Proposing novel ensemble approach of particle swarm optimized and machine learning algorithms for drought vulnerability mapping in Jharkhand, India

Abstract

Drought, a natural and very complex climatic hazard, causes impacts on natural and socio-economic environments. This study aims to produce the drought vulnerability map (DVM) considering novel ensemble machine learning algorithms (MLAs) in Jharkhand, India. Forty, drought vulnerability determining factors under the categories of exposure, sensitivity, and adaptive capacity were used. Then, four machine learning and four novel ensemble approaches of particle swarm optimized (PSO) algorithms, named random forest (RF), PSO-RF, multi-layer perceptron (MLP), PSO-MLP, support vector regression (SVM), PSO-MLP, Bagging, and PSO-Bagging, were established for DVMs. The receiver operating characteristic curve (ROC), mean-absolute-error (MAE), root-mean-square-error (RMSE), precision, and K-index were utilized for judging the performance of novel ensemble MLAs. The obtained results show that the PSO-RF had the highest performance with an AUC of 0.874, followed by RF, PSO-MLP, PSO-Bagging, Bagging, MLP, PSO-SVM and SVM, respectively. Produced DVMs would be helpful for policy intervention to minimize drought vulnerability.

Keywords: Drought vulnerability; particle swarm optimized; ensemble approaches; exposure index; GIS; adaptive capacity index.

1. Introduction

Drought is one of the most widespread natural climatic hazards and complex phenomena because this hazard affects half of the Earth's surface (Hoque et al. 2020; 2021; Saha et al. 2021). Therefore, drought is a temporary but repetitive phenomenon that is common nowadays. Hence, this phenomenon's Spatio-temporal monitoring and vulnerability assessment is a complex task (Diaz et al., 2020). Drought has been recurring intensely and primarily in all continents, affecting large areas in Africa, Central America, South America, North America, Europe, Oceania, and Asia (Dobrovolski, 2015). Generally, the variability and discontinuity in rainfall patterns aggravate the situations. The arid regions are the most common drought victims, although it can occur under any climate. Climate change and extremes, global warming, and overexploitation are the major causes behind drought and aridity. In simple words, drought can be defined as a condition of dryness causing a shortage of water (Dracup et al., 1980). Drought vulnerability can measure with the help of exposure of the people to drought, sensitivity to drought conditions, and adaptive capacity of the people to recover from drought (Engstrom et al., 2020). Drought is classified into four types: meteorological, hydrological, agricultural, and socioeconomic droughts. Meteorological drought is determined by dry weather conditions in a certain area; hydrological drought is related to lowering water supply; agricultural drought is a situation when the agricultural products are affected, and socioeconomic drought hit the people and their economies (Heim, 2002; Wilhite & Glantz, 1985).

Even if all the drought types are interrelated, India being severely dependent on monsoonal rainfall for agricultural activities is mainly hit by meteorological drought. Drought is a severe topic having deadly impacts, needs concerns and studies. The Tropic of Cancer passes through Jharkhand, and it is landlocked on all sides. Severe droughts have affected the state of Jharkhand in India. According to the Disaster Management Department and Directorate of Statistics and Evaluation, Jharkhand, in the past few years, most of the districts of Jharkhand have been affected by frequent drought occurrences. Jharkhand faced severe droughts in years like 2002 and 2009, when the productivity of rice was decreased remarkably. In the year 2002, the production was 110653 metric tonnes; the yield was 1095 kg/hectare in an area of 136359 hectares and the year 2009, production was 244011 metric tonnes, the yield was 1295 kg/hectare in an area of 188422 hectares which were comparatively very low in comparison to non-drought years. Since 71% of the population of Jharkhand is dependent on agriculture, vulnerability prediction will be an effective measure to combat drought. Drought vulnerability prediction using machine learning is relatively a new concept. No such previous literature was found on vulnerability prediction over Jharkhand using machine learning.

In a study, Adede et al. (2019) used the bagging technique and ensemble with artificial neural network (ANN) and support vector regression (SVR) to evaluate drought severity and vegetation conditions in four northern Kenya counties. Artificial neural network models possess specific characteristics similar to that of neural networks of the human brain. In the Selangor River Basin of Malaysia, drought forecasting was done using drought indices and a multi-layer perception (MLP) neural network (Hong et al., 2015). The models were validated using correlation coefficient, RMSE, and MAE (Hong et al., 2015). In another study by Zahraie et al., 2011, meteorological drought forecasting was done using SPI and Support Vector Machine (SVM) in four basins, including Latian, Karaj, Taleghan, and Mamloo of Iran. In 12 districts of western Korea, the severe drought-affected areas were predicted using Random Forest and land surface factors like vegetation, topography, thermal and water (Park et al., 2019). In an article by Nabipour et al. (2020), hydrological drought was forecasted in the Dez dam in Iran using ANN, biogeography-based optimization (BBO), salp swarm optimization (SSO) and grasshopper optimization algorithm. In New South Wales, Australia, Spatio-temporal drought was forecasted using the random forest (RF) method by Dikshit et al. (2020), and the model was validated using the ROC and area under the curve (AUC). In their work, the RF model performed well in forecasting the Spatio-temporal drought scenario. In most recent papers, Dikshit et al. (2021a; 2021b) used the deep learning algorithms for forecasting the drought, and they pointed out better results of these algorithms than the previously used MLAs. Hoque et al. (2020) assessed the drought vulnerability in Bangladesh using an analytical hierarchical process (AHP). In another work, Hoque et al. (2021) used fuzzy logic for analysing the agriculture drought risk in Northern New South Wales, Australia. But still, no ensemble hybrid machine learning algorithms (MLAs) were used for modelling the drought vulnerability.

Ensemble of PSO and MLAs has been used in different fields, including groundwater modelling (Mallick et al. 2021) and gully erosion modelling (Band et al. 2020), and they got better results of MLAs after ensembling with PSO. Such methods are not used in the field of drought vulnerability. In the present study, MLP, RF, SVM, and Bagging models were ensemble with the PSO for mapping the vulnerability of drought in India's Jharkhand state, considering the three indices: including exposure, sensitivity, and adaptive capacity. The main focus of the previously published research works was forecasting and predicting of drought, but rarely emphasized drought vulnerability. In India, no works have been conducted for modelling the drought vulnerability using the ensemble MLAs considering relevant parameters. Very few pieces have been undertaken on drought vulnerability using AHP (Hoque et al. 2020) and fuzzy logic (Hoque et al. 2021). The use of hybrid machine learning models has yielded positive results in mapping the susceptibility of various natural disasters such as landslides (Roy et al. 2019), floods (Tehrany et al. 2014), and gully erosion (Gayen et al. 2019; Roy et al. 2021). This work tries to address the following research questions: (1) are individual and ensemble MLAs applicable to drought vulnerability analysis? and (ii) can these ensembles provide better results than the knowledge-driven models? Therefore, there is a research gap in this regard, and there is no discussion about the differences among the individual machine learning models (MLP, RF, SVM and Bagging) and ensemble models (PSO-MLP, PSO-RF, PSO-SVM and PSO-Bagging) for the drought vulnerability mapping. The method adopted in the present study is still not used not only in India but also in other region for drought vulnerability.

Forty variables were used in the current study, classified as to exposure, sensitivity, and adaptive capacity. Considering the aforesaid research gap this study, therefore, set out: (i) to find key features related to the drought vulnerability, (ii) to predict the spatial drought vulnerability, and (iii) to assess the performance of the Bagging, MLP, SVM, RF, PSO-Bagging, PSO-MLP, PSO-SVM, and PSO-RF in modelling the drought vulnerability using ROC, MAE, RMSE and K-index.

2. Study area

Most of the parts of the Jharkhand state of India are covered by the Chota Nagpur plateau (Khullar, 1999). Tropic of Cancer passing across Kanke, several kilometres away from the capital of Jharkhand. Its coordinates are extended from 22^{0}28’N-25^{0}30’N latitude and 88^{0}22’E-87^{0}40’E longitude (Figure 1). Jharkhand covers a total geographical area of 79.70 lakh hectares. Rivers like Son, Kharkai, Ajay Mayurakshi, Damodar, North Koel, South Koel, Sankh, Brahmani, and Subarnarekha rivers pass through Jharkhand. Jharkhand experiences two types of climates: humid subtropical in the north to tropical wet and dry in the south-east. The state’s average annual rainfall is about 1255mm, and the average temperature is 33^{0} C (Figure 2). The forest cover is about 29% of the total area of Jharkhand. The total cultivable area of Jharkhand contains 38 lakh hectares with a net sown area of 18.04 lakh hectares. The net irrigated area is 1.57 lakh hectares. Under the socioeconomic aspect, many tribal villages in the Jharkhand state have low agricultural productivity, low availability of agricultural infrastructures, limited soil conservation techniques, and irregular water resources. The agricultural activity of the state heavily depends on monsoonal rainfall. However, drought is a recurrent phenomenon of the state which affects the livelihood of the inhabitants. According to the Drought Prone Areas Programme (DPAP), 12 out of 24 districts are very drought-prone districts. In these conditions, spatial mapping of drought sensitivity is crucial for minimizing the effects of drought and devising drought management strategies in the present research area.

3. Materials and Methods

For preparing the overall drought vulnerability using ensemble machine learning methods, this study took into consideration a total of 40 factors based on the previous literature (Hoque et al. 2020; 2021, Saha et al. 2021; 2021a) and the geo-environmental condition of the state, which were separated into three groups: exposure, sensitivity, and adaptive capability. The entire methodology has been illustrated in Figure 3. Data for this project was gathered from a variety of sources like rainfall and temperature for the period of 1901-2020 from the WRIS and the Indian Meteorological Department, population data from Census of India, agricultural worker, cultivable area, net sown area, an area under forest, net irrigated area and cropping intensity from the district statistical handbook, total water use, water requirement, net water availability from groundwater booklet, health, income, and educational index from Jharkhand economical journal and river data was extracted from DEM (Table 1). After collecting the data, thematic layers were prepared using the geospatial tool.

3.1 Drought Inventory data

Drought inventory data are essential for drought vulnerability analyses using various approaches (Hoque et al. 2020). A total of 200 sample villages as inventory datasets were randomly selected from the drought-prone district as mentioned by the Disaster Management Department of Jharkhand (http://disaster.jharkhand.gov.in). An equal number of non-drought sample locations were randomly selected in the research region. In the present study like drought samples after randomly selecting non-drought samples were used for training and validating the drought vulnerability models. For the current investigation, a 70:30 ratio was used to classify sample drought and non-drought locations considering the previous literature (Hoque et al., 2020). 140 (70%) of the 200 drought sample sites and 140 (70%)

of 200 non-drought sites were utilized as training datasets for running the models, while 60 (30%) drought samples and 60 (30%) non-drought samples were used to verify the models.

3.2 Drought vulnerability indicators

Previous studies (Hoque et al. 2020; 2021; Saha et al., 2021; 2021a) have demonstrated that the drought vulnerability of an area is determined by various factors, including topography, climate, hydrology, economics, and so forth. Hence, reviewing previous literatures, considering the geo-environmental conditions, the usability and dependability of factors, and their suitability for drought vulnerability, twenty-one exposure, nine sensitivity and eight adaptive capacity indicator parameters were selected. The spatial layers were built for each parameter with a resolution of 30m by ArcGIS and can effectively reflect the spatial pattern of vulnerability.

3.2.1 Indicators of drought exposure

The dried-up conditions of the research region depend on drought exposure factors. A total of 21 parameters were used as exposure, including extreme and severe drought, drought magnitude, mean intensity and severe and extreme drought return period of 3-month, 6-month and 12-month, rainfall threshold, rainfall trend, and average rainfall (Figure 4). Using the rainfall data from 1901 to 2020, the standardized precipitation index (SPI) was utilized to measure drought. SPI is a drought index with a multi-time scale that needs precipitation data (Wu et al. 2021). In the light of rainfall deviation from a cumulative probability distribution, the SPI shows wet and dry conditions with user-provided precipitation data for a specified period of time (Zamani et al. 2020). The SPI was applied to calculate the rainfall deficit over various time steps of 3, 6, and 12 months (Ghosh 2019). The 3 and 6-month SPIs help evaluate agricultural effect because it reflects short-term seasonal moisture conditions. The 6 and 12-month SPIs, respectively, represent moderate and long-term moisture conditions, and they would be used to assess the effects of drought (Table 2). Usually, SPI was measured for 3, 6, and 12-month time phases in this study. SPI can be determined for any time scale dependent on the probability of distribution. SPI was computed as (McKee et al. 1993):

$$SPI = \frac{P_i - P_m}{SD} \quad (1)$$

where, precipitation is represented by P_i, average precipitation is represented by P and SD means standard deviation.

Using equation 2, the frequency of severe and extreme droughts was estimated in percentages (Anderson 2018):

$$D_{i,100} = \frac{D_i}{i.t} \times 100 \tag{2}$$

where, in a time scale of I in 100 years, the numbers of drought is represented by $D_{i,\,100}$, for a time scale of i in the t year set, i the number of months with droughts (3, 6, 12, and 24 months) depicted by D_i. Drought magnitude (MD) refers to the total amount of water scarcity experienced throughout the drought period, while drought intensity (MID) is computed by dividing magnitude by time (Aladaileh et al. 2019). The formulas of MD and MID are as follows:

$$MD = \sum\nolimits_{i=1}^{m} SPI_{ij} \tag{3}$$

$$MDI = \frac{\sum\nolimits_{i=1}^{m} SPI_{ij}}{m} \quad \text{or} \quad MDI = \frac{MD}{m} \tag{4}$$

where, for a drought period on the *j* time scale SPI value is denoted by *SPIij*, and the number of months by m. The critical rainfall/rainfall threshold was taken as an important parameter because its help us to unentertaining how a drought starts (Alsumaiei, 2020). Using the equation 5 critical rainfall or threshold rainfall was calculated.

$$CR = \sigma SPI + \overline{X} \tag{5}$$

where, standard deviation is donited by σ, $\overline{X}$ is mean value. The SPI value "-1.5" was chosen as the critical rainfall value in this study. The linear regression was performed to calculate the rainfall trend (Staal et al., 2018). The monthly average precipitation was measured for each district by using total monthly precipitation data (Mutti et al. 2020). Finally, the return period was calculated by applying SPI values. All of the SPI values were first sorted in ascending order. After that, a rank was assigned.

Following that, the return period (RP) was determined by dividing the total year by each rank (Aladaileh et al. 2019).

$$RP = \frac{n}{Er} \tag{6}$$

where, number of droughts is denoted by n and rank of drought by Er.

3.2.2 Indicators of drought sensitivity

The sensitivity indicators used in this study are as follows: the population density, agricultural worker, cultivable area, net sown area, cropping intensity, total water use, water requirement, average temperature, and aridity index, slope and soil texture (Figure 5). These factors affect the region's potential threats of being exposed. For example, increased population pressure in a region exposes more people to drought, increasing the area's vulnerability (Cooley et al., 2019). Water requirements would be higher in areas with a high population density. As a result, population density is directly linked to the severity of the drought since more people will be affected. Drought would be more severe if the temperature rises and vice versa (Balaganesh et al. 2020). The net sown area is also an important factor for identifying drought vulnerability. Water requirements will increase as the net sown area grows, and if needed rainfall does not occur, drought severity will increase dramatically, and vice versa (Balaganesh et al. 2020). In this research, the cultivable area was taken as an important parameter for identifying drought vulnerability. With the increasing cultivable area, the vulnerability of drought will be more in the case of agriculture (Meza et al. 2020).

The trend of temperature is an essential parameter in determining drought susceptibility (Yang et al. 2020). Drought will increase as the temperature rises. The temperature trend has been measured using linear regression in this study. Total water usage was chosen as a sensitivity indicator because it is linked to drought vulnerability since high water use regions have greater water needs during dry years (Chuah et al. 2018). As a result, regions with substantially higher water use zones would experience more severe drought than those with lower water use zones. Drought can wreak havoc in areas with a strong demand for water (Edalat & Stephen, 2019). The total water demand was calculated by adding domestic water demand, crop water demand, livestock water demand, and industrial water demand.

Cropping intensity is defined as the ratio of gross cropped area to net cropped area. Drought intensity will rise in tandem with increased cropping intensity. Drought vulnerability would be greater in areas with high cropping intensity (Kamruzzaman et al. 2019). Owing to lack of water, a greater number of crops would be lost. Despite the fact that there are more small and marginal farmers, the drought will have a greater effect on them (Kuwayama et al. 2019). Drought circumstances would significantly impact small farmers than large farmers since most farmers have a limited quantity of land for agriculture and do not employ high-tech production methods. Therefore, agricultural workers are considered a significant parameter in this study. Soil texture determines the water holding capacity and infiltration rate of an area that is important for the measuring the drought vulnerability situation. Slope controls the surface runoff and groundwater recharge. As a result, slope indirectly influence the drought vulnerability condition of a region. With the increasing aridity index value, the dryness will also increase. In contrast, with the declining value of the aridity index, the dryness will decrease (Deng et al. 2020), so it has gate influence efficiency in occurrences of drought vulnerability. The formula of aridity index is as follows:

$$Aridity\ index(AI) = \frac{PET - AET}{PET} \times 100 \quad (7)$$

where, the PET represents the crop's water demand/need. AET is the actual evapo-transpiration calculated using the water balance methodology and the soil's AWC (available water capacity). The water shortage is denoted by (PET-AET).

3.2.3 Indicators of drought adaptive capacity

Vulnerability is generally defined as a population group's failure to respond appropriately to a certain harmful, stressful event (Rygel et al., 2006). As a result, the socioeconomic indicators of adaptation capability, such as net groundwater availability, net irrigated area, health, income and educational index, area under forest, distance from dam, and river, were incorporated in research (Figure 6). This index indicates the population's ability to recover from drought. For example, drought cannot quickly impact areas with a constant supply of large amounts of water throughout the year. And if there is a year, with lack of rainfall in a specific area, and the region can handle the water deficit by using other sources of water (Xu et al. 2019). In this regards groundwater availbility play an important role in reducing the vulnerability of drought. Drought, on the other hand, will wreak havoc on such places if there are no alternative water supplies other than rainfall.

Drought susceptibility is mostly defined by the health, income, and education of local populations. Drought risk will decrease when health, education, and income possibilities are good in a given region. The health index, education index, and income index indicate human development status (Borja 2020). As the values of those indices increase, so does their adaptive strength; for those reasons, those parameters were taken in this research. The drought events are strongly linked to vegetation cover. Drought sensitivity is low in densely vegetated areas, while the bare ground is more vulnerable to drought since there is little or no vegetation to shield the soil from dryness (dos Santos et al. 2020).

Also, with the increasing distance from dams, rivers, and wetlands, the vulnerability of drought events would be greater in a particular area (Guo et al. 2019; Cavus & Aksoy 2019; Rodríguez et al. 2017). On the other hand, when the distance to a dam, a river, and wetlands decreases, the vulnerability of a specific region to drought events decreases. As the water supply gets reduced with distance increase from dam and river.

3.3 Information Gain Ratio (IGR)

Drought is a natural and complex climatic phenomenon where several variables play a crucial role in drought occurrences. But, in every case, all variables don't have equal responsibility to making the area vulnerable to drought. So, in DVM cases, selecting the responsible factors is a challenging task (Yu et al. 2019). For that reason's IGR was implemented in this research. The IGI is a crucial and well-performed method for selecting drought vulnerability determining factors in predicting drought vulnerability (Mandal et al. 2021). Quinlan, in 1993, first proposed this method and expressed that a more excellent IGR value indicates higher predictive capacity. The IGR was calculated by applying Eq. (8).

$$Information\ Gain\ Ratio = \frac{Info(S) - Info(S,A)}{SplitInfo(S,A)} \quad (8)$$

where, split Info (S, A) denotes the information acquired from training data sets.

3.4 Machine learning methods

3.4.1 Random forest (RF)

In 2001, Breiman introduced the model random forest, an effective ensemble-learning approach (Zhang et al. 2021). For regression, grouping, clustering, and interaction detection, the RF model is used. This method has been used widely in a variety of fields and has shown to be more efficient. Because of the bias and high variance, a decision based on a single tree offers a very poor classification. However, since the RF model employs ensemble trees, it can resolve these issues (Jiang et al. 2020). To generate a forest, the RF model employs hundreds of random binary trees. Any tree is built using the CART (classification and regression trees) approach and randomly subset and selected variables at each node, depending on a bootstrap sample (Khan et al. 2020). When solving classification problems, the RF model considers the vote of the unweighted majority class.

3.4.2 Support vector machine (SVM)

An SVM is a binary classifier for supervised learning in data mining based on the law of structural risk reduction (Dou et al. 2020). The hyperplane construction is distinguished from the training subset of inventory data using this technique. Under the original space of n coordinates, hyperplane distinction was given between the points of two separate classes (xi factor in vector x). SVM creates a classification hyperplane in the centre of the highest margin since it indicates the maximum limit of differentiation among the groups (Chen et al. 2020). Drought presence pixels are designated by the +1 (point over hyperplane), while drought absence pixels are designated by the 1 (point under the hyperplane). The training subsets that are close to the ideal hyperplane are selected. Following the acquisition of a decision surface, label pairs (XiYi) with Xi € Rn, Yi € +1,-1, and I = 1....,m. can be prepared for the classification of new data, including the training subset. Forty physical as well as soci-economic parameters were used to create the drought vulnerability in this study. SVM aims is to discover the optimal distinguishing hyperplane that divides the training subset into two groups: non-drought and drought [0, 1]. This study employed the radial basis function (RBF) kernel to create a drought vulnerability model with SVM (He et al. 2019). More information on the SVM function can be found in the work of Roy et al. (2019).

3.4.3 Bagging

Breiman invented a technique called bagging in 1996, which uses the bootstrap aggregation technique to obtain findings with high prediction precision centred on a dependent classifier by combining numerous examples from a training dataset (Kumar et al. 2021). It was used to create an accurate drought vulnerability mapping. Bagging produces excellent results for huge ensembles; using a more significant number of estimators increases the precision of these approaches (Hong et al. 2020). This ensemble is chosen because a slight shift in the training data reflects and enhances estimate capacity. The three main processes in bagging are the random selection of bootstrap samples to construct a range of training subsets, the development of classifiers for different models, and integrating classifier development in the final model. In the early test stage, one-third of cases in bootstrap tests are not exterminated. The bagging classifier uses the displacement technique to create a bootstrap sample from the current training data (Hong et al. 2018). The bagging hybrid ensemble solutions increase each range of classifiers' success by connecting each classifier to the original feature system for the bagging categorising method (Saha et al. 2021). These situations are known as off-bag checks, according to Breiman (1996). A Bagging fits each base classifier on random subsets of the original dataset, then combines their unique predictions to produce a final prediction (either by average or voting).

3.4.4 Multi-layer perceptron (MLP)

The MLP is the most widely used and researched artificial neural network (ANN), with computer science, simulation, and engineering (Zhu & Heddam, 2020). MLP is made up of multiple layers, including an input layer (where data is fed into the network), one or more hidden layer(s), and an output layer (where the network's simulated/predicted values are displayed) (Ahmadi et al. 2020). Each layer contains some processing elements that are all linked together (neurons). Synaptic weights bind neurons in the next layer to neurons in the previous layer, and these weights are changed over time as the training progresses. Move (activation) functions stimulate neurons. The hyperbolic tangent transfer function was used for both the hidden and output layers in this analysis. The backpropagation training technique was utilized in order to minimize errors in the relationship weights of the neurons between targets and anticipated outputs. The trained network is utilized after the training phase to forecast performance based on unseen input data for the test phase.

3.4.5 Particle Swarm Optimization (PSO)

This study employs PSO, a new learning technique introduced by Kennedy and Eberhart in 1995 (Deng et al. 2019). PSO is a novel population-based global problem optimization approach applied to a wide range of optimization issues (Bangyal et al. 2021). The PSO hypothesis is based on the biological analogy of a swarm of birds and its social eating behaviours. In the context of predictive modelling, the PSO method is first adjusted by random solutions in the search for an ideal scenario via the flying problem space (Pradhan & Bhende, 2019). The best-known situation (i.e., the personal best situation, or pbest) as well as the best-known situation of the entire population (i.e., the global best solution, or gbest) then constantly govern each particle's flight. A larger, more significant similarity exists between PSO and another evolutionary calculation technique, such as the genetic algorithm (Moslehi & Haeri, 2020). The PSO method is a superior alternative to the genetic algorithm. It is typically faster for convergence and has no evolutionary operators such as transverse and selection operations found in other algorithms (Mohammadi et al. 2020). The PSO method is also less subject to adjustment than the derivative knowledge since it directly depends on task values. Thus, the PSO method may be kept at optimal location (rather than global optimum) despite its rapid convergence speeds (Zhang et al. 2020).

3.5 Validation and accuracy assessment of the indices

Six statistical methods such as receiver operating characteristic curve (ROC), mean-absolute-error (MAE), root-mean-square-error (RMSE), precision, K-fold cross-validation, and Friedman rank test were used to evaluate the performance of drought vulnerability models in this study.

3.5.1 Receiver operating characteristic curve (ROC)

In evaluating of the appearance of a classification result, the ROC curve is widely and expertly used (Wang et al. 2021). The area confirms the functioning of the model under the curve (AUC). AUC has a value between 0 to 1 (Luque-Fernandez et al. 2019). It plots the drought-affected location (true predictions) and the non-drought-affected location (false predictions) on the Y-axis and X-axis to illustrate model sensitivity (Saha et al., 2020a). The AUC value of 1 indicates perfect prediction ability. On the other hand, a model's poor output is expressed by a value less than 0.5.

3.5.2 MAE and RMSE

The ROC was used to measure the predictive capability of the implemented models, while the MAE and RMSE were used to examine the inaccuracy in predictive models (Abedinpour et al., 2012). The mean absolute error (MAE) value is calculated by adding all the difference values between the realistic and enumerated values that are distant from each other (Sathishkumar et al. 2020). Eq. 9 has been used for this purpose:

$$MAE = \frac{1}{N}\sum_{i=1}^{n}\left|V_{pred} - V_{obs}\right| \qquad (9)$$

When the difference between the enumerated and the actual executed values is expressed as a ratio, RMSE is defined as the square root of that ratio (Razzaq et al. 2018). The formula of RMSE is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}[V_{pred} - V_{obs}]^2}{N}} \qquad (10)$$

The sample size is defined by N, the expected values are defined by V_{pred}, and the observed values are defined by V_{obs}.

3.5.3 Precision and kappa index (K-index)

Precision was also measured to evaluate the prediction capacity of the implemented models. The degree to which the calculated values compare when the same data is analysed repeatedly (Dennis et al. 2019). The magnitude of random deviations in the measurement process is defined by precision.

$$Precision = \frac{TP}{FP + TP} \qquad (11)$$

where, false positive denotes as FP and true positive denotes as TP.

Several researchers applied the kappa coefficient (K-index) to measure disagreement or reliability between categorical items (McHugh and Mary, 2012). The formula of the K-index is as follows (Phong et al., 2019):

$$K - index = \frac{P_{obs} - P_{\exp}}{1 - P_{\exp}} \quad (12)$$

where, P denotes the total number of drought occurrences pixels, P_{exp} and P_{obs} denotes envisaged and measured of models.

3.5.4 Friedman rank test

To evaluate the major differences among the models, a Friedman ranking test was finally employed. This sub-section looked at the results of a novel ensemble approach of particle swarm optimized and machine learning classifiers using statistical tests on various data sets. Using the same random samples, the novel ensembles were assessed (Marshall et al. 2018). The main goal of these experiments was to see which of the methods used had statistically significant differences in results. In this situation, the Friedman rank test is suitable since the normal distributions are not assumed to be uniform or variance is constant (Craig & Fisher 2019). The major differences among model outputs were investigated using Friedman (1937) signed-rank tests. In deciding, it has been the likelihood of hypotheses (p-value); if the p-value is accurate, the model differences significant and vice versa. In order to calculate the major differences between models for this research, the p-value and z value were employed. If the p value is below 0.05, then the alternative hypothesis would be accepted and null hypothesis would be rejected and the results of the drought vulnerability models are statistically different.

4. Results

4.1 Selection of drought vulnerability indicators by information-gain ratio (IGR)

IGR depicted that all the indicators selected for the present study area suitable for mapping the drought vulnerability of the Jharkhand state (Table 3). Maximum value of IGR was achieved by the rainfall trend (0.326) and the lowest value was found in case of health index (0.006).

4.2 Spatial association of exposure, sensitivity, and adaptive capacity models with drought vulnerability models

The exposure, sensitivity, and adaptive capacity indices were calculated by applying four machine learning and four novel ensemble approaches of particle swarm optimized (PSO) algorithms shown in figures 7-12. Index values varied between 0 and 1. The resulting indices were then divided into five categories: very-low, low, moderate, high, and very high. Drought vulnerability models are associated with adaptive capacity, sensitivity, and exposure indexes, which greatly influence vulnerability occurrences (Hoque et al. 2021; Saha et al. 2021). From the exposure and sensitivity maps, it is observed that the Gridih, Deoghar, Chatra, Hazaribagh, Bokaro, Simdega, Garhwa, and Chatra districts fall under very high exposure to vulnerable conditions as well as very highly sensitive to drought vulnerability. In contrast, West Singhbhum, Khunti, Ranchi, some parts of Palamu, Lather, and Lohardage districts have very-low sensitivity to drought vulnerable conditions for random forest (RF), RF-PSO, multi-layer perceptron (MLP), MLP-PSO, support vector regression (SVM), SVM-PSO, Bagging, and Bagging-PSO models, respectively. The areas of very-high exposure class were 53.95%, 55.87%, 40.94%, 41.96%, 39.57%, 31.81%, 51.05%, and 46.19% for RF, PSO-RF, MLP, PSO-MLP, SVM, PSO-SVM, Bagging, PSO-Bagging models, respectively. For very-high sensitivity index cases, the corresponding values were 63.29%, 62.67%, 44.77%, 33.69%, 40.99%, 45.38%, 63.97%, and 63.80% for RF, PSO-RF, MLP, PSO-MLP, SVM, PSO-SVM, Bagging, PSO-Bagging models, respectively (Figures 7-12). The adaptive capacity map also shows that the Giridih, Deoghar, Sindega, Gumla have very high, Hazaribagh, Chatra, and Garhwa districts of Jharkhand have a high capability of adaptation. In contrast, the West Singhbhum, Khunti, East Singhbhum, Palamu, and Ranchi districts of Jharkhand are the least capable (Figure 11). The substantial geographical connection between exposure, sensitivity, and adaptive capacity plays an important role in drought vulnerability prevalence.

4.3 Drought vulnerability modelling by using RF, PSO-RF, MLP, PSO-MLP, SVM, PSO-SVM, Bagging, PSO-Bagging models

The four-machine learning and four ensemble models were used to create the drought vulnerability maps. Drought vulnerability was classified into five vulnerability groups, as forecasted by each model, using the natural-break method (Hoque et al. 2020; Saha et al. 2021). The developed models' values range from 1.0 to 0.0. Eight models reveal that the Garhwa, Simdega, Deoghar, Giridih, Koderma, Jamtara, Dhanbad, Pakur, and Chatra districts fall under the very-high drought vulnerable conditions. In contrast, Hazaribagh and Gumla districts fall under highly vulnerable conditions despite that West and East Singhbhum, Khunti, Latehar, Ranchi, Palamu, Lohardage, and Saraikela-Kharsawan districts are very-low and low drought vulnerable as estimated by eight implemented models (Figure 13). The very-high drought vulnerability zones cover 41.40%, 39.15%, 44.45%, 47.25%, 47.09%, 38.50%, 40.48%, and 42.64% for RF, PSO-RF, MLP, PSO-MLP, SVM, PSO-SVM, Bagging, PSO-Bagging models, respectively and 24.28%, 32.50%, 11.89%, 12.61%, 7.71%, 16.36%, 21.00%, and 22.28% of the study area falls under very-low drought vulnerability (Figure 14). So, we can conclude that near to 50% of the total study area is under threat. All drought vulnerability maps show that the drought-affected areas are in the state's north-eastern, western, and south-western regions, while the least vulnerable areas are in the state's central and south-eastern regions. Drought vulnerability maps that are smooth patterns are generated concurrently with the results from the eight models. In November 2018, the Jharkhand Government announced 18 districts were drought-prone. As per the IMD report, the rainfall amount was more than 40% less than a normal year in the monsoon season of 2018. Again, in 2019, the Government declared ten districts in Jharkhand were drought-affected. People in seven districts, including Deoghar, Garhwa, Khunti, Pakur, Koderma, Bokaro, and Chatra, were compelled to abandon their agricultural pursuits to work in adjacent mica and stone cutting mines. According to the forecast result, these districts are among the most vulnerable.

4.4 Validation and comparisons of applied models

Model validation is essential to evaluate the model’s prediction capacity (Mandal et al., 2021). The ROC, MAE, RMSE, precision, and K-index were utilized to evaluate the ability of novel ensemble approaches of PSO and MLAs. Also, the Friedman rank test was applied to comparing the applied models. The ROC results shows that the Bagging, PSO-Bagging, MLP, PSO-MLP, SVM, PSO-SVM, RF, and PSO-RF models had accuracy 85.1%, 85.5%, 78.1%, 85.6%, 75.0%, 75.6%, 86.3% and 87.4% in cases of success rate curve and 86.3%, 86.9%, 79.9%, 87.2%, 75.5%, 78.7%, 88.2%, and 88.9% in cases of prediction rate curve, and the significance level (P-value) is 0.00 (Figure 15). The AUC values vary from 0.750 to 0.889, indicating that all models can predict drought vulnerability (Table 4 and 5).

The results of RMSE reveals that 0.291, 0.272, 0.325, 0.241, 0.391, 0.360, 0.223 and 0.185 for Bagging, PSO-Bagging, MLP, PSO-MLP, SVM, PSO-SVM, RF, and PSO-RF models, respectively. The MAE, RMSE, precision, and K-index results are presented in Tables 6 and 5 for the eight applied models.

Finally, we can summarize that the PSO-RF model had the most prediction efficiency despite other applied models. For the justification of the second important model, a single model can't become a suitable candidate. The Friedman test reveals that the used models had significant differences between the models (Table 7). The validation results showed slight differences between the models, so all the models expressed satisfactory results. Finally, the RMSE, MAE, precision, and K-index values are nearly equal for the cases of all applied models. So, we can conclude that all models have more or less uniform prediction capacity for assessing drought vulnerability. But among the selected models, PSO-RF has the highest accuracy as per the validation statistics.

6. Discussion

Drought is a detrimental climatic hazard that causes lots of negative impacts on the natural and socio-economic environment of affected regions. For management purposes and reducing the effects of drought mapping of drought vulnerability is a good way (Hoque et al. 2020; Saha et al. 2021a). In a variety of studies, various set of factors were used to predict future drought conditions and drought susceptibility and the relationship between the factors used (Hoque et al. 2020; 2021; Saha et al. 2021a; Ghosh, 2019; Jiang et al., 2019; Siebert et al., 2019; Ebi et al., 2016; Sridevi et al., 2014; Lindner et al., 2010; Chandrasekar et al., 2009). Few studies have developed DVMs for strategic adaptation, drought risk reduction, and long-term planning (Hoque et al., 2020; Saha et al., 2021). Recently, the various field of researchers also applied novel ensemble approaches and machine learning approaches (MLAs) like RF, Bagging, SVM and PSO in various disciplines, like a landslide, gully erosion, flood hazard, and deforestation evaluation (Tehrany et al., 2014; Huang and Zhao, 2018; Park and Kim, 2019; Band et al., 2020; Saha et al., 2021). The findings show that the ability for the prediction of drought vulnerability by the models applied is acceptable. The PSO-RF model has the highest prediction ability, followed by RF, PSO-MLP, PSO-Bagging, Bagging, MLP, PSO-SVM, and SVM. Roy et al. (2019) and Bui et al. (2016) have established that the novel ensemble model had more predictive accuracy in spite of the individual model for the natural hazard prediction mapping. In our cases, the accuracy of RF, MLR, Bagging, and SVM models have increased the success accuracy by 1.1%, 7.5%, 0.4%, and 0.6%, respectively, after making an ensemble with the PSO model. Chen et al. (2018) applied weight-of-evidence (WoE), and evidential belief function (EBF) models and their ensemble with logistic model tree (LMT) for landslide assessment and the accuracy was increased more than 1% after made ensemble as like our work.

Mallick et al. (2020) applied five ensemble models of M5P, RF, ANN, radial basis function (RBF), PSO for assessing groundwater potential zones. Their study found PSO-RF model performed better result than other models. Hong et al. (2018) used the J48 decision tree model and its ensemble with Bagging, AdaBoost, and RTF meta classifiers for the landslide susceptibility map. The RTF with the J48 model is shown to be the best predictive model, followed by AdaBoost and Bagging ensemble models. The present study shows that the Bagging model had comparatively less prediction capacity. In other studies, like landslides, gully erosion, and floods susceptibility prediction, the random forest and PSO-RF model has good prediction capacity (Gayen et al., 2019; Kong et al., 2020). In addition, the PSO-RF model has stable performance and strong robustness in all prediction cases (Kong et al., 2020). Liu et al. (2012) and Hoang and Tien Bui (2018) concluded that PSO has powerful global parameter adjustment that is easy and simple to implement, and parameter search ability proved that the PSO model had applicability prediction and hazard evaluations.

The application of new ensemble MLAs is thus not new, but the implementation for forecasting drought vulnerability is unique. Several researchers, including Gayen and Pourghasemi (2019), Pham et al. (2019), Saha et al. (2020), and Di Napoli et al. (2020) found that, despite machine learning as well as binary statistical models, a novel ensemble model produced superior outcomes. The difference in exposure to vulnerability may be continentality effects, with inadequate irrigation facilities and delayed monsoon, as the significant reason for climatic and agricultural drought occurrences. For example, in 2019, the Jharkhand state government was announced ten districts as drought-prone. Bokaro has been recorded as the worst-hit rest of other districts like Chatra, Deoghar, Garhwa, Giridih, Godda, Hazaribag, Jamtara, within ten districts Koderma, and Pakur affected by deficit rainfall. In 2018, seven districts like Koderma, Garhwa, Khunti, Bokaro, Pakur, Chatra, and Deoghar recorded rain shortages of above 40%. Koderma and Pakur received less than 55% rainfall than their normal quota. As a result, it is home to a huge population, making it the most vulnerable. Water shortages have been induced as a result of extensive concretization and urbanization, increasing sensitivity. Because of the differences in forest cover, the capacity to adapt differs as well. Finally, we might infer that the North-East and South-West districts of Jharkhand are the least exposed and sensitive to drought, but they are also the most adaptable. In contrast, the centre section of the state is highly exposed and sensitive to drought and the least adaptable. As a result, it may be inferred that Jharkhand's declining rainfall pattern impacted on drought occurrences.

7. Conclusion

In the current drought vulnerability study in Jharkhand, a new ensemble and MLAs based methods were used. The nature of the drought is different, so there are simultaneous variations evaluating vulnerability of drought throughout space and time. A total of 40 conditioning factors were integrated with the ensemble and MLAs. Out of the total area 41.40%, 39.15%, 44.55%, 48.25%, 47.10%, 38.50%, 40.48%, and 42.74% area for RF, PSO-RF, MLP, PSO-MLP, SVM, PSO-SVM, Bagging, PSO-Bagging models, respectively, is covered by the high to very-high drought vulnerability zones. The results show that if adequate drought management strategies are not adopted, the region might likely face a vulnerable scenario in the future. Therefore, the design of irrigation, vegetation, and soil water conservation measures should focus on these drought vulnerable areas. Validation results showed that PSO-RF was the most efficient and precise model, followed by RF, PSO-MLP, PSO-Bagging, Bagging, MLP, PSO-SVM, and SVM. Due to a lack of funds and time, the study is limited in operationalising field observations and capturing farmers' and other local people's perspectives. The absence of agricultural dependence data is another drawback of work. Such a deficiency does not, however, affect the precision of the utilized models. Finally, by switching the related information and factors for future planning and development in different regions, this study technique offers a lot of potential for identifying drought vulnerability zones. Jharkhand, located on the Chota Nagpur Plateau, comprises Precambrian rocks and isolated hills that play an adverse role in groundwater recharge, and extensive urbanization and concretization exacerbate the situation. Private and local governmental organizations should prioritize different water resource management, environmental protection, and land use planning in drought-prone parts of India's Jharkhand state.

Funding: No fund was received for this work.

Conflict of Interest: None

Data availability statement: All the relevant data have been provided in the tables. Sources of all the data have been describe properly. Derived data supporting the findings of this study are available from the corresponding author on request.

References

Abedinpour M, Sarangi A, Rajput TBS, Singh M, Pathak H, Ahmad T. 2012. Performance evaluation of AquaCrop model for maize crop in a semi-arid environment. Agr. Water Manage 110, 55-66.

Ahmadi N, Akbarizadeh G. 2020. Iris tissue recognition based on GLDM feature extraction and hybrid MLPNN-ICA classifier. Neural. Comput. Appl., 32(7), 2267-2281.

Aladaileh H, Al Qinna M, Karoly B, Al-Karablieh E, Rakonczai J. 2019. An investigation into the spatial and temporal variability of the meteorological drought in Jordan. Clim., 7(6), 82.

Alsumaiei AA. 2020. Monitoring Hydrometeorological Droughts Using a Simplified Precipitation Index. Clim., 8(2), 19.

Anderson D. 2018. Depression, dust bowl, demography, and drought: The colonial state and soil conservation in East Africa during the 1930s. In The Colonial Epoch in Africa (pp. 209-231). Routledge.

Balaganesh G, Malhotra R, Sendhil R, Sirohi S, Maiti S, Ponnusamy K, Sharma AK. 2020. Development of composite vulnerability index and district level mapping of climate change induced drought in Tamil Nadu, India. Ecol. Indic., 113, 106197.

Band SS, Janizadeh S, Chandra Pal S, Saha A, Chakrabortty R, Shokri M, Mosavi A. 2020. Novel ensemble approach of Deep Learning Neural Network (DLNN) model and Particle Swarm Optimization (PSO) algorithm for prediction of gully erosion susceptibility. Sensors, 20(19), p.5609.

Bangyal WH, Hameed A, Alosaimi W, Alyami H. 2021. A New Initialization Approach in Particle Swarm Optimization for Global Optimization Problems. Comput. Intell. Neurosci., 2021.

Borja K. 2020. Remittances, corruption, and human development in Latin America. Stud. Comp. Int. Dev., 55(3), pp.305-327.Cavus Y, Aksoy H. 2019. Spatial drought characterization for Seyhan River basin in the Mediterranean region of Turkey. Water, 11(7), 1331.

Chaudhari S, Pokhrel Y, Moran E, Miguez-Macho G. 2019. Multi-decadal hydrologic change and variability in the Amazon River basin: understanding terrestrial water storage variations and drought characteristics. Hydrol. Earth Syst. Sci., 23(7), 2841-2862.

Chen QF, Zou TX, Yang ZT, Chen HJ. 2020. Identification of patients with and without minimal hepatic encephalopathy based on gray matter volumetry using a support vector machine learning algorithm. Sci. Rep. 10(1), 1-8.

Chen W, Shahabi H, Shirzadi A, Li T, Guo C, Hong H, Li W, Pan D, Hui J, Ma M, Xi M. 2018. A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment. Geocarto Int. 33(12), pp.1398-1420.

Chuah CJ, Ho BH, Chow WT. 2018. Trans-boundary variations of urban drought vulnerability and its impact on water resource management in Singapore and Johor, Malaysia. Environ. Res. Lett.13(7), 074011.

Craig AR, Fisher WW. 2019. Randomization tests as alternative analysis methods for behavior-analytic data. J Exp Anal Behav . 111(2), 309-328.

Deng H, Yin Y, Han X. 2020. Vulnerability of vegetation activities to drought in Central Asia. Environ. Res. Lett. 15(8), 084005.

Deng W, Yao R, Zhao H, Yang X, Li G. 2019. A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. Soft Comput. 23(7), 2445-2462.

Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT. 2019. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. Lancet Diabetes Endocrinol. 7(6), 442-451.

Di Napoli M, Carotenuto F, Cevasco A, Confuorto P, Di Martire D, Firpo M, Pepe G, Raso E, Calcaterra D. 2020. Machine learning ensemble modelling as a tool to improve landslide susceptibility mapping reliability. Landslides, 17(8), pp.1897-1914.

Diaz V, Perez GAC, Van Lanen HAJ, Solomatine D, Varouchakis EA. 2020. An approach to characterise spatiotemporal drought dynamics. Adv. Water Resour. 137, 103512.

Dikshit A, Pradhan B, Alamri AM. 2021a. Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. Sci. Total Environ. 755, p.142638.

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

Dobrovolski SG. 2015. World droughts and their time evolution: Agricultural, meteorological, and hydrological aspects. Water Resour. 42, 147–158.

dos Santos CA, Mariano DA, Francisco das Chagas A, Dantas FRDC, de Oliveira G, Silva MT, da Silva LL, da Silva BB, Bezerra BG, Safa B, Medeiros SDS., 2020. Spatio-temporal patterns of energy exchange and evapotranspiration during an intense drought for drylands in Brazil. Int. J. Appl. Earth Obs. Geoinf. 85, p.101982.

Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu Z, Chen CW, Han Z, Pham BT. 2020. Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. Landslides, 17(3), pp.641-658.

Ebi KL, Bowen K. 2016. Extreme events as sources of health vulnerability: Drought as an example. Weather. Clim. Extremes . 11, 95-102.

Edalat MM, Stephen H. 2019. Socio-economic drought assessment in Lake Mead, USA, based on a multivariate standardized water-scarcity index. Hydrol. Sci. J. 64(5), 555-569.

Gayen A, Pourghasemi HR. 2019. Spatial modeling of gully erosion: a new ensemble of CART and GLM data-mining algorithms. In Spatial Modeling in GIS and R for Earth and Environmental Sciences (pp. 653-669). Elsevier.

Ghosh KG. 2019. Spatial and temporal appraisal of drought jeopardy over the Gangetic West Bengal, eastern India. Geoenvironmental Disasters, 6(1), 1.

Guo Y, Huang S, Huang Q, Wang H, Wang L, Fang W. 2019. Copulas-based bivariate socioeconomic drought dynamic risk assessment in a changing environment. J. Hydrol. 575, 1052-1064.

He Q, Shahabi H, Shirzadi A, Li S, Chen W, Wang N, Chai H, Bian H, Ma J, Chen Y, Wang X, 2019. Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms. Sci. Total Environ. 663, pp.1-15.

Hoang ND, Bui DT. 2018. Spatial prediction of rainfall-induced shallow landslides using gene expression programming integrated with GIS: a case study in Vietnam. Nat.l Hazards, 92(3), pp.1871-1887.

Hong H, Liu J, Zhu AX. 2020. Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. Sci. Total Environ. 718, 137231.

Hong H, Liu J, Bui DT, Pradhan B, Acharya TD, Pham BT, Zhu AX, Chen W, Ahmad BB. 2018. Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). Catena, 163, pp.399-413.

Hong H, Tsangaratos P, Ilia I, Liu J, Zhu AX, Xu C. 2018. Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. Sci. Total Environ. 630, 1044-1056.

Hoque MAA, Pradhan B, Ahmed N. 2020. Assessing drought vulnerability using geospatial techniques in north-western part of Bangladesh. Sci. Total Environ. 705, 135957.

Hoque MAA, Pradhan B, Ahmed N, Sohel MSI. 2021. Agricultural drought risk assessment of Northern New South Wales, Australia using geospatial techniques. Sci. Total Environ. 756, p.143600.

Huang Y, Zhao L. 2018. Review on landslide susceptibility mapping using support vector machines. Catena, 165, pp.520-529.

Jiang M, Liu J, Zhang L, Liu, C. 2020. An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. Physica A: Statistical Mechanics and its Applications, 541, 122272.

Jiang W, Wang L, Feng L, Zhang M, Yao R. 2020. Drought characteristics and its impact on changes in surface vegetation from 1981 to 2015 in the Yangtze River Basin, China. Int. J. Climatol. 40(7), pp.3380-3397.

Kamruzzaman M, Jang MW, Cho J, Hwang S. 2019. Future changes in precipitation and drought characteristics over Bangladesh under CMIP5 climatological projections. Water, 11(11), 2219.

Khan Z, Gul A, Perperoglou A, Miftahuddin M, Mahmoud O, Adler W, Lausen B. 2020. Ensemble of optimal trees, random forest and random projection ensemble classification. ADAC, 14(1), 97-116.

Khullar DR. 1999. India: A comprehensive geography. Kalyani Publishers.

Kong C, Xu K, Wang J, Wu C, Liu G. 2020. Landslide susceptibility assessment based on different machine-learning methods in Zhaoping County of eastern Guangxi. Nat. Hazards Earth Syst. Sci. Disc., pp.1-41.

Kumar PS, Behera HS, Nayak J, Naik B. 2021. Bootstrap aggregation ensemble learning-based reliable approach for software defect prediction by using characterized code feature. Innov. Syst. Softw. Eng. 1-25.

Kuwayama Y, Thompson A, Bernknopf R, Zaitchik B, Vail P. 2019. Estimating the impact of drought on agriculture using the US Drought Monitor. Am. J. Agric. Econ. 101(1), 193-210.

Lindner M, Maroschek M, Netherer S, Kremer A, Barbati A, Garcia-Gonzalo J, Lexer MJ. 2010. Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. For. Ecol. Manag. 259(4), 698-709.

Liu Y, Yang X, Na F, Ying W. 2012. Method of particle swarm optimization neural network on geological hazards comprehensive evaluation and its application. J. Seismol. Res. Lett. 35(4), pp.572-577.

Luque-Fernandez MA, Redondo-Sánchez D, Maringe, C. 2019. cvauroc: command to compute cross-validated area under the curve for ROC analysis after predictive modeling for binary outcomes. Stat. J. 19(3), 615-625.

Mallick J, Talukdar S, Alsubih M, Ahmed M, Islam ARMT, Shahfahad, Thanh NV. 2021. Proposing receiver operating characteristic-based sensitivity analysis with introducing swarm optimized ensemble learning algorithms for groundwater potentiality modelling in Asir region, Saudi Arabia. Geocarto Int.pp.1-28.

Mandal K, Saha S, Mandal S. 2021. Applying deep learning and benchmark machine learning algorithms for landslide susceptibility modelling in Rorachu river basin of Sikkim Himalaya, India. Geosci. Front. 12(5), 101203.

Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. 2018. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. Res. Synth. Methods 9(4), 602-614.

McHugh ML. 2012. Interrater reliability: the kappa statistic. Biochem. Med. 22(3), pp.276-282.

Meza I, Siebert S, Döll P, Kusche J, Herbert C, Eyshi Rezaei E, Nouri H, Gerdener H, Popat E, Frischen, J, Naumann G. 2020. Global-scale drought risk assessment for agricultural systems. Nat. Hazards Earth Syst. Sci. 20(2), pp.695-712.

Mohammadi FG, Amini MH, Arabnia HR. 2020. Evolutionary computation, optimization, and learning algorithms for data science. In Optimization, Learning, and Control for Interdependent Complex Networks (pp. 37-65). Springer, Cham.

Moslehi F, Haeri A. 2020. A novel hybrid wrapper–filter approach based on genetic algorithm, particle swarm optimization for feature subset selection J. Ambient Intell. Humaniz. Comput. 11(3), 1105-1127.

Mutti PR, de Abreu LP, de MB Andrade L, Spyrides MHC, Lima KC, de Oliveira CP, Dubreuil V, Bezerra BG. 2020. A detailed framework for the characterization of rainfall climatology in semiarid watersheds. Theor. Appl. Climatol. 139(1), pp.109-125.

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

Park S, Kim J. 2019. Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. Appl. Sci. 9(5), p.942.

Pham BT, Prakash I, Singh SK, Shirzadi A, Shahabi H, Bui DT. 2019. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. Catena, 175, pp.203-218.

Phong TV, Phan TT, Prakash I, Singh SK, Shirzadi A, Chapi K, Ly HB, Ho LS, Quoc NK, Pham BT. 2019. Landslide susceptibility modeling using different artificial intelligence methods: A case study at Muong Lay district, Vietnam. Geocarto Int.pp.1-24.

Pradhan C, Bhende CN. 2019. Online load frequency control in wind integrated power systems using modified Jaya optimization. Eng. Appl. Artif. Intell. 77, 212-228.

Razzaq M, Paulevé L, Siegel A, Saez-Rodriguez J, Bourdon J, Guziolowski C. 2018. Computational discovery of dynamic cell line specific Boolean networks from multiplex time-course data. PLoS Comput. Biol. 14(10), e1006538.

Rodríguez JF, Saco PM, Sandi S, Saintilan N, Riccardi G. 2017. Potential increase in coastal wetland vulnerability to sea-level rise suggested by considering hydrodynamic attenuation effects. Nat. Commun. 8(1), 1-12.

Roy J, Saha S, Arabameri A, Blaschke T, Bui DT. 2019. A novel ensemble approach for landslide susceptibility mapping (LSM) in Darjeeling and Kalimpong districts, West Bengal, India. Remote Sens. 11(23), 2866.

Roy J, Saha S. 2021. Integration of artificial intelligence with Meta classifiers for the gully erosion susceptibility assessment in Hinglo river basin, Eastern India. Adv. Space Res. 67 (1), 316-333

Saha A, Saha S. 2020. Comparing the efficiency of weight of evidence, support vector machine and their ensemble approaches in landslide susceptibility modelling: A study on Kurseong region of Darjeeling Himalaya, India. Remote Sens. Appl.: Society and Environ. 100323.

Saha S, Gogoi P, Gayen A, Paul GC. 2021a. Constructing the machine learning techniques based spatial drought vulnerability index in Karnataka state of India. J. Clean. Prod. p.128073. http://dx.doi.org/10.1016/j.jclepro.2021.128073

Saha S, Kundu B, Paul GC, Mukherjee K, Pradhan B, Dikshit A, Abdul Maulud KN, Alamri AM. 2021. Spatial assessment of drought vulnerability using fuzzy-analytical hierarchical process: a case study at the Indian state of Odisha. Geomatics, Nat. Hazards Risk 12(1), pp.123-153.

Saha S, Paul GC, Pradhan B, Abdul Maulud KN, Alamri AM. 2021. Integrating multilayer perceptron neural nets with hybrid ensemble classifiers for deforestation probability assessment in Eastern India. Geomatics, Nat. Hazards Risk 12(1), 29-62.

Saha S, Saha A, Hembram TK, Pradhan B, Alamri AM. 2020. Evaluating the performance of individual and novel ensemble of machine learning and statistical models for landslide susceptibility assessment at Rudraprayag District of Garhwal Himalaya. Appl. Sci. 10(11), p.3772.

Sathishkumar VE, Park J, Cho Y. 2020. Seoul bike trip duration prediction using data mining techniques. IET Intelligent Transport Systems, 14(11), pp.1465-1474.Sathishkumar, V. E., Park, J., & Cho, Y. (2020). Seoul bike trip duration prediction using data mining techniques. IET Intell. Trans. Systems, 14(11), 1465-1474.

Siebert J, Sünnemann M, Auge H, Berger S, Cesarz S, Ciobanu M, Eisenhauer N. 2019. The effects of drought and nutrient addition on soil organisms vary across taxonomic groups, but are constant across seasons. Sci. Rep. 9(1), 1-12.

Singh TP, Nandimath P, Kumbhar V, Das S, Barne P. 2020. Drought risk assessment and prediction using artificial intelligence over the southern Maharashtra state of India. Modeling Earth Syst. and Environ. 1-9.

Sridevi GD, Jyotishi A, Mahapatra S, Jagadeesh G, Bedamatta S. 2014. Climate Change Vulnerability in Agriculture Sector: Indexing and Mapping of Four Southern Indian States. Quaderni - Working Paper DSE, No. 966, Alma Mater Studiorum - Università di Bologna, Dipartimento di Scienze Economiche (DSE), Bologna, http://dx.doi.org/10.6092/unibo/amsacta/4080

Staal A, Tuinenburg OA, Bosmans JH, Holmgren M, van Nes EH, Scheffer M, Zemp DC, Dekker SC. 2018. Forest-rainfall cascades buffer against drought across the Amazon. Nat. Clim. Chang. 8(6), pp.539-543.

Tehrany MS, Pradhan B, Jebur MN. 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J. Hydrol., 512, pp.332-343.

Wang D, Zhang T, Li M, Bueno, R, Jayender J. 2021. 3D deep learning based classification of pulmonary ground glass opacity nodules with automatic segmentation. Comput. Med. Imaging Graph. 88, 101814.

Wu J, Chen X, Yao H, Zhang D. 2021. Multi-timescale assessment of propagation thresholds from meteorological to hydrological drought. Sci. Total Environ. 765, 144232.

Xu H, Tian Z, He X, Wang J, Sun L, Fischer G, Fan D, Zhong H, Wu W, Pope E, Kent C. 2019. Future increases in irrigation water requirement challenge the water-food nexus in the northeast farming region of China. Agr. Water Manage 213, pp.594-604.

Yang M, Mou Y, Meng Y, Liu S, Peng C, Zhou X. 2020. Modeling the effects of precipitation and temperature patterns on agricultural drought in China from 1949 to 2015. Sci. Total Environ. 711, 135139.

Yu L, Cao Y, Zhou C, Wang Y, Huo Z. 2019. Landslide susceptibility mapping combining information gain ratio and support vector machines: a case study from Wushan segment in the three gorges reservoir area, China. Appl. Sci. 9(22), 4756.

Zamani H, Bazrafshan O. 2020. Modeling monthly rainfall data using zero-adjusted models in the semi-arid, arid and extra-arid regions. Meteorol. Atmos. Phys. 132(2), 239-253.

Zhang F, Chen Y, Zhang J, Guo E, Wang R, Li D. 2019. Dynamic drought risk assessment for maize based on crop simulation model and multi-source drought indices. J. Clean. Prod. 233, 100-114.

Zhang G, Wang M, Liu K. 2019. Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China. Int. J. Disaster Risk Sci. 10(3), 386-403.

Zhang W, Wu C, Zhong H, Li Y, Wang L. 2021. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci. Front. 12(1), 469-477.

Zhang Y, Liu X, Bao F, Chi J, Zhang C, Liu P. 2020. Particle swarm optimization with adaptive learning strategy. Knowl.-Based Syst. 196, 105789.

Zhu S, Heddam S. 2020. Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN). Water Qual. Res. J. 55(1), 106-118.