is a lack of research exploring the use of these algorithms to assess the strength or importance of input features in drought prediction.
b) Many studies have focused on predicting the SPI or SPEI values from their past values and based on those value meteorological drought severity level is predicted [43]. However, there is limited research on the binary classification (Yes/No) of meteorological drought occurrence using climatic indicators.
c) The literature commonly employs climatic indicators such as El Nino, Southern Oscillation, Indian Ocean Dipole Mode, and Atlantic Multidecadal Oscillation. However, there is a dearth of studies examining the impact of climatic indicators suggested by GCO, such as pollution factors, sea level, shortwave radiation, dew point, etc., on drought occurrences.
d) Using a variety of indices are critical for effective drought detection, monitoring and management. Until date, there has been no universally accepted drought index among the scientists worldwide. Hence, researchers are still working to alter and reconstruct a comprehensive, simple, and robust drought indicator for effective water resource management and planning [48]. There are uncertainties in the drought projection done in the last decade using PDSI [49].

III. PROPOSED SYSTEM

The meteorological drought occurrence prediction model was designed with two primary objectives:

1. Predicting meteorological drought occurrences using drought indices and climatic indicators.
2. Assessing the impact and contribution of climatic indicators such as pollution and sea level on drought occurrence prediction.

To enhance the prediction performance, a global wrapper-based feature weighting approach is employed. Additionally, a novel algorithm called SOS-based MPSO algorithm is proposed to determine the feature weights.

A. DATA USED

The data utilized in this study to predict the occurrence of meteorological drought were gathered from multiple sources. The data covers a time period from 1995 to 2020, spanning a total of 26 years. The monthly data was collected and hence there 312 instances and 21 attributes values. Within this timeframe, the years 2002, 2007, and 2019 are declared as meteorological drought years by the Government of Tamil Nadu. Refer to Table 1 for a list of the data and their respective sources.

To give a clearer picture about the drought situation of the state SPI6 values for 312 months taken in the study period is plotted and given in Figure 3. The statistical analysis of the dataset is given in Table 2.

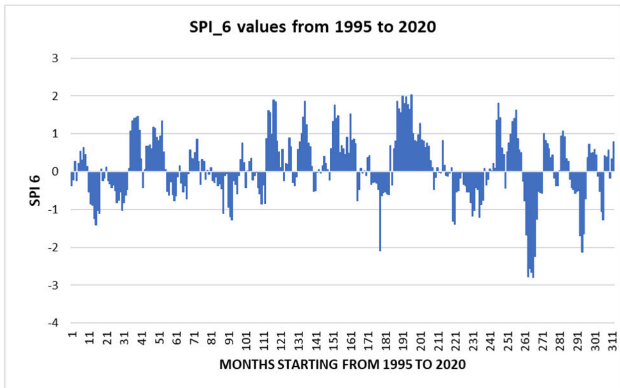

FIGURE 3. Monthwise SPI_6 value from 1995 to 2020.

B. METEOROLOGICAL DROUGHT OCCURRENCE PREDICTION MODEL

The proposed meteorological drought prediction model utilizes 21 input features, which include drought indices and climatic indicators. It performs a binary classification, with the output target variable indicating either Drought or Non-Drought. The weights for these input features are determined using the SOS-based MPSO algorithm. In this implementation, the position vector of the SOS-based MPSO holds the feature weights, and the fitness function aims to maximize precision and recall, with the use of multi-objective optimization.

TABLE 1. List of the data used and its sources.

Data	Source	units
Max temperature, mean temperature, Min temperature,	Climate Engine APP with support from Google Earth Engine and National Oceanic and Atmospheric Administration (NOAA), Desert Research Institute, and US Forest Service. (app.climateengine.com) [50] https://app.climateengine.com/climateEngine	Deg C
Vapour Pressure		kPa
wind speed		Meter per second
precipitation		millimeter
shortwave radiation		W/m^2
Mean Dew point.		Deg C
Cloud Amount.	National Aeronautics and Space Administration (NASA) developed this portal under its Earth Science research program. **(power.larc.nasa.gov)** [51] https://power.larc.nasa.gov/data-access-viewer/	percentage
Annual mean sea level change- Arabian Sea, Bay of Bengal, and Indian Ocean.	Climate Change Indicators Dashboard by IMF's Statistics Department (STA) in association with Economic Co-operation and Development (OECD), the World Bank Group (WBG), the United Nations (UN), the European Commission, and the National Oceanic and Atmospheric Administration (NOAA). **(climatedata.imf.org)**[52] https://climatedata.imf.org/	millimeter
$PM_{2.5}$	Atmospheric Composition Analysis Group of Washington University in St. Louis.[53] **(wustl.app.box.com)** https://wustl.app.box.com/v/ACAG-V5GL01-GWRPM25	micrograms per cubic meter of air (μg/m3).
CO_2	EDGAR v6.0 Greenhouse Gas Emissions published by European Commission, Joint Research Centre.[54] **(data.jrc.ec.europa.eu)** https://data.jrc.ec.europa.eu/dataset/97a67d67-c62e-4826-b873-9d972c4f670b#dataaccess	Tons
SPI	Precipitation data from climateEngine is taken and using the SPI package of R the SPI values are calculated. [55]**(rdrr.io)** https://rdrr.io/cran/spi/man/spi.html	-
SPEI	The Global SPEI database, SPEIbase, is based on monthly precipitation and potential evapotranspiration from the Climatic Research Unit of the University of East Anglia. [56](spei.csic.es) https://spei.csic.es/spei_database/	-

TABLE 2. Statistical test results of the dataset.

S. no	Parameters	Mean	Standard Deviation	Minimum	Maximum
1	Average of Max temp	33.06	2.865	27.170	38.848
2	Average of mean	27.36	2.235	23.006	31.961
3	Average of Min	23.22	2.129	18.824	26.820
4	Average of Vapour Pressure	2.29	0.217	1.7336	2.692
5	Average of wind speed	1.59	2.811	1.020	4.973
6	Average of precitation	2.79	2.338	0.016	13.643
7	Average of shortwave radiation	222.97	32.213	144.149	287.335
8	Average of Mean Dew point	19.56	1.627	15.059	22.331
9	Cloud Amount	63.47	16.779	22.372	92.050
10	PM2.5	26.30	3.899	18.23	31.674
11	CO2	42444.86	23421.53	14268.76	80564
12	Mean Sea level change of Bay of Bengal	26.31	49.171	-78.266	157.696
13	Mean sea level change of Arabian sea	26.48	41.902	-107.94	139.981
14	Mea sea level change of Indian ocean	24.88	26.486	-32.703	77.03
15	SPI_3	0.064	0.880	-3.299	2.730
16	SPI_6	0.095	0.857	-2.788	2.021
17	SPI_12	0.101	0.837	-2.459	2.303
18	SPEI_3	-0.069	0.881	-2.777	1.886
19	SPEI_6	-0.022	0.854	-2.453	1.939
20	SPEI_12	0.022	0.854	-2.353	1.747

During each iteration, the proposed algorithm calculates a new particle position value, which corresponds to the feature weights. Using these weights, a weighted dataset is constructed, and Gradient Boosting is employed as the evaluating classifier to determine the performance metrics of precision and recall. Based on the current precision and recall values, the proposed SOS-based MPSO Model seeks new position values that maximize both precision and recall while minimizing the difference between them.

The weighted dataset is constructed for two reasons:

a. The input dataset suffers from imbalance, leading to poor prediction of the minority drought class. By assigning weights to the features and constructing a weighted dataset, the prediction accuracy can be improved.

b. Incorporating relevant, weighted drought indices and climatic indicators specific to the geographical region as input features enhances the prediction of meteorological drought occurrences.

The reason behind the fitness function design is:

A balance between precision and recall has to be attained, and for that, we can use multi-objective optimization to tune the boundary value that determines whether the binary classification is positive or negative. Because, when the false positive rate rises, so does the precision, and when the false negative rate rises, so does the recall.

The workflow of the overall meteorological drought occurrence prediction process using the SOS-based MPSO is illustrated in Figure 4.

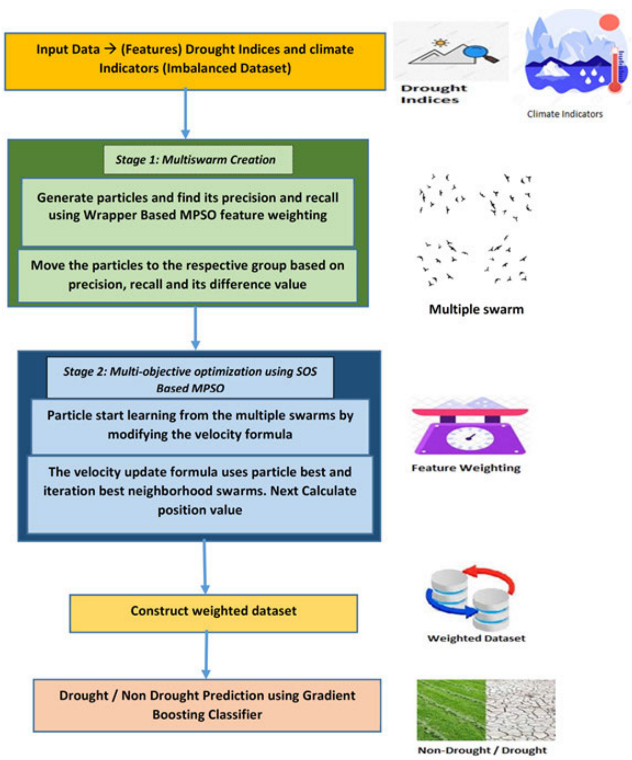

FIGURE 4. Meteorological drought occurrence prediction model using SOS-based MPSO.

C. SINGLE OBJECTIVE SCORER (SOS) BASED MULTIOBJECTIVE PSO ALGORITHM

The proposed SOS-based MPSO algorithm is based on the concept that particles with good performance in a multi-objective solution will be situated between the particles that score well in individual (single) objectives. By learning from these particles, we can approach the best multi-objective solution. The term "individual objective" typically refers to a specific objective within a set of multiple objectives. In most multiswarm techniques, population formation is based on the swarm's fitness value achieved for the multiple objectives, rather than the individual objective. The populations are usually categorized as elite/best/extraordinary and shoddy/ordinary/inferior, and learning is primarily focused on the elite group. However, in our proposed SOS-based MPSO algorithm, population formation is based on swarms that perform well in achieving individual objectives from the set of multiple objectives. Neighborhood learning is conducted using the best population in terms of the individual objective achievement and the best population in terms of the multi-objective achievement. The effectiveness of neighborhood learning methods has been demonstrated by Jinquan et al. [57] and Kennedy et al. [58]. In our (Sundararajan and Kathiravan) previous work on feature weighting with the Two-stage PSO Algorithm [59], modifications were made to PSO population initialization. In this proposed system, however, changes were made to PSO neighborhood learning. The proposed method depends on the single objective best scorers for multi-population generation and learning.

The proposed SOS-based MPSO algorithm initially identifies the populations that excelling in single objectives alone and multi-objectives alone until the iteration reaches MAX_ITER/2. The target population group for improvement after MAX_ITER/2 is the multi-objective group population. In the standard PSO, the position update formula is based on two parameters: the particle's own best and the global best. To prevent particles from converging too early by falling into the same global best, the global best is replaced with the reference particle from other populations. Therefore, in each iteration, the calculation of new particle velocity is performed using two parameters: the particle's own best, and the best particles given each objective within the set of objectives (Objective 1, Objective 2, Objective n).

In this work, we have three single objectives: Objective 1 is to maximize precision; Objective 2 is to maximize recall; and Objective 3 is to minimize the difference between precision and recall. Our multi-objective is to maximize precision and recall.

D. PROPOSED SYSTEM IMPLEMENTATION

The implementation consists of two distinct steps. The first step involves creating the Multiswarm, while the second step focuses on conducting multi-objective optimization using Single Objective scorers.

1) STEP 1 - MULTISWARM CREATION

The input parameters (features) include drought indices such as SPI3, SPI6, SPI12, SPEI3, SPEI6, and SPEI12, as well as prominent climate indicators like sea level, temperature, and CO2. Additionally, essential climate variables related to precipitation, cloud cover, wind speed, and shortwave radiation are considered. The classification process utilizes weighted feature values. Let 'wi' represent the weight for the feature 'i' obtained using the standard PSO. Let 'Fij' denotes the actual feature value, while 'F'ij 'represents the newly calculated weighted feature value. To determine F'ij, the weight value is multiplied by the feature value, if $Fij > 0$ ($F'ij = Fij * wi$), and if Fij < 0, the feature value is divided by the weight value (F'ij = Fij / wi).

Initially, a random particle population is created, and fitness values, namely precision and recall, are calculated. At the end of each iteration, based on the precision value, recall value, and their difference, the particles are assigned to their respective swarm group, as illustrated in Figure 5.

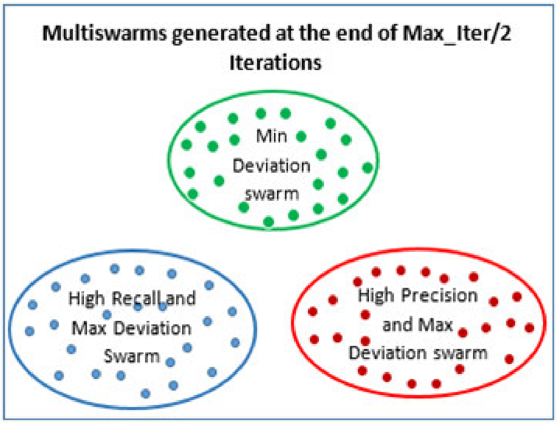

FIGURE 5. **Multiswarm creation.**

The three swarm groups created are:
1. High Precision and Max Deviation (HPMD) swarm
2. High Recall and Max Deviation (HRMD) swarm
3. Min Deviation group

Particles with high precision and maximum deviation between precision and recall values are grouped into one population group called the "High Precision and Max Deviation swarm." Similarly, particles with high recall and maximum deviation with precision belong to the second population group known as the "High Recall and Max Deviation Swarm." The third population group, called the "Min Difference group," consists of particles with minimal difference between precision and recall. The process is halted when the number of iterations reaches half of the Max_Iter value.

2) STAGE 2 - MULTI-OBJECTIVE OPTIMIZATION WITH SINGLE OBJECTIVE SCORERS

Once the number of iterations reaches Max_Iter/2, the focus of the fitness function shifts towards multi-objective optimization. The objective is to maximize precision and recall while minimizing the difference between them. At this stage, the learning process is conducted from the multiswarm, as depicted in Figure 6. Consequently, the modified velocity update formula specified in Equation 6 is utilized.

In accordance with Equation 2, the velocity updating process relies on the particle's personal best (Xpbest) and the global best particle (Xgbest). However, in the SOS-based MPSO approach, reference is made to the best performers in terms of a single objective, rather than the particle's global best. Consequently, the equation is modified, and the new velocity is determined based on the Iteration best particle from the other swarm group, as indicated in Equation 6.

$$\begin{aligned} \mathbf{V(t+1)} &= \mathbf{V(t)} + \mathbf{c1} * \mathbf{rand()} * (\mathbf{Xpbest} - \mathbf{X(t)}) \\ &\quad + \mathbf{c2} * \mathbf{rand()} * (\mathbf{IbestHPMD} - \mathbf{X(t)}) \\ &\quad + \mathbf{c3} * \mathbf{rand()} * (\mathbf{IbestHRMD} - \mathbf{X(t)}) \end{aligned} \tag{6}$$

IbestHPMD - Iteration best particle from High Precision and Max Deviation swarm
IbestHRMD - Iteration best particle from High Recall and Max Deviation swarm

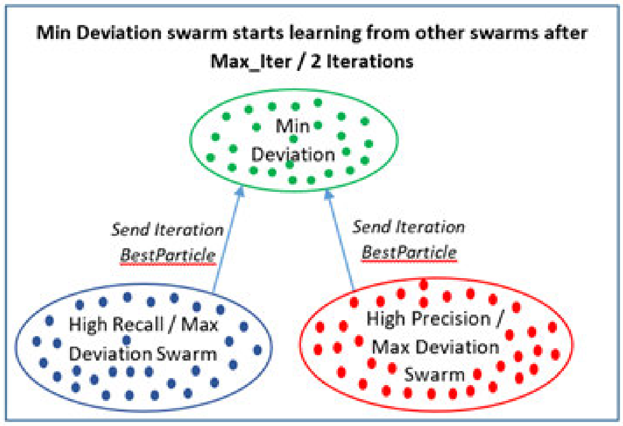

FIGURE 6. **Learning from multiswarm.**

The calculation of the new position value is determined by Equation 1. If the new position value falls below the minimum value or exceeds the maximum value, the particle's position is updated using Equation 7, as presented below. The new position value is assigned as the average of the particle position values of IbestHPMD and IbestHRMD. The pseudocode outlining the proposed work is depicted in Figure 7.

$$\begin{aligned} &X\,(t) = (\mathbf{IbestHPMD} + \mathbf{IbestHRMD})/2 \\ &\text{If X(t)} < \text{min} \ \text{ or X(t)} > \text{max} \end{aligned} \tag{7}$$

E. EXPERIMENTAL SETUP

The implementation is carried out in Python using the sklearn library, and default hyperparameter values are used for all classifiers. The feature weights obtained from these algorithms for the 21 features highlight the importance of the features (drought indices and climatic indicators) in predicting meteorological drought.

Our objective is to assess the performance of the SOS-based MPSO feature weighting approach in predicting meteorological drought. The input dataset consists of 21 features, including drought indices and climatic indicators. The target variable is binary, with 0 indicating "Non Drought" and 1 indicating "Drought." The government has declared the years 2002, 2009, and 2019 as meteorological drought years. To train the model, we utilized the input features from 1995 to 2010. For testing, the years from 2010 to 2020 are used.

```
Pseudocode for SOS-based MPSO
Let the Population Size = S, Dimension=D, Maximum Iteration=
Max_Iter, Particle P = [P1, P2, P3. .PS], Velocity V = [V1, V2, V3...VD],
Position X= [X1, X2…XD]  //Position - represents the weightage of
the features.
Initialize PBest, Iter=0
Randomly Initialize Particle P = [P1, P2, P3 … PS] // Construction of
                                                    multi swarms
 While (Iter<Max_Iter/2)
     For each particle Pi in P
        Calculate weight 'w' using Eq. [3]
        Compute particle velocity 'Vi' using Eq. [2]
        update particle position 'Xi'according to Eq. [1]
          fval_precision,fval_recall=Find_fitness(Pi)
            If  fval_precision > 0.5 and  (fval_precision-fval_recall)
> 0.4 then
            Move Pi to High Precision and Max Deviation swarm
        Elseif  fval_recall > 0.4 and (fval_precision-fval_recall) >
0.4 then
            Move Pi to High Recall and Max Deviation swarm
        Else
            Move Pi to Min Deviation swarm
     End For
        Increment Iter
  End While
  While (Iter<Max_Iter) // Multi-objective optimization with
                     learning from multi swarms
      Do population generation in the HPMD swarm group and
HRMD swarm group.
      IBestHPMD= Max (fval_precision of Pi in P) from the HPMD
swarm group
      IBestHRMD= Max (fval_precision of Pi in P) from the  HRMD
swarm group
    For each particle Pi in P of MinDeviation Swarm group
      Calculate weight 'w' using Eq. [ 3]
      Compute particle velocity Vi using Eq. [6]
      Update particle position Xi according to Eq. [1]
      If (Xi <min or Xi >max) then
          Update particle position  Xi using Eq. [7]
       fval_precision, fval_recall=Find_fitness(Pi)
       pBest=Particle with min deviation and Max precision and Max
Recall
    End For
    Increment Iter
  End While
def find_fitness (Particle Pi )
   Construct a Weighted Dataset using weights from Position X of
particle Pi
   Do Classification on weighted Dataset with GradientBoosting
Classifier
   Calculate Precision, Recall, MCC, and Accuracy
   Return Precision and Recall Value
```

FIGURE 7. Pseudocode for the proposed SOS - based MPSO algorithm.

In this study, we evaluate the performance of the SOS-based MPSO weighting algorithm using various classifiers such as Random Forest, Gradient Boosting, and Decision Tree. We also compare its performance with our previous work on the Two-Stage PSO algorithm. The evaluation is based on precision, recall, and Mathew Correlation Coefficient (MCC) metrics. To showcase the differences between accuracy, precision, and recall, we calculate the accuracy metric.

The MCC is calculated using all four categories of the confusion matrix. It yields a high value only when the model performs well across all categories. The formula for MCC is provided in Equation. 8.

$$\mathbf{MCC} = \frac{\mathbf{TN\ X\ TP - FN\ x\ FP}}{\sqrt{\mathbf{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}} \tag{8}$$

TN - True Negative
TP - True Positive
FN - False Negative
FP - False Positive

Precision and recall measures are calculated using Eq. 4 and Eq. 5, respectively. These measures assess the model's performance for each class (Drought and Non Drought). The accuracy of the model indicates the proportion of correct predictions for the entire dataset. However, it does not provide information about how well the model learned the class boundaries. In imbalanced datasets, the prediction for the minority class tends to be poor, even if the overall accuracy is high. Therefore, we evaluate the class-wise prediction performance using precision and recall.

IV. RESULTS AND DISCUSSION

We compare the performance of Drought and NonDrought class prediction using the Random Forest, Decision Tree, and Gradient Boosting classifiers without feature weighting and with SOS-based MPSO feature weighting using metrics precision, recall, accuracy and MCC values. The metrics are presented in Table 3 and Table 4.

TABLE 3. Drought class prediction performance using various classifiers without feature weighting.

S. No	Method	Drought Class		Non Drought class		MCC	Accuracy
		Precision	Recall	Precision	Recall		
1.	Random Forest	0.9	0.3	0.91	1.0	0.46	0.96
2.	Gradient Boosting	0.81	0.6	0.94	0.97	0.72	0.96
3.	Decision Tree	0.61	0.48	0.94	0.96	0.54	0.95

From Tables 3 and 4, it is evident that all the methods have an accuracy measure above 0.9. However, the precision and recall measures provide a clearer picture of each method's performance in predicting the Drought class.

Table 3 reveals that the prediction results for the Drought class are significantly poorer compared to the Non Drought class. For the Non Drought class, all classifiers achieve precision and recall scores above 0.9. However, these scores

TABLE 4. Prediction performance of SOS - based MPSO feature weighting with various classifiers.

S. No	Method	Precision	Recall	MCC	Accuracy
1.	SOS-based MPSO with Random Forest	0.75	0.64	0.66	0.96
2.	SOS-based MPSO with Gradient Boosting	0.90	0.81	0.76	0.96
3.	SOS-based MPSO with Decision Tree	0.75	0.69	0.68	0.94

are lower for the Drought class. On the other hand, Table 3 demonstrates that without feature weighting, there is a significant difference between precision and recall, particularly noticeable in the Random Forest classifier with a difference of 0.6 (precision: 0.9, recall: 0.3).

In Table 4, the proposed model's performance in predicting the Drought class is presented. Since the Non Drought class does not require any improvement, it is not considered in this analysis. The proposed model successfully enhances the recall scores, particularly for classifiers with low recall without feature weighting. The Gradient Boosting classifier shows the most significant improvement, increasing its recall score from 0.6 to 0.81. The Decision Tree classifier also experiences improvement, with a recall score of 0.69 (up from 0.48), followed by the Random Forest classifier with a recall score of 0.64 (up from 0.3). The high recall values signify the model's excellence in predicting the minority class. Additionally, there is an improvement in precision scores, with Random Forest and Gradient Boosting achieving a high precision value of 0.9. All classifiers successfully achieve the objective of maximizing both precision and recall through the feature weighting approach. After applying feature weighting, the difference between precision and recall values re-mains below 0.1 for all three classifiers. For a visual representation, please refer to Figure 8 which illustrates the impact of feature weighting on drought occurrence prediction.

The second comparison study involves our previous work on Two-Stage PSO feature weighting. Table 5 presents the performance metrics, including precision, recall, MCC, and accuracy, achieved by the Two-Stage PSO algorithm as a feature weighting technique in combination with various classifier algorithms. To visualize the performance comparison between the proposed method and our previous work using the Two-Stage PSO algorithm for feature weighting, please refer to Figure 9, which provides a graph representation.

Based on the performance results, it is evident that the Gradient Boosting classifier outperforms the other classifiers when used in conjunction with the SOS-based MPSO and Two-Stage PSO algorithms. The proposed algorithm demonstrates an improvement in recall values, averaging at 0.1 higher compared to the Two-Stage PSO Algorithm. Consequently, the Gradient Boosting classifier is the preferred

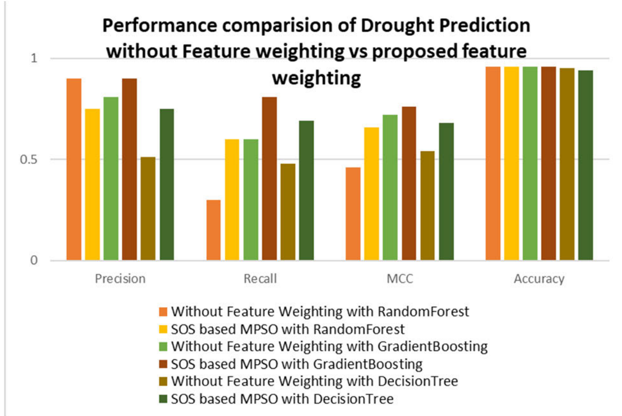

FIGURE 8. Performance comparison of drought prediction without feature weighting and with the proposed feature weighting algorithm.

TABLE 5. Prediction performance of two-stage PSO feature weighting with various classifiers.

S. No.	Method	Precision	Recall	MCC	Accuracy
1.	Two-stage PSO with Random Forest.	0.6	0.3	0.45	0.92
2.	Two-stage PSO with Gradient Boosting.	0.82	0.69	0.72	0.95
3.	Two-stage PSO with Decision Tree.	0.64	0.59	0.62	0.93

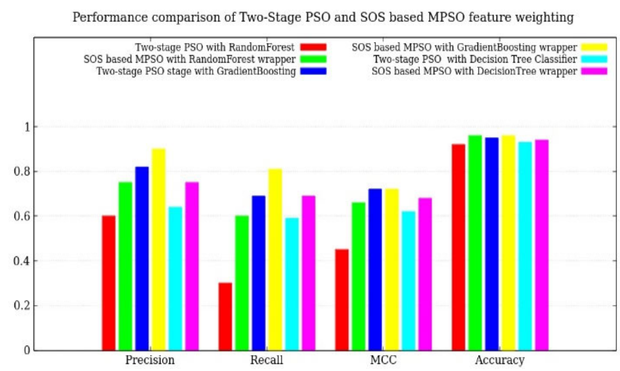

FIGURE 9. Performance comparison of drought prediction with proposed algorithm and two-stage PSO algorithm.

choice over other classifiers. The algorithm is executed with 50 iterations and utilizes 10 particles. Figure 8 showcases the Precision-Recall curve achieved, with an AUC score of 0.85. Figure 10 illustrates the progression of precision and recall values across iterations from 0 to 50. The graph reveals a significant deviation between precision and recall in the initial stages of iteration, which gradually reduces in later stages, resulting in a difference of only 0.09.

The next objective of our analysis is to examine the influence of drought indices and climatic indicators on drought occurrence prediction. This can be observed through the

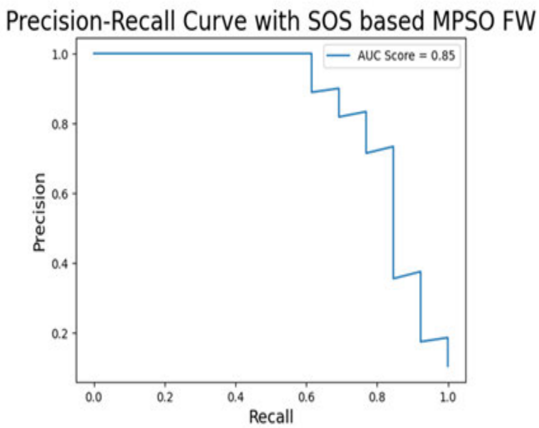

FIGURE 10. Precision-recall curve for SOS-based MPSO with gradientboosting classifier.

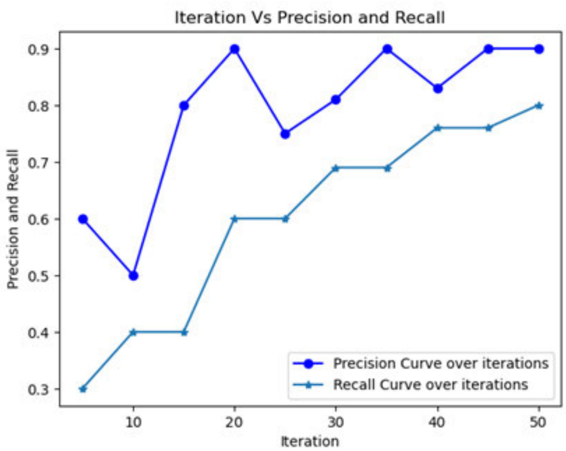

FIGURE 11. Plot of precision and Recall changes over iterations for SOS-based MPSO with gradient boosting classifier.

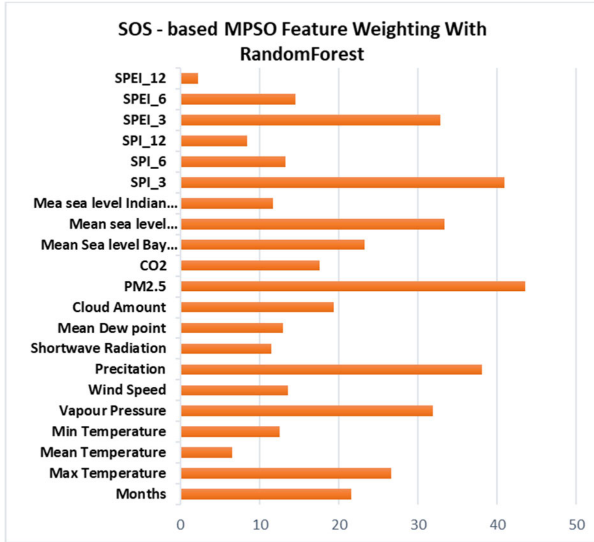

FIGURE 12. Feature Weights given by SOS-based MPSO with random forest.

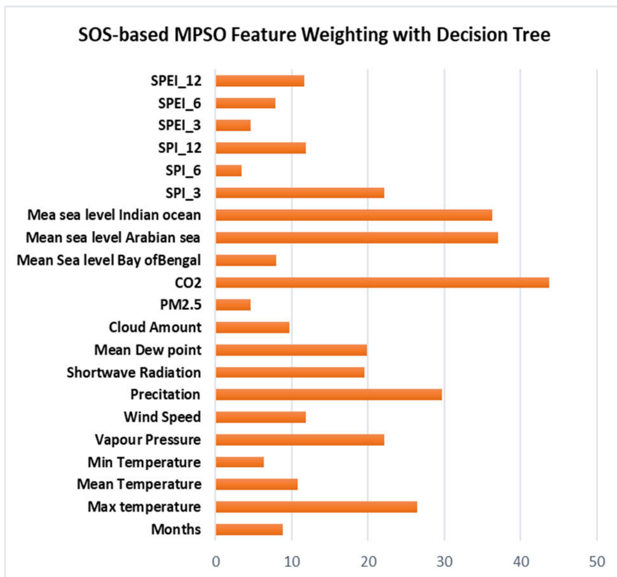

FIGURE 13. Feature Weights given by SOS-based MPSO with decision tree.

feature weight values obtained from the SOS-based MPSO feature weighting. The importance of these features is represented by weight values ranging from 0 to 50. The feature weights obtained from SOS-based PSO weighting with different classifiers, namely Random Forest, Decision Tree, and GradientBoosting, are presented in Figure 12, Figure 13, and Figure 14, respectively.

The top 5 weighted features by each classifier are given in Table 6.

Classification results of the Gradient Boosting and Decision Trees suggest that the mean sea level of the Arabian Sea and CO2 are significant indicators for detecting meteorological drought occurrences in Tamil Nadu. Other important factors in predicting drought include Cloud Amount, Vapor Pressure, PM2.5, Maximum Temperature, and precipitation. Among the drought indices, SPI3, SPI6, and SPEI6 are considered the most reliable indicators for identifying drought events.

To assess the importance of input features in drought classification, the Explainable AI technique known as SHAP

TABLE 6. List of Top 5 weighted Features returned by various classifiers.

S. No	Classifiers	Top 5 weighted features
1.	Random Forest	$PM_{2.5}$, SPI3, Precipitation, Mean Sea level changes of Arabian Sea, and SPEI3
2.	Gradient Boosting	SPI6, SPEI6, CO_2, $PM_{2.5}$, Mean Sea level changes of Arabian Sea, and cloud amount.
3.	Decision Tree	Mean sea level changes of Indian Ocean, Mean sea level changes of Arabian Sea, CO_2, Precipitation, and Max Temperature

(SHapley Additive exPlanations) is utilized. Figure 15 illustrates the SHAP summary plot, revealing the significance of

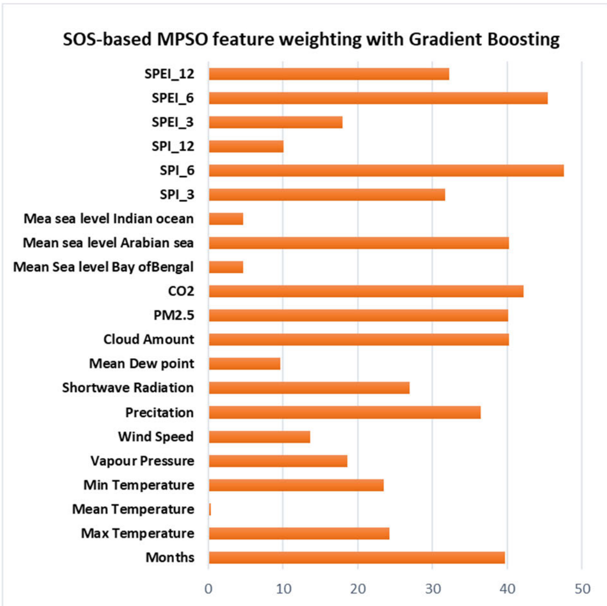

FIGURE 14. Feature Weights given by SOS-based MPSO with gradient boosting.

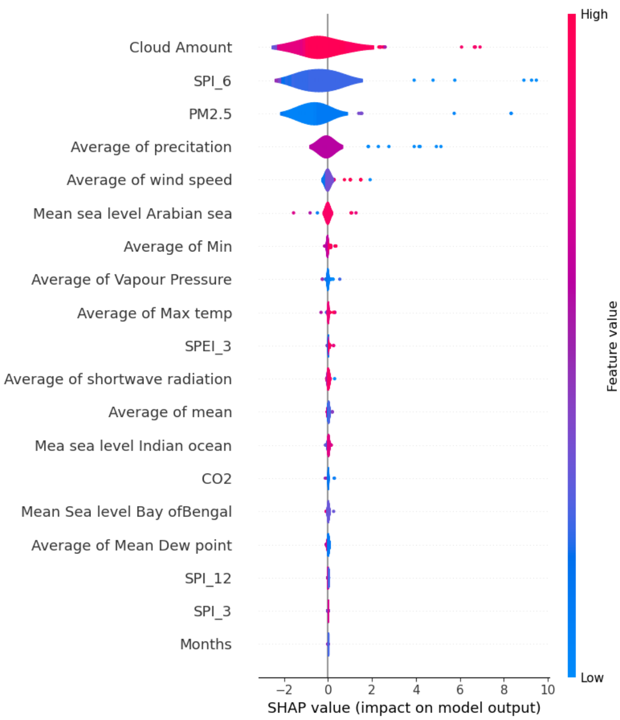

FIGURE 15. SHAP summary plot.

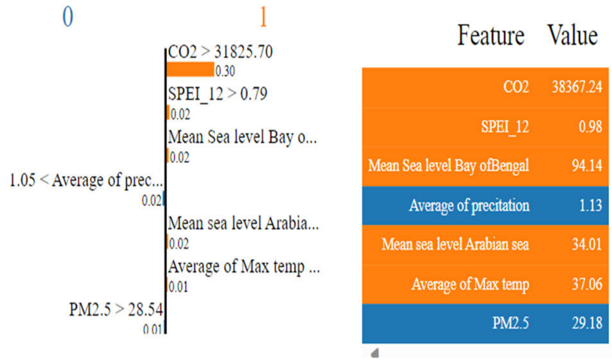

FIGURE 16. LIME plot.

various features. According to the plot, SPI12, PM2.5, and CO2 are the most influential factors, with their contributions to predicting Drought or NonDrought outcomes depicted by pink and blue bars, respectively.

The attribute importance results given by SOS MPSO and SHAP show that to understand the impacts of climate change on hydrological processes, the study of CO2 and sea level changes is crucial.

SPI, coupled with enviro-met (air pollutants and meteorological) parameters, used to measure the drought severity over the Vidarbha region using Random Forest by Kumar et al. [60]. Climatic signals like NINO 3.4, NINO 4, NINO W and SOI are used in addition to lagged SPEI and rainfall as inputs to predict drought [61]. Results have shown that climatic signals alone are giving the best results in drought prediction. In our review paper [62], it can be seen that most of the works use historic SPI and SPEI as input variables, only very few works have used other input factors like air temperature, net radiation, relative humidity, and volumetric soil moisture content. This work uses the combination of two drought indices and climatic indicators to find the drought severity of Tamil Nadu. The relationship existing between PM2.5 and meteorological variables, mainly surface wind and humidity are clearly pointed out by Zhang et al. [24] for their study area China.

The machine learning blackbox was assessed using the LIME (Local Interpretable Model-Agnostic Explanations) technique, enabling us to identify the crucial attributes and their corresponding values that contribute to the decision of classifying a year as either Drought or Non Drought. In Figure 16, the feature and its value are explained, providing insights into the prediction of drought for a specific instance in the year 2017. As 2017 is a drought year, the model accurately classifies it as such, represented by the Drought class marked as '1' (orange), while the Non Drought class is marked as '0' (blue). The features that favor drought classification include CO2, SPEI_12, the mean sea level of the Bay of Bengal and the Arabian Sea, and the Average Maximum Temperature. The value of the attributes for this specific instance is given in the table present inside figure 15. For example, the CO2 value of that instance is 38367.24 and SPEI_12 value is 0.98. The attribute value conditions that favour predicting the drought class CO2 should be greater than 31825.70 and SPEI_12 > 0.79. The orange color bar below it with value 0.30 indicates the feature importance score.

V. CONCLUSION

This paper proposes a novel technique for improving the precision and recall of an imbalanced dataset by combining attribute weighting with multi-objective optimisation. It is the first time that climatic indicators specified by GCOS, in addition to SPI and SPEI, have been used to predict meteorological drought in our research area. The research also indicates that wrapper-based feature weighting methods yield superior results, leading to improved classification accuracy. Another aspect of the study is the incorporation of climatic indicators recommended by GCOS, in addition to SPI and SPEI, to predict meteorological drought. Considering the significant role of pollution elements in climate change, CO2 and PM2.5 are also included in this investigation of drought occurrences. Among various classifiers, Gradient Boosting demonstrates the best performance, with a maximum precision value of 0.9 and a recall of 0.81. By employing the SOS-based MPSO with Gradient Boosting classifier to weight the features, the prediction of meteorological drought occurrences is enhanced. The SOS-based MPSO algorithm utilizes particles that excel in achieving the single objective, in multi-swarm generation, and in learning. The proposed algorithm effectively improves neighborhood learning, resulting in a minor 0.09 difference between precision and recall. The influence of pollution factors on drought occurrences can be elucidated through the assigned weights. An Explainable AI technique is employed to analyze the highly contributing features for drought prediction and to decode how instances predict drought classes. In the future, instance weighting can be performed, and the monthly influence of climatic indicator values on drought prediction can be calculated.

REFERENCES

[1] N. Jankowski and K. Usowicz, "Analysis of feature weighting methods based on feature ranking methods for classification," in *Proc. 18th Int. Conf.*, 2011, pp. 238–247.

[2] I. Niño-Adan, D. Manjarres, I. Landa-Torres, and E. Portillo, "Feature weighting methods: A review," *Expert Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115424, doi: 10.1016/j.eswa.2021.115424.

[3] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. ICNN'95 Int. HY Neural Netw.*, vol. 4, 1995, pp. 1942–1948.

[4] M. Jain, V. Saihjpal, N. Singh, and S. B. Singh, "An overview of variants and advancements of PSO algorithm," *Appl. Sci.*, vol. 12, no. 17, p. 8392, Aug. 2022, doi: 10.3390/app12178392.

[5] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: An overview," *Soft Comput.*, vol. 22, no. 2, pp. 387–408, Jan. 2018.

[6] T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh, and S. Mirjalili, "Particle swarm optimization: A comprehensive survey," *IEEE Access*, vol. 10, pp. 10031–10061, 2022.

[7] H.-T. Chen, W.-C. Wang, X.-N. Chen, and L. Qiu, "Multi-objective reservoir operation using particle swarm optimization with adaptive random inertia weights," *Water Sci. Eng.*, vol. 13, no. 2, pp. 136–144, Jun. 2020.

[8] K. Honghai, S. Fuqing, C. Yurui, W. Kai, and H. Zhiyi, "Reactive power optimization for distribution network system with wind power based on improved multi-objective particle swarm optimization algorithm," *Electr. Power Syst. Res.*, vol. 213, Dec. 2022, Art. no. 108731.

[9] A. Sharma and S. J. Nanda, "A multi-objective chimp optimization algorithm for seismicity de-clustering," *Appl. Soft Comput.*, vol. 121, May 2022, Art. no. 108742.

[10] I. Ahmadianfar, Z. Khajeh, S.-A. Asghari-Pari, and X. Chu, "Developing optimal policies for reservoir systems using a multi-strategy optimization algorithm," *Appl. Soft Comput.*, vol. 80, pp. 888–903, Jul. 2019.

[11] J. Lei, Y. Guo, D. Luo, Z. Xu, and R. Wang, "Fault location of distribution network based on multi-population particle swarm optimization algorithm," *J. Phys. Conf. Ser.*, vol. 2360, no. 1, Nov. 2022, Art. no. 012024.

[12] D. Panday, R. Cordeiro de Amorim, and P. Lane, "Feature weighting as a tool for unsupervised feature selection," *Inf. Process. Lett.*, vol. 129, pp. 44–52, Jan. 2018.

[13] L. Li, G. Li, and L. Chang, "A many-objective particle swarm optimization with grid dominance ranking and clustering," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106661.

[14] A. Rashno, M. Shafipour, and S. Fadaei, "Particle ranking: An efficient method for multi-objective particle swarm optimization feature selection," *Knowl.-Based Syst.*, vol. 245, Jun. 2022, Art. no. 108640.

[15] H. Ma, S. Shen, M. Yu, Z. Yang, M. Fei, and H. Zhou, "Multi-population techniques in nature inspired optimization algorithms: A comprehensive survey," *Swarm Evol. Comput.*, vol. 44, pp. 365–387, Feb. 2019.

[16] *WMO's Global Climate Indicators*. Accessed: Aug. 2023. [Online]. Available: https://gcos.wmo.int/en/global-climate-indicators

[17] *Essential Climate Variables*. Accessed: Aug. 2023. [Online]. Available: https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables

[18] Ø. Hodnebrog et al., "Water vapour adjustments and responses differ between climate drivers," *Atmos. Chem. Phys.*, vol. 19, no. 20, pp. 12887–12899, Oct. 2019.

[19] A. Rovere, P. Stocchi, and M. Vacchi, "Eustatic and relative sea level changes," *Current Climate Change Rep.*, vol. 2, no. 4, pp. 221–231, Dec. 2016, doi: 10.1007/s40641-016-0045-7.

[20] H. Salimi, E. Asadi, and S. Darbandi, "Meteorological and hydrological drought monitoring using several drought indices," *Appl. Water Sci.*, vol. 11, no. 2, pp. 1–11, Feb. 2021.

[21] *TanDEM-X—Digital Elevation Model (DEM)—Global, 30m*. Accessed: May 2023. [Online]. Available: https://gdk.gdi-de.org/geonetwork/srv/api/records/8545a026-2e0c-466f-b6de-99faa639e3c0

[22] *Classification: Precision and Recall*. Accessed: Sep. 2023. [Online]. Available: developers.google.com/machine-learning/crash-course/classification/precision-and-recall

[23] E. K. Sahin, "Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest," *Social Netw. Appl. Sci.*, vol. 2, no. 7, p. 1308, Jul. 2020, doi: 10.1007/s42452-020-3060-1.

[24] X. Zhang, X. Xiao, F. Wang, G. Brasseur, S. Chen, J. Wang, and M. Gao, "Observed sensitivities of PM2.5 and O_3 extremes to meteorological conditions in China and implications for the future," *Environ. Int.*, vol. 168, Oct. 2022, Art. no. 107428.

[25] G. Naumann, L. Alfieri, K. Wyser, L. Mentaschi, R. A. Betts, H. Carrao, J. Spinoni, J. Vogt, and L. Feyen, "Global changes in drought conditions under different levels of warming," *Geophys. Res. Lett.*, vol. 45, no. 7, pp. 3285–3296, Apr. 2018.

[26] A. Dai, "Increasing drought under global warming in observations and models," *Nature Climate Change*, vol. 3, no. 1, pp. 52–58, Jan. 2013, doi: 10.1038/nclimate1633.

[27] L. Jiang, L. Zhang, C. Li, and J. Wu, "A correlation-based feature weighting filter for naive Bayes," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 2, pp. 201–213, Feb. 2019.

[28] L. Jiang, L. Zhang, L. Yu, and D. Wang, "Class-specific attribute weighted naive Bayes," *Pattern Recognit.*, vol. 88, pp. 321–330, Apr. 2019.

[29] L. Jiang, C. Li, S. Wang, and L. Zhang, "Deep feature weighting for naive Bayes and its application to text classification," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 26–39, Jun. 2016.

[30] Y. Cui, X. Meng, and J. Qiao, "A multi-objective particle swarm optimization algorithm based on two-archive mechanism," *Appl. Soft Comput.*, vol. 119, Apr. 2022, Art. no. 108532.

[31] X. Xia, L. Gui, G. He, B. Wei, Y. Zhang, F. Yu, H. Wu, and Z.-H. Zhan, "An expanded particle swarm optimization based on multi-exemplar and forgetting ability," *Inf. Sci.*, vol. 508, pp. 105–120, Jan. 2020.

[32] B. Wei, X. Xia, F. Yu, Y. Zhang, X. Xu, H. Wu, L. Gui, and G. He, "Multiple adaptive strategies based particle swarm optimization algorithm," *Swarm Evol. Comput.*, vol. 57, Sep. 2020, Art. no. 100731.

[33] A. de Campos, A. T. R. Pozo, and E. P. Duarte, "Parallel multi-swarm PSO strategies for solving many objective optimization problems," *J. Parallel Distrib. Comput.*, vol. 126, pp. 13–33, Apr. 2019.

[34] W. Li, X. Meng, Y. Huang, and Z.-H. Fu, "Multipopulation cooperative particle swarm optimization with a mixed mutation strategy," *Inf. Sci.*, vol. 529, pp. 179–196, Aug. 2020.

[35] W. Ye, W. Feng, and S. Fan, "A novel multi-swarm particle swarm optimization with dynamic learning strategy," *Appl. Soft Comput.*, vol. 61, pp. 832–843, Dec. 2017.

[36] W. Sun, A. Lin, H. Yu, Q. Liang, and G. Wu, "All-dimension neighborhood based particle swarm optimization with randomly selected neighbors," *Inf. Sci.*, vol. 405, pp. 141–156, Sep. 2017.

[37] F. Kılıç, Y. Kaya, and S. Yildirim, "A novel multi population based particle swarm optimization for feature selection," *Knowl.-Based Syst.*, vol. 219, May 2021, Art. no. 106894.

[38] H. Wang, H. Sun, C. Li, S. Rahnamayan, and J.-S. Pan, "Diversity enhanced particle swarm optimization with neighborhood search," *Inf. Sci.*, vol. 223, pp. 119–135, Feb. 2013.

[39] X. Li, Z. Wang, Y. Ying, and F. Xiao, "Multipopulation particle swarm optimization algorithm with neighborhood learning," *Sci. Program.*, vol. 2022, pp. 1–20, Jan. 2022.

[40] X. Zhang, H. Liu, and L. Tu, "A modified particle swarm optimization for multimodal multi-objective optimization," *Eng. Appl. Artif. Intell.*, vol. 95, Oct. 2020, Art. no. 103905.

[41] H. Golkar Hamzee Yazd, N. Salehnia, S. Kolsoumi, and G. Hoogenboom, "Prediction of climate variables by comparing the knearest neighbor method and MIROC5 outputs in an arid environment," *Climate Res.*, vol. 77, no. 2, pp. 99–114, Feb. 2019.

[42] N. Nabipour, M. Dehghani, A. Mosavi, and S. Shamshirband, "Short-term hydrological drought forecasting based on different nature-inspired optimization algorithms hybridized with artificial neural networks," *IEEE Access*, vol. 8, pp. 15210–15222, 2020.

[43] A. Dikshit, B. Pradhan, and A. M. Alamri, "Short-term spatio-temporal drought forecasting using random forests model at new South Wales, Australia," *Appl. Sci.*, vol. 10, no. 12, p. 4254, Jun. 2020.

[44] A. Danandeh Mehr, R. Tur, C. Çalıgkan, and E. Tas, "A novel fuzzy random forest model for meteorological drought classification and prediction in ungauged catchments," *Pure Appl. Geophysics*, vol. 177, no. 12, pp. 5993–6006, Dec. 2020.

[45] M. Ali, R. C. Deo, N. J. Downs, and T. Maraseni, "Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting," *Comput. Electron. Agricult.*, vol. 152, pp. 149–165, Sep. 2018.

[46] M. Behifar, A. A. Kakroodi, M. Kiavarz, and G. Azizi, "Satellite-based drought monitoring using optimal indices for diverse climates and land types," *Ecol. Informat.*, vol. 76, Sep. 2023, Art. no. 102143.

[47] M. K. Nematchoua, J. A. Orosa, and M. Afaifia, "Prediction of daily global solar radiation and air temperature using six machine learning algorithms; a case of 27 European countries," *Ecol. Informat.*, vol. 69, Jul. 2022, Art. no. 101643.

[48] M. A. Faiz, Y. Zhang, N. Ma, F. Baig, F. Naz, and Y. Niaz, "Drought indices: Aggregation is necessary or is it only the researcher's choice?" *Water Supply*, vol. 21, no. 8, pp. 3987–4002, Dec. 2021, doi: 10.2166/ws.2021.163.

[49] J. Sheffield, E. F. Wood, and M. L. Roderick, "Little change in global drought over the past 60 years," *Nature*, vol. 491, no. 7424, pp. 435–438, Nov. 2012, doi: 10.1038/nature11575.

[50] Accessed: Aug. 2023. [Online]. Available: https://app.climateengine.com/climateEngine

[51] *NASA Power Data Access Viewer (DAV)*. Accessed: Aug. 2023. [Online]. Available: https://power.larc.nasa.gov/data-access-viewer/

[52] *Climate Change Indicators Dashboard*. Accessed: Aug. 2023. [Online]. Available: https://climatedata.imf.org/

[53] *Satellite-Derived* $PM_{2.5}$. Accessed: Aug. 2023. [Online]. Available: https://wustl.app.box.com/v/ACAG-V5GL01-GWRPM25

[54] *EDGAR v6.0 Greenhouse Gas Emissions*. Accessed: Aug. 2023. [Online]. Available: https://data.jrc.ec.europa.eu/dataset/97a67d67-c62e-4826-b873-9d972c4f670b

[55] *R Package Standardized Precipitation Index (SPI)*. Accessed: Jun. 2023. [Online]. Available: https://rdrr.io/cran/spi/man/spi.html

[56] *Global SPEI Database*. Accessed: Jun. 2023. [Online]. Available: https://spei.csic.es/spei_database/

[57] J. Xu, H. Lin, and H. Guo, "Dynamic neighborhood genetic learning particle swarm optimization for high-power-density electric propulsion motor," *Chin. J. Aeronaut.*, vol. 35, no. 12, pp. 253–265, Dec. 2022.

[58] J. Kennedy, "Small worlds and mega-minds: Effects of neighborhood topology on particle swarm performance," in *Proc. Congr. Evol. Comput. (CEC)*, 1999, pp. 1931–1938.

[59] K. Sundararajan and K. Srinivasan, "Feature-weighting-based prediction of drought occurrence via two-stage particle swarm optimization," *Sustainability*, vol. 15, no. 2, p. 929, Jan. 2023, doi: 10.3390/su15020929.

[60] N. Kumar and A. Middey, "Extreme climate index estimation and projection in association with enviro-meteorological parameters using random forest-ARIMA hybrid model over the vidarbha region, India," *Environ. Monitor. Assessment*, vol. 195, no. 3, p. 380, Mar. 2023, doi: 10.1007/s10661-022-10902-2.

[61] M. H. Le, G. C. Perez, D. Solomatine, and L. B. Nguyen, "Meteorological drought forecasting based on climate signals using artificial neural network—A case study in Khanhhoa province Vietnam," *Proc. Eng.*, vol. 154, pp. 1169–1175, Dec. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877705816319178

[62] K. Sundararajan, L. Garg, K. Srinivasan, A. Kashif Bashir, J. Kaliappan, G. P. Ganapathy, S. K. Selvaraj, and T. Meena, "A contemporary review on drought modeling using machine learning approaches," *Comput. Model. Eng. Sci.*, vol. 128, no. 2, pp. 447–487, 2021.

KARPAGAM SUNDARARAJAN received the B.E. and M.E. degrees in computer science and engineering from Anna University, India, in 2005 and 2010, respectively. She is currently pursuing the Ph.D. degree with the School of Information Technology and Engineering, VIT University, Vellore, India. She has presented and published more than 15 papers in conferences and journals. Her current research interests include machine learning, bio-inspired optimization technique, feature engineering, and deep learning. She is a Life Member of Indian Society for Technical Education.

KATHIRAVAN SRINIVASAN (Senior Member, IEEE) received the B.E. degree in electronics and communication engineering, the M.E. degree (Hons.) in communication systems engineering, and the Ph.D. degree in information and communication engineering from Anna University, Chennai, India. Previously, he was a Faculty/Lecturer with the Department of Computer Science and Information Engineering and also the Deputy Director with the Office of International Affairs, National Ilan University, Taiwan. He is currently an Associate Professor Senior with the School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore, India. He has around 15 years of research experience in the area of machine learning and artificial intelligence and its applications. His research interests include machine learning, artificial intelligence, deep learning, communication systems and networks, computer vision and multimedia, data analytics, and feature engineering. Moreover, he received the Best Service Award as the Deputy Director with the Office of International Affairs, National Ilan University, in 2016, where he also received the Best Service Award from the Department of Computer Science & Information Engineering; the Best Paper Award at the IEEE International Conference on Applied System Innovation, Sapporo, Japan, in May 2017; the Best Paper Award at International Conference on Communication, Management and Information Technology (ICCMIT 2017), Warsaw, Poland; and the Best Conference Paper Award at the IEEE International Conference on Applied System Innovation, Chiba, Tokyo, in April 2018. He is currently serving as an Editor for the IEEE Future Directions and *KSII Transactions on Internet and Information Systems* (TIIS); an Associate Editor for IEEE Access, *IET Networks*, and the *Journal of Internet Technology*; and an editorial board member and a reviewer for various IEEE Transactions, SCI, SCIE, and Scopus indexed journals. He has played an active role in organizing several international conferences, seminars, and lectures. He has been a keynote speaker in many international conferences and IEEE events.

JAYAKUMAR KALIAPPAN received the B.E. degree in computer science and engineering from M. K. University, India, in 2002, and the M.E. degree in computer science and engineering and the Ph.D. degree in intrusion detection systems from Anna University, India, in 2005 and 2018, respectively. He is currently an Associate Professor Senior with the School of Computing Science and Engineering, VIT University, Vellore, India. He has presented and published more than 45 papers in conferences and journals. His current research interests include intrusion detection systems, machine learning, and the IoT. He is a Life Member of Indian Society for Technical Education.